The Lamperti Transform

Applications to Stochastic Local Volatility Models

S.G. de Boer

$$\psi(X(t),t) := \int \frac{1}{\sigma(x,t)}\,\mathrm{d}x\Big|_{x=X(t)}$$

TU Delft

The Lamperti Transform

Applications to Stochastic Local Volatility Models

by

S.G. de Boer

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Friday August 24, 2020 at 11:00 AM.

Made in collaboration with Optiver Services B.V.

Student number:	4376803	
Project duration:	December 1, 2019 – August 24, 2020	
Thesis committee:	Prof. Dr. Ir. C.W. Oosterlee,	Delft University of Technology
	Dr. D. Kurowicka,	Delft University of Technology
	Dr. Ir. L.A. Grzelak,	Delft University of Technology
	Dr. Ir. P. Chebolu,	Optiver Services B.V.

This thesis is confidential and cannot be made public until February 24, 2021.

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

TU Delft

optiver

Preface

Before you lies the result of over eight months of research into the intricacies of the Lamperti transform and its application to local stochastic volatility models. This document has been written to meet the final set of requirements for completion of the Applied Mathematics MSc programme at the Delft University of Technology. I have written this thesis in close collaboration with the derivatives pricing team at Optiver Services B.V., as they are ever searching for cutting-edge methods to help them solve problems. I am lucky enough to join Optiver as a full-time employee after having completed this MSc track.

The Lamperti transform is a transformation that can be applied to certain stochastic differential equations in order to remove state-dependent diffusion terms. These diffusion terms typically enforce boundary conditions and correlation structures present within the system of stochastic differential equations. Getting rid of such state-dependent terms often leaves the system at hand with a simple unit diffusion term. In turn, a wide variety of opportunities presents itself with respect to parameter estimation, numerical simulation, extension of state spaces and even fully solving systems of equations. This thesis covers the details of the transformation, and explores a solution to a general local stochastic volatility model made possible by the Lamperti transform.

I want to express gratitude to my supervisor, Dr. Ir. L.A. Grzelak, for advising me on the main structure of this thesis and steering me in the right direction whenever I lost sight of the final product. I am also indebted to a number of the great minds at Optiver. Without the help of Dr. I. Petursson, Dr. P. Chebolu, Dr. K. Hartmann, Dr. A. Swiech and Dr. Y. Zhao, this thesis would not have been possible. I must give special thanks to Dr. P. Chebolu for being my daily supervisor within Optiver. Finally, even though our contact was brief, I want to thank Dr. D. Bang, for the clarifications of one of his papers that inspired many derivations in this thesis.

S.G. de Boer
Delft, July 2020

Contents

1

Introduction

In this day and age, financial markets and the players operating on them are more advanced than ever. Every day, more progressions are made in every aspect coming in to play when trading on these markets. This can be progression in the accuracy of pricing of financial derivatives, the speed at which they can be traded, information infrastructure, information handling etc. Companies all over the world pride themselves on the speed at which they can process incoming and outgoing information, but also on the quality of the trades that they make. The quality of every trade is likely directly related to the pricing model that a trader has in place for pricing the traded product. Quality over quantity is not an unpopular saying, and it is a partial explanation for the desire that financial institutions have for accurate, fast and tractable pricing models.

Pricing models exist in excess, though many possess inherent flaws. These flaws are often the result of assumptions, which in turn tend to lack financial or mathematical soundness. Unlike the pricing functions, the models from which they are derived are rather well established and often remain the same. In other words, researchers taking a stab at deriving a new set of state-of-the-art pricing functions often start from the same point, which is a basic model. Examples of such models are the SABR model [31], Heston model [21], CEV model [33] etc. The idea behind most new approaches is to end up with fast, closed-form and computationally tractable pricing functions that can either price options as fast as is required in today's market circumstances, or possess other desirable qualities such as e.g. fast and easy calibration procedures. Typical problems that the pricing functions resulting from these approaches tend to face are over-parameterization, too little room for adjustments to propagate a trader's opinion on market circumstances, slowness in evaluation etc. It is therefore no surprise that new papers discussing how to deal with these issues are consistently published, see e.g. Cui *et al.* [21] on how to speed up the calibration of the Heston model or Gatheral *et al.* [28] on novel methods for modeling processes with financial applications.

This thesis showcases a rather contemporary method of solving a generalized system of stochastic differential equations (SDE's) comparable to the SABR model. The solution is derived from a stochastic-local volatility (SLV) model in which the local volatility (LV) component is kept general. This generality is maintained throughout all derivations, eventually yielding a model containing an undefined LV function. This function can then be specified however the user of the model deems suitable, as long as minor constraints are satisfied. Obviously, this is a very valuable quality of the model as it is highly customizable. The solution is inspired by Bang [9, 10], and its performance and the way of obtaining it will be discussed in great detail in this thesis. The solution consists of a set of pricing functions that seemingly possess all the aforementioned desirable properties, i.e. fast in evaluation, computational tractability, flexibility etc., with little disadvantages. The generalized SLV model that is used is typically denoted in the form of two SDE's, though in the majority of this thesis an atypical three SDE form is used. This extended system is used to isolate the LV component, in turn enabling for appropriate application of an SDE transformation called the Lamperti transform, which will provide the key to solving the entire system. The Lamperti transform is a highly versatile method for transforming SDE's into new equations typically more suitable for simulation and parameter estimation procedures, and its inner-workings and various applications will be the main focus of this thesis.

1.1. The Lamperti transform

While SDE's need not be used in financial settings only, financial modeling is typically done using a system of SDE's that describes the dynamics of some underlying on which financial products are defined. These SDE's are supposed to capture behaviors of processes through a random noise component added to an ordinary differential equation (ODE). As one can imagine, these models are often subjected to some boundary conditions, and sometimes even correlation components in case of multi-dimensional systems. These constraints and additional components such as correlation are mainly captured by the so-called diffusion term in the equation, with little influence from its drift component. The diffusion term often depends on the state of the process, as it has to enforce the aforementioned boundary conditions. However, SDE's with state-dependent diffusion terms come with a multitude of complexities when trying to obtain solutions or even approximations to model parameters. For example, the Extended Kalman filter (EKF) [41], which is one of the standard methods for parameter approximation in the field of statistical inference, requires higher order terms for the approximation of variables present in a model with state-dependent diffusion (see Kristensen *et al.* [37]). As Baadsgaard *et al.* [6] discuss, these higher order terms cause accuracy and tractability to get lost, resulting in lower consistency of the results. Not only in statistics can these terms cause problems with simulations, though. Financial modeling is often done with SLV models or variations thereof. A simple example is the Black-Scholes (BS) model [13] for pricing options. Solutions to this model only lie on the positive axis, not including zero. Simulation of such a system can be complicated in a numerical setting as it might result in values outside of the relatively limited state-space of the underlying assumed by the model, e.g. simulation may result in zero, which is an output not permitted by the lognormal distribution assumed for the underlying process in the BS model. However, the state-space of the underlying could be increased by reducing or removing the impact of the boundary conditions, i.e. by removing state-dependency from the diffusion term, consequently increasing the numerical stability of estimation procedures performed on this model. Therefore, means of getting rid of these types of diffusion terms by shifting the state-dependency to the drift term, are required.

The transformation that can transform an SDE to unit diffusion, i.e. a diffusion coefficient equal to one, is called the Lamperti transform. The transform is not only useful in approximation of the solution of an SDE, but should also be applied before performing any type of computational modeling. This is because removal of state-dependency in the diffusion term also removes the necessity of dealing with boundary conditions, making simulation more stable [18, 33, 40, 43]. How exactly the Lamperti transform can be used to bypass the computational issues related to state-dependent diffusion terms, and how to extend these methods into actually solving a system of SDE's, is one of the main topics of this thesis.

1.2. Outline

Even though the structure of this thesis was hinted at in the introduction, the more detailed format is the following. In Chapter 2, the reader is introduced to the numerous applications and intricacies of the Lamperti transform. This chapter covers the main theorem statement and proof, establishing the validity of the transform. Some examples are given, showcasing the large range of cases to which the transform can be applied. Finally, a sneak peek is given into how it can be used in the setting of financial modeling.

The method briefly mentioned in Chapter 2, is made complete in Chapter 3. Three sets of pricing functions are derived from a generalized SLV model, and numerical tests are performed to check the validity of the assumptions made in the derivations. The chapter is rounded off with a summary.

Chapter 4 showcases the output of the pricing functions derived in the previous chapter, and how they compare to results realized from other, more established models. Various local volatility (LV) functions, both parametric and non-parametric, are considered, and it is verified that the output option prices do not admit arbitrage. Finally, the chapter discusses the calibration algorithm that is used for calibration to a generated set of market prices. Results are shown for different procedures, and intermediate conclusions are drawn regarding the optimal methods for selecting a custom LV function.

Chapter 5 explores the possibilities regarding application of the Lamperti transform generalized to multi-dimensional form. Examples will be given and a conclusion is drawn on the relevance of such a transformation in a financial setting. Finally, the thesis is concluded in Chapter 6.

2

The Lamperti transform

Many non-trivial SDE's contain a state-dependent diffusion term. Such terms enforce an interaction between the state of the process and the random noise component. The following generic SDE is a typical example of a process with a state-dependent diffusion term

$$\mathrm{d}X(t) = \mu\,\mathrm{d}t + \sigma(X(t), t)\,\mathrm{d}W(t),$$

where μ is a constant, $\sigma : A \to \mathbb{R}$ and $W(t)$ is a Brownian motion [46] under some arbitrary measure. The Lamperti transform provides a way of getting rid of state-dependent diffusion terms. Using the SDE above as an example, the Lamperti transform is of the following form

$$\psi(X(t), t) := \int \frac{1}{\sigma(x, t)}\,\mathrm{d}x\bigg|_{x=X(t)}.$$

If the transform above defines a one-to-one mapping from the state-space of $X(t)$ to $\mathbb{R}$ for all $t > 0$, then it can be used to transform the dynamics of $X(t)$ through an application of Itô's lemma [46] to $\psi(X(t), t)$ (examples shown in Section 2.2). However, it is possible that the above definition for the transform is not a one-to-one mapping, in which case some minor adjustments have to be made that are covered in the next section. As discussed by Møller-Madsen [43] and Iacus [33], this transformation can essentially expand the space in which the solutions to a problem lie. For example, solutions that live only on the positive axis due to boundary conditions enforced by the state-dependent diffusion term in the SDE, can live also on the negative (and zero) axis after application of the Lamperti transform to obtain unit diffusion. This means that the Lamperti transform is essentially a way of mapping one state-space to another. In fact, for appropriate application of the transform, this mapping has to be one-to-one.

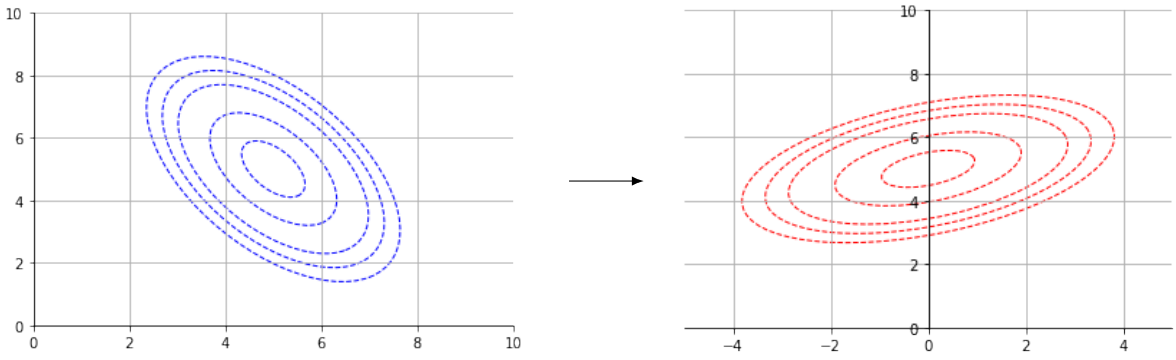

Figure 2.1: Exemplary figure to display the mapping of a process' state-space to a new space, that occurs when applying the Lamperti transform.

Applications to the Extended Kalman filter

The brief introduction gives rise to the idea that the Lamperti transform is a very useful tool for improving the stability of estimation procedures. One such procedure is the EKF, which is the non-linear extension of

the normal Kalman filter [35]. Estimation procedures using the EKF are generally considered the norm in the field of state-estimation. The EKF uses multiple measurements taken over a certain amount of time to produce estimates of unknown variables related to the distribution from which the samples are taken. This is done by constructing with a joint probability distribution for the unknown variables at each step in time. The algorithm consists of two steps, namely the prediction step and the update step. In the prediction step, an estimation of the unknown variables is made based on prior knowledge of the state of the process for which said variables are being estimated. This estimation typically follows from some model that is in accordance with the EKF framework. After the prediction follows the update step, in which the prediction is compared to actual measurements. Based on these measurements, and their accompanied likelihood, the prediction is updated. This process is recursive, and runs until some preset requirements are met.

Disadvantages of the EKF estimation procedure are the high importance of an accurate initial guess of the state of the process and that an incorrectly modeled process causes the algorithm to rapidly diverge. This second point is the most important, especially in cases of state-dependent diffusion. Typically, the process of which parameters are being estimated is an SDE which gets modelled by using a discretization scheme [47]. This discretization has the underlying assumption that there is a one-to-one mapping between the continuous time parameters and the discrete time parameters. However, as confirmed by numerical tests performed in Baadsgaard [5], the estimates produced under this assumption are highly biased. Therefore, the estimation procedure must be in line with the process that is being modeled, i.e. both must be living in a continuous time setting. However, for processes with state-dependent diffusion terms, this means higher order filter approximations are required (see Baadsgaard *et al.* [6]). Since the EKF algorithm consists of multiple estimations of the process parameters, these higher order terms pose a real problem, because the accuracy and consistency of the estimations becomes dubious. Baadsgaard *et al.* [6] discuss a consistent method for shifting the properties of the state-dependent diffusion term over to the drift term, by transforming the process to which the EKF will be applied. This method for removing problematic diffusion terms was later dubbed the Lamperti transform.

Applications in financial modeling

This thesis focuses mostly on financial applications of the Lamperti transform, hence highlighting a few exemplary applications from previous research is appropriate. As mentioned earlier, removal of state-dependent diffusion terms in a process can result in an increased state-space. Not only is this a desirable quality for estimation procedures, but it also presents computational benefits. The diffusion term typically enforces certain boundary conditions and correlation constructions, which can force the process to, for example, only take on positive values. If these types of restrictions are coupled with an observation equation that uses the same constraints, instability of the estimation ensues (see Møller-Madsen [43] and Baadsgaard *et al.* [6]). A simple example here is the BS model [13], which assumes that the underlying follows a Geometric Brownian motion (lognormal dynamics). Simulation of these dynamics can be problematic as one has to account for the boundary condition at the origin. Removal of this boundary condition would extend the state-space of the underlying, thereby simplifying and stabilizing simulation frameworks. As this idea is most easily supported through examples, Section 2.3 showcases a number of clarifying exemplary applications.

Another purely computational application is discussed by Chassagneux *et al.* [18] who continue on findings originally discussed by Neuenkirch-Szpruch [40] on application of the Lamperti transform to obtain strong approximations of SDE systems with non-globally Lipschitz [46] diffusion coefficients. Such coefficients often play a role in enforcing boundary conditions in financial models, e.g. the CIR process [33], 3/2 model [7], Heston model [21] etc. The general idea of applying the Lamperti transform to bypass simulation related issues stemming from state-dependent diffusion terms remains unchanged, though Chassagneux *et al.* [18] propose a modification of the standard Euler-Marayuma discretization scheme [47]. This scheme, applied to the transformed process containing constant diffusion coefficients, allows for improved strong convergence while remaining tractable in Monte Carlo frameworks.

State-space extensions aside, Bang [10] shows that in some situations the transform can simply be utilized to rewrite processes into something more suitable for the current setting. This can be particularly useful when attempting to price certain derivatives, as alternative expressions for underlying process of these derivatives can be defined. An elegant example of using the transform for this reason is a large focus of this thesis and a glimpse of how it works is revealed in Section 2.5. First, the upcoming sections focus on appropriately defining the framework in which the transform is applied and showcase numerous examples highlighting the benefits of its application.

2.1. Establishing the framework

As discussed in the introduction of this chapter, the Lamperti transform is a useful tool for improving the numerical stability of SDE systems in a simulation/estimation framework by providing a method for bypassing the issues typically connected to state-dependent diffusion coefficients. In this section, the setting in which the Lamperti transform can be applied is generalized. Starting off, some preliminary knowledge is discussed, as well as a proof that the Lamperti transform indeed does what it is made out to do. Second, application of the transform to one-dimensional diffusion processes is discussed, after which the section is finished off with an example and a brief discussion on the multi-dimensional version of the transform.

The applicability of the Lamperti transform is dependent on certain conditions that must be satisfied by the diffusion component of the SDE at hand. Properly defining these conditions requires some preliminary knowledge regarding the relation between the type of mapping a function facilitates between its domain and codomain, and its invertibility.

Definition 2.1.1. Let the sets A and B be the domain and codomain of a function f, respectively. The function $f : A \to B$ is injective[1] if distinct elements in the domain are mapped to distinct elements in the codomain, i.e.

$$\forall a_1, a_2 \in A, f(a_1) = f(a_2) \implies a_1 = a_2.$$

An injective function is typically called an injection.

Injectivity is a rather strong property that is not trivially possessed by most arbitrary functions unless their domain is restricted. For example, simple functions definitions such as $f(x) = x^2$, $f(x) = \sin(x)$ etc., are not injective on the domain $\mathbb{R}$. A definition in the same line as injectivity, is surjectivity.

Definition 2.1.2. Let the sets A and B be the domain and codomain of a function f, respectively. The function $f : A \to B$ is surjective[2] if every element in B is mapped to by at least one element in A, i.e.

$$\forall b \in B, \exists a \in A \quad \text{s.t.} \quad f(a) = b.$$

A surjective function is typically called a surjection.

As expected, functions that possess both injectivity and surjectivity are part of a particularly special class of functions.

Definition 2.1.3. Let the sets A and B be the domain and codomain of a function f, respectively. The function $f : A \to B$ is bijective if it is both injective and surjective. Bijective functions are typically called bijections.

The definitions above can be considered a build-up for the upcoming theorem, which yields a powerful result regarding the invertibility of bijections.

Corollary 2.1.3.1. *For two sets A and B, the function $f : A \to B$ is bijective if and only if it is invertible, i.e. $f^{-1} : B \to A$ exists.*

Proof. See Lay [39]. □

In light of the precise definition of the Lamperti transform later on in this section, the following corollary will prove to be very important. This is because it will provide insight in what constraints allow a function defined as an integral of another function to be a bijection.

Corollary 2.1.3.2. *Let $\{X(t), t \geq 0\}$ be an arbitrary random process with state-space A, and let the function $f : A \to B$ be defined by*

$$f(X(t)) := \int g(x)\,\mathrm{d}x \Big|_{x=X(t)},$$

where $g : A \to C$ is some continuously differentiable function on $\mathbb{R}$, i.e. $g(x) \in \mathscr{C}^1(\mathbb{R})$. The following holds

1. *If the function g can take on both positive and negative values, or has zero in its codomain, the function f is not invertible.*

[1] Injective is often used synonymously with one-to-one.
[2] Surjective is often used synonymously with onto.

2. If the function g is strictly positive or strictly negative, f^{-1} exists and is differentiable.

Proof. (*sketch*) The first point follows easily from the relation between bijective and invertible functions. Note that if g is continuous and capable of taking on both positive and negative values then f, while also continuous, is not strictly monotonic. In turn, this means that f is not an injection and hence not a bijection by Definition 2.1.3. From Corollary 2.1.3.1 it can then be concluded that f is not invertible. Analogously follows that f is not an injection if g can take on the value zero.

The second item is an immediate consequence of the inverse function theorem. See Munkres [44] for additional details. □

These preliminaries regarding basic function theory are obviously included here for a reason. Recall from the introduction of this chapter that a generic SDE with state-dependent diffusion could be of the form

$$\mathrm{d}X(t) = \mu\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t). \tag{2.1}$$

The diffusion term is the main, and essentially also the only, component of the Lamperti transform as it was showcased at the start of this chapter. However, in order for the Lamperti transform to be well-defined, this function σ must possess some simple properties, e.g. it must be nonzero and strictly negative or strictly positive on the domain of the underlying process. Moreover, in order for the transform to actually be applicable, σ must be defined in such a way that the inverse of the transform exists. This is precisely where Corollary 2.1.3.1 and 2.1.3.2 come in handy.

The final key piece of theory in the application of the Lamperti transform is Itô's lemma, of which a generalized version can be found in Lemma 2.1.5. Note that the processes discussed in the upcoming lemmas/theorems exist under an arbitrary measure unless otherwise mentioned. Moreover, the following definition will prove helpful with regards to the adopted terminology.

Definition 2.1.4. An Itô process is a random process $X(t)$ that forms a solution to an SDE of the form

$$\mathrm{d}X(t) = \mu(X(t),t)\,\mathrm{d}t + \sigma(X(t),t)\,\mathrm{d}W(t),$$

where $\mu, \sigma : A \to \mathbb{R}$ and $W(t)$ a Brownian motion.

Using this definition for Itô processes will prove helpful in stating the following lemmas/theorems.

Lemma 2.1.5 (Itô's lemma)**.** *Given an Itô process $X(t)$ with dynamics*

$$\mathrm{d}X(t) = \mu(X(t),t)\,\mathrm{d}t + \sigma(X(t),t)\,\mathrm{d}W(t),$$

and a function $\psi(X(t),t) \in \mathscr{C}^2(\mathbb{R}\times[0,+\infty))$. Then it holds that

$$Z(t) \coloneqq \psi(X(t),t),$$

is again an Itô process, with dynamics

$$\mathrm{d}Z(t) = \frac{\partial\psi(X(t),t)}{\partial t}\,\mathrm{d}t + \frac{\partial\psi(X(t),t)}{\partial X}\,\mathrm{d}X(t) + \frac{1}{2}\frac{\partial^2\psi(X(t),t)}{\partial X^2}(\mathrm{d}X(t))^2. \tag{2.2}$$

Where $(\mathrm{d}X(t))^2$ is computed by stating that

$$(\mathrm{d}t)^2 = \mathrm{d}W(t)\,\mathrm{d}t = 0, \quad (\mathrm{d}W(t))^2 = \mathrm{d}t.$$

Proof. See Øksendal [46]. □

It will be useful to rewrite Equation (2.2) so that it is properly set up for the upcoming theorem covering the Lamperti transform. Rewriting to appropriately distinguish between the drift and diffusion terms yields

$$\begin{aligned}
\mathrm{d}Z(t) &= \frac{\partial\psi(X(t),t)}{\partial t}\,\mathrm{d}t + \frac{\partial\psi(X(t),t)}{\partial X}\,\mathrm{d}X(t) + \frac{1}{2}\frac{\partial^2\psi(X(t),t)}{\partial X^2}(\mathrm{d}X(t))^2 \\
&= \frac{\partial\psi(X(t),t)}{\partial t}\,\mathrm{d}t + \frac{\partial\psi(X(t),t)}{\partial X}\big(\mu(X(t),t)\,\mathrm{d}t + \sigma(X(t),t)\,\mathrm{d}W(t)\big) + \frac{1}{2}\frac{\partial^2\psi(X(t),t)}{\partial X^2}\sigma^2(X(t),t)\,\mathrm{d}t \\
&= \left(\frac{\partial\psi(X(t),t)}{\partial t} + \frac{\partial\psi(X(t),t)}{\partial X}\mu(X(t),t) + \frac{1}{2}\frac{\partial^2\psi(X(t),t)}{\partial X^2}\sigma^2(X(t),t)\right)\mathrm{d}t + \frac{\partial\psi(X(t),t)}{\partial X}\sigma(X(t),t)\,\mathrm{d}W(t),
\end{aligned} \tag{2.3}$$

which will come in handy later. The following theorem discusses the main result of the Lamperti transform, and immediately showcases its practicality.

Theorem 2.1.6 (Lamperti transform)**.** *Let $X(t)$ be an Itô process with dynamics as in Lemma 2.1.5 containing a diffusion term that is continuously differentiable and strictly positive or strictly negative on the domain of $X(t)$, and define*

$$\psi(X(t),t) \coloneqq \int \frac{1}{\sigma(x,t)}\,\mathrm{d}x\bigg|_{x=X(t)}, \quad \psi(X(0),0)=0.$$

If ψ is a one-to-one mapping from the state-space of $X(t)$ to $\mathbb{R}$ for $t\in[0,+\infty)$, let $Z(t)\coloneqq\psi(X(t),t)$. If this is not the case, define

$$Z(t)\coloneqq\psi(X(t),t)\coloneqq\int_{\xi}^{x}\frac{1}{\sigma(u,t)}\,\mathrm{d}u\bigg|_{x=X(t)},$$

where ξ is an arbitrary point inside the state-space of $X(t)$ selected such that σ does not change sign in the resulting integration interval.

Then $Z(t)$ is a one-to-one mapping from the state-space of $X(t)$ to $\mathbb{R}$ for $t\in[0,+\infty)$, and has the following dynamics with unit diffusion

$$\mathrm{d}Z(t)=\left(\frac{\partial\psi(\psi^{-1}(Z(t),t),t)}{\partial t}+\frac{\mu(\psi^{-1}(Z(t),t),t)}{\sigma(\psi^{-1}(Z(t),t),t)}-\frac{1}{2}\frac{\partial\sigma(\psi^{-1}(Z(t),t),t)}{\partial X}\right)\mathrm{d}t+\mathrm{d}W(t).$$

Proof. Note that the proof is analogous for either definition of ψ mentioned in the theorem. From Equation (2.3) it can be easily seen that unit diffusion can be obtained by choosing

$$Z(t)\coloneqq\psi(X(t),t)\coloneqq\int_{\xi}^{x}\frac{1}{\sigma(u,t)}\,\mathrm{d}u\bigg|_{x=X(t)},$$

because, by using the Leibniz integration theorem [49], this selection yields

$$\frac{\partial\psi(X(t),t)}{\partial X}=\frac{1}{\sigma(X(t),t)}.$$

Furthermore,

$$\frac{\partial\psi(X(t),t)}{\partial t}=\frac{\partial}{\partial t}\int_{\xi}^{x}\frac{1}{\sigma(u,t)}\,\mathrm{d}u\bigg|_{x=X(t)}, \quad \frac{\partial^2\psi(X(t),t)}{\partial X^2}=-\frac{\frac{\partial\sigma(X(t),t)}{\partial X}}{\sigma^2(X(t),t)}.$$

Combining these results with the fact that $X(t)=\psi^{-1}(Z(t),t)$, provided this inverse exists, and filling everything in into Equation (2.3), gives

$$\begin{aligned}\mathrm{d}Z(t)&=\left(\frac{\partial}{\partial t}\int_{\xi}^{x}\frac{1}{\sigma(u,t)}\,\mathrm{d}u\bigg|_{x=X(t)}+\frac{\mu(X(t),t)}{\sigma(X(t),t)}-\frac{1}{2}\frac{\partial\sigma(X(t),t)}{\partial X}\right)\mathrm{d}t+\mathrm{d}W(t)\\&=\left(\frac{\partial\psi(\psi^{-1}(Z(t),t),t)}{\partial t}+\frac{\mu(\psi^{-1}(Z(t),t),t)}{\sigma(\psi^{-1}(Z(t),t),t)}-\frac{1}{2}\frac{\partial\sigma(\psi^{-1}(Z(t),t),t)}{\partial X}\right)\mathrm{d}t+\mathrm{d}W(t),\end{aligned}$$

which finalizes the proof. □

Theorem 2.1.6 shows that for a general Itô process with state-dependent diffusion, the Lamperti transform can be used to obtain a new process with unit diffusion. A useful property of this theorem is that it works off a non-parametric definition for a process with only minor constraints on its diffusion term. This fact allows for major flexibility in application of the transform, which is more explicitly showcased later in this chapter.

Important to note is that the function ψ always has to contain the entire diffusion term, else it might not properly transform to unit diffusion. For example, given an Itô process $X(t)$ with time-independent diffusion and dynamics defined by

$$\mathrm{d}X(t)=\mu X(t)\,\mathrm{d}t+\sigma X(t)\,\mathrm{d}W(t), \quad X(0)=x_0$$

with μ and σ real constants, then the appropriate Lamperti transform would be

$$\psi(X(t))\coloneqq\int\frac{1}{\sigma x}\,\mathrm{d}x\bigg|_{x=X(t)}, \quad \psi(X(0))=0,$$

which can be evaluated to become

$$\psi(X(t))=\frac{\log(X(t))-\log(x_0)}{\sigma}.$$

Analogously, the same approach can be followed for time-dependent diffusion. If the Lamperti transform in this example would have been defined in a way such that it did not contain either σ or $X(t)$, full transformation to unit diffusion would not have been possible. It could be stated that constants need not be included in the transform as they generally do not lead to difficulties in either simulation or estimation frameworks, but without any additional difficulties a full transformation to unit diffusion is possible and one is ensured of having bypassed any boundary conditions present on the original system. Some more examples of these types of applications are shown in the next section, including additional details.

The Lamperti transform is far from restricted to one-dimensional definitions, as it can also be applied in vector form to certain types of multi-variate Itô processes. The majority of the upcoming chapters focus on its one-dimensional definitions, while the reader is referred to Chapter 5 for multi-dimensional applications. This is because examples covering multi-dimensional transformations can quickly lead to complicated derivations that require additional preliminary knowledge.

2.2. Examples

It will be useful to discuss some concrete exemplary applications of the Lamperti transform to popular types of stochastic processes. The Geometric Brownian motion was briefly mentioned in the previous section, but the example was kept slightly incomplete. Typically, one would finish by applying Itô's lemma to find the dynamics of $\psi(X(t))$. These dynamics would then contain unit diffusion, indicating that the transformation was successful. This section will cover examples of some other processes and their transformation to unit diffusion. Note that simulation of the dynamics of $\psi(X(t))$ will result in information about its distribution. Through the inverse of the function ψ, this information can be converted into information about the distribution of the original process $X(t)$.

It becomes clear that the availability of the Lamperti transform depends on whether or not an explicit expression for its inverse exists. If this is not the case, one is forced to approximate it numerically, resulting in complicated dynamics and potential inaccuracies. The final example in this section will cover a process for which an explicit inverse does not exist, resulting in improperly written dynamics of the transformed process. Generally, the processes for which the transform cannot be inverted explicitly are customized variations of existing processes that, as of now, see little real-life application. This is in line with the final example shown in this section, which covers a modified version of a process popular in financial modeling.

Driftless Geometric Brownian motion

One of the most basic examples of a process with state-dependent diffusion would be a driftless Geometric Brownian motion described by the dynamics

$$\mathrm{d}\alpha(t) = \nu\alpha(t)\,\mathrm{d}W(t), \quad \alpha(0) = \alpha_0, \tag{2.4}$$

where ν is a positive constant and $W(t)$ is a Brownian motion under an arbitrary measure. More details on what this process is typically used for in real-life applications are omitted here, as the objective of this example is simply to showcase an application of the Lamperti transform. Similar to the short example shown earlier, the transform is defined as

$$Z(t) \coloneqq \psi(\alpha(t)) \coloneqq \int \frac{1}{\nu\alpha}\,\mathrm{d}\alpha\bigg|_{\alpha=\alpha(t)}, \quad Z(0) = 0. \tag{2.5}$$

In this case the transform has a well-defined inverse, which means that Theorem 2.1.6 can be applied. This yields

$$\mathrm{d}Z(t) = -\frac{1}{2}\nu\,\mathrm{d}t + \mathrm{d}W(t). \tag{2.6}$$

This process can now be simulated and, since the inverse of the transform exists, the mapping back and forth between $\alpha(t)$ and $Z(t)$ can be done by using the fact that $\alpha(t) = \psi^{-1}(Z(t))$. This mapping makes it possible to obtain information about the distribution of $\alpha(t)$ through only simulating the transformed process. More information on distribution recovery in a computational setting is shown in the next section.

Interesting to note is that the condition for the boundary at the origin enforced by the state-dependent diffusion in Equation (2.4), is no longer present in the transformed process shown in Equation (2.6). In fact, the process resulting from application of the Lamperti transform is free from any boundary conditions. This application is therefore a nice example of the earlier discussed state-space extension that can be achieved by using the Lamperti transform.

3/2 model volatility

The standard Heston model [21] can be modified to become the 3/2 model [7]. This modification consists of altering the stochastic volatility process of the Heston model to instead be an inverse of a square-root process, rather than the typical square root process. The new volatility process is then described by the dynamics

$$\mathrm{d}v(t) = \kappa v(t)(\bar{v} - v(t))\,\mathrm{d}t + \gamma v^{3/2}(t)\,\mathrm{d}W^{\mathbb{Q}}(t), \quad v(0) = v_0 \tag{2.7}$$

where the speed of mean-reversion is now governed by $\kappa v(t)$, $\bar{v}$ denotes the long-term mean of the process, γ is the positive volatility-of-volatility parameter and $W^{\mathbb{Q}}(t)$ is a standard Brownian motion under the risk-neutral measure $\mathbb{Q}$. As Hambly-Vaicenavicius [32] discuss, modeling the volatility in this way allows for a more natural method for pricing certain financial products. Clearly, the diffusion term in Equation (2.7) enforces an unattainable boundary condition at the origin somewhat similar to that discussed in the previous example. Application of Theorem 2.1.6 thus provides the Lamperti transform

$$Z(t) \coloneqq \psi(v(t)) \coloneqq \int \frac{1}{\gamma v^{3/2}}\,\mathrm{d}v\bigg|_{v=v(t)}, \quad Z(0) = 0, \tag{2.8}$$

which has a well-defined inverse. Following along with the steps shown in Theorem 2.1.6 shows that, after application of Itô's lemma to Equation (2.8), the process $Z(t)$ is described by the dynamics

$$\mathrm{d}Z(t) = \left(\kappa \frac{\psi^{-1}(Z(t))(\bar{v} - \psi^{-1}(Z(t)))}{\gamma(\psi^{-1}(Z(t)))^{3/2}} - \frac{1}{2}\left(\frac{3}{2}\gamma\sqrt{\psi^{-1}(Z(t))}\right)\right)\mathrm{d}t + \mathrm{d}W^{\mathbb{Q}}(t). \tag{2.9}$$

Note that Equation (2.9) makes use of the existence of an explicit expression for the inverse of the Lamperti transform with functional form defined in Equation (2.8). As expected, the dynamics for $Z(t)$ show successful removal of the state-dependent diffusion term and therewith the boundary condition at the origin present in the original process. Once more, the process $Z(t)$ can now be simulated and mapped back to the state-space of the original process $v(t)$ using the inverse of the transform. A showcase of this procedure in a computational setting is done in the next section.

Pearson diffusion

The Pearson diffusions as defined in Forman-Sørensen [25] form a set of stationary solutions to the SDE

$$\mathrm{d}X(t) = \theta(X(t) - \mu)\,\mathrm{d}t + \sqrt{2\theta(aX^2(t) + bX(t) + c)}\,\mathrm{d}W(t), \quad X(0) = x_0, \tag{2.10}$$

where θ is a positive value that controls the speed of the diffusion and a, b, c and μ determine the state-space of $X(t)$, making sure that the square root is well-defined inside of it. Finally, $W(t)$ is a Brownian motion under an arbitrary measure. This class of diffusions allows for parameter estimation through statistical inference, as discussed by Forman-Sørensen [25]. In light of this, it might be useful to improve the tractability of these estimation procedures by removing the state-dependent diffusion term. The Lamperti transform is defined as

$$Z(t) \coloneqq \psi(X(t)) \coloneqq \int \frac{1}{\sqrt{2\theta(ax^2 + bx + c)}}\,\mathrm{d}x\bigg|_{x=X(t)}, \quad Z(0) = 0, \tag{2.11}$$

which has a well-defined inverse, provided that a, b, c and μ are defined in the earlier discussed appropriate way. Since the inverse exists, a simple application of Itô's lemma results in

$$\mathrm{d}Z(t) = \left(\frac{\theta(\psi^{-1}(Z(t)) - \mu)}{\sqrt{2\theta(a(\psi^{-1}(Z(t)))^2 + b\psi^{-1}(Z(t)) + c)}} - \frac{1}{2}\frac{\theta(2a\psi^{-1}(Z(t)) + b)}{\sqrt{2\theta(a(\psi^{-1}(Z(t)))^2 + b\psi^{-1}(Z(t)) + c)}}\right)\mathrm{d}t + \mathrm{d}W(t). \tag{2.12}$$

These rewritten dynamics are quite cumbersome compared to the original dynamics for $X(t)$, but they do allow for extra possibilities in hypothesis testing, partially because the attainable reflecting boundary enforced by the diffusion term in the original process is no longer present in Equation (2.12). The exact benefits of this transformed process in an estimation setting go beyond the scope of this example, and the reader is referred to Møller-Madsen [43] or Forman-Sørensen [25] for more details.

Modified CIR process

As discussed in the introduction to this section, the applicability of the Lamperti transform somewhat relies on the existence of an explicit expression for its inverse, else the dynamics of the transformed process can in

general not be simulated without still having to account for the boundary conditions on the original process. The three previous examples showcased processes for which this inverse was well-defined, but it will be useful to showcase a rather unnatural process for which an explicit expression does not exist. The typical CIR process [33] can be modified to become

$$\mathrm{d}X(t) = \kappa(\bar{x} - X(t))\,\mathrm{d}t + \left(1 + \sigma\sqrt{X(t)}\right)\mathrm{d}W(t), \quad X(0) = x_0, \tag{2.13}$$

where κ is the speed at which the process reverts back to its mean $\bar{x}$, σ is a positive constant and $W(t)$ is a Brownian motion under an arbitrary measure. For this process, the transform would be defined as

$$Z(t) := \psi(X(t)) := \int \frac{1}{(1+\sigma\sqrt{x})}\,\mathrm{d}x\Big|_{x=X(t)}, \quad Z(0) = 0, \tag{2.14}$$

which evaluates to

$$Z(t) = \psi(X(t)) = \frac{2\sigma\sqrt{X(t)} - 2\log(\sigma\sqrt{X(t)}+1) - 2\sigma\sqrt{x_0} + 2\log(\sigma\sqrt{x_0}+1)}{\sigma^2}. \tag{2.15}$$

The function ψ as shown in Equation (2.15), does not admit an explicit expression for its inverse. If one insists on still using the transform, resorting to numerical methods for approximating the inverse will be necessary. As mentioned earlier, the dynamics for $Z(t)$ can now not be simulated without still having to account for the boundary conditions on the process $X(t)$. Therefore, these dynamics are not immediately convenient for numerical simulations. Again, while it is possible to resort to numerical techniques to bypass these issues, e.g. numerically approximating $\psi^{-1}(Z(t))$, this would defeat the simplicity of the Lamperti transform and is not the intention behind showing this example.

The examples shown in this section only considered time-independent diffusion types. Transforming processes with time-dependent diffusion terms works analogously, since Theorem 2.1.6 is defined for both cases. However, there are some intricacies one has to keep in mind, which also holds for multi-dimensional applications of the Lamperti transform. These topics will be discussed in more detail in Chapter 5.

2.3. Numerical testing

It was stated in the previous section that simulation of the dynamics of the transformed process $Z(t)$ can give information about the distribution of the original process $X(t)$ through the inverse of the Lamperti transform. This section aims to extend on two of the aforementioned examples by performing some numerical simulations.

Driftless Geometric Brownian motion simulation

It is now time to come back to the first example shown in the previous section and extend it using some numerical simulations. Something that was left out when initially covering this example, is the fact that $Z(t)$ can be found explicitly. Indeed, integrating the left and right hand side of Equation (2.6) gives

$$Z(t) = -\frac{1}{2}\nu t + W(t). \tag{2.16}$$

From the above expression it can be easily seen that $Z \sim N(-\frac{1}{2}\nu t, t)$. Moreover, from the fact that $Z(t) = \psi(\alpha(t))$, it follows that $\alpha(t) = \alpha_0\,\mathrm{e}^{\nu Z(t)}$. Therefore, it can be concluded that $\alpha \sim LN\left(\log(\alpha_0) - \frac{1}{2}\nu^2 t, \nu^2 t\right)$. It turns out that this example is a special case in which the Lamperti transform actually leads to a solution of the original process. Even though this makes numerical simulation relatively trivial, the upcoming figure following from such a simulation still nicely showcases the fact that the process $Z(t)$ is easier to simulate, as one does not have to account for the boundary at zero. The resulting distribution for $Z(t)$ can then be easily translated back to the distribution for $\alpha(t)$ by making use of the inverse of the transform.

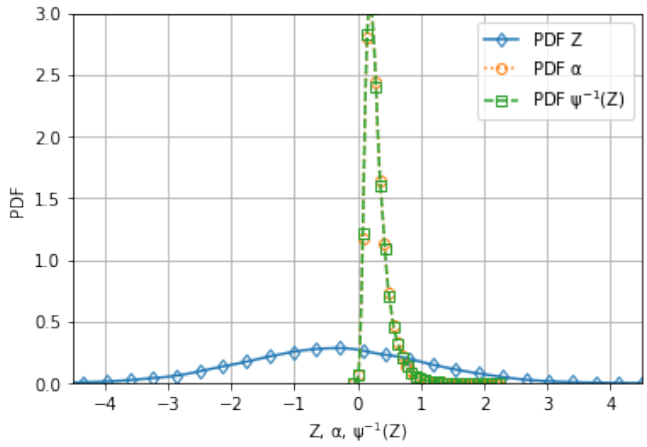

Figure 2.2: Probability density functions of a process before and after application of the Lamperti transform, and the resulting function after translating back to the original process by making use of the inverse of the transform.

The figure above confirms the earlier statement regarding recovery of the distribution of $\alpha(t)$ through the inverse of the Lamperti transform. Note that this figure was constructed by making use of a simple Monte Carlo (MC) simulation in which the Euler-Marayuma scheme was applied to discretize the dynamics of the processes (see Chapter 9 of Oosterlee-Grzelak [47] for more details). First, paths for $X(t)$ and $Z(t)$ are simulated. After that, the inverse transform can be applied to the simulated paths of $Z(t)$. This process results in three different sets of paths, each for which the density can be computed and plotted. More specifically, for the construction of Figure 2.2, 10000 paths with 1000 steps in time were simulated with the parameter set

$$\alpha_0 = 0.3, \quad \nu = 0.4, \quad T = 2.$$

Figure 2.3 shows 100 of these paths and makes it clear that the boundary at the origin was indeed not taken into account when simulating the transformed process $Z(t)$.

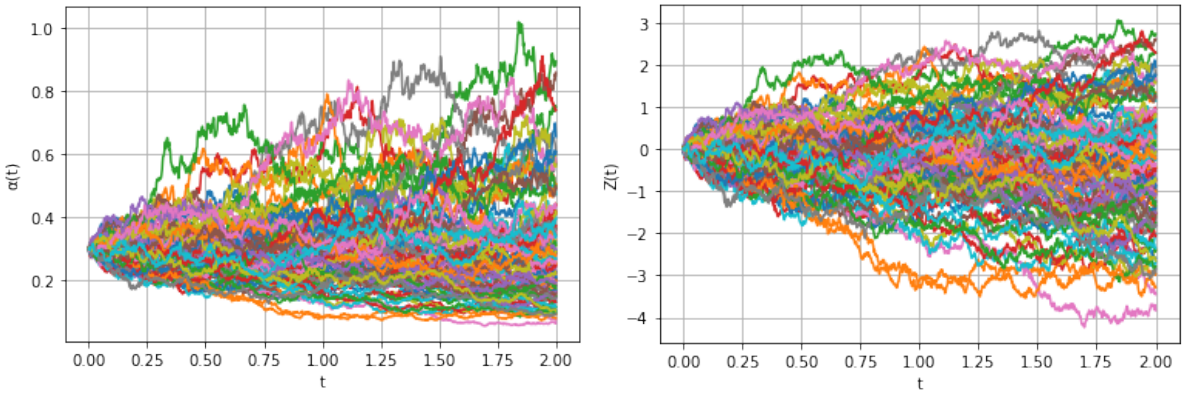

Figure 2.3: Paths obtained with an MC simulation of the dynamics shown in Equation (2.4), and the dynamics resulting from application of the Lamperti transform to this equation.

3/2 model volatility simulation

Application of the Lamperti transform to the dynamics of the volatility process in the 3/2 model yielded Equation (2.8). This equation can be evaluated, yielding

$$Z(t) = \psi(v(t)) = -\frac{2}{\gamma\sqrt{v(t)}} + \frac{2}{\gamma\sqrt{v_0}}, \tag{2.17}$$

where the initial condition $Z(0) = 0$ was used. Equation (2.17) can now be inverted, allowing for simulation of the dynamics of $Z(t)$ later on. This operation results in

$$v(t) = \psi^{-1}(Z(t)) = \frac{4v_0}{(\gamma v_0 Z(t) - 2)^2}. \tag{2.18}$$

Provided with an arbitrary set of parameters, i.e.

$$v_0 = 0.2, \quad \kappa = 0.1, \quad \bar{v} = 0.02, \quad \gamma = 1.0, \quad T = 2,$$

a figure similar to Figure 2.2 can be obtained. The dynamics for $v(t)$ and $Z(t)$ are simulated using the method described earlier, after which the inverse transformation is applied to the resulting paths of $Z(t)$. This procedure will once again showcase the fact that the distribution of the original process can be recovered by applying the inverse of the transform to the paths of the transformed process. Doing exactly this, yields Figure 2.4.

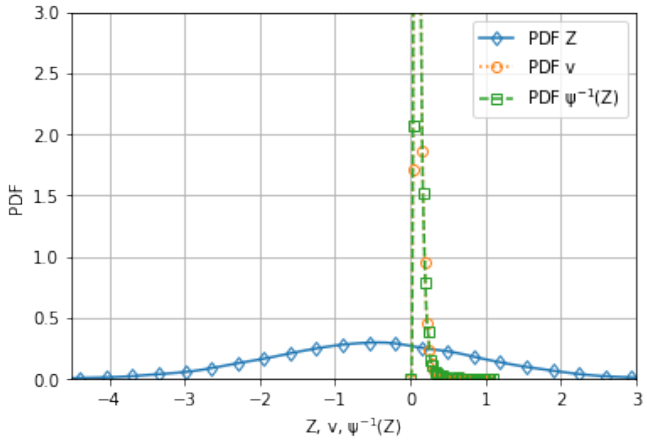

Figure 2.4: Probability density functions of the 3/2 model volatility process before and after application of the Lamperti transform, and the resulting function after translating back to the original process by making use of the inverse of the transform.

Indeed, it can be seen from the figure above that application of the Lamperti transform once again results in a process with a state-space that is not limited by the boundary conditions enforced on $v(t)$. In order to highlight this fact even better, Figure 2.5 shows 100 of the paths simulated for both $v(t)$ and the transformed process $Z(t)$.

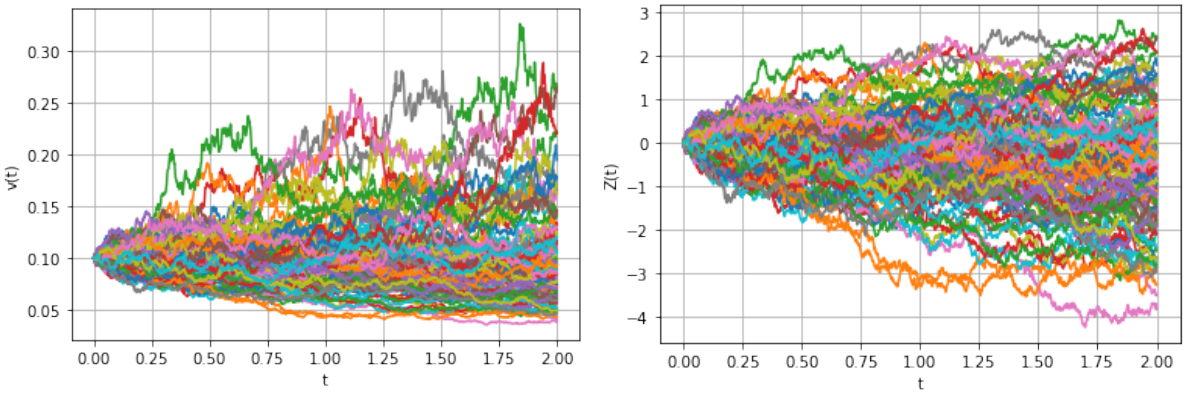

Figure 2.5: Paths obtained by with an MC simulation of the dynamics shown in Equation (2.7), and the dynamics resulting from application of the Lamperti transform to this equation.

Reasons for applying the Lamperti transform to the 3/2 model volatility process are that simulation of the dynamics of $v(t)$ requires attention to the boundary at the origin, especially when simulating with higher

values for T. This is because discretization of the dynamics can lead to numerical complexities (in this case values of zero) breaking the simulation. The transformed dynamics are not subjected to these problems and can be simulated freely. There is no computational overhead when using the transform as the expression for the inverse transform is explicit, hence the paths for $Z(t)$ can be mapped back to the state-space of $v(t)$ virtually instantaneously.

2.4. Rough volatility and the Lamperti transform

Recall from the introduction to this chapter that the Lamperti transform regularly sees applications in the field of financial mathematics. This section describes some recent developments in this field and where the transform comes into play. Note that the contents of this section are kept relatively general, as the advanced parts of this topic are beyond the scope of this thesis.

The increased availability of high frequency trading data has led to the rise of a new method for modeling volatility. These so called Rough volatility (RV) models are based on the widely accepted idea that volatility has a long memory. Comte-Renault [20] published a paper pioneering this concept, in which they proposed to simulate log volatility with a model containing a fractional Brownian motion (fBM) component. This fBM component could ensure that the long memory of the volatility process was appropriately captured.

Definition 2.4.1. Let $W_H(t)$ be a fractional Brownian motion. Then $W_H(t)$ is a continuous-time Gaussian process for $t \in [0, T]$ that starts at zero, and has expectation zero and covariance function

$$\mathbb{E}[W_H(t)W_H(s)] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right),$$

where $s, t \in [0, T]$ and $H \in (0, 1)$ being the Hurst parameter.

The fBM is a generalized version of the regular Brownian motion (BM), as its increments are not necessarily uncorrelated. This correlation is controlled by the Hurst parameter, H. If $H = \frac{1}{2}$, the original BM is recovered, while $H > \frac{1}{2}$ and $H < \frac{1}{2}$ yield positive and negative correlation between the increments, respectively. Consequently, the Hurst parameter is responsible for the smoothness of the resulting process, where a higher value results in a smoother motion.

The results of Comte-Renault [20] were further generalized by Gatheral *et al.* [28], who chose to adopt a rough fractional stochastic volatility model based on the fractional stochastic volatility model proposed by Comte-Renault, after results showed that log volatility behaves very similar to a fBM with a Hurst parameter of 0.1. This model is defined as

$$\sigma(t) := \mathrm{e}^{X(t)}, \quad t \in [0, T], \tag{2.19}$$

where $X(t)$ satisfies

$$X(t) = \nu \int_{-\infty}^{t} \mathrm{e}^{-\kappa(t-s)}\,\mathrm{d}W_H(s) + \mu, \tag{2.20}$$

which is an explicit solution to the fractional Ornstein-Uhlenbeck process (see Cheridito *et al.* [19])

$$\mathrm{d}X(t) = \kappa(\mu - X(t))\,\mathrm{d}t + \nu\,\mathrm{d}W_H(t), \tag{2.21}$$

where κ and ν are positive constants, $\mu \in \mathbb{R}$ and $H \in (0, \frac{1}{2})$. This model is stationary, but behaves locally as an fBM for very small values for κ. For a much more detailed description of the model the reader is referred to Gatheral *et al.* [28].

These RV models are interesting within the scope of this thesis mainly because the Lamperti transform can be used to derive prediction formulae to linearly estimate the fBM that forms the main component in modeling the log volatility. Recall from the introduction to this chapter that it is not new that the Lamperti transform can play a vital role in estimation procedures. In order to appropriately apply the transform in this setting, Nuzman-Poor [45] propose an alternative definition for it. This new definition is more appropriate for the aim of showcasing a number of prediction formulae for fBM related variables. Definition 2.1 in Nuzman-Poor [45] reads

Definition 2.4.2. Let J be a set of positive real numbers that allows multiplication, and let I be a positive real index set. Then, for each $H \in \mathbb{R}$, the Lamperti transform denoted by L_H is an invertible map between J-valued

functions on I, and J-valued functions on $\ln(I)$. For each $y : I \to J$, the transformed function $L_H y : \ln(I) \to J$ is defined as

$$(L_H y)(\tau) := \mathrm{e}^{-H\tau} y(\tau), \quad \tau \in \ln(I),$$

and for each $j : \ln(I) \to J$, the inverse transformation is defined as

$$(L_H^{-1} j)(t) := t^H j(\ln(t)), \quad t \in I.$$

Defining the transform through Definition 2.4.2 is useful because it allows the main result of Lamperti [38] to be more straightforwardly interpreted. This result yields the property that for an H-self-similar process $Y(t)$, the transformed process $L_H Y(\tau)$ is stationary. Consequently, this means that the transformed process is a generating process for $Y(t)$. Before the significance of this fact is described, it will be necessary to repeat Definition 2.2 stated in Nuzman-Poor [45].

Definition 2.4.3. For some $H > 0$ and all $a > 0$, the random process $\{Y(t), t \geq 0\}$ is called wide-sense H-self-similar if

1. $\mathbb{E}[Y^2(t)] < \infty$;
2. $\mathbb{E}[Y(at)] = a^H \mathbb{E}[Y(t)]$;
3. $\mathbb{E}[Y(at_1)Y(at_2)] = a^{2H} \mathbb{E}[Y(t_1)Y(t_2)]$.

The most important point now is that the fBM process is an H-self-similar process. Fractional Brownian motion has many other characteristics, but this particular property leads to the possibility of constructing a generator through application of the Lamperti transform. Availability of this generator leads to easier derivations of certain predictor functions that can be used for estimation of variables related to the fBM. The exact definitions and methods for using such predictor functions are beyond the scope of this thesis, but the reader can turn to Section 4.3 of Nuzman-Poor [45] for detailed descriptions. Finally, the availability of the aforementioned functions clearly helps in modeling rough volatility, let it be through the earlier discussed RSFV model, or any other model using fBM components.

2.5. Transforming a generalized SLV model

In this thesis, solving a general SLV model with an application of the Lamperti transform plays a major role. The solution will show the versatility of the transform and its power in financial modeling, as the SLV model that will be solved allows for great customization. The transform is used in precisely the same fashion as in the examples shown in Section 2.2, though the motivation behind its application in this setting is different from that in the typical simulation and estimation procedures discussed earlier. This section covers an important result obtained from application of the transformation, while Chapter 3 will go more into the details of the exact derivation of the solution to the upcoming SLV model. In order to appropriately apply the Lamperti transform, a general SLV model of the following form is defined

$$\mathrm{d}F(t) = \sigma(F(t))\,\mathrm{d}\eta(t), \tag{2.22}$$

$$\mathrm{d}\eta(t) = \alpha(t)\,\mathrm{d}W_\eta^T(t), \tag{2.23}$$

$$\mathrm{d}\alpha(t) = \nu\alpha(t)\,\mathrm{d}W_\alpha^T(t), \tag{2.24}$$

where the LV function σ controls the backbone of the distribution of $F(t)$. Moreover, $W_\eta^T(t)$ and $W_\alpha^T(t)$ are Brownian motions under the T forward measure, ν is a positive volatility-of-volatility parameter, and $\mathrm{d}W_\eta^T(t)\,\mathrm{d}W_\alpha^T(t) = \rho_{\eta,\alpha}\,\mathrm{d}t$ where $\rho \equiv \rho_{\eta,\alpha} \in [-1,1]$. Defining this system under the T forward measure will allow natural comparisons to other more established models in Chapter 4. The integral concept in solving this system is that the Lamperti transform can be used to construct a connection between the state-space of $F(T)$ and that of $\eta(T)$, i.e. to construct an expression for $F(T)$ in terms of $\eta(T)$. This will prove to be the crux to solving the entire model, and will be extensively discussed in Chapter 3.

In the setting of this general SLV model, it is clear that the system contains a product of the LV component $\sigma(F(t))$, and the stochastic volatility (SV) component $\eta(t)$. This means that application of the Lamperti transform to obtain unit diffusion can alternatively be seen as a means of splitting the LV and SV components. According to Bang [9], it is "a tool of choice to disentangle local from pure stochastic volatility". These points are summarized by the following theorem.

Lemma 2.5.1. *For a local volatility function σ satisfying the conditions necessary for application of Theorem 2.1.6, i.e. σ must be continuous and strictly positive or strictly negative on the domain of $F(t)$, consider the Lamperti transform:*

$$G(F(t)) \coloneqq \int_{F(0)}^{a} \frac{1}{\sigma(u)}\,\mathrm{d}u\bigg|_{a=F(t)},$$

with the associated process $g(t) \coloneqq G(F(t))$. Applying Itô's lemma as shown in Lemma 2.1.5, results in

$$\begin{aligned} g(T) &= \eta(T) - \mu(T), \\ \mu(T) &\coloneqq \eta(0) + \frac{1}{2}\int_0^T \alpha^2(t)\frac{\partial\sigma(G^{-1}(g(t)))}{\partial F}\,\mathrm{d}t. \end{aligned}$$

Proof. It holds that $g \coloneqq G \in \mathscr{C}^2([0,+\infty))$. Therefore, Itô's lemma can be applied, yielding

$$\begin{aligned} \mathrm{d}g(t) &= \frac{\partial G(F(t))}{\partial t}\,\mathrm{d}t + \frac{\partial G(F(t))}{\partial F}\,\mathrm{d}F(t) + \frac{1}{2}\frac{\partial^2 G(F(t))}{\partial F^2}(\mathrm{d}F(t))^2 \\ &= \frac{\partial G(F(t))}{\partial t}\,\mathrm{d}t + \frac{\partial G(F(t))}{\partial F}\sigma(F(t))\alpha(t)\,\mathrm{d}W_F^T(t) + \frac{1}{2}\frac{\partial^2 G(F(t))}{\partial F^2}\sigma^2(F(t))\alpha^2(t)\,\mathrm{d}t. \end{aligned} \tag{2.25}$$

In order to get rid of state-dependent diffusion in these dynamics, Lamperti's transform can be applied. Setting the following

$$G(F(t)) \coloneqq \int_{F(0)}^{a} \frac{1}{\sigma(u)}\,\mathrm{d}u\bigg|_{a=F(t)},$$

gives rise to

$$\frac{\partial G(F(t))}{\partial F} = \frac{1}{\sigma(F(t))}, \quad \frac{\partial G(F(t))}{\partial t} = \frac{\partial}{\partial t}\int_{F(0)}^{F(t)} \frac{1}{\sigma(u)}\,\mathrm{d}u = 0, \quad \frac{\partial^2 G(F(t))}{\partial F^2} = -\frac{\frac{\partial\sigma(F(t))}{\partial F}}{\sigma^2(F(t))}.$$

Filling these expressions in into Equation (2.25), gives

$$\mathrm{d}g(t) = \alpha(t)\,\mathrm{d}W_\eta^T(t) - \frac{1}{2}\frac{\partial\sigma(F(t))}{\partial F}\alpha^2(t)\,\mathrm{d}t,$$

which, after integrating, gives the desired result. □

Lemma 2.5.1 has proven an important fact, namely that it is possible to write

$$F(T) = G^{-1}(\eta(T) - \mu(T)), \tag{2.26}$$

which is a well-defined expression for $F(T)$ in terms of $\eta(T)$ and the LV component. Recall from the previous section that the Lamperti transform could be applied to diffusion terms satisfying minor conditions. In this setting, this allowed for generality of the LV function σ to be preserved. In the next chapter, the LV function will remain general, and it is shown that a new expression for $F(T)$ indeed provides the key to approximation a solution to the earlier defined general SLV model.

Finally, note that the focus of the application of the Lamperti transform in this setting was not at all on increasing the state-space of the process or estimating any parameters. As briefly mentioned at the start of this chapter, in some situations one only needs to find a clever expression for a certain variable, and the Lamperti transform offers a method for doing so.

Practical assumption and justification

The general SLV model defined at the start of this chapter will not be used in fully dynamic fashion. In other words, any pricing functions derived using Lemma 2.5.1 are not meant for pricing path-dependent options, and the model will need the necessary extensions if one intends to price those types of derivatives. This means that the expression defined in Equation (2.26) can be further simplified by setting $\mu(T)$ to be a constant, i.e. $\mu(T) = \mu$. Recall from Lemma 2.5.1 that

$$\mu(T) \coloneqq \eta(0) + \frac{1}{2}\int_0^T \alpha^2(t)\frac{\partial\sigma(G^{-1}(g(t)))}{\partial F}\,\mathrm{d}t. \tag{2.27}$$

Since the focus will solely be on pricing vanilla European options, the fact that assuming $\mu(T) = \mu$ alters the dynamics and therewith the paths of $F(t)$, is not relevant in this setting. Moreover, while changing the distribution slightly, the resulting density for $F(T)$ will still be valid, i.e. non-negative and integrating to 1 on $\mathbb{R}$. This is proven by Corollary 2.5.1.1.

Corollary 2.5.1.1. *By construction, $G^{-1}(M(T)-\mu)$ is a random variable associated with a valid density, i.e. a density that is non-negative and integrates to 1 on $\mathbb{R}$.*

Proof. (*sketch*) This follows directly from the fact that G^{-1} is a measurable function that is applied to the random variable $M(T)-\mu$, thus yielding a new random variable in turn associated with a well-defined density, i.e. one that is non-negative and integrates to 1 on $\mathbb{R}$. □

Continuing, by the martingale representation theorem (see Øksendal [46]) it holds, due to the driftless dynamics of $F(t)$, that $F(T)$ is a martingale under the T forward measure. This means that

$$\mathbb{E}^T[F(T)] = \mathbb{E}^T[G^{-1}(\eta(T)-\mu(T))] = F(0). \tag{2.28}$$

Based on this fact, it is natural to calibrate μ such that this first moment is preserved, i.e. such that $\mathbb{E}^T[G^{-1}(\eta(T)-\mu)] = F(0)$. Clearly, calibrating μ to this constraint does not mean that higher order moments of $F(T)$ are also preserved. In order to test the accuracy of assuming $\mu(T) = \mu$ and the impact this has on the underlying distribution of $F(T)$, a simple numerical experiment can be done.

First, a non-constant LV function is selected such that the partial derivative of this function present in the expression of $\mu(T)$ does not equal zero. Secondly, an MC framework is used to simulate 10000 paths with 1000 time-steps for both $\eta(t)$ and $\alpha(t)$ by discretizing Equation (2.23) and (2.24) according to the Euler-Marayuma scheme (see Oosterlee-Grzelak [47]). This allows for computation of $\mu(T)$ through Equation (2.27) and therewith evaluation of $G^{-1}(\eta(T)-\mu(T))$ for each simulated path. Thirdly, once more using the simulated paths for $\eta(T)$, the constant μ is calibrated such that $\mathbb{E}^T[G^{-1}(\eta(T)-\mu)] = F(0)$. Finally, the values $\mathbb{E}^T[(G^{-1}(\eta(T)-\mu))^m]$ and $\mathbb{E}^T[(G^{-1}(\eta(T)-\mu(T)))^m]$ are compared for different values of m. This comparison will showcase how accurately the calibrated μ allows $(G^{-1}(\eta(T)-\mu))^m$ to reproduce the higher order moments of $F(T)$. Using the parameter set

$$F(0) = 0.11, \quad \eta(0) = 0.11, \quad \alpha(0) = 0.02, \quad \sigma(F(t)) = F(t), \quad \nu = 0.3, \quad \rho = -0.8,$$

the results of this numerical experiment are showcased in Table 2.1, where

$$\varepsilon_{\text{absolute}} := \left|\mathbb{E}^T\left[\left(G^{-1}(M(T)-\mu(T))\right)^m\right] - \mathbb{E}^T\left[\left(G^{-1}(M(T)-\mu)\right)^m\right]\right|, \tag{2.29}$$

and

$$\varepsilon_{\text{relative}} := \left|\frac{\mathbb{E}^T\left[\left(G^{-1}(M(T)-\mu(T))\right)^m\right]}{\mathbb{E}^T\left[\left(G^{-1}(M(T)-\mu)\right)^m\right]} - 1\right|. \tag{2.30}$$

T	m	$\mathbb{E}^T\left[\left(G^{-1}(M(T)-\mu(T))\right)^m\right]$	$\mathbb{E}^T\left[\left(G^{-1}(M(T)-\mu)\right)^m\right]$	$\varepsilon_{\text{absolute}}$	$\varepsilon_{\text{relative}}$
5	1	0.11	0.11	7.11e-6	6.47e-5
	2	0.012	0.012	4.38e-7	3.61e-5
	3	0.0013	0.0013	3.92e-7	2.92e-4
	4	0.00015	0.00015	1.04e-7	6.99e-4
30	1	0.11	0.11	9.06e-5	8.24e-4
	2	0.012	0.012	5.67e-5	4.62e-3
	3	0.0014	0.0014	1.78e-5	1.29e-2
	4	0.00016	0.00016	3.74e-6	2.39e-2

Table 2.1: Results of a numerical experiment set up to analyze the assumption that $\mu(T) = \mu$, in which various moments of $F(T)$ are computed using both $\mu(T)$ and μ.

Note that the differences between the first moments are caused purely by computational errors stemming from the MC simulation and numerical integration procedures. For low T values, Berestycki *et al.* [12] show that $F(T)$ will converge to the exact solution of the SLV model defined earlier, indicating that taking μ as a constant will not lead to issues. This result is supported by Table 2.1. However, this leaves the problem of maintaining sufficient accuracy when T rises, for which Table 2.1 shows that the validity of the assumption starts to decay slightly as T increases substantially. More specifically, the assumption that $\mu(T) = \mu$ seems to affect mainly the third moment and fourth moment of the distribution of $F(T)$ for high T values. After standardization, these moments are called the skewness and kurtosis, respectively, of the distribution.

Additional argumentation can be given regarding the benefit of setting $\mu(T) = \mu$ from both a practical and computational point of view. Firstly, the constraint to which μ is calibrated can be straightforwardly replaced with the identical constraint that the put-call parity must hold for the at-the-money put and call options. This replacement is made because, once a solution to the defined general SLV model has been obtained, calibrating to this new constraint omits the necessity of performing an MC simulation to compute the value for μ that ensures preservation of the time zero forward. In turn this will make computing μ rather than $\mu(T)$ more tractable from a computational standpoint and clearly faster. The method for approaching this new procedure will be extensively discussed in Section 4.1 in Chapter 4.

Secondly, due to the risk of not appropriately capturing the skewness and kurtosis of the distribution of $F(T)$ when using μ rather than $\mu(T)$, it is important that these inaccuracies are dealt with accordingly. Recall that σ was kept undefined in the specification of the general SLV model defined earlier in this section, and that this was done so that the user of the model is essentially free in specifying a custom LV function with only minor restrictions. This topic will be extensively investigated later in this thesis, though it is now an appropriate time to briefly discuss how the generality of σ strengthens the soundness of the assumption that $\mu(T) = \mu$. Assuming the user of the model defines σ in parametric fashion, it is likely that some of the parameters in this function will end up in the calibration algorithm eventually used to calibrate the model. In turn, as Bang [9] briefly mentions, this means that in most cases the LV function is capable of absorbing minor inaccuracies stemming from the assumption that $\mu(T) = \mu$. The same cannot be said, however, for fully non-parametric LV functions or simpler functional forms such as $\sigma(F(t)) = F(t)$. In these cases, other model parameters will be responsible for the absorption of inaccuracies. Nevertheless, extra attention must be paid to minor inaccuracies in the model output. Again, LV function and calibration related topics will be individually highlighted in Chapter 4 of this thesis.

Concluding, numerous arguments and a number of numerical tests have been discussed on why the assumption $\mu(T) = \mu$ is practical for future derivations and computational procedures. Based on these points and the results shown in Table 2.1, unless otherwise mentioned, the following definition for $F(T)$ will be used from here on out

$$F(T) \coloneqq G^{-1}(\eta(T) - \mu). \tag{2.31}$$

3

General SLV model solution

The approximative solution to the general SLV model that will be discussed in this thesis is inspired by one originally proposed by Bang [10]. In this chapter the results and points made in Section 2.5 of the previous chapter will be extended and used to obtain the aforementioned solution. Section 2.5 discussed the validity of the assumption behind defining Equation (2.31). One of the arguments mentioned was that the goal of solving the general SLV model is to derive pricing functions for vanilla European options, i.e. not for path dependent products, which is precisely what will be done in this chapter. Recall that the SLV model defined in Section 2.5 reads

$$\begin{aligned} \mathrm{d}F(t) &= \sigma(F(t))\,\mathrm{d}\eta(t), \\ \mathrm{d}\eta(t) &= \alpha(t)\,\mathrm{d}W_\eta^T(t), \\ \mathrm{d}\alpha(t) &= \nu\alpha(t)\,\mathrm{d}W_\alpha^T(t), \end{aligned}$$

where the local volatility function σ controls the backbone of the distribution of $F(t)$. Moreover, it still holds that $W_\eta^T(t)$ and $W_\alpha^T(t)$ are regular Brownian motions under the T forward measure, ν is a positive volatility-of-volatility parameter, and $\mathrm{d}W_\eta^T(t)\,\mathrm{d}W_\alpha^T(t) = \rho\,\mathrm{d}t$, where $\rho \in [-1,1]$.

The approach to approximating a solution to this model to obtain approximative pricing functions for vanilla European options on $F(T)$ is as follows. Equation (2.31) can be used to express prices of call options on $F(T)$, with a strike K and maturity T, in terms of an expectation term $\mathbb{E}^T[(\eta(T)-K)^+]$ under the T forward measure. Provided that, for simplicity, the risk-free rate r is set to zero throughout this chapter, this term is clearly equal to the time zero value of a call option with strike K and maturity T on the underlying $\eta(T)$. Note, this is a purely theoretical concept as options on stochastic volatility processes are generally not a tradable product, though for simplicity this terminology can be adopted. In other words, call options on $F(T)$ with strike K and maturity T, denoted by $C_F(K)$, can be expressed in terms of call options on $\eta(T)$ with the same strike K and maturity T, denoted by $C_\eta(K)$. These steps can then be repeated once more to express $C_\eta(K)$ in terms of another expectation term similar in form, though now $\eta(T)$ is replaced by an auxiliary process $\xi(T)$. Again, this expectation term will represent the value of a vanilla European option on $\xi(T)$ with strike K and maturity T, hence this terminology will once more be adopted with similar notation. This final step requires increased attention to detail beyond the scope of this introduction, and the reader is referred to Section 3.3 to see the full derivations. Analogously, all these steps can be repeated for deriving pricing functions for put options on $F(T)$.

The beauty of this three-step approach can perhaps be more easily visualized with a metaphor. Imagine the problem space is essentially set up as a Matryoshka doll[1] with three other, smaller dolls inside of it. The initial doll represents the act of pricing options on $F(T)$. Inside of this act are the three smaller dolls, each representing pricing functions for options on a different underlying. The final doll represents the existence of pricing functions for (theoretical) options on $\xi(T)$. In other words, in order to price options on $F(T)$, pricing functions on $F(T)$ are required. In order for pricing functions on $F(T)$ to be evaluated, pricing functions on

[1] More information: `https://en.wikipedia.org/wiki/Matryoshka_doll`

$\eta(T)$ are required, etc. This is how the complicated issue of directly deriving pricing functions for options on $F(T)$ will be shifted around, chipping away at the issue at each step. Finally, the terminology that will be used in this chapter is that pricing functions on any one of the three underlyings, i.e. $F(T), \eta(T)$ and $\xi(T)$, will be called a set of functions. In other words, three different sets of functions are derived in this chapter. Moreover, as the route to obtaining these functions contains a number of assumptions, this chapter will cover the validity and significance of said assumptions.

3.1. Pricing functions on $F(T)$

In this section, the first set of pricing functions for options on $F(T)$ is derived. Lemma 2.5.1 plays an important role in this section, as similar notation and functions will be used here. For clarity, recall that this lemma stated that for a continuous and strictly positive LV function σ, the functional form for the Lamperti transform corresponding to the earlier defined SLV model was defined as

$$G(F(t)) \coloneqq \int_{F(0)}^{a} \frac{1}{\sigma(u)}\,\mathrm{d}u\bigg|_{a=F(t)}. \tag{3.1}$$

Moreover, recall from Section 2.5 that the new expression for $F(T)$ in terms of its stochastic volatility, following from the above Lamperti transform and an additional assumption also discussed in Section 2.5, reads

$$F(T) \coloneqq G^{-1}(\eta(T) - \mu). \tag{3.2}$$

For future derivations it is important that the function G can take variable input, without running the risk of becoming ill-defined. Recall from the definition of the Lamperti transform that this risk generally does not exist, as the interval of the integral present in the transform is always within the state-space of the underlying process. In order to prevent computational breakdowns when subjecting the transform to an integration interval that is partially outside of the state-space of $F(t)$, a new version can be defined. Let $\mathscr{X}$ denote the state-space of $F(t)$. Then the generalized Lamperti transform reads

$$G(a) \coloneqq \int_{F(0)}^{a} \mathbb{1}_{u\in\mathscr{X}} \frac{1}{\sigma(u)}\,\mathrm{d}u\bigg|_{a=F(t)}. \tag{3.3}$$

The extension of Equation (3.1) to Equation (3.3) can essentially be seen as building on a computational safety net. Unless otherwise mentioned, Equation (3.3) is the function that will be used from here on out. With an eye on computational tractability, the following theorem covers the pricing functions for options on $F(T)$. Note that the terminology of (theoretical) options on $\eta(T)$ defined in the introduction of this chapter is adopted here and recall that for simplicity it was assumed that the risk-free rate r is equal to zero throughout all derivations in this chapter.

Theorem 3.1.1. *For an arbitrary strike K and maturity T, prices of call and put options on $F(T)$, denoted by $C_F(K)$ and $P_F(K)$, can be written in terms of (theoretical) call/put options on $\eta(T)$ with the same strike K and maturity T, denoted by $C_\eta(K)$ and $P_\eta(K)$. These pricing functions read*

$$C_F(K) = \sigma(K)C_\eta(G(K)+\mu) + \int_K^{+\infty} \sigma'(k)C_\eta(G(k)+\mu)\,\mathrm{d}k, \tag{3.4}$$

$$P_F(K) = \sigma(K)P_\eta(G(K)+\mu) - \int_{-\infty}^{K} \sigma'(k)P_\eta(G(k)+\mu)\,\mathrm{d}k, \tag{3.5}$$

or, in terms of the CDF of $\eta(T)$ denoted by Φ_η, these functions can be expressed as

$$C_F(K) = \int_K^{+\infty} (1-\Phi_\eta(G(k)+\mu))\,\mathrm{d}k, \tag{3.6}$$

$$P_F(K) = \int_{-\infty}^{K} \Phi_\eta(G(k)+\mu)\,\mathrm{d}k. \tag{3.7}$$

Proof. In this proof the integral representation discussed in Lemma A.1.1 is used, as well as the results from Lemma A.2.1. Lemma A.1.1 states that the following holds for a twice differentiable function $f:\mathbb{R}\to\mathbb{R}^+$

$$f(S) = f(a) + f'(a)(S-a) + \int_a^{+\infty} f''(v)(S-v)^+\,\mathrm{d}v + \int_{-\infty}^{a} f''(v)(v-S)^+\,\mathrm{d}v,$$

where $\mathbb{R}^+$ denotes the set of strictly positive real numbers. By setting $f = G^{-1}$, $S = \eta(T) - \mu$, $a = G(K)$, using the notation $g'(x) = \frac{\mathrm{d}g(x)}{\mathrm{d}x}$ and $g''(x) = \frac{\mathrm{d}^2 g(x)}{\mathrm{d}x^2}$ for an arbitrary function g and dummy variable x, and observing from Lemma A.2.1 that $(G^{-1})'(G(k)) = \sigma(k)$ and $(G^{-1})''(G(k)) = \sigma'(k)\sigma(k)$ for a dummy variable k, everything can be filled in to obtain

$$G^{-1}(\eta(T)-\mu) = G^{-1}(G(K)) + \sigma(K)(\eta(T)-\mu-G(K)) + \int_{-\infty}^{G(K)} (G(k)-\eta(T)+\mu)^+ \sigma'(k)\sigma(k)\,\mathrm{d}k \\ + \int_{G(K)}^{+\infty} (\eta(T)-\mu-G(k))^+ \sigma'(k)\sigma(k)\,\mathrm{d}k,$$

to which a variable transformation can be applied, yielding

$$F(T) - K = \sigma(K)(\eta(T)-\mu-G(K)) + \int_{-\infty}^{K} (G(k)-\eta(T)+\mu)^+ \sigma'(k)\,\mathrm{d}k + \int_{K}^{+\infty} (\eta(T)-\mu-G(k))^+ \sigma'(k)\,\mathrm{d}k.$$

Using Equation (3.2) and multiplying both sides of the equation above with $\mathbb{1}_{K\leq F(T)}$, gives

$$(F(T)-K)^+ = \sigma(K)(\eta(T)-\mu-G(K))^+ + \int_{K}^{+\infty} (\eta(T)-\mu-G(k))^+ \sigma'(k)\,\mathrm{d}k.$$

Finally, taking the expectation under the T forward measure on both sides of this equation and moving it inside the integral by applying Fubini's theorem [26], yields

$$C_F(K) = \sigma(K) C_\eta(G(K)+\mu) + \int_{K}^{+\infty} \sigma'(k) C_\eta(G(k)+\mu)\,\mathrm{d}k. \tag{3.8}$$

Analogously, the above result can be derived for put options, concluding the first part of the proof.

For the second part of the proof, the following observation can be made

$$\begin{aligned}(F(T)-K)^+ &= \int_K^{+\infty} \mathbb{1}_{k\leq F(T)}\,\mathrm{d}k \\ &= \int_K^{+\infty} \mathbb{1}_{G(k)\leq G(F(T))}\,\mathrm{d}k\end{aligned}$$

Using the expression for $F(T)$ shown in Equation (3.2) then gives

$$\begin{aligned}&= \int_K^{+\infty} \mathbb{1}_{G(k)+\mu\leq \eta(T)}\,\mathrm{d}k \\ &= \int_K^{+\infty} (1 - \mathbb{1}_{\eta(T)\leq G(k)+\mu})\,\mathrm{d}k.\end{aligned}$$

Once more, application of Fubini's theorem when taking the expectation on both sides of the equation gives the final result

$$C_F(K) = \int_K^{+\infty} (1 - \Phi_\eta(G(k)+\mu))\,\mathrm{d}k. \tag{3.9}$$

Finally, the above result can be derived analogously for put options. □

These functions are exact for pricing options on the underlying $F(T)$ as stated by Equation (3.2), free of butterfly arbitrage within the same expiry as per Corollary 2.5.1.1, and will be fast in evaluation provided appropriate pricing functions exist for options on $\eta(T)$. Taking Equation (3.9) as an example, it can be seen that computation of $C_F(K)$ requires one main integral evaluation, and potentially a high number of additional evaluations of G. In the numerical simulation, the amount of evaluations of the function G relies on the amount of integration points used. Luckily, these evaluations can be vectorized, and integration packages in, for example, Python, are highly optimized. Provided that the LV function is fast in its evaluation, the potentially large amount of G evaluations will therefore not pose a problem. Moreover, note that these pricing functions make use of the existence of Equation (3.3), as the put option prices require integration down to $-\infty$, which means that the integration interval is likely to partially lie outside of the state space of $F(t)$ for certain LV functions.

Even though the functions are now well-defined on paper, the fact that the integrals integrate over an interval going up (or down) to (minus) infinity, is still an issue. This is because the integration interval has to be bounded in order to perform numerical integration. A heuristic approach to truncating these integration intervals is discussed in the last subsection of Section 3.4.

3.2. Pricing functions on $\eta(T)$

This section makes use of the earlier defined terminology regarding theoretical options on the SV component of the earlier defined general SLV model. The aim is to find fast closed-form pricing functions for options on this SV component such that the pricing functions derived in the previous section can be appropriately evaluated. Since the first set of pricing functions has already been defined, the general SLV model can be reduced to the normal SABR model

$$\mathrm{d}\eta(t) = \alpha(t)\,\mathrm{d}W_\eta^T(t), \tag{3.10}$$

$$\mathrm{d}\alpha(t) = \nu\alpha(t)\,\mathrm{d}W_\alpha^T(t), \tag{3.11}$$

where the same constraints as before hold for the parameters, i.e. $\nu > 0$, correlation between the two Brownian motions is captured by $\rho \in [-1,1]$ etc. The power of the Lamperti transform approach shines through here as the LV function does not have to be dealt with in these derivations, while it still plays a role in the pricing of options on the main underlying $F(T)$. The system above can be written out easily to yield explicit expressions for $\eta(t)$ and $\alpha(t)$

$$\eta(t) = \eta(0) + \int_0^t \alpha(t)\,\mathrm{d}W_\eta^T(t), \tag{3.12}$$

$$\alpha(t) = \alpha(0)\,\mathrm{e}^{-\frac{1}{2}\nu^2 t + \nu W_\alpha^T(t)}. \tag{3.13}$$

Important to note is that an exact solution to the normal SABR model as defined through Equation (3.10) and (3.11) has already been showcased in earlier research. Indeed, Antonov *et al.* [1] show that such an exact solution expressed in multiple nested integrals can be derived. However, while this solution is exact, these nested integrals make its evaluation computationally heavy. This fact justifies an attempt at deriving a faster solution. Additional details on this solution proposed by Antonov *et al.* [1] will be given later, when this topic is revised to compare its output to the output of the new solution to the normal SABR model derived in this section.

Besides the existence of a heavy exact solution, Hagan *et al.* [31] showcase faster asymptotical approximation formulae for pricing options with the SABR model. While very fast and easy to implement, the downside of these functions is that they are asymptotical in time and therefore tend to become inaccurate when trying to price options with large maturities. Moreover, due to divisions by the strike in specific terms inside the formulae, results break down when pricing options with strikes close to zero, leading to arbitrageable output. These facts justify attempting to derive an approximation capable of pricing vanilla European options that is more accurate and remains accurate even when pricing options with large maturities. Once again, additional details are omitted here as this method will be more extensively discussed in comparison procedures later on.

In preparation of solving the normal SABR model, the upcoming additional lemma will prove to be very useful. Lemma 3.2.1 introduces a new process $z(t)$. The dynamics of this process might seem strange at first, but the reason for defining them in this form will become clear in the main theorem statement of this section, where the result of this lemma is used to introduce some auxiliary functions, eventually resulting in the pricing functions for options on $\eta(T)$.

Lemma 3.2.1. *Consider the process $z(t)$, with dynamics*

$$\mathrm{d}z(t) = \nu\sqrt{1 + 2\rho z(t) + z^2(t)}\,\mathrm{d}W_\eta^T(t), \quad z(0) = 0.$$

Then for any $t \geq 0$, $\tilde{\eta}(t) \coloneqq \eta(0) + \frac{\alpha(0)}{\nu} z(t)$ has the same marginal distributions as $\eta(t)$.

Proof. (*sketch*) Provided with the normal SABR model as described by Equations (3.10) and (3.11), it follows from Gyöngy's lemma [30] that there exists a process $\tilde{\eta}(t)$ that has equal marginal distributions to $\eta(t)$ (described by Equation (3.10)) and dynamics described by

$$\mathrm{d}\tilde{\eta}(t) = \sqrt{V(t,\tilde{\eta}(t))}\,\mathrm{d}W_\eta^T(t),$$
$$\mathrm{d}\alpha(t) = \nu\alpha(t)\,\mathrm{d}W_\alpha^T(t),$$

where

$$V(t,x) \coloneqq \mathbb{E}^T\left[\alpha^2(t) | \eta(t) = x\right].$$

Lemma B.1.1 in Section B.1 of Appendix B then shows through a series of measure changes that the function $V(t,x)$ as defined above is quadratic. Moreover, it is shown that this eventually results in the possibility to omit the time dependence and that the following expression can be obtained

$$\sqrt{V(K)} = \alpha(0)\sqrt{1+2\rho\nu f(K)+\nu^2 f^2(K)}, \tag{3.14}$$

where

$$f(K) \coloneqq \frac{1}{\alpha(0)}(K-\eta(0)).$$

Substituting Equation (3.14) into the expression for the dynamics of $\tilde{\eta}(t)$ shown earlier then yields that

$$\mathrm{d}\tilde{\eta}(t) = \alpha(0)\sqrt{1+2\rho\nu\left(\frac{\tilde{\eta}(t)-\eta(0)}{\alpha(0)}\right)+\nu^2\left(\frac{\tilde{\eta}(t)-\eta(0)}{\alpha(0)}\right)^2}\,\mathrm{d}W_\eta^T(t). \tag{3.15}$$

It becomes clear that the variable transform $\tilde{\eta}(t) = \eta(0) + \frac{\alpha(0)}{\nu}z(t)$ can be applied to Equation (3.15), yielding the result stated in the lemma and therewith concluding the proof. □

The rewriting process of the dynamics of $\tilde{\eta}(t)$ mentioned above is relatively complicated and too unnecessarily detailed to include here in the main text, hence the reader was referred to Appendix B, Section B.1 to read it. The result might seem unnecessary, but the alternative process $\tilde{\eta}(t)$ allows for clever rewriting in the upcoming theorem. Moreover, this result is checked using a numerical experiment in Section 3.4.

As was stated in the introduction to this chapter, the pricing functions for options on $\eta(T)$ will be expressed in terms of prices of options on an auxiliary process/underlying, $\xi(T)$. Let it be emphasized once more that these are purely theoretical options defined through the terminology framework discussed earlier. Finally, the following theorem constitutes the body of this section by showing that pricing functions for options on $\eta(T)$, expressed in prices of options on $\xi(T)$, exist.

Theorem 3.2.2. *For a given strike K, let $y(K) \coloneqq \frac{\nu}{\alpha(0)}(K-\eta(0))$ and let*

$$\chi(K) \coloneqq \frac{\sqrt{1-\rho^2+(y(K)+\rho)^2}+y(K)+\rho}{1+\rho}.$$

Functional forms for the CDF of $\eta(T)$ denoted by Φ_η, and prices of (theoretical) call options on $\eta(T)$ with strike K and maturity T, denoted by $C_\eta(K)$, then read

$$C_\eta(K) = \frac{\alpha(0)}{2\nu}\left[(1+\rho)C_\xi(\chi(K))+(1-\rho)P_{1/\xi}\left(\frac{1}{\chi(K)}\right)\right], \tag{3.16}$$

$$\Phi_\eta(K) = \Phi_\xi(\chi(K)), \tag{3.17}$$

in which $C_\xi(\chi(K))$ and $P_{1/\xi}(\chi(K))$ denote the prices of (theoretical) options on $\xi(T)$ and $1/\xi(T)$, respectively, while Φ_ξ denotes the CDF of $\xi(T)$, where $\xi(T)$ is an auxiliary process defined through

$$\xi(T) \coloneqq \mathrm{e}^{\nu h(T)}, \tag{3.18}$$

with

$$h(T) \coloneqq \int_0^{z(T)} \frac{1}{\nu\sqrt{1+2\rho x+x^2}}\,\mathrm{d}x, \tag{3.19}$$

where the process $z(T)$ is defined as $z(T) = \frac{\nu}{\alpha(0)}(\tilde{\eta}(T)-\eta(0))$.

Proof. Equation (3.19) can be solved explicitly, yielding

$$h(T) = \frac{\log\left(\sqrt{1+2\rho z(T)+z^2(T)}+\rho+z(T)\right)}{\nu} - \frac{\log(1+\rho)}{\nu},$$

which can be inverted to obtain the following alternative explicit expression for $z(T)$

$$z(T) = -\rho + \frac{1}{2}\left((1+\rho)\xi(T) - \frac{1-\rho}{\xi(T)}\right). \tag{3.20}$$

Moreover, the equation for $\chi(K)$ as stated in the theorem can be easily manipulated to obtain

$$y(K) = -\rho + \frac{1}{2}\left((1+\rho)\chi(K) - \frac{1-\rho}{\chi(K)}\right). \tag{3.21}$$

Important to note is that the definition for $y(K)$ was specifically chosen such that its form matches that of $z(T)$. This selection makes it straightforward to subtract $y(K)$ and $z(T)$. Performing this step gives

$$\frac{\nu}{\alpha(0)}(\tilde{\eta}(T) - K) = \frac{1}{2}\left((1+\rho)(\xi(T) - \chi(K)) + (1-\rho)\left(\frac{1}{\chi(K)} - \frac{1}{\xi(T)}\right)\right).$$

Lemma 3.2.1 can now be used to revert back to the original process $\eta(T)$, though special attention has to be paid the fact that the equality of $\tilde{\eta}(T)$ and $\eta(T)$ holds only in distribution. Switching back to $\eta(T)$ notation and simplifying further yields

$$\begin{aligned}
\frac{\nu}{\alpha(0)}(\eta(T) - K) &\overset{d}{=} \frac{1}{2}\left((1+\rho)(\xi(T) - \chi(K)) + (1-\rho)\left(\frac{1}{\chi(K)} - \frac{1}{\xi(T)}\right)\right), \\
\eta(T) - K &\overset{d}{=} \frac{\alpha(0)}{2\nu}\left((1+\rho)(\xi(T) - \chi(K)) + (1-\rho)\left(\frac{1}{\chi(K)} - \frac{1}{\xi(T)}\right)\right), \\
(\eta(T) - K)^+ &\overset{d}{=} \left(\frac{\alpha(0)}{2\nu}\left((1+\rho)(\xi(T) - \chi(K)) + (1-\rho)\left(\frac{1}{\chi(K)} - \frac{1}{\xi(T)}\right)\right)\right)^+.
\end{aligned}$$

It is important to observe that the right hand side of the equation above consists of a sum of two increasing functions of $\xi(T)$, which both vanish at $\chi(K)$. This means that the expression above, containing the maximum of the sum of two increasing functions, can be written as a sum of maxima as per application of Jamshidian's trick [34]. This yields

$$(\eta(T) - K)^+ \overset{d}{=} \frac{\alpha(0)}{2\nu}\left((1+\rho)(\xi(T) - \chi(K))^+ + (1-\rho)\left(\frac{1}{\chi(K)} - \frac{1}{\xi(T)}\right)^+\right). \tag{3.22}$$

Since the right hand side in the equation above is non-zero if and only if $\xi(T) > \chi(K)$, and the left hand side is non-zero if and only if $\eta(T) > K$, it is clear that

$$\mathbb{1}_{\eta(T)>K} \overset{d}{=} \mathbb{1}_{\xi(T)>\chi(K)},$$

from which consequently follows that

$$\mathbb{1}_{\eta(T)\leq K} \overset{d}{=} \mathbb{1}_{\xi(T)\leq\chi(K)}. \tag{3.23}$$

Finally, taking the expectation of the left and right hand side of the Equation (3.22) and (3.23) yields the desired result stated in the theorem. □

As per Theorem 3.2.2, closed form pricing functions for options on $\eta(T)$ can indeed be derived and expressed in terms of option prices on $\xi(T)$. The theorem also shows that the call to the CDF of $\eta(T)$ in Equation (3.6) and (3.7) can be elegantly replaced with a call to the CDF of $\xi(T)$, as per Equation (3.17), without adding computational overhead.

If one so desired, it would be possible to round off the model at this point. The underlying $\xi(T)$ is defined as an exponential of some other well-defined process, and can thus be approximated through moment-matching. However, as Bang [9] confirms, the accuracy of the results can be improved without gaining substantial computational overhead, by deriving explicit pricing functions for options on $\xi(T)$ also. Moreover, since the reader was promised that the model would be covered in full detail, the final set of functions is derived in the next section.

3.3. Finalizing the solution

The previous section showcased the derivation of pricing functions capable of pricing options on $\eta(T)$, expressed in terms of prices of options on an auxiliary process $\xi(T)$. In more detail, it was shown that the price of a call option on $\eta(T)$ with strike K and maturity T, denoted by $C_\eta(K)$, could be expressed in terms of a call on $\xi(T)$ with maturity T and strike $\chi(K)$, and a put option on $1/\xi(T)$ with maturity T and strike $1/\chi(K)$. In

this notation the functional form for $\chi(K)$ was defined in the statement of Theorem 3.2.2. Moreover, recall that $\xi(t)$ was defined as

$$\xi(t) \coloneqq \mathrm{e}^{\nu h(t)}, \tag{3.24}$$

where $\nu > 0$ was the volatility-of-volatility parameter in the normal SABR model and

$$h(t) \coloneqq \int_0^{z(t)} \frac{1}{\nu\sqrt{1+2\rho x+x^2}}\,\mathrm{d}x, \quad h(0)=0, \tag{3.25}$$

$$z(t) = \frac{\nu}{\alpha(0)}(\tilde{\eta}(T)-\eta(0)), \tag{3.26}$$

where $\tilde{\eta}(t)$ is a process with equal marginal distributions to $\eta(t)$ defined in Equation (3.12). Details regarding these two processes were discussed in Lemma 3.2.1. In this section the aim is to derive the final set of pricing functions, or, in reference to the earlier discussed metaphor, cover the smallest Matryoshka doll. In other words, the aim is find explicit functional expressions for $C_\xi(\chi(K))$ and $P_{1/\xi}(1/\chi(K))$.

As was mentioned at the end of Section 3.2, Bang [9] states that deriving these functions can result in higher accuracy of the output of Equation (3.16) and (3.17) than that which would be obtained when approximation $\xi(T)$ through a moment-matching approach. However, to make sure that the advantage of greater accuracy does not get outweighed by the loss in computational performance from having to evaluate these additional pricing functions rather than approximating $\xi(T)$, close attention must be paid to the computational tractability of said pricing functions.

Equation (3.24) makes it clear that finding an explicit expression for the density of $h(t)$ enables pricing of options on $\xi(T)$. Using a call option as an example, this follows naturally from the following

$$C_\xi(\chi(K)) = \mathbb{E}^T[(\xi(T)-\chi(K))^+] = \mathbb{E}^T[(\mathrm{e}^{\nu h(T)}-\chi(K))^+] = \int_{-\infty}^{+\infty}(\mathrm{e}^{\nu h(T)}-\chi(K))^+ p^T(x)\,\mathrm{d}x \tag{3.27}$$

where p^T is the density of $h(T)$ under the T forward measure. In light of this, a measure change to capture the dynamics of $h(T)$ more meticulously will be useful. The definition for such a measure follows naturally after an application of Itô's lemma to $h(t)$, of which the results are neatly summarized by the following corollary.

Corollary 3.3.0.1. *Let $h(t)$ be defined as in Equation (3.25) with $z(t)$ as shown in Equation (3.26). Moreover, recall that $z(t)$ is described by the dynamics*

$$\mathrm{d}z(t) = \nu\sqrt{1+2\rho z(t)+z^2(t)}\,\mathrm{d}W_\eta^T(t), \quad z(0)=0.$$

An application of Itô's lemma to $h(t)$ then yields

$$\mathrm{d}h(t) = \mathrm{d}W_\eta^T(t) - \frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\,\mathrm{d}t, \tag{3.28}$$

where

$$\bar{h} \coloneqq \frac{1}{2\nu}\ln\left(\frac{1+\rho}{1-\rho}\right). \tag{3.29}$$

Proof. See Section B.2 in Appendix B. □

Corollary 3.3.0.1 states that under the T forward measure $h(T)$ can be described by a Brownian motion with drift. Having an expression for the dynamics of $h(t)$ will prove to be useful in the following lemma statement, where the earlier discussed approach for obtaining the density of $h(t)$ is discussed in detail.

Lemma 3.3.1. *Call and put options on $\xi(T)$ and $1/\xi(T)$ with strike $\chi(K)$ and $1/\chi(K)$ and maturity T, denoted by $C_\xi(\chi(K))$ and $P_{1/\xi}\left(1/\chi(K)\right)$, respectively, can be priced with*

$$C_\xi(\chi(K)) \approx \frac{3}{2\gamma}C(\nu,\nu,k,\bar{h},T) - \frac{1}{2\gamma}C(\nu,3\nu,k,\bar{h},T), \tag{3.30}$$

$$P_{1/\xi}\left(\frac{1}{\chi(K)}\right) \approx \frac{1}{2\gamma}C(-\nu,3\nu,k,\bar{h},T) - \frac{3}{2\gamma}C(-\nu,\nu,k,\bar{h},T), \tag{3.31}$$

$$\tag{3.32}$$

and the CDF of $\xi(T)$, denoted by Φ_ξ, can be expressed in functional form as

$$(1-\Phi_\xi(\chi(K))) \approx \frac{3}{2\gamma}\Upsilon(\nu,k,\bar{h},T) - \frac{1}{2\gamma}\Upsilon(3\nu,k,\bar{h},T). \tag{3.33}$$

For the functions above the following expressions are used

$$\bar{h} \coloneqq \frac{1}{2\nu}\ln\left(\frac{1+\rho}{1-\rho}\right) \tag{3.34}$$

$$y(K) \coloneqq \frac{\nu}{\alpha(0)}(K-M(0)), \tag{3.35}$$

$$\chi(K) \coloneqq \frac{\sqrt{1-\rho^2+(y(K)+\rho)^2}+y(K)+\rho}{1+\rho}, \tag{3.36}$$

$$k \coloneqq \frac{1}{\nu}\ln\big(\chi(K)\big), \tag{3.37}$$

$$\gamma \coloneqq \frac{3}{2}\Upsilon^*(\nu,\bar{h},T) - \frac{1}{2}\Upsilon^*(3\nu,\bar{h},T), \tag{3.38}$$

using λ and σ as dummy variables, the above expressions use the supplementary functions

$$\Upsilon^*(\lambda,\bar{h},T) \coloneqq \mathrm{e}^{\lambda^2 T/8}\left[\mathrm{e}^{-\lambda\bar{h}/2}\Phi\left(\frac{\bar{h}}{\sqrt{T}}-\frac{\lambda\sqrt{T}}{2}\right)+\mathrm{e}^{\lambda\bar{h}/2}\Phi\left(-\frac{\bar{h}}{\sqrt{T}}-\frac{\lambda\sqrt{T}}{2}\right)\right],$$

$$C(\sigma,\lambda,k,\bar{h},T) \coloneqq \mathrm{e}^{\sigma^2 T/2}\Upsilon(\lambda,k-\sigma T,\bar{h}+\sigma T,T) - \mathrm{e}^{\sigma k}\Upsilon(\lambda,k,\bar{h},T),$$

$$\Upsilon(\lambda,k,\bar{h},T) \coloneqq \begin{cases} \mathrm{e}^{(\lambda^2 T/8)-(\lambda\bar{h}/2)}\Phi\left(-\frac{k}{\sqrt{T}}-\frac{\lambda\sqrt{T}}{2}\right) & \text{if} \quad k \geq -\bar{h}, \\ \Upsilon^*(\lambda,\bar{h},T) - \mathrm{e}^{(\lambda^2 T/8)+(\lambda\bar{h}/2)}\Phi\left(\frac{k}{\sqrt{T}}-\frac{\lambda\sqrt{T}}{2}\right) & \text{if} \quad k < -\bar{h}. \end{cases}$$

Proof. Based on Equation (3.28), it is straightforward to define a new measure under which $h(T)$ can be described by a driftless Brownian motion. This will simplify the steps necessary for deriving its density under the T forward measure. Considering these points, let $\theta^*(t)$ be an exponential martingale [46] under the T forward measure described by

$$\frac{\mathrm{d}\theta^*(t)}{\theta^*(t)} = \frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\,\mathrm{d}W_\eta^T(t),$$

such that

$$\theta^*(t) = \theta^*(0)\,\mathrm{e}^{-\frac{1}{2}\left(\frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\right)^2 t+\frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))W_\eta^T(t)}. \tag{3.39}$$

The new measure under which $h(t)$ is a driftless Brownian motion can then be defined through

$$\frac{\mathrm{d}\mathbb{Q}^*}{\mathrm{d}\mathbb{Q}^T} = \theta^*(T).$$

Let δ_a denote the Dirac distribution at an arbitrary point a. This new measure combined with basic characteristics of the expectation operator then allows for writing

$$\mathbb{E}^T[\delta_a(h(T))] = \mathbb{E}^{\mathbb{Q}^*}\left[\frac{1}{\theta^*(T)}\delta_a(h(T))\right] = \mathbb{E}^{\mathbb{Q}^*}\left[\mathbb{E}^{\mathbb{Q}^*}\left[\frac{1}{\theta^*(T)}\middle|h(T)\right]\delta_a(h(T))\right]. \tag{3.40}$$

Continuing, let

$$f(x) \coloneqq \sqrt{\cosh(\nu(x+\bar{h}))}, \tag{3.41}$$

for which Lemma B.3.1 in Appendix B shows that $\mathrm{d}\left(\frac{f(h(t))}{\theta^*(t)}\right) = \mathscr{O}(\mathrm{d}t)$, which means that this form for f neutralizes the diffusion term of $\theta^*(t)$, i.e. working out $\mathrm{d}\left(\frac{f(h(t))}{\theta^*(t)}\right)$ shows all terms of randomness will cancel out leaving only $\mathscr{O}(\mathrm{d}t)$ terms. Moreover, from Lemma B.3.2 it follows that through a second order Taylor expan-

sion of $f(x)$ in $\mathrm{e}^{-\nu|x+\bar{h}|}$, it is possible to write

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}^*}\left[\frac{1}{\theta^*(T)}\middle|h(T)\right] &= \frac{1}{f(h(T))}\mathbb{E}^{\mathbb{Q}^*}\left[\frac{f(h(T))}{\theta^*(T)}\middle|h(T)\right] \\ &= \mathrm{e}^{-(\nu/2)|h(T)+\bar{h}|}\frac{3-\mathrm{e}^{-\nu|h(T)+\bar{h}|}}{2}\frac{1}{\gamma}+\mathscr{O}(\mathrm{e}^{-\nu|h(T)+\bar{h}|}-1) \\ &\approx \mathrm{e}^{-(\nu/2)|h(T)+\bar{h}|}\frac{3-\mathrm{e}^{-\nu|h(T)+\bar{h}|}}{2}\frac{1}{\gamma} \\ &=: \frac{1}{\theta^\dagger(h(T))},\end{aligned} \tag{3.42}$$

where letting $\mathbb{E}^{\mathbb{Q}^*}\left[\frac{f(h(T))}{\theta^*(T)}\middle|h(T)\right] := \frac{1}{\gamma}$ is justified by fact that $\mathrm{d}\left(\frac{f(h(t))}{\theta^*(t)}\right) = \mathscr{O}(\mathrm{d}t)$. Moreover, this constant γ will enforce the property that the Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{Q}^T}{\mathrm{d}\mathbb{Q}^*} = \frac{1}{\theta^*(T)}$ integrates to one, i.e.

$$\mathbb{E}^{\mathbb{Q}^*}\left[\frac{1}{\theta^\dagger(h(T))}\right] = 1.$$

Substituting Equation (3.42) into Equation (3.40), and recalling that $h(T)$ is a driftless Brownian motion under the earlier defined measure $\mathbb{Q}^*$, yields

$$\mathbb{E}^T[\delta_a(h(T))] \approx \int_{\mathbb{R}}\frac{\delta_a(x)}{\theta^\dagger(x)}p^{\mathbb{Q}^*}(x)\,\mathrm{d}x = \int_{\mathbb{R}}\frac{\mathrm{e}^{-(\nu/2)|x+\bar{h}|}\left(3-\mathrm{e}^{-\nu|x+\bar{h}|}\right)}{2\gamma}\frac{\mathrm{e}^{-x^2/(2T)}}{\sqrt{2\pi T}}\,\mathrm{d}x, \tag{3.43}$$

where $p^{\mathbb{Q}^*}$ denotes the density function of $h(T)$ under $\mathbb{Q}^*$. By basic properties of the expectation operator, this result shows that the density function of $h(T)$ under T can be approximated by

$$\zeta(x) := \frac{\mathrm{e}^{-(\nu/2)|x+\bar{h}|}\left(3-\mathrm{e}^{-\nu|x+\bar{h}|}\right)}{2\gamma}\frac{\mathrm{e}^{-x^2/(2T)}}{\sqrt{2\pi T}}. \tag{3.44}$$

where an expression for the constant γ can be derived through standard integral operations applied to the no-arbitrage constraint that $\int_{-\infty}^{+\infty}\zeta(x)\,\mathrm{d}x = 1$, resulting in Equation (3.38). Equation (3.44) is an approximation for the density due to the error arising from the steps necessary for derivation of Equation (3.42). Concluding, since $\xi(T) = \mathrm{e}^{\nu h(T)}$, it is natural to introduce a variable k and define it as in Equation (3.37), i.e.

$$k := \frac{1}{\nu}\ln\left(\chi(K)\right), \tag{3.45}$$

because this choice enables the derivation of an explicit expression for the call option price, i.e.

$$\begin{aligned}C_\xi(\chi(K)) = \mathbb{E}^T[(\xi(T)-\chi(K))^+] &= \mathbb{E}^T[(\mathrm{e}^{\nu h(T)}-\mathrm{e}^{\nu k})^+] \\ &= \mathbb{E}^T[(\mathrm{e}^{\nu h(T)}-\mathrm{e}^{\nu k})\mathbb{1}_{h(T)\geq k}] \\ &\approx \int_k^{+\infty}(\mathrm{e}^{\nu x}-\mathrm{e}^{\nu k})\zeta(x)\,\mathrm{d}x,\end{aligned} \tag{3.46}$$

which can be written as Equation (3.30) through basic integral operations. Moreover, through similar operations the following explicit expression for the put option price can be found

$$P_{1/\xi}\left(\frac{1}{\chi(K)}\right) = \mathbb{E}^T\left[\left(\frac{1}{\chi(K)}-\frac{1}{\xi(T)}\right)^+\right] \approx \int_k^{+\infty}(\mathrm{e}^{-\nu k}-\mathrm{e}^{-\nu h(T)})\zeta(x)\,\mathrm{d}x, \tag{3.47}$$

which can be rewritten into Equation (3.31), again by using standard integral operations. Finally, using Equation (3.45) and the alternative expression for $\chi(K)$ admitted by this equation, it turns out that

$$(1-\Phi_\xi(\chi(K))) = \mathbb{E}^T\left[\mathbb{1}_{\xi(T)>\chi(K)}\right] \approx \int_k^{+\infty}\zeta(x)\,\mathrm{d}x, \tag{3.48}$$

and this expression can be rewritten in straightforward fashion to become Equation (3.33). Note that due to ζ being an approximation for the density of $h(T)$ under the T forward measure, these pricing functions are also approximations with an error term equal the one arising in definition of Equation (3.42). □

Lemma 3.3.1 showcased a functional expression for the CDF of $\xi(T)$ and functions that can be used to approximate the prices of options on $\xi(T)$ and $1/\xi(T)$. These functions can be used in combination with the results of Section 3.2 to price options on $\eta(T)$, which in combination with the results on Section 3.1 allows for approximating option prices on $F(T)$. Whether this approximation is accurate enough to produce competitive output will be discovered in the next section and discussed in more detail in Chapter 4. Note that the derived pricing functions require no integral evaluations, and only use a few calls to the normal CDF per evaluation. Moreover, by construction the pricing functions for options on $\eta(T)$, $\xi(T)$ and $1/\xi(T)$ output non-arbitrageable prices for any parameter set.

3.4. Numerical experiments

This section reflects on some of the results discussed in the previous two sections. More specifically, numerical verification is done on the result stated by Lemma 3.2.1, and investigation into the accuracy of the pricing functions derived in both Section 3.2 and 3.3 is done through comparing their output with existing normal SABR model solutions and approximations. Finally, the section is finished with a discussion on how to handle the infinite integration intervals present in the pricing functions derived in Section 3.1. For now, recall that the normal SABR model was defined through

$$\mathrm{d}\eta(t) = \alpha(t)\,\mathrm{d}W_\eta^T(t), \tag{3.49}$$

$$\mathrm{d}\alpha(t) = \nu\alpha(t)\,\mathrm{d}W_\alpha^T(t), \tag{3.50}$$

where $W_\eta^T(t)$ and $W_\alpha^T(t)$ are regular Brownian motions under the T forward measure, ν is a positive volatility-of-volatility parameter, and $\mathrm{d}W_\eta^T(t)\,\mathrm{d}W_\alpha^T(t) = \rho\,\mathrm{d}t$, where $\rho \in [-1,1]$. Moreover, it was shown that Equation (3.49) and (3.50) could be straightforwardly solved to yield

$$\eta(t) = \eta(0) + \int_0^t \alpha(t)\,\mathrm{d}W_\eta^T(t), \tag{3.51}$$

$$\alpha(t) = \alpha(0)\,\mathrm{e}^{-\frac{1}{2}\nu^2 t + \nu W_\alpha^T(t)}. \tag{3.52}$$

Verifying Lemma 3.2.1

Section 3.2 contained a lemma that stated the existence and derivation of the dynamics of a process $\tilde{\eta}(t)$ that has equal marginal distributions to the process $\eta(t)$ as stated in Equation (3.51). These dynamics for $\tilde{\eta}(t)$ read

$$\mathrm{d}\tilde{\eta}(t) = \alpha(0)\sqrt{1 + 2\rho\nu\left(\frac{\tilde{\eta}(t)-\eta(0)}{\alpha(0)}\right) + \nu^2\left(\frac{\tilde{\eta}(t)-\eta(0)}{\alpha(0)}\right)^2}\,\mathrm{d}W_\eta^T(t), \tag{3.53}$$

As this lemma was substantial, and directly enabled the derivation of functions capable of pricing options on $\eta(T)$, it is valuable to verify its validity and robustness in a numerical setting. This can be straightforwardly done using an MC simulation to generate paths for both $\eta(t)$ and $\tilde{\eta}(t)$, such that the distributions of $\eta(T)$ and $\tilde{\eta}(T)$ can be compared for various times T. In this MC framework, a single seed is used to generate 10000 paths with 1000 steps in time for both $\eta(t)$ and $\tilde{\eta}(t)$, using their dynamics discretized according to the Euler-Marayuma scheme (see Oosterlee-Grzelak [47]). This process can be repeated for multiple times T, which will allow for a good overview of the similarities of the distributions. In the upcoming figures these similarities are showcased using plots of the probability density and cumulative distribution functions of both $\eta(T)$ and $\tilde{\eta}(T)$. For the set of parameters

$$\eta(0) = \tilde{\eta}(0) = 1.0, \quad \alpha(0) = 0.1, \quad \nu = 0.3, \quad \rho = -0.7, \quad T \in \{2, 10, 30\},$$

the results of this procedure are showcased in Figure 3.1. In order to judge the similarity of the showcased CDF's in a more quantitative way, the Kolmogorov-Smirnov (KS) test [42] can be used. The null-hypothesis is that the two test samples come from the same distribution, so for a low p-value this hypothesis can be rejected. Tests showed that the dependency of the two samples through the Brownian motion $W_\eta^T(t)$ is strongly diminished by the high amount of MC paths that are used, so for now it will be ignored that the two-sample KS test is supposed to be applied to two independent samples. In Table 3.1 the results of the KS test for each time T are shown. Recall that the KS-statistic is the supremum of the difference between the two input samples.

T	KS-statistic	p-value
2	0.0037	1.00
10	0.0047	0.98
30	0.0047	0.98

Table 3.1: Results of two-sample Kolmogorov-Smirnov tests for comparing the CDF's shown in Figure 3.1, with a null-hypothesis stating that the samples come from the same distribution.

The table above supports the statement made by Lemma 3.2.1 as the obtained p-values do not allow for rejection of the null-hypothesis. In fact, since the proof of the lemma was free of approximations, the small differences between the distributions of $\eta(T)$ and $\tilde{\eta}(T)$ can be connected to the fact that only a limited amount of paths was simulated.

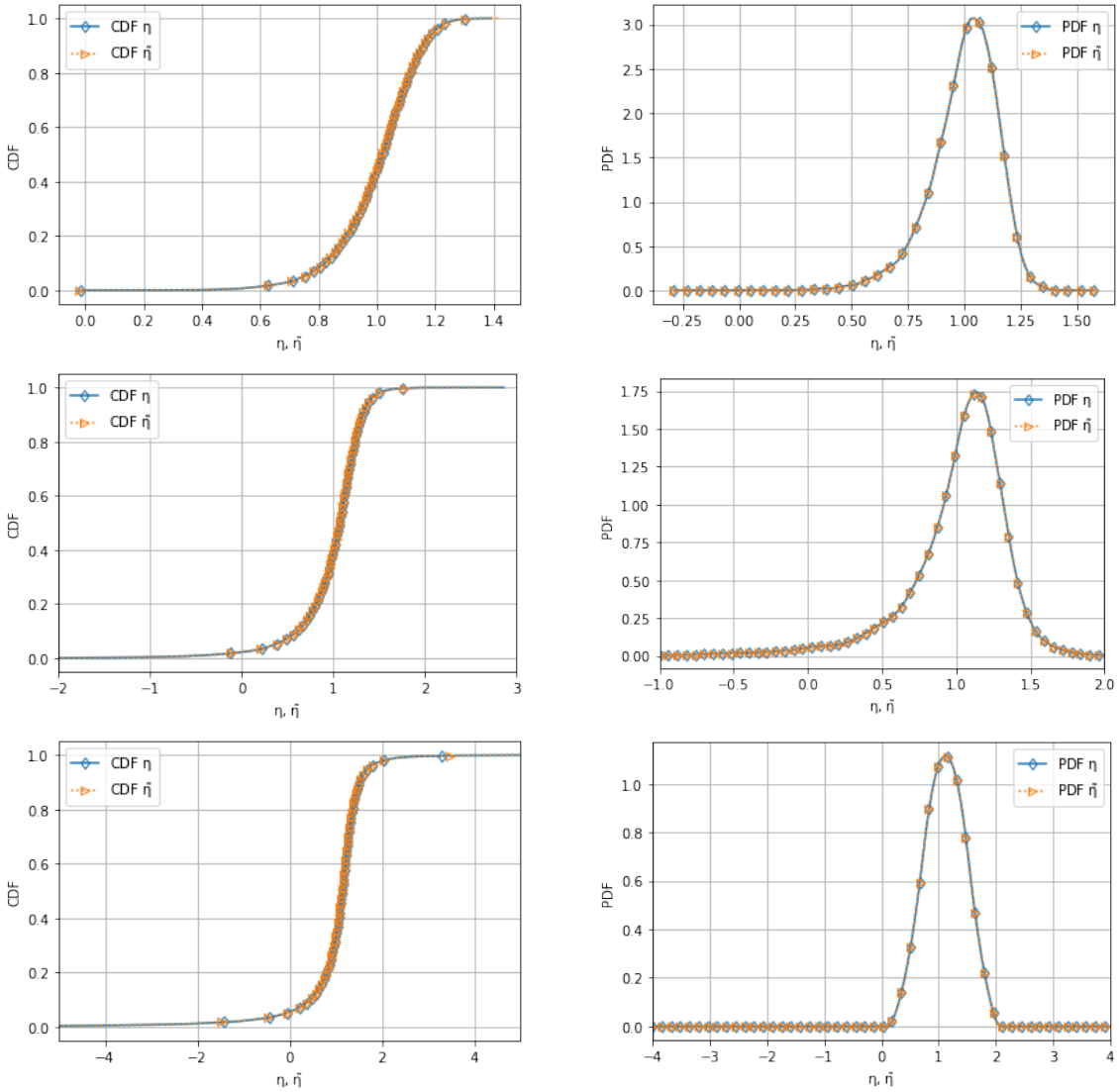

Figure 3.1: Showcase of the CDF and PDF comparison for $\eta(T)$ and $\tilde{\eta}(T)$, obtained with an MC simulation to generate 10000 paths and 1000 steps in time for both $\eta(t)$ and $\tilde{\eta}(t)$. Values for T that are shown: $T = 2$ (top), $T = 10$ (middle), $T = 30$ (bottom).

Testing normal SABR pricing functions

The main theorem of Section 3.2, i.e. Theorem 3.2.2, stated the existence and definition of functions capable of pricing options on $\eta(T)$, expressed in terms of prices of options on an auxiliary process $\xi(T)$. Combined with the results stated in Section 3.3 these functions essentially formed an approximative solution to the normal SABR model. It was briefly mentioned that exact and approximating solutions to this model have

already been showcased in earlier research, though this new approach aimed to avoid some of the difficulties some of the established methods face. Again, more information on the established methods will be given in Chapter 4 where a more detailed comparison is performed, but for now a quick glance at the accuracy and execution time of the new method is appropriate.

The following set of parameters can be used to generate prices of call options on $\eta(T)$ using the new method, Hagan's asymptotical approximation [31] and Antonov's exact approach [1]

$$\eta(0) = 0.2, \quad \alpha(0) = 0.01, \quad \rho = -0.6, \quad \nu = 0.3, \quad T = 30.$$

Figure 3.2 showcases the implied volatility curves obtained from the different models using these parameters.

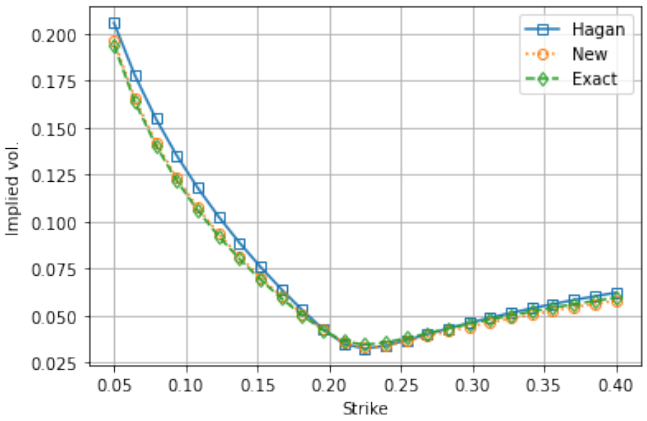

Figure 3.2: Comparison of implied volatility curves admitted by the solution to the normal SABR derived in this chapter to those admitted by the existing approximation by Hagan *et al.* [31] and the exact solution by Antonov *et al.* [1].

From a first glance at Figure 3.2, the following can remarks can be made. The figure above supports the statement that the new solution does not suffer from inaccuracies when pricing options with high maturities like Hagan's approximation does. Moreover, production of the above implied volatility curve using the new method was done within a single second of execution time, while the pricing procedure with Antonov's approach took over three minutes. This was clearly expected as Antonov's approach requires evaluation of multiple double integrals while the new method only requires a few calls to the standard normal CDF.

Looking more critically at Figure 3.2, it shows that the new solution starts to deviate from Antonov's exact approach in the area of strikes significantly larger than the ATM strike. Recall from Lemma 3.3.1 that an error term came up in the derivation of functions for pricing option on $\xi(T)$. Obviously, as prices for options on $\eta(T)$ are expressed in terms of prices of options on $\xi(T)$, this error inevitably plays a role in the final output of the solution. Luckily, it shows that the inaccuracies are only minor, even when pricing options with a relatively high maturity of 30 years. Finally, additional showcases of the inaccuracy of the output and other technicalities, e.g. absence of arbitrage, calibration etc., will be discussed in Chapter 4.

Heuristic integration interval truncation

Recall from Theorem 3.1.1 in Section 3.1 that pricing functions for options on $F(T)$ were obtained and expressed in terms of functions of the stochastic volatility component $\eta(T)$. These call and put option pricing functions for options with arbitrary strike K and maturity T expressed in terms of the CDF of $\eta(T)$, denoted by Φ_η, read

$$C_F(K) = \int_K^{+\infty} (1 - \Phi_\eta(G(k) + \mu))\,\mathrm{d}k, \tag{3.54}$$

$$P_F(K) = \int_{-\infty}^{K} \Phi_\eta(G(k) + \mu)\,\mathrm{d}k, \tag{3.55}$$

where the generalized Lamperti transform was defined as

$$G(a) := \int_{F(0)}^{a} \mathbb{1}_{u\in\mathscr{X}} \frac{1}{\sigma(u)}\,\mathrm{d}u\Big|_{a=F(t)}, \tag{3.56}$$

with $\mathscr{X}$ denoting the state-space of the process $F(t)$. In order to provide the integrals in Equation (3.54) and (3.55) with finite support in a computational setting, i.e. bounded integration intervals, a heuristic yet straightforward approach for appropriate truncation can be taken.

Taking Equation (4.8) as an example, the integration interval can be truncated by making use of the following. Recall from Section 2.5 that

$$F(T) \coloneqq G^{-1}(\eta(T) - \mu). \tag{3.57}$$

Denoting the probability operator under the T forward measure by $\mathbb{P}^T$, it follows straightforwardly that for $x \in \mathscr{X}$ it holds that

$$\mathbb{P}^T(F(T) \leq x) = \mathbb{P}^T(G^{-1}(\eta(T) - \mu) \leq x), \tag{3.58}$$

which in turn yields that

$$\Phi_F(x) = \mathbb{P}^T(\eta(T) \leq G(x) + \mu), \tag{3.59}$$

$$= \Phi_\eta(G(x) + \mu), \tag{3.60}$$

where Φ_F and Φ_η denote the CDF of $F(T)$ and $\eta(T)$ under the T forward measure, respectively. The idea behind the heuristic approach to bounding the infinite integration interval is to now find a value $\hat{x}$ such that $\Phi_F(\hat{x}) = 1 - \epsilon$ for ϵ small, e.g. ϵ <1e-8. Through similar reasoning as used in Corollary 2.5.1.1 to state that $F(T)$ is associated with a valid density function, it is known that such a value for $\hat{x}$ exists by construction. An expression for $\hat{x}$ can then be obtained as follows

$$\Phi_F(\hat{x}) = \Phi_\eta(G(\hat{x}) + \mu) = 1 - \epsilon, \tag{3.61}$$

$$G(\hat{x}) + \mu = \Phi_\eta^{-1}(1 - \epsilon), \tag{3.62}$$

$$\hat{x} = G^{-1}(\Phi_\eta^{-1}(1 - \epsilon) - \mu). \tag{3.63}$$

Note that G^{-1} exists by construction (see Section 2.5), and Φ_η^{-1}, being the inverse of a valid CDF, can easily be computed numerically. Finally, applying this truncation to Equation (3.54) yields the following function for approximating prices of vanilla European call options on $F(T)$ with strike K and maturity T

$$C_F(K) \approx \int_K^{G^{-1}(\Phi_\eta^{-1}(1-\epsilon)-\mu)} (1 - \Phi_\eta(G(k) + \mu))\,\mathrm{d}k. \tag{3.64}$$

Analogously, Equation (3.55) can be approximated through

$$P_F(K) \approx \int_{G^{-1}(\Phi_\eta^{-1}(\epsilon)-\mu)}^K \Phi_\eta(G(k) + \mu)\,\mathrm{d}k. \tag{3.65}$$

It is clear that the error of this approximation, denoted by $\varepsilon_{\text{call}}$ for Equation (3.54) and ε_{put} for Equation (3.55), reads

$$\varepsilon_{\text{call}} \coloneqq \int_{G^{-1}(\Phi_\eta^{-1}(1-\epsilon)-\mu)}^{+\infty} (1 - \Phi_\eta(G(k) + \mu))\,\mathrm{d}k, \tag{3.66}$$

$$\varepsilon_{\text{put}} \coloneqq \int_{-\infty}^{G^{-1}(\Phi_\eta^{-1}(\epsilon)-\mu)} \Phi_\eta(G(k) + \mu)\,\mathrm{d}k. \tag{3.67}$$

Important to note is that these error terms are essentially the effects of extreme tail events on the prices of options on $F(T)$. In other words, only the impact of realizations of $F(T)$ occurring with a probability of less than ϵ is excluded from the option prices. As discussed by Kotzé *et al.* [36] and Bliss-Panigirtzoglou [14], excluding these extreme tail events is, due to their scarcity, unlikely to have a significant impact on the prices of options with strikes within the strike range as observable in the market. Since pricing options that are very far in-the-money or out-of-the-money is not a focus in the thesis, it can for now be safely assumed that Equation (3.54) and (3.55) form a sufficiently accurate approximation to Equation (3.64) and (3.65), respectively. Nevertheless, the upcoming chapter will make extensive use these functions, which in turn will further highlight the significance of this approximation in a more practical setting.

4

Numerical experiments and calibration

In the preceding chapter, functions were derived that can be used for approximating prices of vanilla European options on an underlying $F(T)$ with dynamics described by the general SLV model

$$\mathrm{d}F(t) = \alpha(t)\sigma(F(t))\,\mathrm{d}W_F^T(t), \tag{4.1}$$

$$\mathrm{d}\alpha(t) = \nu\alpha(t)\,\mathrm{d}W_\alpha^T(t), \tag{4.2}$$

where $W_F^T(t)$ and $W_\alpha^T(t)$ are Brownian motions under the T forward measure, $\mathrm{d}W_F^T(t)\,\mathrm{d}W_\alpha^T(t) = \rho\,\mathrm{d}t$, and the same conditions as before hold for the volatility-of-volatility parameter ν and the correlation ρ. Going forward, these pricing functions will be referred to as the *Transform solution*. Moreover, the earlier made assumption for the risk-free rate r, i.e. that $r = 0$, will be used for simplicity once more.

This chapter will focus on performing a number of numerical experiments that will put the robustness and quality of the Transform solution to the test. These experiments include the following.

- Setting $\sigma(F(t)) = 1$ and comparing output of the Transform solution to that following from the Hagan *et al.* [31] asymptotical approximation and the Antonov *et al.* [1] exact solution to the normal SABR model [8] to verify if its output for basic LV specifications is competitive to already established models (see Section 4.2).
- Examining model assumptions first discussed in Section 2.5 of Chapter 2 by showcasing their impact on model output and discussing the significance of this impact (see Section 4.3).
- Using the fact that the Transform solution is defined for general σ specifications satisfying minor requirements to specify a custom flexible LV function, and verifying that the Transform solution when combined with this LV function also enjoys additional flexibility (see Section 4.4).
- Repeating the above procedure using a non-parametric LV specification based on the famous LV models proposed by Dupire [24] and Derman-Kani [22] (see Section 4.6).

Besides these experiments all divided over individual sections, a powerful method for calibration of the Transform solution will be discussed. This calibration is essentially a two-step process, since the Transform solution requires both preliminary and intermediary calibration of two parameters present in the model. While this discussion on the main calibration algorithm is done in the penultimate section of this chapter, the approach for the preliminary calibration, as well as a small introduction into Hagan's approximating formula and Antonov's exact approach, is discussed in the upcoming section

4.1. Approach

While the general SLV model stated above sees a multitude of applications, in the field of financial modeling it is typically used for pricing interest rate derivatives. The typical interest rate derivatives are interest rate caps and floors, swaps and swaptions. Exact definitions of these products are left out of this thesis as they are not important for the tasks at hand and additional details can easily be found anywhere. Instead, this discusses the set-up required for working with the Transform solution, as well as some preliminaries on other models typically used for the same purpose of pricing interest rate derivatives.

Asymptotical approximation

As mentioned earlier, one of the models that will be used for comparisons is the one proposed in the paper that originally introduced the SABR model. The main result of this 2002 paper by Hagan *et al.* [31], was the set of analytical approximation formulae for the lognormal implied volatility of options priced with the BS model [13], obtained through application of perturbation techniques. The formulae were easy to implement and vanilla European options could easily be priced by plugging the resulting implied volatility back into the standard BS model.

According to Hagan *et al.* [31], given a forward price f and strike K, the implied volatility of a European option with maturity T can be approximated with

$$\hat{\sigma}_{\text{Hagan}}(f,K) \approx \frac{\alpha(0)}{(fK)^{(1-\beta)/2}\left(1+\frac{(1-\beta)^2}{24}\log^2(f/K)+\frac{(1-\beta)^4}{1920}\log^4(f/K)\right)}\left(\frac{z}{x(z)}\right)\cdot \left(1+\left(\frac{(1-\beta)^2}{24}\frac{\alpha(0)^2}{(fK)^{1-\beta}}+\frac{1}{4}\frac{\rho\beta\nu\alpha(0)}{(fK)^{(1-\beta)/2}}+\frac{2-3\rho^2}{24}\nu^2\right)T\right), \quad (4.3)$$

where

$$z := \frac{\nu}{\alpha(0)}(fK)^{(1-\beta)/2}\log(f/K),$$
$$x(z) := \log\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right).$$

Note that for the at-the-money (ATM) implied volatility, Equation (4.3) reduces significantly.

Unfortunately, there are two considerable issues with Hagan's approximating formula for the implied volatility. Firstly, Equation (4.3) contains numerous divisions by the strike K, which means that its output will start to break down as K approaches zero. The deterioration around zero means that the density of the underlying following from option prices computed with Hagan's formula, can take on negative values near zero. In turn, this means that the prices obtained with Equation (4.3) might not be arbitrage-free. Luckily, this problem can be avoided with certain refinements of the formula (see for example Berestycki *et al.* [12]), e.g. a shift introduced to the underlying rate. Secondly, Equation (4.3) is an asymptotical approximation in time for the implied volatility. This means that, unless extra terms are taken into account, the approximation is likely to lose accuracy when pricing options with high maturities. The severity of this breakdown is showcased in the next section, where implied volatilities are computed for options with maturities up to 50 years.

Antonov's exact approach

The second model used for comparisons is actually an exact solution to the SABR model for specific β and ρ values. Antonov *et al.* [4], discuss a number of parameter sets for which exact analytical pricing functions can be derived from the SABR model, followed by detailed derivations of said functions. The case considered in this thesis is the free-boundary normal ($\beta = 0$) with general correlation. In case one is interested in using CIR type LV functions [33], i.e. $\beta = \frac{1}{2}$, the free-boundary normal formulas can be used as a control variate. Note that Antonov *et al.* [4] also derive explicit formulas for the purely lognormal case, but set the requirement of zero correlation. Zero correlation is an unusual occurrence in most situations where the SABR model is typically used, hence this case is omitted from the tests.

The system of formulas following from the numerous papers of Antonov *et al.* [1–4], is very extensive and requires great attention to detail with its implementation. The formulas are merely used for comparison of their output, and thus the large system of equations is omitted from this thesis. For a detailed explanation and showcase of the approach implemented in this thesis, the reader is referred to Section 3 and 4 in Antonov *et al.* [1].

Preliminary calibration

So far, prices of call and put options on $F(T)$ with strike K and maturity T following from the Transform solution have been denoted by $C_F(K)$ and $P_F(K)$, respectively. However, when discussing the preliminary calibration procedure necessary for calibrating μ and $\alpha(0)$, it will be useful to adopt slightly different notation to improve the clarity of the description of said procedure. In light of this, this subsection adopts the notation $C_F(K) \equiv C_F(K,\alpha(0),\mu)$ and $P_F(K) \equiv P_F(K,\alpha(0),\mu)$.

It was briefly discussed in both Chapter 2 and the introduction to this chapter that the parameter μ, which arose from application of the Lamperti transformation, has to be calibrated in order to maintain preservation of the forward. Recall that this can be done by finding the value for μ such that the put-call parity holds for the ATM put and call option, i.e. such that

$$C_F(F(0),\alpha(0),\mu) = P_F(F(0),\alpha(0),\mu). \tag{4.4}$$

Simultaneously, the initial value for the stochastic volatility (SV) process, $\alpha(0)$, has to be calibrated to appropriately reprice the ATM options as seen in the market. This can be done by making use of the formula for the value of the ATM straddle[1] as discussed by Brenner-Subrahmanyam [16], which reads

$$C_F(F(0),\alpha(0),\mu) + P_F(F(0),\alpha(0),\mu) = \hat{\sigma}_{\text{market}}(F(0),F(0))F(0)\sqrt{2T/\pi}, \tag{4.5}$$

where $\hat{\sigma}_{\text{market}}(F(0),F(0))$ is the implied volatility of the ATM option with maturity T and underlying $F(T)$, as shown in the market.

It becomes clear that simultaneously calibrating μ and $\alpha(0)$ reduces to a two-dimensional root-finding problem, which can be seen as finding a solution to the system

$$C_F(F(0),\alpha(0),\mu) - P_F(F(0),\alpha(0),\mu) = 0, \tag{4.6}$$

$$C_F(F(0),\alpha(0),\mu) + P_F(F(0),\alpha(0),\mu) = \hat{\sigma}_{\text{market}}(F(0),F(0))F(0)\sqrt{\frac{2T}{\pi}}, \tag{4.7}$$

in μ and $\alpha(0)$. Note that in this two-dimensional system, Equation (4.6) is responsible for obtaining the correct value for μ, while Equation (4.7) takes care of $\alpha(0)$. There exist numerous methods for solving this problem, but in this thesis a modified Powell's method [48] is used. Rather than writing the necessary code for implementation of this method, the `scipy.optimize.root`[2] function is used with the kwarg `method='hybr'`[3]. The function takes as input a vector function and an initial guess, and outputs a special object containing the results of the algorithm. Provided that the system of equations described by Equation (4.6) and (4.7) is sufficiently stable, the initial parameter guesses need not be very accurate.

While μ is a parameter unique to the Transform solution, $\alpha(0)$ is also present in Hagan's asymptotical approximation and Antonov's exact approach. Analogous to the Transform solution, Equation (4.7) can be used directly to calibrate $\alpha(0)$ for Antonov's exact approach, provided that the call and put option prices following from this model are substituted in place of $C_F(F(0),\alpha(0),\mu)$ and $P_F(F(0),\alpha(0),\mu)$. Hagan's approximation formulae output option prices directly in terms of implied volatility, which means that in order to calibrate $\alpha(0)$ appropriately the simple constraint $\hat{\sigma}_{\text{Hagan}}(F(0),F(0)) = \hat{\sigma}_{\text{market}}(F(0),F(0))$ can be used, where $\hat{\sigma}_{\text{market}}(F(0),F(0))$ is the implied volatility of the ATM option as shown in the market.

4.2. Output figures

It is now time to put the Transform solution to the test by visualizing its accuracy through various figures. Since Antonov's approach yields an exact solution for the free-boundary ($\beta = 0$) case, it makes for an excellent benchmark for comparisons of different methods. Recall that setting $\beta = 0$ means the underlying SABR model is reduced to its purely stochastic-normal form, i.e.

$$\begin{aligned} \mathrm{d}F(t) &= \alpha(t)\,\mathrm{d}W_F^T(t), \\ \mathrm{d}\alpha(t) &= \nu\alpha(t)\,\mathrm{d}W_\alpha^T(t). \end{aligned}$$

As was discussed earlier, a great benefit of using the Transform solution is its compatibility with customized LV functions. However, most established models do not share this quality, which means that it would be hard to compare the Transform solution to other models when selecting a custom LV function. For this reason, and to nicely compare to the benchmark provided by Antonov's approach, a constant LV function of $\sigma(F(t)) = 1$ is selected. Since the Transform solution was derived from a generalized SLV model under the T forward measure, this comparison is straightforward. Recall that the derived approximative pricing functions for options on $F(T)$ with maturity T and strike K, last discussed in the final subsection of Section 3.4 read

$$C_F(K) = \int_K^{+\infty} (1-\Phi_\eta(G(k)+\mu))\,\mathrm{d}k \approx \int_K^{G^{-1}(\Phi_\eta^{-1}(1-\epsilon)-\mu)} (1-\Phi_\eta(G(k)+\mu))\,\mathrm{d}k, \tag{4.8}$$

[1] A straddle is a financial product comprised of a long call and long put option with equal strike and maturity, on the same underlying.
[2] Documentation: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.root.html`.
[3] Documentation: `https://docs.scipy.org/doc/scipy/reference/optimize.root-hybr.html#optimize-root-hybr`.

for call options, while for put options it holds that

$$P_F(K) = \int_{-\infty}^{K} \Phi_\eta(G(k)+\mu)\,\mathrm{d}k \approx \int_{G^{-1}(\Phi_\eta^{-1}(\epsilon)-\mu)}^{K} \Phi_\eta(G(k)+\mu)\,\mathrm{d}k, \tag{4.9}$$

where ϵ is a small positive constant, μ is a constant calibrated to maintain preservation of the forward and, for $\mathscr{X}$ state-space of $F(t)$, G is the generalized Lamperti transform

$$G(a) := \int_{F(0)}^{a} \mathbb{1}_{u\in\mathscr{X}} \frac{1}{\sigma(u)}\,\mathrm{d}u\bigg|_{a=F(t)}. \tag{4.10}$$

For simulation of the results a relatively atypical set of SABR model parameters is used, i.e. high positive correlation ρ and high volatility-of-volatility ν. This is done as a way to both test the robustness of the model, as well as to see if its output remains competitive even in extreme market conditions. For varying maturities T, this parameter set is

$$\rho = 0.7, \quad \nu = 0.4, \quad F(0) = 0.13, \quad \hat{\sigma}_{\text{market}}(F(0), F(0)) = 4\%. \tag{4.11}$$

Using this parameter set and performing all the aforementioned steps for appropriate calibration of the necessary model parameters, i.e. $\alpha(0)$ and μ, results in the implied volatility curves showcased in Figure 4.1. Note that for evaluation of Equation (4.8) and(4.9) in this experiment, it was set that ϵ = 10e-6.

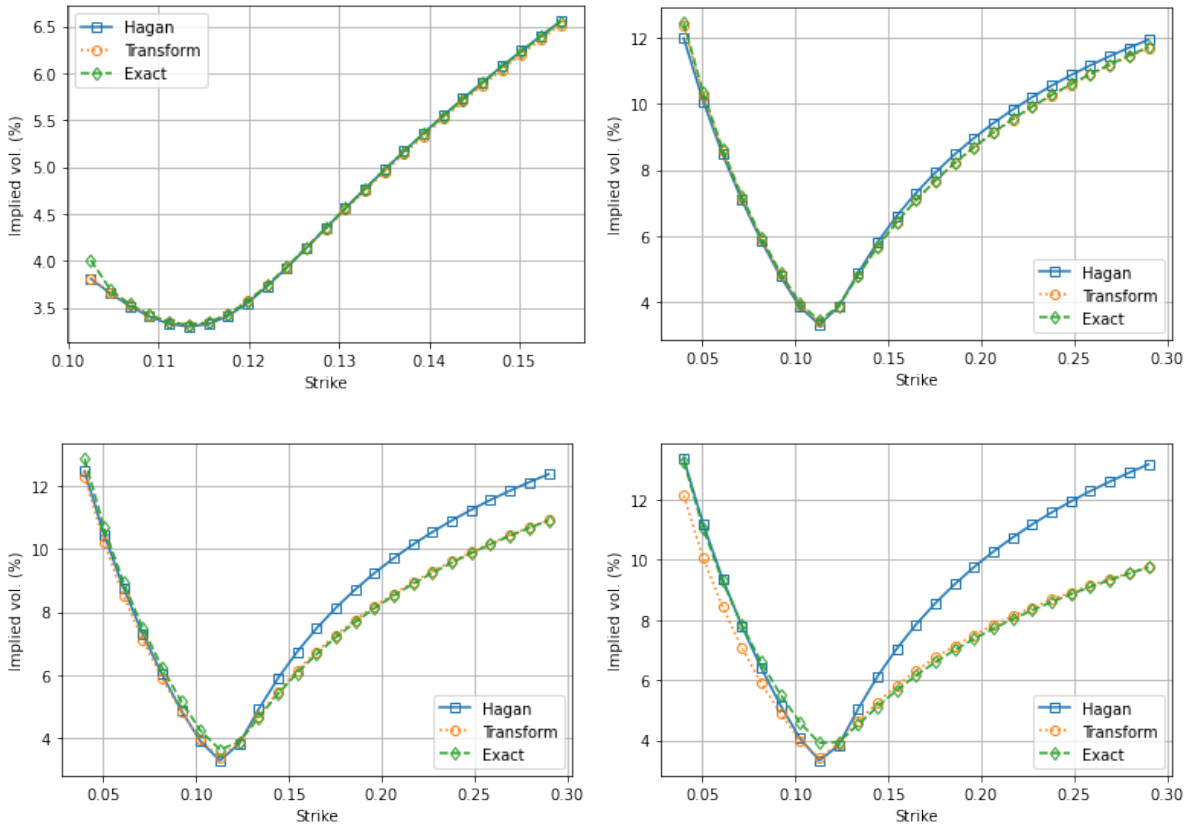

Figure 4.1: Implied volatility curve comparison for the normal SABR model, obtained from the Transform solution, Hagan's asymptotical approximation and Antonov's exact approach, for $T = 0.5$ (top-left), $T = 5$ (top-right), $T = 20$ (bottom-left) and $T = 50$ (bottom-right).

In line with results briefly showcased in Section 3.4, it turns out that the Transform solution is indeed highly competitive for pricing options with maturities up to 50 years, even when subjected to atypical values for the model parameters. Moreover, it turns out that the truncation of the integration intervals in Equation (4.8) and (4.9), according to the approach described in the final subsection of Section 3.4, did not cause significant inaccuracies. However, this experiment is not enough to be conclusive regarding the validity of this approach, hence further tests will be done in the upcoming sections of this chapter.

In order to judge the results shown in Figure 4.1 in a more quantitative fashion, Table 4.1 has been constructed. For four strikes and each of the four maturities, this table shows the differences between the prices of the options priced with the Transform solution or Hagan's approximation formula, and the exact prices obtained with Antonov's approach. The table adopts the notation

$$\varepsilon_{\text{transform}} := |\hat{\sigma}_{\text{market}} - \hat{\sigma}_{\text{transform}}|,$$

and

$$\varepsilon_{\text{Hagan}} := \left|\hat{\sigma}_{\text{market}} - \hat{\sigma}_{\text{Hagan}}\right|.$$

T	K	$\hat{\sigma}_{\text{market}}(\%)$	$\varepsilon_{\text{transform}}(\%)$	$\varepsilon_{\text{Hagan}}(\%)$
0.5	0.0068	3.53	0.013	0.021
	0.018	3.43	0.0059	0.017
	0.039	5.36	0.026	0.0032
	0.050	6.23	0.030	0.0067
5	-0.039	8.64	0.078	0.17
	0.013	3.41	0.049	0.12
	0.12	9.55	0.020	0.29
	0.17	11.20	0.031	0.54
20	-0.039	8.96	0.43	0.18
	0.013	3.65	0.25	0.34
	0.12	8.90	0.051	1.24
	0.17	10.41	0.022	1.43
50	-0.039	9.29	0.85	0.089
	0.013	3.91	0.50	0.59
	0.12	8.03	0.10	2.72
	0.17	9.33	0.037	3.28

Table 4.1: Differences between option prices expressed in implied volatility terms priced with the Transform solution or Hagan's approximation formula, and the exact prices obtained with Antonov's approach. Implied volatility curves corresponding with these results are shown in Figure 4.1.

Finally, it is important to note that, without considering the preliminary calibration of μ and $\alpha(0)$, production of an implied volatility curve for one of the maturities shown in Figure 4.1 took approximately one second. This is in contrast to Hagan's approximation formula, which can output a similar curve in approximately one millisecond. This is because the Transform solution, while requiring only few evaluations of the standard normal CDF, also contains nested integrals in the pricing for options on $F(T)$. As discussed by Bang [9], this procedure can be sped up by computing some of these integrals on a fine grid beforehand and caching the results, though such optimization steps have been omitted in this test setting.

4.3. Examining model assumptions

Recall from Section 2.5 that, in order to make the derivation of approximative pricing functions for vanilla European options on $F(T)$ easier, the practical assumption that $\mu(T) = \mu$, where

$$\mu(T) := \eta(0) + \frac{1}{2}\int_0^T \alpha^2(t)\frac{\partial\sigma(G^{-1}(g(t)))}{\partial F}\,\mathrm{d}t, \tag{4.12}$$

and μ a constant calibrated such that the time zero forward is preserved. Moreover, G is defined as in Equation (4.10), $g(t) := G(F(t))$ and $\eta(0)$ is the initial value of an auxiliary process that was introduced to apply the Lamperti transform appropriately. It was then shown that this assumption affects the distribution of $F(T)$ only mildly. Now that pricing functions for options on $F(T)$ have been derived (the Transform solution), it is possible to examine the impact of this assumption on model output.

The numerical experiment that will be used to check the significance of this assumption is set up as follows.

- Use an MC framework to simulate 20000 paths with 1000 time-steps for $F(t)$ by discretizing the dynamics as in Equation (4.1) according to the Euler-Marayuma scheme (see Oosterlee-Grzelak [47]).

- Compute prices of vanilla European call options on $F(T)$ for various K and T using the simulated paths.
- Compute option prices for various K and T with the Transform solution.
- Compare the two sets of prices through implied volatility curves and the implied density of $F(T)$ admitted by each set.

This experiment will be repeated for three different LV specifications, appropriately showcasing the impact of the assuming $\mu(T) = \mu$ in various scenarios. It has been verified in Section 2.5 that this assumption is justified for low T, hence the experiments in this section focus mainly on highlighting the impact of this assumption on implied volatility curve shapes for high T. The set of model parameters used for the MC simulation is

$$F(0) = 0.13, \quad \alpha(0) = 0.008, \quad \rho = -0.6, \quad \nu = 0.3, \quad T \in \{10, 30\}.$$

The Transform solution will be evaluated using equal parameters, with the exception of μ and $\alpha(0)$, which will be obtained through the preliminary calibration procedure discussed in the second subsection of Section 4.1, using the ATM option price as generated by the MC simulation. This preliminary calibration will be performed for each of the considered T and LV specifications. For evaluation of the approximative pricing functions, i.e. Equation (4.8) and (4.9), this section will set ϵ = 10e-8. The idea is that this small value for ϵ will allow for highlighting mostly the errors coming from setting $\mu(T) = \mu$, with little additional noise coming from the error of truncating the pricing function integrals as discussed in the final subsection of Section 3.4.

Standard case

Starting off, the basic case can be considered. Let $\sigma(F(t)) = 1$, such that the general SLV model denoted by Equation (4.1) and (4.2) reduces to the normal SABR model. Using the experiment described earlier, Figure 4.2 can be produced. Note that the figure adopts the abbreviation IV for implied volatility.

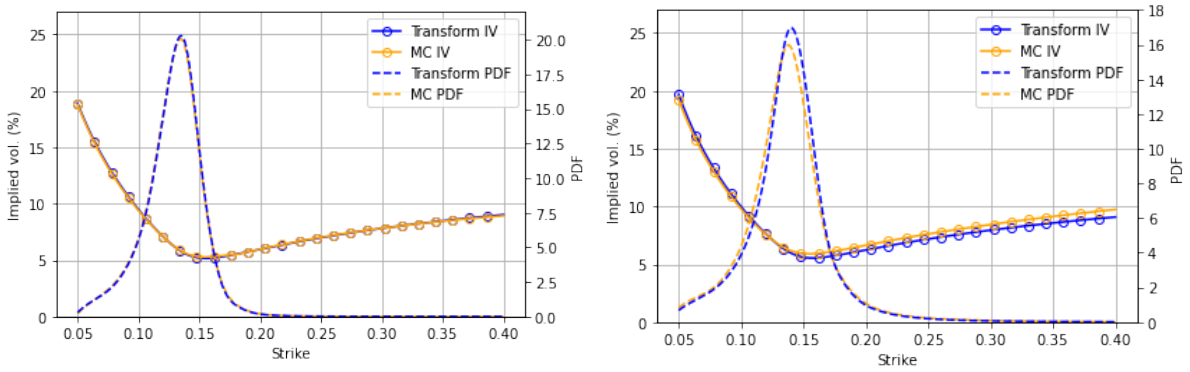

Figure 4.2: Implied volatility curve and implied density of $F(T)$ comparison constructed using the experiment described in the introduction of this section, for $T = 10$ (left) and $T = 30$ (right) and $\sigma(F(t)) = 1$.

As expected, judging from the form of Equation (4.12) when combined with the selected σ, the impact of the assumption that $\mu(T) = \mu$ is low in this case. This is mainly caused by the fact that the stochastic component in the definition of $\mu(T)$ becomes zero, as σ is set to a constant. In the $T = 30$ case, the implied volatility curves start to differ slightly in the area of strikes far above the ATM strike, which is in line once more with the result of the experiment discussed in Section 2.5.

Alternative cases

The second case that will be considered is one in which $\sigma(F(t)) := F^{0.01}(t)$. Recall, it was extensively discussed in Section 2.1 of Chapter 2 that application of the transform is only possible if the LV component is continuously differentiable and strictly positive on the entire state-space of $F(t)$. This constraint is clearly violated with the selected LV specification, hence it must be slightly modified to become $\sigma(F(t)) := F^{0.01}(t) + \delta$, where δ is a small positive constant, e.g. δ = 10e-8, that will shift the LV function up and away from the singularity at zero. Important to note is that this shift is merely an *ad hoc* numerical fix, which, since it is only small, has an

effect on the behavior of the original function that will be close to negligible in practical applications. Once more following the steps described in the introduction to this section yields Figure 4.3.

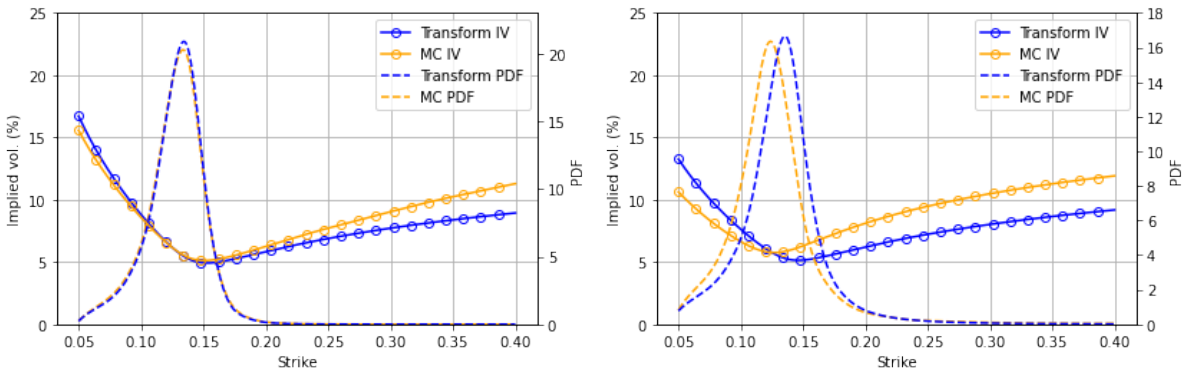

Figure 4.3: Implied volatility curve and implied density of $F(T)$ comparison constructed using the experiment described in the introduction of this section, for $T = 10$ (left) and $T = 30$ (right) and $\sigma(F(t)) := F^{0.01}(t) + \delta$, where δ is a small positive constant, e.g. $\delta = 10e-8$.

As it turns out, for this LV specification, the assumption that $\mu(T) = \mu$ has a significant impact on both the implied volatility curve shape and the implied density of $F(T)$. Recall, however, that this assumption was made mostly for practical reasons, and its impact must therefore be judged from a practical standpoint. In practice, the model is generally first calibrated to market prices. During this procedure, the LV and model parameters are adjusted until a good fit to the market has been obtained, therewith, as discussed in Section 2.5, essentially absorbing some of the inaccuracies stemming from assuming that $\mu(T) = \mu$. However, this need not lead to model dynamics similar to those of the market, thus resulting in a model that requires regular recalibration. It becomes clear that one must anticipate small inaccuracies in the model output, and that the calibrated model must be handled with care. Section 4.4 discusses an alternative approach, in which a custom LV function is designed that allows for quick adjustments to be made to the shape of the implied volatility curve, omitting the need for regular recalibration of the model.

In order to abolish the outside chance that the shift applied to the LV function is what caused the major inaccuracies showcased in Figure 4.3, the experiment can be performed once more, though in this case with a purely lognormal LV specification. To this end, consider $\sigma(F(t)) := F(t)$, to which an ad hoc shift for prevention of numerical issues in the Lamperti transform need not be applied. Results of running the earlier described experiment once more, yields Figure 4.4.

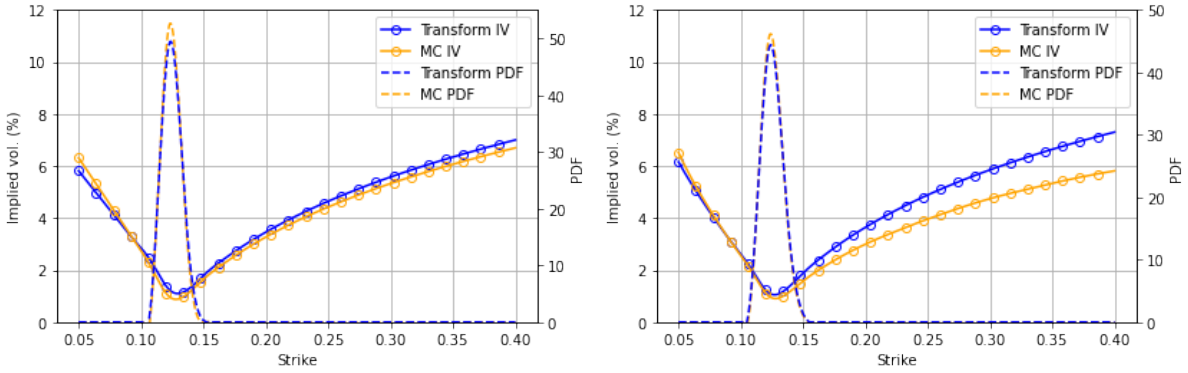

Figure 4.4: Implied volatility curve and implied density of $F(T)$ comparison constructed using the experiment described in the introduction of this section, for $T = 10$ (left) and $T = 30$ (right) and $\sigma(F(t)) := F(t)$.

As expected, similar statements as for the previous LV specification can be made regarding the differences between the two implied volatility curves in the region of strikes far above the ATM strike. In conclusion, while assuming that $\mu(T) = \mu$ helped significantly in both the derivation and the implementation of the Transform

solution, it's impact on model output cannot be ignored. In the upcoming sections of this chapter, various other experiments will be performed, during which the Transform solution will be calibrated to market prices. In turn, this will highlight the impact of letting $\mu(T) = \mu$ in a more practical setting.

4.4. Custom parametric LV

Traders at trading-desks often have an opinion on market conditions and require the possibility to translate this opinion into their pricing model. This is typically done through fine-tuning the LV function in the model by changing parameters that affect the shape of the implied volatility curve obtained with said model. Due to application of the Lamperti transform, the Transform solution is compatible with any arbitrary LV specification meeting basic requirements (see Section 2.5). It is therefore an interesting endeavor to design a parametric LV function that can provide the Transform solution with valuable flexibility. In this setting, flexibility of the model is determined by the control the model user has over isolated parts of the output implied volatility curve. Flexible models are valuable for traders as they omit the necessity for full recalibration when market conditions change only slightly. This section first focuses on the designing process of a custom LV function, after which the behavior of the Transform solution when combined with this function is showcased.

Constructing the function

Recall that in the general SLV model used throughout this thesis

$$\mathrm{d}F(t) = \alpha(t)\sigma(F(t))\,\mathrm{d}W_F^T(t), \tag{4.13}$$

$$\mathrm{d}\alpha(t) = \nu\alpha(t)\,\mathrm{d}W_\alpha^T(t), \tag{4.14}$$

the LV component is encapsulated in the function σ. In this setting, the designing process of the LV function is done with four main requirements in mind to ensure sufficient model flexibility and control over output implied volatility curves. These requirements are as follows.

- The LV curve must not contain singularities, ensuring that the process $F(t)$ can take on both positive and negative values without having to introduce a shift parameter.
- The final option pricing functions must be flexible enough to be calibrated to a large set of options without being over-parameterized.
- The backbone of the distribution of the underlying must change when its level and volatility change.
- Control must be maintained over isolated sections of the output implied volatility curve.

It is possible to set many more constraints, but this will ultimately depend on trading-desk specific preferences. This section aims to provide an exemplary function that can be used when pricing interest rate derivatives, showcasing the possibilities one has in designing a custom LV function, while simultaneously satisfying the aforementioned requirements.

The first requirement can be met in numerous ways, some of which are rather trivial. For example, a constant LV function $\sigma(F(t)) = 1$ could be selected, reducing the general SLV model to the purely stochastic normal SABR model [31] and thus ensuring that the distribution of $F(t)$ allows for negative values. However, with such a selection the model user will not have control over the shape of the implied volatility curve through the LV curve. In order to both maintain control over curve shapes as well as allowing $F(t)$ to take on negative values, the following approximation can be used

$$x \approx \ln(1 + \mathrm{e}^x), \quad x \in [3, +\infty). \tag{4.15}$$

To allow the user of the model to manipulate the domain in which this approximation is sufficiently accurate, a smoothing parameter can be introduced. This results in the first component of the custom LV function

$$g(f, a) \coloneqq a\ln(1 + \mathrm{e}^{f/a}), \quad a > 0. \tag{4.16}$$

By controlling the domain in which Equation (4.15) is accurate, the parameter a essentially manipulates the slope of the admitted curve at the origin. This translates to an effect showcased in Figure 4.5.

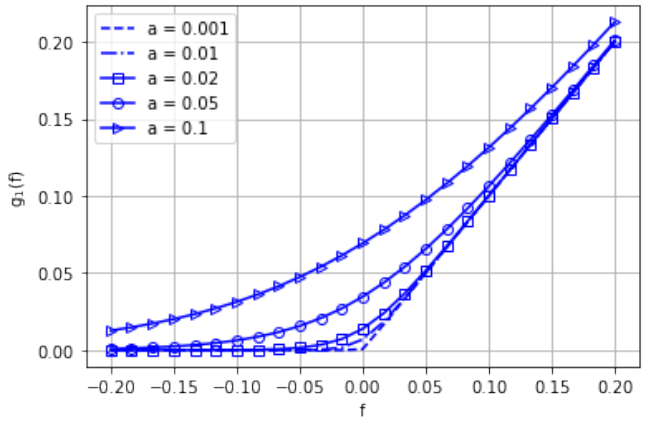

Figure 4.5: Shape of the curve admitted by Equation (4.16) for different values for the smoothing parameter.

The figure above indeed shows that using only Equation (4.16) as the LV function would enable $F(t)$, described by the dynamics shown in Equation (4.13), to take on negative values without issues, as it remains well-defined on the entire real line. However, flexibility in the shape of the curve is lacking and must be taken care of by meeting the remaining requirements.

The second requirement will be automatically met by not introducing too many parameters that must be calibrated, and making sure that those that do get introduced have a significant, non-overlapping effect on the shape of the resulting curve. In light of this, and to simultaneously cover the third requirement, a level-dependent power specification can be built onto Equation (4.16). This component will ensure that the backbone of the distribution of $F(t)$, i.e. the path that the ATM implied volatility traverses as the ATM value changes, is dependent on the level of $F(t)$. Recall that for the standard SABR model, with LV specification $\sigma(F(t)) := F^{\beta}(t)$, it holds that $\beta \in [0,1]$. Keeping this in mind, and aiming to avoid complex distributions for $F(t)$ with $t \in [0,T]$, it is a natural decision to select a level-dependent power that varies between 0 and 1. The most straightforward function that possesses this property is the sigmoid function $y : \mathbb{R} \to [0,1]$ where

$$y(x) := \frac{1}{1+e^{-x}}. \tag{4.17}$$

In order to construct a natural connection between the speed at which y changes value around the origin and the volatility of the process $F(t)$, the function input can be scaled by $\alpha(0)$. This will essentially ensure that the most relevant part of the curve admitted by y is squeezed into an interval appropriate for the process $F(t)$. Enhancing y with this scaling parameter results in the following specification for the level-dependent power

$$h \equiv h(f) := \frac{1}{1+e^{-f/\alpha(0)}}, \tag{4.18}$$

which, for various values for $\alpha(0)$, admits the curves shown in Figure 4.6.

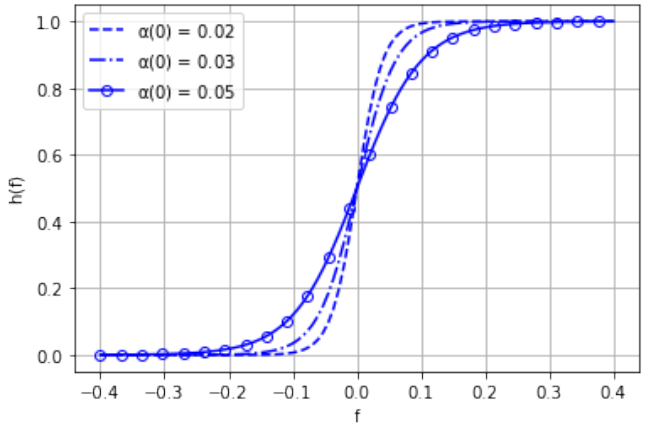

Figure 4.6: Shape of the curve admitted by the sigmoid function specified in Equation (4.18) for varying values of the scaling parameter $\alpha(0)$.

Having established Equation (4.16) and (4.18), it is possible to make use of their variable input and construct an LV function which allows the model user to have isolated control over the shape of the LV curve in the range of strikes above the ATM strike. This function reads

$$\sigma(f) := g^h(f, a_1) + bg(f - Z, a_2), \tag{4.19}$$

where Z is a strike fixed at a value larger than the ATM strike around which the LV curve can be controlled using $b \in \mathbb{R}$, and $a_1, a_2 > 0$ are two smoothing parameters. Prior to testing the behavior of the function above, the following points can be made regarding the reasoning behind its definition.

- The final LV function consists of two terms, where, for a_2 sufficiently small, the effect of the second term only becomes significant right before and beyond Z. To ensure that this remains the case, a_2 can be fixed at a small value such that the LV after Z, i.e. the upside, can be controlled by b.
- Judging from the effect of a_1 on the shape of the curve admitted by g showcased in Figure 4.5, the shape of the LV curve in the area of strikes below the ATM strike, i.e. the downside, can be controlled by this parameter.
- Construction of the LV function using two terms should (provided a_1 and a_2 are small) allow the model user to alter isolated sections of the curve, e.g. the upside can be controlled using b without affecting curve shapes in the downside. This ensures that the earlier discussed fourth requirement is also met.

In the following subsection these points will be verified through multiple numerical experiments. Moreover, while flexibility of the LV function is desirable, option prices are generally shown in implied volatility terms, hence it is important that these changes in the LV curve propagate appropriately to the shape of the implied volatility curve. This will also be tested in the next subsection.

Flexibility showcase

It will be useful to first show the effects of the flexible parameters in Equation (4.19), i.e. a_1 and b, on the shape of the LV curve, before moving on to showcasing their effects on the shape of the implied volatility curve admitted by the Transform solution when combined with this custom LV function. Figure 4.7 showcases the effects of varying a_1 and b on the shape of the LV curve.

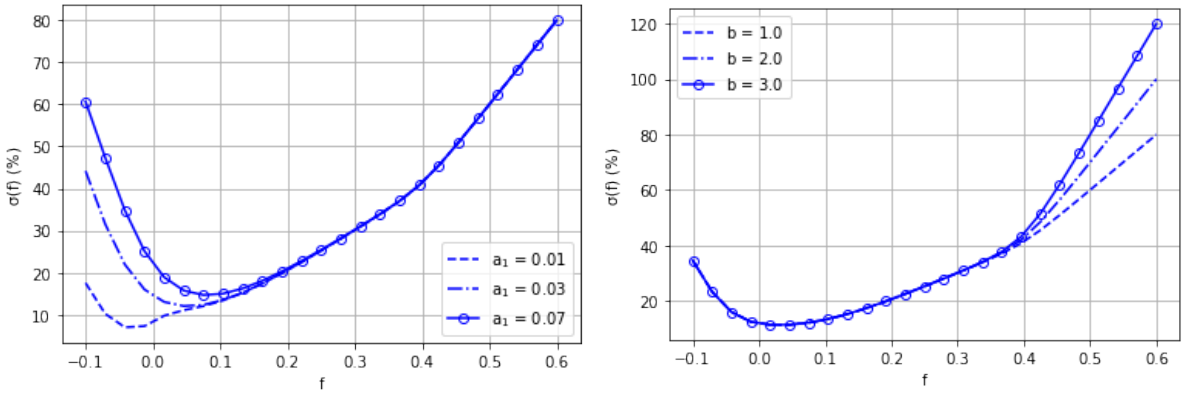

Figure 4.7: Showcase of the LV curve admitted by Equation (4.19) for different parameter sets. Fixed parameters used: $F(0) = 0.11, Z = 0.4, a_2 = 0.02$.

The figures above show that the desired effect of obtaining control over isolated areas of the LV curve has been successfully obtained. Indeed, the parameter a_1 is in control of the shape of the curve in the downside while b controls the upside. However, LV and implied volatility do not map one-to-one, hence these effects might not be similar on the shape of an implied volatility curve. In order to verify this, let the LV function in the Transform solution be governed by Equation (4.19). Using the parameter set

$$F(0) = 0.11, \quad \alpha(0) = 0.05, \quad \nu = 0.4, \quad \rho = -0.7, \quad Z = 0.4, \quad a_2 = 0.02 \quad T = 10,$$

the following figures show the effect of a_1 and b on the shape of the admitted implied volatility curve. Note that for the production of these figures the Transform solution parameter μ is calibrated such that the time-zero forward is preserved, i.e. such that $\mathbb{E}^T[F(T)] = F(0)$ (see Section 4.1). Since there is no set of comparable market prices in this case, the approach to truncation of the integrals in Equation (4.8) and (4.9), discussed in the final subsection of Section 3.4, cannot be used. Considering that this is a test setting, a simple method is to just select a rather large integration interval, e.g. $[-5,5]$, and numerically verify if it is large enough, i.e. if it captures a large enough section of the density of $F(T)$. For the implied volatility curves below this approach is taken, resulting in the integration interval being set to $[K,3.0]$ for Equation (4.8) and $[-0.6,K]$ for Equation (4.9).

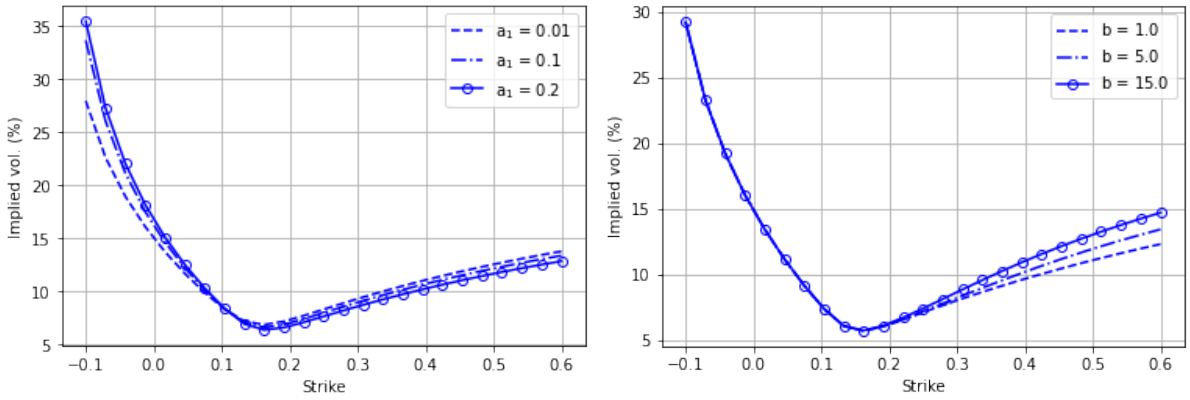

Figure 4.8: Showcase of the IV curve admitted by the Transform solution combined with Equation (4.19) for different parameter sets.

It shows that only the effect of b on the shape of the implied volatility curve is comparable to that on the LV curve shown on the right in Figure 4.7. The smoothing parameter a_1 seems to have a very marginal effect on the shape of the implied volatility curve unless it is set to such high values that it starts affecting also the upside of the curve. It is therefore a good idea to fix a_1 similar to a_2, and use b to control merely the upside of the implied volatility curve. If one requires isolated control over the downside also, the LV function must be extended with additional terms. Another option would be to change a_1 and b simultaneously, in a way that b corrects for the changes a_1 causes in the upside, though this approach is not entirely intuitive nor fast as Z would have to be adjusted often, depending on which part of the curve must be controlled.

Important to note is that, contrary to ρ and ν, the parameter b, in combination with Z, is capable of changing the implied volatility curve shape in isolated areas. Figure 4.9 showcases the effect of ρ and ν on the model output, highlighting that they affect curve shape in both the downside and upside.

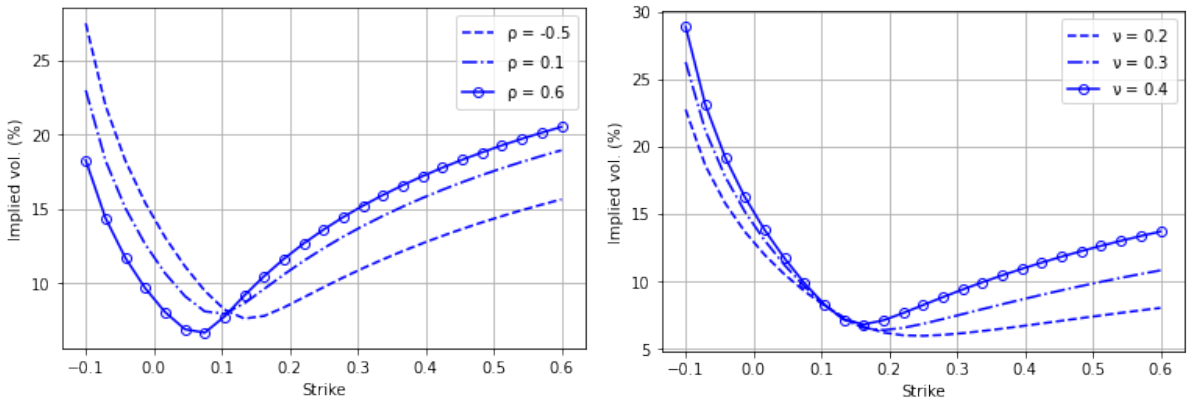

Figure 4.9: Showcase of the IV curve admitted by the Transform solution combined with Equation (4.19) for different parameter sets.

Finally, it was stated in the previous subsection that the level-dependent power specification defined in

Equation (4.18), ensures that the backbone of the distribution of $F(t)$ changes as the value of $F(t)$ changes. Recall that the backbone of this distribution can be seen as the path the ATM implied volatility traverses as the ATM strike changes. A straightforward numerical experiment to check the shape of this backbone is to use the Transform solution combined with the new LV function to compute the price of an option with strike $F(0)$, for various values for $F(0)$. Results of such an experiment are shown in Figure 4.10.

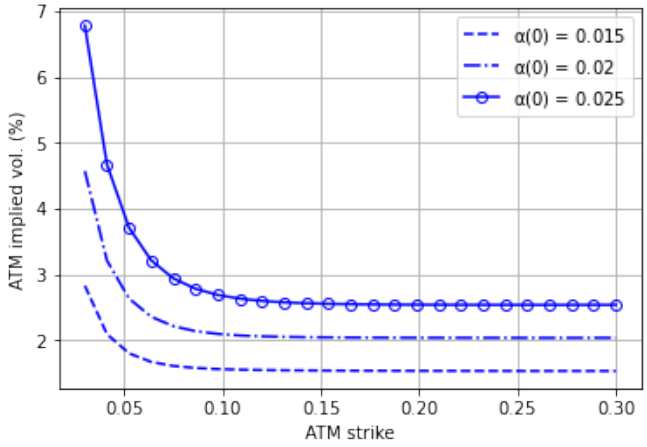

Figure 4.10: Shape of the curve admitted by the sigmoid function specified in Equation (4.18) for varying values of the scaling parameter $\alpha(0)$. Model parameters used: $\nu = 0.3, \rho = -0.5, T = 10, a_1 = 0.02, a_2 = 0.02, b = 1.0, Z = 0.4$.

Figure 4.10 shows that the backbone starts off with high curvature and becomes flatter as $F(0)$ increases. This is an expected result as the power specification in the custom LV function ensures that the distribution of $F(t)$ becomes more lognormal as $F(t)$ increases. These results are in line with the findings discussed by Hagan *et al.* [31] regarding the backbone of $F(t)$ as predicted by the SABR model for different LV specifications.

Application in practice

The final test of model flexibility is to calibrate the Transform solution with Equation (4.19) as the LV function to a generated set of market prices. Assuming the preliminary calibration of $\alpha(0)$ and μ has been performed, a stable approach to this procedure can be described by the following diagram.

Select values for a_1, a_2, b based on market/model experience and a value for Z based on preference.	→	Fit the Transform solution combined with custom LV to market prices by calibrating only ρ and ν.

→

Use b in combination with Z to quickly adjust the implied volatility curve as market conditions or opinions change.

This approach is preferred over attempting to calibrate b, ρ and ν in a single procedure, as a combination of changes to both ρ and ν can have similar effects to changing only b. Therefore, calibrating only ρ and ν with all other parameters fixed during the procedure increases stability by reducing the chance of over-fitting. The initial values for the parameters of the LV function can be selected based on market experience and, with regard to Z, preference.

Finally, it is time to put the quality of the Transform solution with Equation (4.19) as its LV component, to the test. Generating a set of market prices with Hagan's asymptotical approximation to the SABR model [31] for which the LV component is $\sigma(F(t)) := F^{\beta}(t)$ with $\beta \in [0,1]$, and attempting to calibrate the Transform solution to this set of prices, will show whether or not the custom LV function provides enough flexibility. Moreover, it is a good test for showcasing the inherent flexibility of the Transform solution as a whole. The parameters used for generation of the market prices are

$$\alpha(0) = 0.1, \quad \beta = 0.9, \quad \nu = 0.3, \quad \rho = -0.6, \quad T = 10, \quad F(0) = 0.11.$$

Setting once more that ϵ = 10e-6 for evaluation of Equation (4.8) and (4.9), it is time to move on to the calibration procedure used to fit the model to the generated market prices. The details of this procedure, which uses the Levenberg-Marquardt algorithm [21], are discussed in the next section, though for now the results of the model calibration are showcased in Figure 4.11.

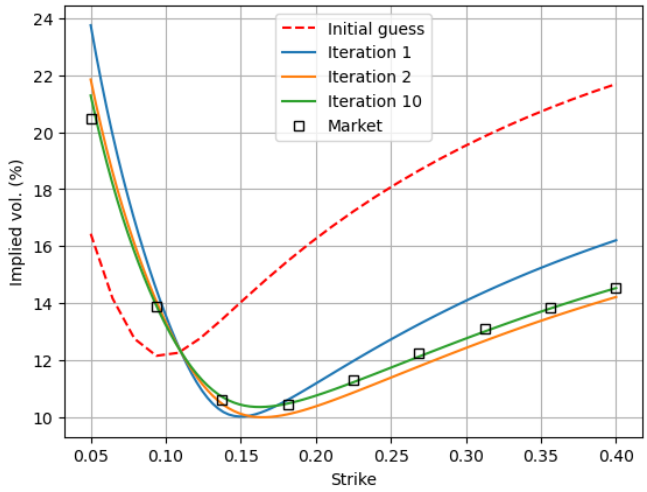

Figure 4.11: Results of ten iterations of the Levenberg-Marquardt calibration algorithm applied to the Transform solution with custom LV shown in Equation (4.19). Set of model parameters after ten iterations is: $\mu = 0.12, \alpha(0) = 0.078, \nu = 0.26, \rho = -0.52, a_1 = 0.01, a_2 = 0.01, b = 1.0, Z = 0.3$.

It turns that with only ten iterations of the Levenberg-Marquardt calibration algorithm (see Section 4.5) a good fit can be obtained. For a better overview of the quality of the fit, Table 4.2 is constructed, where $\varepsilon_{\text{iter.}i} := |\hat{\sigma}_{\text{market}} - \hat{\sigma}_{\text{transform}}|$ in which $\hat{\sigma}_{\text{transform}}$ is the implied volatility of an option obtained with the Transform solution, and i indicates the iteration from which the data was obtained. This table shows that an average error of approximately 10e-4 remains after ten iterations of the calibration algorithm. From these results it can be concluded that the Transform solution combined with the customized LV function possesses decent flexibility. If market conditions cause the market prices to change in the upside, the parameter b can be used to adjust the model output quickly without having to fully recalibrate. Using a two core Intel(R) Core(TM) i7-6560U processor, completion of ten iterations takes approximately two minutes. This is the run time when making use of parallel computing, more commonly referred to as multiprocessing in the Python programming language.

As a final observation, note that the usage of Equation (4.16) in the construction of the custom LV function makes it possible for the Transform solution to price options with negative strikes freely. This is contrary to, e.g. Hagan's approximation [31] or Antonov's approach [1], where a shift parameter has to be introduced to shift negative strikes into the positive domain (see Antonov *et al.* [3]).

K	$\hat{\sigma}_{\text{market}}(\%)$	$\varepsilon_{\text{iter. }0}(\%)$	$\varepsilon_{\text{iter. }10}(\%)$
0.050	20.47	4.04	0.82
0.094	13.88	1.74	0.022
0.14	10.60	2.83	0.10
0.18	10.41	5.06	0.054
0.23	11.27	5.94	0.059
0.27	12.22	6.44	0.10
0.31	13.08	6.77	0.084
0.36	13.85	7.01	0.045
0.40	14.54	7.15	0.015

Table 4.2: Result of calibration procedure plotted in Figure 4.11.

4.5. Calibration method

Provided the Transform solution is used with the LV function defined in Equation (4.19), there are two model parameters that require an initial guess before the calibration process can be started. These parameters are ρ and ν, and their initial values can be detrimental for rapid convergence. Note that $\alpha(0)$ and μ are not mentioned, as their values follow from a calibration to either market information or the model itself. In other words, both of these parameters are implied from the values of the other model parameters or market infor-

mation. This is precisely the two-dimensional root-finding problem discussed in the preliminary calibration subsection of Section 4.1 and described by Equation (4.6) and (4.7).

The two model parameters that require calibration do not come with simple methods for obtaining accurate initial guesses. For example, it is tough to appropriately guess ρ, other than the fact that it is likely possible to decide on its sign based on market conditions. This means that, based on market conditions, one is likely to set the initial value for ρ to something within the interval $[-1.0, 0.5]$. Similar ideas hold for the approximation of ν, where one must base the initial guess on values following from earlier experiments, and judge the quality of the guess using experience with calibration procedures. Finally, it is likely that the designer/user of a custom LV function has a feel for what common values for its parameters are. In this setting, the previous section discussed and showcased typical values for the parameters in the custom LV function through the effects they have on the shape of the LV and implied volatility curves.

The method for calibration used in this thesis is the Levenberg-Marquardt (LM) algorithm, which is a tool for solving non-linear least squares problems, and it is combined with an auxiliary nested algorithm to calibrate the μ and $\alpha(0)$ parameters in each iteration. The LM algorithm dynamically changes some damping factor, which makes it so that the algorithm alternates between the steepest descent method and the Gauss-Newton method. More specifically, the method gives more weight to the Gauss-Newton approximation when the iterate is close to the optimum value, and vice versa for the steepest descent approximation. The algorithm and its application in this setting will be briefly discussed here, but for a more detailed discussion the reader is referred to Cui *et al.* [21] or Gavin [29]. Important to note is that the LM algorithm is a local optimization algorithm. This emphasizes the value of a decent initial guess as not every arbitrary local minimum leads to a good fit.

The problem setting is as follows. Let the market price of a call option on $F(T)$, with strike K and maturity T, be denoted by $C_F^\dagger(K)$. Moreover, for the vector of parameters $\boldsymbol{\theta} := [\rho, \nu]^\intercal$, the price for the aforementioned call option following from the Transform solution is denoted as $C_F(\boldsymbol{\theta}, K)$. The vector of residuals is then defined as the vector containing the differences between the model and market option prices for different strikes. This means that for n different strikes, the following holds

$$r_i(\boldsymbol{\theta}) := C_F(\boldsymbol{\theta}, K_i) - C_F^\dagger(K_i), \quad i = 1, \ldots, n, \tag{4.20}$$

$$\boldsymbol{r}(\boldsymbol{\theta}) := [r_1(\boldsymbol{\theta}), \ldots, r_n(\boldsymbol{\theta})]^\intercal \in \mathbb{R}^n. \tag{4.21}$$

Finally, the aim of the calibration procedure can then be defined as the following minimization problem

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^2} f(\boldsymbol{\theta}), \tag{4.22}$$

where $f(\boldsymbol{\theta}) := \frac{1}{2}\boldsymbol{r}^\intercal(\boldsymbol{\theta})\boldsymbol{r}(\boldsymbol{\theta})$. Having defined the outline of the problem, the Jacobian matrix $\boldsymbol{J}$ is constructed of elements defined by

$$J_{ji} := \left[\frac{\partial C_F(\boldsymbol{\theta}, K_i)}{\partial \theta_j} \right]. \tag{4.23}$$

Using this Jacobian matrix, the gradient of the function f is defined as

$$\nabla f := \boldsymbol{J}\boldsymbol{r}. \tag{4.24}$$

The search step, i.e. the parameter adjustment computed in each iteration of the LM algorithm, can be written as

$$\Delta\boldsymbol{\theta} := (\boldsymbol{J}\boldsymbol{J}^\intercal + \lambda \boldsymbol{I})^{-1} \nabla f, \tag{4.25}$$

where λ is the aforementioned damping factor determining the weight given to the Gauss-Newton and steepest descent approximation. The full LM algorithm is then

Algorithm 1 Levenberg-Marquardt algorithm to calibrate the Transform solution.

1: For the initial guess $\boldsymbol{\theta}_0$, calibrate μ and $\alpha(0)$ using the method described in the third subsection of Section 4.1, and compute $||\boldsymbol{r}(\boldsymbol{\theta}_0)||_2$ and $\boldsymbol{J}_0$.
2: Select a weight factor $\omega > 0$. Set the initial damping factor to $\lambda_0 = \omega \max(\text{diag}(\boldsymbol{J}_0))$, and let $\gamma_0 = 2$.
3: **for** k = 0, 1, 2, ... **do**
4: Compute the parameter adjustment $\Delta\boldsymbol{\theta}_k$ using Equation (4.25).
5: Calculate new parameter set and residual vector using $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \Delta\boldsymbol{\theta}_k$ and $||\boldsymbol{r}(\boldsymbol{\theta}_{k+1})||_2$.
6: Re-calibrate μ and $\alpha(0)$ using the method described in Section 4.1.
7: Compute constraint factors $\delta_L := \Delta\boldsymbol{\theta}_k^\top(\lambda_k\Delta\boldsymbol{\theta}_k + \boldsymbol{J}_k\boldsymbol{r}(\boldsymbol{\theta}_k))$ and $\delta_F := ||\boldsymbol{r}(\boldsymbol{\theta}_k)||_2 - ||\boldsymbol{r}(\boldsymbol{\theta}_{k+1})||_2$.
8: **if** $\delta_F > 0$ and $\delta_L > 0$ **then**
9: New parameter set is an improvement. Calculate $\boldsymbol{J}_{k+1}$, set $\lambda_{k+1} = \lambda_k$ and $\gamma_{k+1} = \gamma_k$.
10: **else**
11: Set $\lambda_k = \lambda_k\gamma_k$ and $\gamma_{k+1} = 2\gamma_k$. Go back to line 4.
12: **if** $||\boldsymbol{r}(\boldsymbol{\theta}_k)||_2 \leq \epsilon$ or $||\boldsymbol{J}||_\infty \leq \epsilon$ or $||\Delta\boldsymbol{\theta}_k||_2/||\boldsymbol{\theta}_k||_2 \leq \epsilon$ **then**
13: Break.

Additional measures for increasing numerical stability can be taken to prevent the algorithm from explosively diverging in cases of, e.g. a poor initial guess or an unlucky local minimum. An example of such a measure is introducing a new step in Algorithm 1, between the current line 4 and 5, that will move the program to line 11 if application of the parameter adjustment $\Delta\boldsymbol{\theta}_k$ would result in $\rho > 1$, $\rho < -1$ or $\nu < 0$, as these parameter values are not allowed. Experimenting with the algorithm showed that such measures can at times provide a correction large enough to prevent divergence.

Finalizing, the constant ϵ is typically set to be no smaller than 10e-10, both for computational reasons and because in most cases the slight gain in accuracy does not make up for the loss in speed. The Jacobian matrix can easily be computed using finite difference methods, and, as mentioned at the end of the previous section, the entire calculation can be severely sped up by making use of parallel computing.

4.6. Non-parametric LV

The emphasis in the final few sections of this chapter has been on the fact that one has the ability to combine the Transform solution with customized LV functions. These parametric functions can be designed to trader specifications and tweaked to let the opinions of the user influence the output of the model. However, it is still to be determined whether the Transform solution can also be combined with non-parametric LV functions. While these types of functions being used in a model made powerful by its versatility is rather counter intuitive due to their general lack of flexibility, testing whether or not they result in output that fits nicely to market prices is still an interesting endeavor. Moreover, it might spark ideas regarding custom LV functions comprised of both parametrical and non-parametrical components.

Perhaps the most well-known non-parametric LV function is the one proposed by Dupire [24] and Derman-Kani [22]. This function provides a way of implying an LV surface directly from an implied volatility surface corresponding to a set of vanilla European options. In turn, this allows for exact calibration of the LV model to any other set of arbitrage-free vanilla European options. However, this so-called LV model is not entirely free of problems. Besides the complexities in a numerical setting that will be discussed later, another issue is that the dynamics of implied volatility curves obtained with from this model are opposite of those typically observed in the market (see Hagan *et al.* [31]). There are a few other limitations to this approach, though the focus in this section will mainly be on the numerical issues that one might run into when using the Dupire LV function in the Transform solution. In other words, the objective of this section is to acquire an LV curve through Dupire's formula and perform numerical experiments on the Transform solution once combined with this non-parametric function.

Dupire's formula is generally expressed in terms of call option prices, but for increased stability an alternative expression in terms of implied volatilities can be derived (see Oosterlee-Grzelak [47]). This formula, yielding the local volatility at time T at a value of the underlying equal to K, reads

$$\sigma_{LV}^2(T,K) = \frac{\frac{\partial\omega}{\partial T}}{1 + K\frac{\partial\omega}{\partial K}\left(\frac{1}{2} - \frac{y}{\omega}\right) + \frac{1}{2}K^2\frac{\partial^2\omega}{\partial K^2} + \frac{1}{2}K^2\left(\frac{\partial\omega}{\partial K}\right)^2\left(-\frac{1}{8} - \frac{1}{2\omega} + \frac{y^2}{2\omega^2}\right)}, \tag{4.26}$$

where

$$y(T,K) \coloneqq \log\left(\frac{K}{S(0)}\right), \tag{4.27}$$

$$\omega(T,K) \coloneqq \hat{\sigma}^2(T,K)T, \tag{4.28}$$

$$\frac{\partial\omega(T,K)}{\partial T} = \hat{\sigma}^2(T,K) + 2T\hat{\sigma}(K,T)\frac{\partial\hat{\sigma}(K,T)}{\partial T}, \tag{4.29}$$

$$\frac{\partial\omega(T,K)}{\partial K} = 2T\hat{\sigma}(K,T)\frac{\partial\hat{\sigma}(K,T)}{\partial K}, \tag{4.30}$$

$$\frac{\partial^2\omega(T,K)}{\partial K^2} = 2T\left(\frac{\partial\hat{\sigma}(K,T)}{\partial K}\right)^2 + 2T\hat{\sigma}(K,T)\frac{\partial^2\hat{\sigma}(K,T)}{\partial K^2}, \tag{4.31}$$

where $\hat{\sigma}(K,T) \equiv \hat{\sigma}_{\text{market}}(K,T)$ is the implied volatility as observed in the market of an option with strike K and maturity T, and $S(0)$ is the time zero value of the underlying on which the options in the market are defined. Obviously, in case the underlying is a forward rate as in the current setting, appropriate measures have to be taken whenever $r \neq 0$. However, recall that in this thesis it is assumed without loss of generality that $r = 0$. Furthermore, note that in Equation (4.26), the following simplified notation was used

$$y \equiv y(T,K), \quad \omega \equiv \omega(T,K), \quad \frac{\partial\omega}{\partial T} \equiv \frac{\partial\omega(T,K)}{\partial T}, \quad \frac{\partial\omega}{\partial K} \equiv \frac{\partial\omega(T,K)}{\partial K}, \quad \frac{\partial^2\omega}{\partial K^2} \equiv \frac{\partial^2\omega(T,K)}{\partial K^2}.$$

One might expect that using an LV function perfectly implied from market prices, and letting it govern the LV component in an arbitrary generalized SLV model, will result in model output that fits perfectly to those same market prices. However, as van der Stoep *et al.* [50] show, this need not be the case when the general SLV model takes on the form of a Heston type model [21]. Nonetheless, in this section the aim is to find the set of model parameters that will result in the most accurate fit of the output of the Transform solution with Equation (4.26) as the LV specification, to a set of market prices. It will be attempted to find these parameters using a calibration procedure similar to that used in Section 4.4.

Analogous to the approach taken in Section 4.4, a set of arbitrary market prices, to be used in an interest rate setting, can be generated. This time, prices are generated with Antonov's exact solution to the normal SABR model [1] using the parameter set

$$\alpha(0) = 0.02, \quad \nu = 0.3, \quad \rho = -0.5, \quad T = 30, \quad F(0) = 0.11,$$

which yields an implied volatility skew rather than a smile as generated in Section 4.4. For the rest of this section, assume no additional market prices can be obtained or generated, similar to a real-world setting. Figure 4.12 then shows the LV curve obtained from application of Equation (4.26) to this generated set of prices.

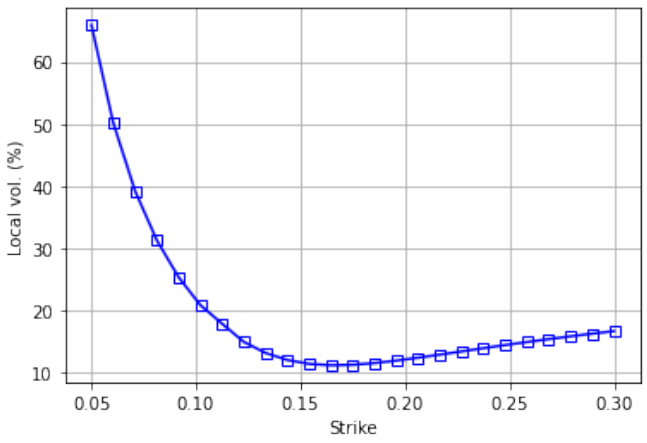

Figure 4.12: Local volatility curve obtained from the market prices used in Section 4.4, using Dupire's local volatility formula shown in Equation (4.26).

The fact that the market prices are generated with an analytical formula allows for easier calculation of the partial derivatives present in Dupire's formula through finite differencing. Computing the LV curve using Equation (4.26) with only a limited amount of market prices might require complicated interpolation or

fitting techniques in practice to make up for the discrete strike and maturity ranges (see Gatheral [27]), but these complexities play a much lesser role in this test setting.

Recall that the approximative pricing used for approximating prices of options on $F(T)$ with a strike K and maturity T, read

$$C_F(K) \approx \int_K^{G^{-1}(\Phi_\eta^{-1}(1-\epsilon)-\mu)} (1-\Phi_\eta(G(k)+\mu))\,\mathrm{d}k, \tag{4.32}$$

$$P_F(K) \approx \int_{G^{-1}(\Phi_\eta^{-1}(\epsilon)-\mu)}^{K} \Phi_\eta(G(k)+\mu)\,\mathrm{d}k, \tag{4.33}$$

where μ is a constant calibrated to maintain preservation of the forward and, for $\mathscr{X}$ state-space of $F(t)$, G is the generalized Lamperti transform

$$G(a) \coloneqq \int_{F(0)}^{a} \mathbb{1}_{u\in\mathscr{X}} \frac{1}{\sigma(u)}\,\mathrm{d}u\Big|_{a=F(t)}. \tag{4.34}$$

It is important to note at this stage that a selection for ϵ, necessary for evaluation of Equation (4.32) and Equation (4.33), might lead to a situation in which σ will have to be evaluated at points outside of its domain, i.e. the range of strikes as observable in the market. To combat this issue, it is possible to follow along with the reasoning of Kotzé *et al.* [36] and Bliss-Panigirtzoglou [14], who state that, due to the relative inadequacy of tail outcomes, it is not critically important how one extrapolates the implied density of $F(T)$ in the tails as long as the result is reasonable. Similar reasoning can be applied here for extrapolation of Dupire's LV curve, though rather than forcing the curve to extrapolate flat outside its domain, simple linear extrapolation will be used. Behvand [11] states that it might be preferable for computational reasons to extrapolate the LV flat outside the interval $[0.01F(0), 4F(0)]$, though in the current setting minimization of execution time is a not a priority.

While the LV curve is relatively flat in the area of strikes above the ATM, the same cannot be said about its steepness at the lower strikes. Therefore, the linear extrapolation of the LV on the left side of the strike range, i.e. below 0.05, has to be handled with the possibility of explosions in mind. Nevertheless, setting ϵ = 10e-6 and applying Equation (4.32) and (4.33) in the calibration framework as described in Section 4.5 is expected to yield a decent fit to the generated market prices. The results of this operation for different iterations are showcased in Figure 4.13.

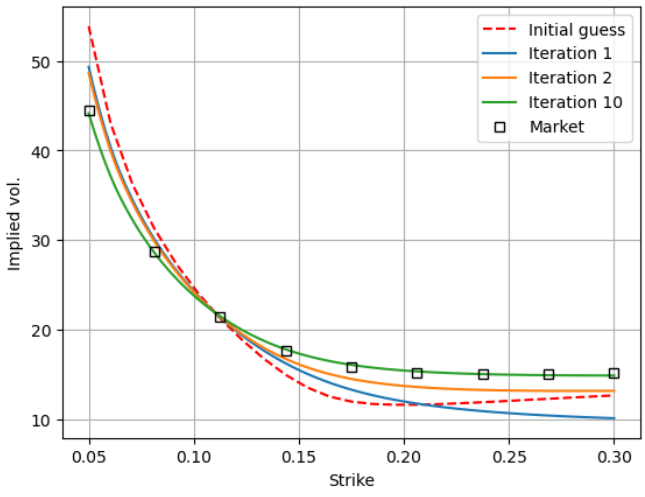

Figure 4.13: Results of only ten iterations of the LM calibration algorithm applied to the Transform solution with Dupire's LV computed using Equation (4.26). Set of model parameters after ten iterations is: $\mu = -0.15, \alpha(0) = 0.10, \rho = 0.95, \nu = 0.12$

In order to get a quantitative overview of the quality of the obtained fit, Table 4.3 is constructed. Note that in this table it holds once more that $\varepsilon_{\text{iter}.i} \coloneqq |\hat{\sigma}_{\text{market}} - \hat{\sigma}_{\text{transform}}|$ with i indicating the iteration from which the data was obtained.

K	$\hat{\sigma}_{\text{market}}(\%)$	$\varepsilon_{\text{iter. }0}(\%)$	$\varepsilon_{\text{iter. }10}(\%)$
0.050	44.49	9.37	0.35
0.081	28.65	2.60	0.078
0.11	21.48	0.21	0.0090
0.14	17.64	2.71	0.14
0.18	15.86	3.92	0.19
0.21	15.20	3.60	0.12
0.24	15.02	3.14	0.0042
0.27	15.04	2.78	0.15
0.30	15.15	2.52	0.29

Table 4.3: Result of calibration procedure plotted in Figure 4.13.

While the fit looks accurate at first glance, it turns out that something unexpected has happened. As mentioned in the description of Figure 4.13, the set of parameters obtained after ten iterations of the calibration algorithm is

$$\mu = -0.15, \quad \alpha(0) = 0.10, \quad \rho = 0.95, \quad \nu = 0.12.$$

It becomes clear immediately that these parameters are considerably different from those used to generate the market prices. This is a result of the extrapolation of the LV curve that had to be performed to properly evaluate the pricing functions. Let it be clear that this is not a flaw of Dupire's LV curve, but of the Transform solution when combined with LV functions not defined on the entire state-space of $F(t)$. This statement might not seem significant, as a good fit to the market has still been obtained, but the act of using a pricing model with dynamics severely different from market dynamics, e.g. model dynamics use $\rho = 0.95$ while the market can be modelled with dynamics using $\rho = -0.5$, is clearly problematic.

5

Multi-dimensional application

The main focus in the thesis so far has been on the various applications of the Lamperti transform in one-dimensional fashion, i.e. defining a transform dependent on a single state-variable. In this chapter, the possibilities of a multi-dimensional generalization of the typical univariate Lamperti transform will be explored. Such a multi-dimensional version would enable the transformation of, for example, the SABR model [31] or Heston model [21], directly, without having to introduce an auxiliary process that isolates the diffusion component, as done in Chapter 3 for a general SLV model. This will not only more efficient, as isolating diffusion terms in a multi-dimensional system becomes more complicated in the presence of drift terms, but also enable the transformation of processes with more than a single diffusion term. For example, provided with the exemplary system

$$\mathrm{d}X(t) = \alpha(t)\,\mathrm{d}W_1(t) + \sqrt{\alpha(t)}X(t)\,\mathrm{d}W_2(t), \tag{5.1}$$
$$\mathrm{d}\alpha(t) = \gamma\alpha(t)\,\mathrm{d}W_1(t) + \mathrm{d}W_2(t), \tag{5.2}$$

where $W_1(t)$ and $W_2(t)$ are two correlated Brownian motions and γ is a real constant, application of the Lamperti transform as it has been discussed so far need not be consistent nor straightforward. This is simply because the two diffusion components make it unclear how to appropriately apply the univariate Lamperti transform such that constant diffusion coefficients are obtained. It becomes clear that a multi-dimensional generalization will be helpful in expanding the family of SDE's to which the transform can be applied. Before going deeper into these topics, some preliminary knowledge on multi-dimensional systems is discussed in Section 5.1. Furthermore, recall that the primary focus when applying the Lamperti transformation is obtaining constant diffusion coefficients. While this might lead to more complicated drift terms, the transformed process will be more stable in estimation [6] and simulation procedures [18, 33, 40, 43].

5.1. Preliminaries

Provided the entire chapter will require appropriate handling of multi-dimensional SDE systems, the main preliminary knowledge is once more Itô's lemma, though this time its multi-dimensional variation is required. Before stating this lemma, a short discussion on using the Cholesky decomposition [47] to form connections between SDE systems written in terms of correlated Brownian motions and those written in terms of uncorrelated Brownian motions, will prove useful. Provided with an arbitrary system of correlated SDE's

$$\mathrm{d}\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t), t)\,\mathrm{d}t + \boldsymbol{\sigma}(\mathbf{X}(t), t)\,\mathrm{d}\mathbf{W}(t), \tag{5.3}$$

where $\boldsymbol{\mu} : A \to \mathbb{R}^n$, $\boldsymbol{\sigma} : A \to \mathbb{R}^{n\times n}$ and $\mathbf{W} \in \mathbb{R}^n$ a column vector of correlated Brownian motions with $\mathrm{d}W_i(t)\,\mathrm{d}W_j(t) = \rho_{i,j}\,\mathrm{d}t$ if $i \neq j$, $\rho_{i,j} \in [-1,1]$, and $\mathrm{d}W_i(t)\mathrm{d}W_i(t) = \mathrm{d}t$, for $i, j = 1,2,\ldots,n$. For now, let $\boldsymbol{\mu} \equiv \boldsymbol{\mu}(\mathbf{X}(t), t)$ and

$\boldsymbol{\sigma} \equiv \boldsymbol{\sigma}(\mathbf{X}(t), t)$. Applying the Cholesky decomposition to Equation (5.3) then yields

$$\begin{aligned}
\begin{bmatrix} \mathrm{d}X_1(t) \\ \vdots \\ \mathrm{d}X_n(t) \end{bmatrix} &= \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \mathrm{d}t + \begin{bmatrix} \sigma_{1,1} & \dots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \dots & \sigma_{n,n} \end{bmatrix} \begin{bmatrix} \mathrm{d}W_1(t) \\ \vdots \\ \mathrm{d}W_n(t) \end{bmatrix} \\
&= \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \mathrm{d}t + \begin{bmatrix} \sigma_{1,1} & \dots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \dots & \sigma_{n,n} \end{bmatrix} \begin{bmatrix} 1 & 0 & \dots \\ \rho_{1,2} & \sqrt{1-\rho_{1,2}^2} & \dots \\ \vdots & \ddots & \vdots \\ \rho_{1,n} & \dots & \dots \end{bmatrix} \begin{bmatrix} \mathrm{d}\widetilde{W}_1(t) \\ \vdots \\ \mathrm{d}\widetilde{W}_n(t) \end{bmatrix} \\
&= \boldsymbol{\mu}\,\mathrm{d}t + \boldsymbol{\sigma}\mathbf{L}\,\mathrm{d}\widetilde{\mathbf{W}}(t) =: \boldsymbol{\mu}\,\mathrm{d}t + \bar{\boldsymbol{\sigma}}\,\mathrm{d}\widetilde{\mathbf{W}}(t), \qquad (5.4)
\end{aligned}$$

where $\widetilde{\mathbf{W}}$ is a column vector of independent Brownian motions and $\bar{\boldsymbol{\sigma}} := \boldsymbol{\sigma}\mathbf{L}$. Summarizing, Equation (5.3) has been rewritten to become

$$\mathrm{d}\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t), t)\,\mathrm{d}t + \bar{\boldsymbol{\sigma}}(\mathbf{X}(t), t)\,\mathrm{d}\widetilde{\mathbf{W}}(t), \qquad (5.5)$$

with $\bar{\boldsymbol{\sigma}}$ as in Equation (5.4). Decomposing a system of correlated SDE's in this way will enable showcasing more clearly how application of the Lamperti transform moves the correlation structure from the diffusion term to the drift term, later on in this chapter. First, the multi-dimensional version of Itô's lemma will be stated. Note that the processes discussed in the following exist under an arbitrary measure and that original notation is used once more.

Lemma 5.1.1. *(Multi-dimensional Itô's lemma) Let* $\mathbf{X}$ *be a state vector of Itô processes given by Equation (5.5), i.e.*

$$\mathrm{d}\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t), t)\,\mathrm{d}t + \bar{\boldsymbol{\sigma}}(\mathbf{X}(t), t)\,\mathrm{d}\widetilde{\mathbf{W}}(t),$$

where $\boldsymbol{\mu} : A \to \mathbb{R}^n$, $\bar{\boldsymbol{\sigma}} : A \to \mathbb{R}^{n\times n}$ *and* $\widetilde{\mathbf{W}} \in \mathbb{R}^n$ *a column vector of independent Brownian motions. Moreover, let* $\mathbf{Z} \in \mathbb{R}^n$ *be a column vector of transformations defined by*

$$\mathbf{Z}(t) := \boldsymbol{\psi}(\mathbf{X}(t), t) := [\psi_1(\mathbf{X}(t), t), \psi_2(\mathbf{X}(t), t), \dots, \psi_n(\mathbf{X}(t), t)]^\mathsf{T}, \qquad (5.6)$$

where $\boldsymbol{\psi} : A \to \mathbb{R}^n$ *is a continuous twice differentiable function. Then it holds that* $\mathbf{Z}$ *is again a state vector of Itô processes, with*

$$\mathrm{d}Z_k(t) = \frac{\partial \psi_k(\mathbf{X}(t), t)}{\partial t}\,\mathrm{d}t + \sum_{i=1}^{n} \frac{\partial \psi_k(\mathbf{X}(t), t)}{\partial X_i}\,\mathrm{d}X_i(t) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \psi_k(\mathbf{X}(t), t)}{\partial X_i \partial X_j}\,\mathrm{d}X_i(t)\,\mathrm{d}X_j(t). \qquad (5.7)$$

for $k = 1, 2, \dots, n$. *In further evaluation of the above expression it can be used that for* $i, j = 1, 2, \dots, n$ *with* $i \neq j$

$$\mathrm{d}\widetilde{W}_i(t)\,\mathrm{d}\widetilde{W}_j(t) = (\mathrm{d}t)^2 = \mathrm{d}\widetilde{W}_i(t)\,\mathrm{d}t = 0, \quad (\mathrm{d}\widetilde{W}_i(t))^2 = \mathrm{d}t.$$

Proof. See Øksendal [46]. □

It becomes clear from Equation (5.7) that the process of defining the appropriate form for the Lamperti transform will require the solution to a system of partial differential equations (PDE's) containing the diffusion terms of each SDE in the multi-variate system.

Briefly examining Equation (5.7) shows that the diffusion term in the dynamics of $Z_k(t)$ will be determined by the $\sum_{i=1}^{n} \frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial X_i}\,\mathrm{d}X_i(t)$ term. For a non-trivial state vector of Itô processes $\mathbf{X}$ and non-diagonal noise matrix $\bar{\boldsymbol{\sigma}}$, the aforementioned term will yield a system of non-linear inhomogeneous first-order PDE's. Existence of solutions to such systems generally requires stringent initial/boundary conditions, and even then the process of formulating an expression for either a particular or general solution can be very cumbersome. However, this system does define the crux in obtaining a non-parametric expression for the Lamperti transform. The upcoming sections will cover the expansion of the full system of PDE's and discuss the issues with obtaining a solution.

5.2. Fully generalized transform difficulties

An appropriate first step in exploring the possibilities of deriving a generalized form for the Lamperti transform in a multi-dimensional setting will be applying Lemma 5.1.1 to a fully generalized non-parametric SLV model. This will result in an expression that can be rewritten to distinguish between drift and diffusion terms, revealing the system of PDE's to which the Lamperti transform must provide a solution. Van der Stoep *et al.* [50] present such a general SLV model and define it as

$$\begin{aligned}
\mathrm{d}S(t) &= rS(t)\,\mathrm{d}t + S(t)\sigma(S(t),t)\phi(v(t))\,\mathrm{d}\widetilde{W}_S^{\mathbb{Q}}(t),\\
\mathrm{d}v(t) &= a_v(v(t),t)\,\mathrm{d}t + b_v(v(t),t)\left(\rho_{S,v}\,\mathrm{d}\widetilde{W}_S^{\mathbb{Q}}(t) + \sqrt{1-\rho_{S,v}^2}\,\mathrm{d}\widetilde{W}_v^{\mathbb{Q}}(t)\right),
\end{aligned}$$

where $\widetilde{W}_S^{\mathbb{Q}}(t)$ and $\widetilde{W}_v^{\mathbb{Q}}(t)$ are two independent Brownian motions under the risk neutral measure $\mathbb{Q}$, $\rho_{S,v}$ denotes the correlation parameter, σ is the LV component, ϕ controls the stochastic volatility and r is a constant interest rate parameter. In order to apply the multi-dimensional Itô's lemma, define the non-parametric transformation

$$\mathbf{Z}(t) \coloneqq \boldsymbol{\psi}(\mathbf{X}(t),t) \coloneqq [\psi_1(\mathbf{X}(t),t), \psi_2(\mathbf{X}(t),t)]^\intercal,$$

where $\mathbf{X}(t) \coloneqq [S(t), v(t)]^\intercal$. By Lemma 5.1.1 it then follows that for $k \in \{1,2\}$

$$\begin{aligned}
\mathrm{d}Z_k(t) = \frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial t}\,\mathrm{d}t + \frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial S}\,\mathrm{d}S(t) + \frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial v}\,\mathrm{d}v(t) + \frac{\partial^2 \psi_k(\mathbf{X}(t),t)}{\partial S \partial v}\,\mathrm{d}S(t)\,\mathrm{d}v(t)\\
+ \frac{1}{2}\frac{\partial^2 \psi_k(\mathbf{X}(t),t)}{\partial S^2}(\mathrm{d}S(t))^2 + \frac{1}{2}\frac{\partial^2 \psi_k(\mathbf{X}(t),t)}{\partial v^2}(\mathrm{d}v(t))^2. \quad (5.8)
\end{aligned}$$

Simplifying Equation (5.8) and appropriately isolating the diffusion terms eventually results in

$$\begin{aligned}
\mathrm{d}Z_k(t) = \Bigg(&\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial t} + rS(t)\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial S} + a_v(X(t),t)\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial v} + \frac{1}{2}S^2(t)\sigma^2(S(t),t)\phi^2(v(t))\frac{\partial^2 \psi_k(\mathbf{X}(t),t)}{\partial S^2}\\
&+ \frac{1}{2}b_v^2(v(t),t)\frac{\partial^2 \psi_k(\mathbf{X}(t),t)}{\partial v^2} + S(t)\sigma(S(t),t)\phi(v(t))b_v(v(t))\rho_{S,v}\frac{\partial^2 \psi_k(\mathbf{X}(t),t)}{\partial S \partial v}\Bigg)\mathrm{d}t\\
&+ \left(S(t)\sigma(S(t),t)\phi(v(t))\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial S} + b_v(v(t),t)\rho_{S,v}\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial v}\right)\mathrm{d}\widetilde{W}_S^{\mathbb{Q}}(t)\\
&+ b_v(v(t),t)\sqrt{1-\rho_{S,v}^2}\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial v}\,\mathrm{d}\widetilde{W}_v^{\mathbb{Q}}(t). \quad (5.9)
\end{aligned}$$

It becomes clear that in order for this expression to be transformed into one with constant diffusion coefficients, the following equations must be satisfied

$$S(t)\sigma(S(t),t)\phi(v(t))\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial S} = c_1, \quad (5.10)$$

$$b_v(v(t),t)\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial v} = c_2, \quad (5.11)$$

where c_1 and c_2 are two non-zero constants. Equation (5.10) implies that it must hold that

$$\psi_k(\mathbf{X}(t),t) = \int \frac{c_1}{s\sigma(s,t)\phi(v(t))}\,\mathrm{d}s\Big|_{s=S(t)} + \vartheta(v(t)), \quad (5.12)$$

in which $\vartheta(v(t))$ denotes an arbitrary function of $v(t)$, arising from the fact that the integration is performed with respect to $S(t)$. Using the equation above, it follows from Equation (5.11) that

$$c_2 = b_v(v(t),t)\frac{\partial}{\partial v}\left(\int \frac{c_1}{s\sigma(s,t)\phi(v(t))}\,\mathrm{d}s\Big|_{s=S(t)} + \vartheta(v(t))\right). \quad (5.13)$$

As it turns out, the system described by Equation (5.10) and (5.11) does not admit a solution due to $\vartheta(v(t))$ not depending on $S(t)$. This is quite an interesting result and it is a consequence of the application of the transform to a system containing an equation with a diffusion term dependent on multiple state-variables, i.e. the product between σ and ϕ in the dynamics of $S(t)$. As a result, the transform function has to satisfy

two equations at the same time, resulting in a system with too many unknowns to yield any solution. It is generally not possible to work around the absence of a solution to the system of non-linear inhomogeneous PDE's shown in Equation (5.10) and (5.11), unless trivial functional definitions are chosen for ϕ. Indeed, rewriting Equation (5.10) and taking the partial derivative with respect to $v(t)$ on both sides yields

$$\frac{\partial^2 \psi_k(\mathbf{X}(t),t)}{\partial v \partial S} = \frac{c_1}{S(t)\sigma(S(t),t)} \frac{\mathrm{d}}{\mathrm{d}v}\left(\frac{1}{\phi(v(t))}\right), \tag{5.14}$$

while rewriting Equation (5.11) and taking the partial derivative with respect to $S(t)$ on both sides gives

$$\frac{\partial^2 \psi_k(\mathbf{X}(t),t)}{\partial v \partial S} = 0. \tag{5.15}$$

Equation (5.14) can now be combined with Equation (5.15), showing that in order for the PDE system described by Equation (5.10) and (5.11) to have a solution, it must hold that

$$\frac{\mathrm{d}}{\mathrm{d}v}\left(\frac{1}{\phi(v(t))}\right) = 0,$$

which is only the case if the function ϕ is set to be a constant. This result confirms the earlier statement that the system of PDE's described by Equation (5.10) and (5.11) is in general not solvable unless ϕ is set to be a trivial function, e.g. constant.

In fact, applying the multi-dimensional Itô lemma to an arbitrary state vector of Itô processes shows some interesting results. Namely, Equation (5.7) can be rewritten into

$$\begin{aligned}
\mathrm{d}Z_k(t) &= \frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial t}\mathrm{d}t + \sum_{i=1}^{n} \frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial X_i}\mathrm{d}X_i(t) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \psi_k(\mathbf{X}(t),t)}{\partial X_i \partial X_j}\mathrm{d}X_i(t)\,\mathrm{d}X_j(t) \\
&= \left(\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial t} + \sum_{i=1}^{n} \mu_i(\mathbf{X}(t),t)\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial X_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \bar{\sigma}_{i,k}(\mathbf{X}(t),t)\bar{\sigma}_{j,k}(\mathbf{X}(t),t)\frac{\partial^2 \psi_k(\mathbf{X}(t),t)}{\partial X_i \partial X_j}\right)\mathrm{d}t \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{n} \bar{\sigma}_{i,j}(\mathbf{X}(t),t)\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial X_i}\,\mathrm{d}\widetilde{W}_j(t).
\end{aligned}$$

From this expression it becomes clear that a transformation to unit diffusion boils down to finding a solution to a system of partial differential equations defined by

$$\begin{aligned}
\sum_{i=1}^{n} \bar{\sigma}_{i,1}(\mathbf{X}(t),t)\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial X_i} &= c_1, \\
\sum_{i=1}^{n} \bar{\sigma}_{i,2}(\mathbf{X}(t),t)\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial X_i} &= c_2, \\
&\vdots \\
\sum_{i=1}^{n} \bar{\sigma}_{i,n}(\mathbf{X}(t),t)\frac{\partial \psi_k(\mathbf{X}(t),t)}{\partial X_i} &= c_n,
\end{aligned}$$

where $c_1, c_2, \ldots, c_n$ are non-zero constants. This system is in general not solvable for the same reasons as in the earlier example on the general SLV model, concluding that this fully general multi-dimensional approach to the Lamperti transform is not feasible unless stringent conditions are imposed on the system, or trivial function specifications are used. It remains interesting, however, to check if the transform can be of benefit in multi-variate SDE systems where the diffusion terms of different equations do not directly depend on one another. Section 5.3 explores some of these less general possibilities and discusses potential applications through an example.

5.3. Less general transform and examples

The preceding section attempted a fully generalized application of a multi-dimensional Lamperti transform, but ran into trouble with mixed diffusion terms inside a single SDE. A less general version of the transform can be defined, strongly resembling the original definition as stated in Theorem 2.1.6. This version will form the body of this section and therefore plays an essential role in the upcoming examples. Note that in this section there is a distinction being made between state-dependent diffusion terms and time-dependent terms. Time-dependent terms are terms that only depend on time, i.e. not the state of the process.

Theorem 5.3.1 (Multi-dimensional Lamperti transform)**.** *Given a state vector of Itô processes* $\mathbf{X}$ *with*

$$\mathrm{d}\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t),t)\,\mathrm{d}t + \boldsymbol{\sigma}(\mathbf{X}(t),t)\bar{\mathbf{M}}(t)\,\mathrm{d}\widetilde{\mathbf{W}}(t),$$

where $\boldsymbol{\mu}: A \to \mathbb{R}^n$, $\bar{\mathbf{M}} := \mathbf{ML}: A \to \mathbb{R}^{n\times n}$ *a matrix of functions of* t *in which* $\mathbf{L} \in \mathbb{R}^{n\times n}$ *is a lower triangular matrix with positive diagonal entries originating from the Cholesky decomposition of correlated Brownian motions,* $\widetilde{\mathbf{W}} \in \mathbb{R}^n$ *a column vector of independent Brownian motions and* $\boldsymbol{\sigma}: A \to \mathbb{R}^{n\times n}$ *a diagonal matrix with each element continuously differentiable and strictly negative or strictly positive on its domain, and with each element containing at most one* $X_i(t)$ *term for* $i = 1,2,\ldots,n$*, i.e.*

$$\sigma_{i,i}(\mathbf{X}(t),t) = \sigma_{i,i}(X_i(t),t), \quad i = 1,2,\ldots,n. \tag{5.16}$$

Let

$$\psi_i(X_i(t),t) := \int \frac{1}{\sigma_{i,i}(x,t)}\,\mathrm{d}x\bigg|_{x=X_i(t)}, \quad i = 1,2,\ldots,n.$$

If ψ_i *is a one-to-one mapping of the state space of* $X_i(t)$ *to* $\mathbb{R}$ *for* $t \in [0,+\infty)$*, set* $Z_i(t) = \psi_i(X_i(t),t)$*. If not, define*

$$Z_i(t) := \psi_i(X_i(t),t) := \int_\xi^x \frac{1}{\sigma_{i,i}(u,t)}\,\mathrm{d}u\bigg|_{x=X_i(t)}, \quad i = 1,2,\ldots,n,$$

where ξ *is an arbitrary point inside the state space of* $X_i(t)$*.*

Then $Z_i(t)$ *is a one-to-one mapping from* $X_i(t)$ *to* $\mathbb{R}$ *for* $t \in [0,+\infty)$*, and has the following dynamics with time-dependent diffusion*

$$\mathrm{d}Z_i(t) = \left(\frac{\partial\psi_i(\psi_i^{-1}(Z_i(t),t),t)}{\partial t} + \frac{\mu(\boldsymbol{\psi}^{-1}(\mathbf{Z}(t),t),t)}{\sigma_{i,i}(\psi_i^{-1}(Z_i(t),t),t)} - \frac{1}{2}\frac{\partial\sigma_{i,i}(\psi_i^{-1}(Z_i(t),t),t)}{\partial X_i}\right)\mathrm{d}t + \sum_{j=1}^{n}\bar{m}_{i,j}(t)\,\mathrm{d}\widetilde{W}_j(t), \tag{5.17}$$

where $\bar{m}_{i,j}(t)$ *with* $i,j = 1,2,\ldots,n$ *are elements of* $\bar{\mathbf{M}}(t)$*, and, due to the conditions imposed on the diffusion matrix* $\boldsymbol{\sigma}$*, it was possible to use that* $\mathbf{X}(t) = \boldsymbol{\psi}^{-1}(\mathbf{Z}(t),t)$*.*

Proof. The proof of Theorem 2.1.6 can be applied to each individual process $X_i(t)$ for $i = 1,2,\ldots,n$. □

Theorem 5.3.1 shows a less general version of the Lamperti transform applied in a multi-variate setting with state and time-dependent diffusion terms. This additional time-dependency results in the transformation not providing a process with unit diffusion, making it come up short on the true definition of the Lamperti transform. However, there is something to be said about whether it is actually desirable to include the time-dependent component in the transform in order to obtain unit diffusion in Equation (5.17). In a real-world setting, there are two main reasons a time-dependency might be imposed on the diffusion component of a process. First, it is a good way to capture the seasonality of a process, which is very useful when modeling processes like the price of electricity or gas. Second, it might be done to let some real-world observations influence the dynamics of the original process. However, in this case if one were to include the time-dependent diffusion components in the transform, the first term in Equation (5.17) might pose a problem. This is because one would be forced to take a derivative with respect to time of a set of observations, leading to the need for numerical differentiation. It might of course happen that this derivative can also be observed, in which case this differentiation issue does not exist, yet its inclusion in the transform might still result in troublesome expressions down the line. Moreover, this additional time-dependent component might become zero, resulting in a division by zero in the transform immediately halting simulations. A good example of such a transformation can be found in Møller-Madsen [43].

As the previous paragraph made clear, it highly depends on the situation whether or not it is a good decision to include a time-dependent component in the Lamperti transform. However, for completeness, and to avoid any difficult integrals stemming from inclusion of the time-dependent component in the transform (if one decides to do so), the following theorem is stated, showing the steps required for successfully transforming Equation (5.17) to a process with unit diffusion without altering the original transform defined in Theorem 5.3.1.

Theorem 5.3.2. *Given a state vector of Itô processes* $\mathbf{X}$ *with*

$$\mathrm{d}\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t),t)\,\mathrm{d}t + \bar{\mathbf{M}}(t)\,\mathrm{d}\widetilde{\mathbf{W}}(t),$$

where $\boldsymbol{\mu} : A \to \mathbb{R}^n$, $\bar{\mathbf{M}} \coloneqq \mathbf{ML} : A \to \mathbb{R}^{n\times n}$ *an invertible matrix of functions of* t *in which* $\mathbf{L} \in \mathbb{R}^{n\times n}$ *is a lower triangular matrix with positive diagonal entries originating from the Cholesky decomposition of correlated Brownian motions and* $\widetilde{\mathbf{W}}(t) \in \mathbb{R}^n$ *a column vector of independent Brownian motions. Then*

$$\mathbf{Z}(t) \coloneqq \boldsymbol{\psi}(\mathbf{X}(t),t) \coloneqq \bar{\mathbf{M}}^{-1}(t)\mathbf{X}(t), \tag{5.18}$$

defines a transformation vector resulting in a new state vector of Itô processes with dynamics given by

$$\mathrm{d}\mathbf{Z}(t) = \left(\frac{\mathrm{d}\bar{\mathbf{M}}^{-1}(t)}{\mathrm{d}t}\bar{\mathbf{M}}(t)\mathbf{Z}(t) + \bar{\mathbf{M}}^{-1}(t)\boldsymbol{\mu}(\bar{\mathbf{M}}(t)\mathbf{Z}(t),t)\right)\mathrm{d}t + \mathrm{d}\widetilde{\mathbf{W}}(t), \tag{5.19}$$

where the derivative of the matrix function is defined element-wise, and, due to the conditions imposed on $\bar{\mathbf{M}}(t)$*, it was possible to use* $\mathbf{X}(t) = \bar{\mathbf{M}}(t)\mathbf{Z}(t)$.

Proof. The result follows from a straightforward application of Lemma 5.1.1 to an element of the transformation defined in Equation (5.18). Luckily, the cumbersome double and cross derivatives that typically show up in this multi-dimensional setting will not play a role here. This can easily be seen by looking at the i^{th} elements of $\mathrm{d}\mathbf{X}(t)$ and $\mathbf{Z}(t)$, which are

$$\mathrm{d}X_i(t) = \mu_i(\mathbf{X}(t),t)\,\mathrm{d}t + \sum_{j=1}^{n} \bar{m}_{i,j}(t)\,\mathrm{d}\widetilde{W}_j(t), \quad i = 1,2,\ldots,n, \tag{5.20}$$

and

$$Z_i(t) \coloneqq \psi_i(\mathbf{X}(t),t) \coloneqq \sum_{j=1}^{n} \bar{m}^{-1}_{i,j}(t)X_j(t), \quad i = 1,2,\ldots,n, \tag{5.21}$$

where $\bar{m}_{i,j}(t)$ and $\bar{m}^{-1}_{i,j}(t)$ with $i,j = 1,2,\ldots,n$ are elements of $\bar{\mathbf{M}}(t)$ and $\bar{\mathbf{M}}^{-1}(t)$, respectively. Indeed, since Equation (5.21) is a linear combination of $X_i(t)$ terms, it turns out that

$$\frac{\partial^2\psi_i(\mathbf{X}(t),t)}{\partial X_i \partial X_j} = \frac{\partial^2\psi_i(\mathbf{X}(t),t)}{\partial X_i^2} = 0, \quad i,j = 1,2,\ldots,n.$$

Therefore, applying multi-dimensional Itô's lemma to Equation (5.21) results in

$$\mathrm{d}Z_i(t) = \frac{\partial\psi_i(\mathbf{X}(t),t)}{\partial t}\,\mathrm{d}t + \sum_{j=1}^{n}\frac{\partial\psi_i(\mathbf{X}(t),t)}{\partial X_j}\,\mathrm{d}X_j(t).$$

Substituting in Equation (5.20) and (5.21) gives

$$\begin{aligned}
\mathrm{d}Z_i(t) &= \frac{\partial}{\partial t}\sum_{j=1}^{n}\bar{m}^{-1}_{i,j}(t)X_j(t)\,\mathrm{d}t + \sum_{j=1}^{n}\bar{m}^{-1}_{i,j}(t)\left(\mu_j(\mathbf{X}(t),t)\,\mathrm{d}t + \sum_{k=1}^{n}\bar{m}_{j,k}(t)\,\mathrm{d}\widetilde{W}_k(t)\right)\\
&= \left(\frac{\partial}{\partial t}\sum_{j=1}^{n}\bar{m}^{-1}_{i,j}(t)X_j(t) + \sum_{j=1}^{n}\bar{m}^{-1}_{i,j}(t)\mu_j(\mathbf{X}(t),t)\right)\mathrm{d}t + \sum_{j=1}^{n}\bar{m}^{-1}_{i,j}(t)\sum_{k=1}^{n}\bar{m}_{j,k}(t)\,\mathrm{d}\widetilde{W}_k(t)
\end{aligned}$$

Grouping terms and substituting the inverse of Equation (5.21) yields

$$\begin{aligned}
&= \left(\sum_{j=1}^{n}\frac{\mathrm{d}\bar{m}^{-1}_{i,j}(t)}{\mathrm{d}t}\sum_{k=1}^{n}\bar{m}_{j,k}(t)Z_k(t) + \sum_{j=1}^{n}\bar{m}^{-1}_{i,j}(t)\mu_j(\mathbf{X}(t),t)\right)\mathrm{d}t + \sum_{k=1}^{n}\sum_{j=1}^{n}\bar{m}^{-1}_{i,j}(t)\bar{m}_{j,k}(t)\,\mathrm{d}\widetilde{W}_k(t)\\
&= \left(\sum_{k=1}^{n}\sum_{j=1}^{n}\frac{\mathrm{d}\bar{m}^{-1}_{i,j}(t)}{\mathrm{d}t}\bar{m}_{j,k}(t)Z_k(t) + \sum_{j=1}^{n}\bar{m}^{-1}_{i,j}(t)\mu_j(\mathbf{X}(t),t)\right)\mathrm{d}t + \sum_{k=1}^{n}\sum_{j=1}^{n}\bar{m}^{-1}_{i,j}(t)\bar{m}_{j,k}(t)\,\mathrm{d}\widetilde{W}_k(t)
\end{aligned}$$

At this point it is a good idea to revert back to matrix notation, cancelling out multiple summation terms

$$= \left(\sum_{k=1}^{n}\left(\frac{\mathrm{d}\bar{\mathbf{M}}^{-1}(t)}{\mathrm{d}t}\bar{\mathbf{M}}(t)\right)_{i,k}Z_k(t) + \left(\bar{\mathbf{M}}^{-1}(t)\boldsymbol{\mu}(\mathbf{X}(t),t)\right)_i\right)\mathrm{d}t + \sum_{k=1}^{n}\left(\bar{\mathbf{M}}^{-1}(t)\bar{\mathbf{M}}(t)\right)_{i,k}\mathrm{d}\widetilde{W}_k(t)$$

Finalizing the matrix notation and making use of the fact that for an arbitrary invertible matrix $\mathbf{A}$ it holds that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix, results in

$$\mathrm{d}Z_i(t) = \left(\left(\frac{\mathrm{d}\bar{\mathbf{M}}^{-1}(t)}{\mathrm{d}t}\bar{\mathbf{M}}(t)\mathbf{Z}(t)\right)_i + \left(\bar{\mathbf{M}}^{-1}(t)\boldsymbol{\mu}(\mathbf{X}(t),t)\right)_i\right)\mathrm{d}t + \mathrm{d}\widetilde{W}_i(t).$$

The result of the derivation above holds for an arbitrary $i \in \{1,2,\ldots,n\}$ and since $\mathbf{X}(t) = \bar{\mathbf{M}}(t)\mathbf{Z}(t)$, it can be generalized to give the final result

$$\mathrm{d}\mathbf{Z}(t) = \left(\frac{\mathrm{d}\bar{\mathbf{M}}^{-1}(t)}{\mathrm{d}t}\bar{\mathbf{M}}(t)\mathbf{Z}(t) + \bar{\mathbf{M}}^{-1}(t)\boldsymbol{\mu}(\bar{\mathbf{M}}(t)\mathbf{Z}(t),t)\right)\mathrm{d}t + \mathrm{d}\widetilde{\mathbf{W}}(t), \tag{5.22}$$

concluding the proof. □

Now that the required theorems for obtaining unit diffusion in multi-dimensional SDE systems have been stated, it will be appropriate to showcase the application of these theorems with an example.

Multi-dimensional modified CKLS model

The Chan-Karolyi-Longstaff-Sanders (CKLS) models [33] form a family of models that pose solutions to SDE's of the form

$$\mathrm{d}X(t) = (a + bX(t))\,\mathrm{d}t + cX^d(t)\,\mathrm{d}W(t). \tag{5.23}$$

These models are generally used for financial applications to model real-world processes on which options or other derivatives can be priced (see e.g. Chan *et al.* [17]). In order for the Lamperti transform to be applicable to Equation (5.23), the four constant model parameters a, b, c and d must be defined such that $a, b, c > 0$ and $d > \frac{1}{2}$, in which case the state-space of $X(t)$ is strictly positive.

In the setting of this section, it will be an interesting test to extend this process with an additional time-dependent component in the diffusion term, followed by application of the multi-dimensional Lamperti transform to a two-dimensional state vector containing two of these new processes. This additional time-dependent component might increase the flexibility of the process in terms of real-life applications, but will mostly serve to highlight the usage of Theorem 5.3.1 and 5.3.2. Let $\mathbf{X}$ be a state vector of two of these modified CKLS processes under an arbitrary measure with dynamics defined by

$$\mathrm{d}\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t),t)\,\mathrm{d}t + \boldsymbol{\sigma}(\mathbf{X}(t),t)\bar{\mathbf{M}}(t)\,\mathrm{d}\widetilde{\mathbf{W}}(t), \quad \mathbf{X}(0) = x_0 \tag{5.24}$$

which, after fully writing out the vectors and matrices, becomes

$$\begin{aligned}\mathrm{d}\begin{bmatrix}X_1(t)\\X_2(t)\end{bmatrix} &= \begin{bmatrix}(a_1 + b_1X_1(t))\\(a_2 + b_2X_2(t))\end{bmatrix}\mathrm{d}t + \begin{bmatrix}c_1X_1^{d_1}(t) & 0\\0 & c_2X_2^{d_2}(t)\end{bmatrix}\begin{bmatrix}m_{1,1}(t) + \rho_{1,2}m_{1,2}(t) & \sqrt{1-\rho_{1,2}^2}m_{1,2}(t)\\ m_{2,1}(t) + \rho_{1,2}m_{2,2}(t) & \sqrt{1-\rho_{1,2}^2}m_{2,2}(t)\end{bmatrix}\mathrm{d}\begin{bmatrix}\widetilde{W}_1(t)\\\widetilde{W}_2(t)\end{bmatrix}\\ &=: \begin{bmatrix}(a_1 + b_1X_1(t))\\(a_2 + b_2X_2(t))\end{bmatrix}\mathrm{d}t + \begin{bmatrix}c_1X_1^{d_1}(t) & 0\\0 & c_2X_2^{d_2}(t)\end{bmatrix}\begin{bmatrix}\bar{m}_{1,1}(t) & \bar{m}_{1,2}(t)\\ \bar{m}_{2,1}(t) & \bar{m}_{2,2}(t)\end{bmatrix}\mathrm{d}\begin{bmatrix}\widetilde{W}_1(t)\\\widetilde{W}_2(t)\end{bmatrix}\end{aligned} \tag{5.25}$$

where the notation convention adopted in the preceding theorems is used once more and it holds that $W_1(t)$ and $W_2(t)$ are two independent Brownian motions and $\rho_{1,2} \in [-1,1]$. Theorem 5.3.1 states that the first vector of transformations can now be constructed in order to obtain state-independent diffusion. The elements of this transformation vector $\mathbf{Z}$ are defined as

$$Z_1(t) \coloneqq \psi_1(X_1(t)) \coloneqq \int \frac{1}{c_1x^{d_1}}\,\mathrm{d}x\bigg|_{x=X_1(t)}, \quad Z_1(0) = 0, \tag{5.26}$$

$$Z_2(t) \coloneqq \psi_2(X_2(t)) \coloneqq \int \frac{1}{c_2x^{d_2}}\,\mathrm{d}x\bigg|_{x=X_2(t)}, \quad Z_2(0) = 0, \tag{5.27}$$

which can be further evaluated into the invertible expressions

$$Z_1(t) = \psi_1(X_1(t)) = \frac{X_1^{1-d_1}(t) - x_0^{1-d_1}}{c_1 - c_1d_1}, \tag{5.28}$$

$$Z_2(t) = \psi_2(X_2(t)) = \frac{X_2^{1-d_2}(t) - x_0^{1-d_2}}{c_2 - c_2d_2}. \tag{5.29}$$

These transformations are essentially the same as they would be defined for the one-dimensional case, but the time-dependent diffusion component in the original process causes the dynamics of the processes in $\mathbf{Z}$ to take on a different form. Indeed, Theorem 5.3.1 describes that the dynamics of the processes in $\mathbf{Z}$ in this case must be defined as

$$\mathrm{d}\begin{bmatrix} Z_1(t) \\ Z_2(t) \end{bmatrix} = \begin{bmatrix} (a_1 + b_1\psi_1^{-1}(Z_1(t)))\frac{1}{c_1(\psi_1^{-1}(Z_1(t)))^{d_1}} - \frac{1}{2}c_1 d_1(\psi_1^{-1}(Z_1(t)))^{d_1-1} \\ (a_2 + b_2\psi_2^{-1}(Z_2(t)))\frac{1}{c_2(\psi_2^{-1}(Z_2(t)))^{d_2}} - \frac{1}{2}c_2 d_2(\psi_2^{-1}(Z_2(t)))^{d_2-1} \end{bmatrix}\mathrm{d}t + \begin{bmatrix} \bar{m}_{1,1}(t) & \bar{m}_{1,2}(t) \\ \bar{m}_{2,1}(t) & \bar{m}_{2,2}(t) \end{bmatrix}\mathrm{d}\begin{bmatrix} \widetilde{W}_1(t) \\ \widetilde{W}_2(t) \end{bmatrix}. \tag{5.30}$$

It was discussed earlier that including the time-dependent component of the diffusion term in the transform is undesirable as it might lead to having to deal with cumbersome function expressions. However, even though state-independent diffusion has been obtained through application of the first transform, Equation (5.30) still lacks unit diffusion. This is where application of Theorem 5.3.2 becomes appropriate. To this end, note that

$$\bar{\mathbf{M}}^{-1}(t) = \frac{1}{\det(\bar{\mathbf{M}}(t))}\begin{bmatrix} \bar{m}_{2,2}(t) & -\bar{m}_{1,2}(t) \\ -\bar{m}_{2,1}(t) & \bar{m}_{1,1}(t) \end{bmatrix}, \tag{5.31}$$

which allows for explicit definition of the transform

$$\mathbf{Y}(t) \coloneqq \boldsymbol{\varphi}(\mathbf{Z}(t), t) \coloneqq \bar{\mathbf{M}}^{-1}(t)\mathbf{Z}(t). \tag{5.32}$$

According to Theorem 5.3.2, the newly defined transform above can be described by the dynamics

$$\mathrm{d}\mathbf{Y}(t) = \left(\frac{\mathrm{d}\bar{\mathbf{M}}^{-1}(t)}{\mathrm{d}t}\bar{\mathbf{M}}(t)\mathbf{Y}(t) + \bar{\mathbf{M}}^{-1}(t)\boldsymbol{\mu}(\bar{\mathbf{M}}(t)\mathbf{Y}(t), t)\right)\mathrm{d}t + \mathrm{d}\widetilde{\mathbf{W}}(t). \tag{5.33}$$

Equation (5.33) becomes very large when filling in all the separate components explicitly, hence the full expression has been omitted for simplicity. This example showed that the double transformation indeed results in a process described by dynamics containing unit diffusion. The finalized dynamics are more tractable in a computational setting due to the presence of unit diffusion. Similar to the one-dimensional cases, this additional tractability is reflected in the fact that one does not have to take care of any boundary conditions when modeling the dynamics shown in Equation (5.33). Moreover, due to application of the Cholesky decomposition, the correlation structure originally present in the diffusion component has been successfully transferred to the drift term.

Discussion on multi-dimensional applications

At this point it is relevant to note that, as of now, the the multi-dimensional Lamperti transform sees most of its applications in the field of statistical inference and parameter estimation. Unlike its one-dimensional version, the amount of processes with applications in the field of financial modeling to which the multi-dimensional transform can be applied has not been extensively explored yet in current research. While this thesis lays an emphasis on the usage of the Lamperti transform in financial SLV models, and thus does not go into in-depth analysis of its strengths in a parameter estimation, a short discussion on a real-world application of parameter estimation will be appropriate. A rather natural multi-dimensional system that might come to mind is a competition model defined through the competitive Lotka-Volterra equations (see Bomze [15]), more commonly referred to as a predator-prey model. These models see a lot of usage in certain fields of biology, where accurate predictions of the stability of natural equilibria are necessary. Predator-prey models quite obviously consist of at least two dependent equations, typically containing two dependent sources of randomness. Enforcing time-dependency on the diffusion term of these processes is the ideal way to introduce the concept of seasonality into, for example, the survival rate of the predator and/or prey. It becomes clear that Theorem 5.3.1 and 5.3.2 are much more relevant in this setting (see Møller-Madsen [43] for a detailed example). This is because removal of state-dependent diffusion allows for a much easier application of the EKF as it was discussed in Chapter 2, which is an ideal method for variable predictions in these types of models.

6

Conclusion

This thesis explored the many different applications of a rather unpopular transformation of SDE's, capable of improving the stability of simulation and parameter estimation procedures in both one-dimensional and multi-dimensional settings. The transformation, most commonly named the Lamperti transform, is most often used as a one-to-one mapping from the state-space of a process with state-dependent diffusion to that of a process with unit diffusion. Diffusion terms typically enforce boundary conditions imposed on a system, hence the presence of state-dependency brings forth numerous difficulties in parameter estimation procedures such as e.g. the Extended Kalman filter [6]. Moreover, having to deal with potentially stringent boundary conditions can quickly complicate simulation schemes in an MC framework.

In the body of this document, three main applications of the Lamperti transform in a financial setting were discussed. First, numerous examples were given to showcase its ability to map the state-space of a process with state-dependent diffusion to that of a process with unit diffusion. Some of these examples were coupled with numerical experiments to verify that one indeed does not have to take into account any of the original boundary conditions when simulating the transformed process. This is especially beneficial when a process is subjected to a rather stringent boundary condition such as e.g. an absorbing origin.

Second, a brief introduction was given into rough volatility models [28, 45]. These models are relatively novel when it comes to modeling volatility, but, while in a different way, the Lamperti transform still enjoys usage. Rough volatility models make use of a process called fractional Brownian motion, which is essentially a standard Brownian motion with a modified covariance function. This covariance function contains an important parameter, the Hurst parameter, that controls the correlation between the increments of the fractional Brownian motion. In other words, the Hurst parameter is in control of the memory of the process. Nuzman-Poor [45] state an alternative definition of the Lamperti transform under which the main result of Lamperti [38] can be elegantly utilized, and use it to define estimation formulae for H-self-similar processes such as a fractional Brownian motion. Rough volatility models heavily rely on the availability of market data, hence formulae for estimating model parameters are of high value. In this setting the Lamperti transform is virtually exclusively used as a one-to-one mapping, rather than as a transformation to obtain dynamics more suitable for simulation, as was showcased in the first part of Chapter 2.

Third, the most prominent application of the transform in this thesis was discussed. Starting off, a general SLV model was defined, consisting of three SDE's. This model, typically written in two equation form, made use of an auxiliary process to isolate the LV component. In turn, this allowed for application of the Lamperti transform, eventually yielding the definition of an alternative expression of the original process in terms of its volatility. This new expression formed the key to fully solving the model, as there now existed a functional definition for the original process that could be used in future derivations.

Using the aforementioned functional definition for the original process, three sets of tractable arbitrage-free pricing functions were derived using methods inspired by Bang [9]. This strategy of deriving three different sets of functions came from application of the Lamperti transform, as it allowed for the price of an option on a certain underlying to be expressed in prices of options on the volatility of that underlying. Numerous assumptions were made to get to this point, but numerical experiments showed that their significance was severely diminished by the benefit of being able to combine the model with any arbitrary well-defined LV

function. This quality was once more a direct result of the application of the Lamperti transform, further highlighting its strength in this setting. The only potential bottleneck in terms of computational tractability were the improper integrals present in some of the pricing functions. For this reason, a heuristic method was discussed for estimating appropriate truncation points for the infinite integration intervals.

Having defined explicit formulations for the option pricing functions, their output was compared to models already established in the financial industry. Selecting a constant LV function reduced the general SLV model to the normal SABR model, making it straightforward to compare the output of the new solution, the Transform solution, to Hagan's asymptotical approximation formula [31] and Antonov's exact approach [3]. Results showed that the Transform solution is highly competitive with existing models and, unlike Hagan's approximation formula, maintains its accuracy when pricing options with large maturities, i.e. larger than 20 years. However, this accuracy comes with increased numerical costs, resulting in a full calibration scheme of ten iterations taking approximately two minutes to execute.

After some basic comparisons of different models, experimentation was done with the LV function. Recall that the Transform solution enjoys the property of being compatible with any continuously differentiable, strictly positive LV function, hence putting this to the test was a natural continuation of the thesis. The custom LV function was defined based on the requirement that the distribution of the original process had to be able to take on negative values, while simultaneously resulting in a flexible model capable of appropriately calibrating to market prices and propagating user opinions on market conditions. For the calibration procedure, the Levenberg-Marquardt algorithm [21, 29] was used in combination with some additional preliminary and intermediary steps to accurately maintain preservation of the forward as well as reprice the ATM options. It turned out that the selection for the LV function indeed satisfied the aforementioned requirements as the calibration algorithm was capable of fitting to market prices with an accuracy of approximately 10e-4 after only ten iterations.

Up until Section 4.6, not much attention had been paid to the combination of the Transform solution with non-parametric LV functions. This is because their inherent lack of parameters limits the model user in including opinions on market conditions into the model output. However, for completeness of the thesis, and to further put to the test the robustness of the Transform solution, Section 4.6 was created. For the non-parametric function, the famous expression for the LV known as Dupire's formula [24] was used. Dupire's formula allows for the construction of an LV surface by directly implying it from implied volatilities as seen in the market. Generating a set of market prices using Antonov's exact solution to the normal SABR model therefore allowed for the construction of an LV curve according to Dupire's formula. After taking care of minor numerical complexities, the curve could be directly implemented in the Transform solution and the resulting model was calibrated to the earlier generated set of market prices. In the end, a close fit to market prices was obtained once more, where most remaining inaccuracies were caused by the extrapolation of the LV curve necessary for compatibility with the Transform solution. The section did, however, further highlight the flexibility of the Transform solution.

Finalizing the document, an investigation was done into the applications of the Lamperti transform in a multi-dimensional setting. First, it was attempted to define a vector of transformations that could be applied to a generalized SLV model proposed by van der Stoep *et al.* [50]. This approach resulted in a system of non-linear inhomogeneous first-order PDE's always containing one unknown more than there exist equations in the system. Having concluded that a fully general approach is in general not at all straightforward, a less general result was stated. In a multi-variate setting, the Lamperti transform can still be defined and applied in its typical form provided that the multi-dimensional system does not contain diffusion terms dependent on more than a single state-variable. It was shown that the transform could even be applied to systems with an added time-dependency in the diffusion term. While it is usually no problem to have a time-dependent diffusion term in estimation procedures such as the Extended Kalman filter, it was shown for completeness how to fully transform to unit diffusion using a single auxiliary theorem. Lastly, it was stated that the multi-dimensional Lamperti transform currently enjoys only limited applications in financial modeling, simply because there are currently few SDE systems without mixed diffusion terms that are used in this field. However, the final chapter was finished off with a short introduction to an application of the multi-variate transform to predator-prey models [15], and a reference to a fully worked example [43] with biology applications beyond the scope of this thesis.

A

Preliminary lemmas

This appendix serves the purpose of stating and proving lemmas required for deriving the pricing functions discussed in this thesis. Additional details surrounding the lemmas goes beyond the scope of this thesis, and will therefore not be covered. The somewhat indirect relation to the main subject of the thesis is the reason these lemmas have been moved to this appendix.

A.1. Carr-Madan integral representation

Lemma A.1.1. *For any twice-differentiable function $f : \mathbb{R} \to \mathbb{R}^+$ the following identity holds*

$$f(S) = f(a) + f'(a)(S-a) + \int_a^{+\infty} f''(v)(S-v)^+ \,\mathrm{d}v + \int_{-\infty}^{a} f''(v)(v-S)^+ \,\mathrm{d}v, \tag{A.1}$$

where $S > 0$, a is a positive constant, $\mathbb{R}^+$ denotes the set of strictly positive real numbers and for an arbitrary function f and dummy variable x the notation $f'(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x}$ and $f''(x) = \frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2}$ is used.

Proof. Given a twice differentiable function f, positive constant a and $S > 0$, the following holds

$$\begin{aligned}
f(S) - f(a) &= \int_a^S f'(u)\,\mathrm{d}u \\
&= \int_a^S \big(f'(u) - f'(a) + f'(a)\big)\,\mathrm{d}u \\
&= f'(a)(S-a) + \int_a^S \int_a^u f''(v)\,\mathrm{d}v\,\mathrm{d}u \\
&= f'(a)(S-a) + \int_a^S \int_v^S f''(v)\,\mathrm{d}u\,\mathrm{d}v \\
&= f'(a)(S-a) + \int_a^S f''(v)(S-v)\,\mathrm{d}v,
\end{aligned}$$

which follows from basic properties of integration. Continuing

$$\begin{aligned}
f(S) &= f(a) + f'(a)(S-a) + \int_a^S f''(v)(S-v)\,\mathrm{d}v \\
&= f(a) + f'(a)(S-a) + \int_a^S \mathbb{1}_{a\leq S} f''(v)(S-v)\,\mathrm{d}v + \int_a^S \mathbb{1}_{a>S} f''(v)(S-v)\,\mathrm{d}v
\end{aligned}$$

The presence of the indicator function then allows for further rewriting and changing of integral bounds, i.e.

$$\begin{aligned}
&= f(a) + f'(a)(S-a) + \int_a^S \mathbb{1}_{a\leq S} f''(v)(S-v)^+\,\mathrm{d}v + \int_S^a \mathbb{1}_{a>S} f''(v)(v-S)^+\,\mathrm{d}v \\
&= f(a) + f'(a)(S-a) + \int_a^{+\infty} \mathbb{1}_{a\leq S} f''(v)(S-v)^+\,\mathrm{d}v + \int_{-\infty}^a \mathbb{1}_{a>S} f''(v)(v-S)^+\,\mathrm{d}v \\
&= f(a) + f'(a)(S-a) + \int_a^{+\infty} f''(v)(S-v)^+\,\mathrm{d}v + \int_{-\infty}^a f''(v)(v-S)^+\,\mathrm{d}v.
\end{aligned}$$

□

A.2. Function inverse identities

Lemma A.2.1. *Given a strictly positive and continuous function $f : \mathbb{R} \to \mathbb{R}^+$, where $\mathbb{R}^+$ denotes the set of strictly positive real numbers, with the following relation*

$$g(x) := \int_a^x \frac{1}{f(u)}\,\mathrm{d}u.$$

The following relations hold for arbitrary x

$$(g^{-1})'(g(x)) = f(x) \qquad \textit{and} \qquad (g^{-1})''(g(x)) = f'(x)f(x),$$

where the notation $f'(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x}$ and $f''(x) = \frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2}$ for an arbitrary function f and dummy variable x, is used.

Proof. The following identity can be used

$$(g^{-1})'(x) = \frac{1}{g'(g^{-1}(x))}.$$

Rewriting the first relation then yields

$$(g^{-1})'(g(x)) = (g^{-1})'(g(x)) = \frac{1}{g'(g^{-1}(g(x)))} = \frac{1}{g'(x)},$$

and, since it holds that $g'(x) = \frac{1}{f(x)}$, the equation above reduces to

$$(g^{-1})'(g(x)) = \frac{1}{g'(x)} = f(x),$$

which proves the first relation.

For the second relation the chain rule can be applied, which yields

$$((g^{-1})'(g))'(x) = (g^{-1})''(g(x))g'(x) \implies (g^{-1})''(g(x)) = \frac{((g^{-1})'(g))'(x)}{g'(x)}.$$

Applying the first relation stated in the lemma to this equation then gives

$$(g^{-1})''(g(x)) = \frac{((g^{-1})'(g))'(x)}{g'(x)} = \frac{f'(x)}{\frac{1}{f(x)}} = f'(x)f(x),$$

which proves the second relation. □

A.3. Lévy's characterization

Lemma A.3.1. *Given d local continuous martingales, denoted by $W_1, \ldots, W_d$ with $W_i(0) = 0$ for all $i = 1, \ldots, d$ and $\langle W_i, W_j \rangle(t) = \delta_i(j)\,t$, where $\delta_i(j)$ is the Kronecker delta. Then $(W_1, \ldots, W_d)$ is a d-dimensional Brownian motion.*

Proof. See Wang [51] for a proof of a simplified version of the lemma. □

B

Supporting proofs

In this thesis a number of relatively complex theorems and proofs are showcased. In order to keep the main text concise and easy to follow, some statements and proofs have been moved to this appendix. While the contents of this appendix generally play an auxiliary role in the derivation of numerous results discussed in this thesis, they play an important part in obtaining an overview of all details that went into derivation of said results. In case the reader is merely interested in a superficial view of the results covered in this thesis, the main text and the shown sketches of the proofs discussed in this appendix will suffice. Naturally, this appendix makes use of terminology, notation and definitions defined in the main text, without unnecessarily repeating them.

B.1. Application of Gyöngy's lemma

Given the normal SABR model in its standard form

$$\begin{aligned} \mathrm{d}\eta(t) &= \alpha(t)\,\mathrm{d}W_\eta^T(t), \\ \mathrm{d}\alpha(t) &= \nu\alpha(t)\,\mathrm{d}W_\alpha^T(t), \end{aligned}$$

where $W_\eta^T(t)$ and $W_\alpha^T(t)$ are two correlated Brownian motions under the T forward measure with $\mathrm{d}W_\eta^T(t)\,\mathrm{d}W_\alpha^T(t) = \rho\,\mathrm{d}t$ for $\rho \in [-1,1]$. Gyöngy's lemma [30] then states that there exists some process $\tilde{\eta}(t)$ with equal marginal distributions to $\eta(t)$, and dynamics

$$\begin{aligned} \mathrm{d}\tilde{\eta}(t) &= \sqrt{V(t,\tilde{\eta}(t))}\,\mathrm{d}W_\eta^T(t), \\ \mathrm{d}\alpha(t) &= \nu\alpha(t)\,\mathrm{d}W_\alpha^T(t), \end{aligned}$$

where

$$V(t,x) := \phi^2(x)\mathbb{E}^T\left[\alpha^2(t)|\eta(t)=x\right].$$

The following lemma can now be stated.

Lemma B.1.1. *The function $V(t,x)$ as defined above, is quadratic, which means that the dynamics for $\tilde{\eta}(t)$ can be written as*

$$\mathrm{d}\tilde{\eta}(t) = \alpha(0)\sqrt{1+2\rho\nu f(K)+\nu^2 f^2(K)}\,\mathrm{d}W_\eta^T(t),$$

where

$$f(K) := \frac{1}{\alpha(0)}\int_{\eta(0)}^{K}\frac{1}{\phi(u)}\,\mathrm{d}u.$$

Proof. Based on research done by Balland-Tran [8] on deriving an explicit expression for the SABR local volatility, the following identity can be used

$$\mathbb{E}^T\left[\alpha^2(t)|\eta(t)=K\right] = \frac{\mathbb{E}^T\left[\alpha^2(t)\delta(\eta(t)-K)\right]}{\mathbb{E}^T\left[\delta(\eta(t)-K)\right]}, \tag{B.1}$$

where δ is the Dirac delta function [23]. The proof of this lemma will be subdivided into two parts, where each part aims to find an alternative expression for the numerator and denominator of Equation (B.1), respectively. In order rewrite the numerator of the right hand side of the above equation, an auxiliary process $J(t)$ can be introduced. This process is defined as

$$J(t) \equiv J(\eta(t), \alpha(t)) \coloneqq \frac{1}{\alpha(t)}(\eta(t) - K). \tag{B.2}$$

Using the fact that $\alpha(t)$ is a Geometric Brownian motion, the process $J(t)$ can be used to write the following

$$\begin{aligned}\mathbb{E}^T\left[\alpha^2(t)\delta(\eta(t)-K)\right] &= \mathbb{E}^T\left[\alpha(t)\delta(J(t))\right] \\ &= \alpha(0)\mathbb{E}^T\left[\mathrm{e}^{\nu W_\alpha^T(t) - \frac{1}{2}\nu^2 t}\delta(J(t))\right],\end{aligned} \tag{B.3}$$

which holds by characteristics of the Dirac delta function δ when it is composed with a smooth function. Continuing, a new measure $\hat{\mathbb{Q}}$ defined through $\mathrm{d}\hat{\mathbb{Q}}/\,\mathrm{d}\mathbb{Q}^T = \mathrm{e}^{\nu W_\alpha^T(t) - \frac{1}{2}\nu^2 t}$ can be introduced. Applying this to the equation above straightforwardly yields a new expression for the numerator in Equation (B.1)

$$\mathbb{E}^T\left[\alpha^2(t)\delta(\eta(t)-K)\right] = \alpha(0)\mathbb{E}^{\hat{\mathbb{Q}}}\left[\delta(J(t))\right]. \tag{B.4}$$

Moreover, application of Itô's lemma to $J(t)$ yields

$$\begin{aligned}\mathrm{d}J(t) &= \frac{\partial J(t)}{\partial \alpha(t)}\mathrm{d}\alpha(t) + \frac{\partial J(t)}{\partial \eta(t)}\mathrm{d}\eta(t) + \frac{1}{2}\frac{\partial^2 J(t)}{\partial \alpha^2(t)}(\mathrm{d}\alpha(t))^2 + \frac{1}{2}\frac{\partial^2 J(t)}{\partial \eta^2(t)}(\mathrm{d}\eta(t))^2 + \frac{\partial^2 J(t)}{\partial \alpha(t)\partial\eta(t)}\mathrm{d}\alpha(t)\,\mathrm{d}\eta(t) \\ &= \mathrm{d}W_\eta(t) - \nu J(t)\,\mathrm{d}W_\alpha^T(t) + \nu^2 J(t)\,\mathrm{d}t - \nu\rho\,\mathrm{d}t \\ &= \left(\rho\,\mathrm{d}W_\alpha^T(t) + \sqrt{1-\rho^2}\,\mathrm{d}\tilde{W}^T(t)\right) - \nu J(t)\,\mathrm{d}W_\alpha^T(t) + \nu^2 J(t)\,\mathrm{d}t - \nu\rho\,\mathrm{d}t,\end{aligned} \tag{B.5}$$

where the Cholesky decomposition [47] was used to rewrite $W_\eta^T(t)$ in terms of $W_\alpha^T(t)$ and an independent Brownian motion $\tilde{W}^T(t)$. Using the earlier defined measure $\hat{\mathbb{Q}}$, it is possible to write this expression in terms of two independent Brownian motions under $\hat{\mathbb{Q}}$, denoted by $W_\alpha^{\hat{\mathbb{Q}}}(t)$ and $\tilde{W}^{\hat{\mathbb{Q}}}(t)$

$$\begin{aligned}\mathrm{d}J(t) &= \left(\rho(\mathrm{d}W_\alpha^{\hat{\mathbb{Q}}}(t) + \nu\,\mathrm{d}t) + \sqrt{1-\rho^2}\,\mathrm{d}\tilde{W}^{\hat{\mathbb{Q}}}(t)\right) - \nu J(t)(\mathrm{d}W_\alpha^{\hat{\mathbb{Q}}}(t) + \nu\,\mathrm{d}t) + \nu^2 J(t)\,\mathrm{d}t - \nu\rho\,\mathrm{d}t \\ &= \left(\rho\,\mathrm{d}W_\alpha^{\hat{\mathbb{Q}}}(t) + \sqrt{1-\rho^2}\,\mathrm{d}\tilde{W}^{\hat{\mathbb{Q}}}(t)\right) - \nu J(t)\,\mathrm{d}W_\alpha^{\hat{\mathbb{Q}}}(t) \\ &= (\rho - \nu J(t))\,\mathrm{d}W_\alpha^{\hat{\mathbb{Q}}}(t) + \sqrt{1-\rho^2}\,\mathrm{d}\tilde{W}^{\hat{\mathbb{Q}}}(t) \\ &= \sqrt{1 - 2\nu J(t)\rho + \nu^2 J^2(t)}\,\mathrm{d}B^{\hat{\mathbb{Q}}}(t),\end{aligned} \tag{B.6}$$

where Lemma B.1.2 proves that $B^{\hat{\mathbb{Q}}}(t)$ is a Brownian motion under the $\hat{\mathbb{Q}}$ measure, defined by

$$B^{\hat{\mathbb{Q}}}(t) = \int_0^t \frac{\rho - \nu J(s)}{\sqrt{1 - 2\nu J(s)\rho + \nu^2 J^2(s)}}\,\mathrm{d}W_\alpha^{\hat{\mathbb{Q}}}(s) + \int_0^t \frac{\sqrt{1-\rho^2}}{\sqrt{1 - 2\nu J(s)\rho + \nu^2 J^2(s)}}\,\mathrm{d}\tilde{W}^{\hat{\mathbb{Q}}}(s),$$

Continuing, a similar approach can be taken to rewrite the denominator of Equation (B.1). Again making use of characteristics of the Dirac delta function allows for writing

$$\mathbb{E}^T\left[\delta(\eta(t)-K)\right] = \frac{1}{\alpha(0)}\mathbb{E}^T\left[\mathrm{e}^{-\nu W_\alpha^T(t) + \frac{1}{2}\nu^2 t}\delta(J(t))\right].$$

Through introduction of the change of measure $\mathrm{d}\tilde{\mathbb{Q}}/\,\mathrm{d}\mathbb{Q}^T = \mathrm{e}^{-\nu W_\alpha^T(t) - \frac{1}{2}\nu^2 t}$, the next equality is easily obtained

$$\frac{1}{\alpha(0)}\mathbb{E}^T\left[\mathrm{e}^{-\nu W^\alpha(t) + \frac{1}{2}\nu^2 t}\delta(J(t))\right] = \frac{\mathrm{e}^{\nu^2 t}}{\alpha(0)}\mathbb{E}^{\tilde{\mathbb{Q}}}\left[\delta(J(t))\right]. \tag{B.7}$$

Similar to before, an explicit expression for the dynamics of $J(t)$ under this newly defined measure can be obtained. Recall that an application of Itô's lemma to $J(t)$ earlier resulted in

$$\mathrm{d}J(t) = (\rho\,\mathrm{d}W_\alpha^T(t) + \sqrt{1-\rho^2}\,\mathrm{d}\tilde{W}^T(t)) - \nu J(t)\,\mathrm{d}W^\alpha(t) + \nu^2 J(t)\,\mathrm{d}t - \nu\rho\,\mathrm{d}t.$$

Taking similar steps as before, switching to the earlier defined measure $\tilde{\mathbb{Q}}$ allows for the above expression to be written in terms of two independent Brownian motions under $\tilde{\mathbb{Q}}$, denoted by $W_\alpha^{\tilde{\mathbb{Q}}}$ and $\tilde{W}^{\tilde{\mathbb{Q}}}$

$$\begin{aligned}
&= (\rho(\mathrm{d}W_\alpha^{\tilde{\mathbb{Q}}}(t) - \nu\,\mathrm{d}t) + \sqrt{1-\rho^2}\,\mathrm{d}\tilde{W}^{\tilde{\mathbb{Q}}}(t)) - \nu J(t)(\mathrm{d}W_\alpha^{\tilde{\mathbb{Q}}}(t) - \nu\,\mathrm{d}t) + \nu^2 J(t)\,\mathrm{d}t - \nu\rho\,\mathrm{d}t \\
&= (\rho\,\mathrm{d}W_\alpha^{\tilde{\mathbb{Q}}}(t) + \sqrt{1-\rho^2}\,\mathrm{d}\tilde{W}^{\tilde{\mathbb{Q}}}(t)) - \nu J(t)\,\mathrm{d}W_\alpha^{\tilde{\mathbb{Q}}}(t) + 2\nu^2 J(t)\,\mathrm{d}t - 2\nu\rho\,\mathrm{d}t \\
&= (\rho - \nu J(t))\,\mathrm{d}W_\alpha^{\tilde{\mathbb{Q}}}(t) + \sqrt{1-\rho^2}\,\mathrm{d}\tilde{W}^{\tilde{\mathbb{Q}}}(t) + (2\nu^2 J(t) - 2\nu\rho)\,\mathrm{d}t \\
&= \sqrt{1-2\nu J(t)\rho + \nu^2 J^2(t)}\,\mathrm{d}B^{\tilde{\mathbb{Q}}}(t) + \frac{\partial}{\partial J(t)}\left(1 - 2\nu J(t)\rho + \nu^2 J^2(t)\right)\mathrm{d}t, \qquad \text{(B.8)}
\end{aligned}$$

where, analogous to earlier steps, Lemma B.1.2 shows that $B^{\tilde{\mathbb{Q}}}(t)$ is a Brownian motion under the $\tilde{\mathbb{Q}}$ measure, defined by

$$B^{\tilde{\mathbb{Q}}}(t) = \int_0^t \frac{\rho - \nu J(s)}{\sqrt{1-2\nu J(s)\rho + \nu^2 J^2(s)}}\,\mathrm{d}W_\alpha^{\tilde{\mathbb{Q}}}(s) + \int_0^t \frac{\sqrt{1-\rho^2}}{\sqrt{1-2\nu J(s)\rho + \nu^2 J^2(s)}}\,\mathrm{d}\tilde{W}^{\tilde{\mathbb{Q}}}(s).$$

Due to $J(t)$ not following the same dynamics under $\hat{\mathbb{Q}}$ as under $\tilde{\mathbb{Q}}$, division of Equation (B.4) by Equation (B.7) necessary for evaluation of Equation (B.1) does not cancel out the expectation operators, i.e. $\mathbb{E}^{\hat{\mathbb{Q}}}[\delta(J(t))]\,/\mathbb{E}^{\tilde{\mathbb{Q}}}[\delta(J(t))] \neq 1$, limiting simplification of Equation (B.1). In light of this, let $q(J(t)) \coloneqq 1 - 2\nu J(t)\rho + \nu^2 J^2(t)$, and define

$$\frac{\mathrm{d}\bar{\rho}(t)}{\bar{\rho}(t)} = -\frac{q'(J(t))}{\sqrt{q(J(t))}}\,\mathrm{d}B^{\tilde{\mathbb{Q}}}(t), \qquad \text{(B.9)}$$

$$\bar{\rho}(t) = \exp\left(-\int_0^t \frac{q'(J(s))}{\sqrt{q(J(s))}}\,\mathrm{d}B^{\tilde{\mathbb{Q}}}(s) - \int_0^t \frac{1}{2}\frac{(q'(J(s)))^2}{q(J(s))}\,\mathrm{d}s\right), \qquad \text{(B.10)}$$

where $\bar{\rho}(t)$ is the exponential martingale [46] under the $\tilde{\mathbb{Q}}$ measure. It is now possible to introduce the new measure $\bar{\mathbb{Q}}$ through $\mathrm{d}\bar{\mathbb{Q}}/\,\mathrm{d}\tilde{\mathbb{Q}} = \bar{\rho}(t)$. Clearly, it follows easily from Equation (B.8) that $J(t)$ is a driftless Brownian motion under $\bar{\mathbb{Q}}$. Not only that, but it follows easily that under the $\bar{\mathbb{Q}}$ measure, the process $J(t)$ has the same dynamics as under the $\hat{\mathbb{Q}}$ measure, i.e.

$$dJ(t) = \sqrt{q(J(t))}\,\mathrm{d}B^{\hat{\mathbb{Q}}}(t), \qquad \text{(B.11)}$$

$$dJ(t) = \sqrt{q(J(t))}\,\mathrm{d}B^{\bar{\mathbb{Q}}}(t), \qquad \text{(B.12)}$$

In order to appropriately execute the measure change from $\tilde{\mathbb{Q}}$ to $\bar{\mathbb{Q}}$ in Equation (B.7) such that optimal use is made of the earlier discussed characteristics of the Dirac delta function, a final auxiliary process can be introduced. Let $X(t) \coloneqq \mathrm{e}^{\nu^2 t} q(J(0))/q(J(t))$. Application of Itô's lemma to $X(t)$ then yields

$$\begin{aligned}
\mathrm{d}X(t) &= \frac{\partial X(t)}{\partial t}\,\mathrm{d}t + \frac{\partial X(t)}{\partial J(t)}\,\mathrm{d}J(t) + \frac{1}{2}\frac{\partial^2 X(t)}{\partial J^2(t)}(\mathrm{d}J(t))^2 \\
&= \nu^2 X(t)\,\mathrm{d}t - \frac{q'(J(t))}{q(J(t))}X(t)\,\mathrm{d}J(t) + \frac{1}{2}\frac{2(q^2(J(t)) - q(J(t))q''(J(t))}{q^2(J(t))}X(t)(\mathrm{d}J(t))^2 \\
&= \nu^2 X(t)\,\mathrm{d}t - \frac{q'(J(t))}{q(J(t))}X(t)\sqrt{q(J(t))}\,\mathrm{d}B^{\bar{\mathbb{Q}}}(t) + \frac{1}{2}\frac{2(q'(J(t)))^2 - 2\nu^2 q(J(t))}{q^2(J(t))}X(t)q(J(t))\,\mathrm{d}t \\
&= -\frac{q'(J(t))}{\sqrt{q(J(t))}}X(t)\,\mathrm{d}B^{\bar{\mathbb{Q}}}(t) + \frac{(q'(J(t)))^2}{q(J(t))}X(t)\,\mathrm{d}t.
\end{aligned}$$

Performing a change of measure to the earlier defined measure $\bar{\mathbb{Q}}$ gives

$$\begin{aligned}
&= -\frac{q'(J(t))}{\sqrt{q(J(t))}}X(t)\left(\mathrm{d}B^{\bar{\mathbb{Q}}}(t) + \frac{q'(J(t))}{\sqrt{q(J(t))}}\,\mathrm{d}t\right) + \frac{(q'(J(t)))^2}{q(J(t))}X(t)\,\mathrm{d}t \\
&= -\frac{q'(J(t))}{\sqrt{q(J(t))}}X(t)\,\mathrm{d}B^{\bar{\mathbb{Q}}}(t).
\end{aligned}$$

Finally, comparing the above result to Equation (B.9) shows that

$$\frac{\mathrm{d}X(t)}{X(t)} = \frac{\mathrm{d}\bar{\rho}(t)}{\bar{\rho}(t)},$$

which, after integrating the left and right hand side and cancelling out $X(0) = \bar{\rho}(0) = 1$, results in

$$X(t) = \bar{\rho}(t). \tag{B.13}$$

Based on the way $X(t)$ is defined, it will be useful to rewrite the denominator, i.e. Equation (B.7), to become

$$\frac{\mathrm{e}^{\nu^2 t}}{\alpha(0)}\mathbb{E}^{\tilde{\mathbb{Q}}}\left[\delta(J(t))\right] = \frac{\mathrm{e}^{\nu^2 t}}{\alpha(0)q(J(0))}\mathbb{E}^{\tilde{\mathbb{Q}}}\left[\frac{q(J(0))}{q(J(t))}\delta(J(t))\right], \tag{B.14}$$

which follows directly from the identity

$$\mathbb{E}^{\tilde{\mathbb{Q}}}\left[\delta(J(t))\right] = \mathbb{E}^{\tilde{\mathbb{Q}}}\left[1|J(t)=0\right]p^{\tilde{\mathbb{Q}}}(0), \tag{B.15}$$

where $p^{\tilde{\mathbb{Q}}}$ denotes the density of $J(t)$ under the $\tilde{\mathbb{Q}}$ measure. Finally, Equation (B.14) can then be rewritten into

$$\begin{aligned}\frac{\mathrm{e}^{\nu^2 t}}{\alpha(0)q(J(0))}\mathbb{E}^{\tilde{\mathbb{Q}}}\left[\frac{q(J(0))}{q(J(t))}\delta(J(t))\right] &= \frac{1}{\alpha(0)q(J(0))}\mathbb{E}^{\tilde{\mathbb{Q}}}\left[X(t)\delta(J(t))\right] \\ &= \frac{1}{\alpha(0)q(J(0))}\mathbb{E}^{\tilde{\mathbb{Q}}}\left[\bar{\rho}(t)\delta(J(t))\right] \\ &= \frac{1}{\alpha(0)q(J(0))}\mathbb{E}^{\tilde{\mathbb{Q}}}\left[\delta(J(t))\right]. \end{aligned} \tag{B.16}$$

Equation (B.16) concludes the attempt to rewrite the denominator of Equation (B.1), for the earlier discussed reason that $J(t)$ follows the same dynamics under $\bar{\mathbb{Q}}$ as under $\hat{\mathbb{Q}}$, which was the measure used for the alternative expression of the numerator in Equation (B.1), shown in Equation (B.4). Indeed, because of this fact it holds that $\mathbb{E}^{\hat{\mathbb{Q}}}\left[\delta(J(t))\right]/\mathbb{E}^{\bar{\mathbb{Q}}}\left[\delta(J(t))\right] = 1$, which means that division of Equation (B.4) by Equation (B.16) easily yields that

$$V(t,K) = q(J(0))\alpha^2(0). \tag{B.17}$$

This expression can be further simplified, consequently leading to conclusion of the proof

$$\sqrt{V(K)} = \alpha(0)\sqrt{1-2\nu J(0)\rho+\nu^2 J^2(0)},$$

which can be written as

$$\sqrt{V(K)} = \alpha(0)\sqrt{1+2\rho\nu f(K)+\nu^2 f^2(K)}, \tag{B.18}$$

where

$$f(K) \coloneqq \frac{1}{\alpha(0)}(K-\eta(0)).$$

□

Lemma B.1.2. *Let $\mathbb{Q}$ be an arbitrary measure. Then it holds that the process $B^{\mathbb{Q}}(t)$, defined by*

$$B^{\mathbb{Q}}(t) \coloneqq \int_0^t \frac{\rho-\nu J(s)}{\sqrt{1-2\nu J(s)\rho+\nu^2 J^2(s)}}\mathrm{d}W_\alpha^{\mathbb{Q}}(s) + \int_0^t \frac{\sqrt{1-\rho^2}}{\sqrt{1-2\nu J(s)\rho+\nu^2 J^2(s)}}\mathrm{d}\tilde{W}^{\mathbb{Q}}(s),$$

where $J(t)$ is defined as in Equation (B.2), and $W_\alpha^{\mathbb{Q}}(t), \tilde{W}^{\mathbb{Q}}(t)$ are two independent Brownian motions under $\mathbb{Q}$, is a Brownian motion under the $\mathbb{Q}$ measure.

Proof. This can be proven by first noting that $B^{\mathbb{Q}}(t)$ is the sum of two local martingales under $\mathbb{Q}$ and thus also a local martingale under the standard filtration. Moreover, it holds that

$$\begin{aligned}\mathrm{d}\langle B^{\mathbb{Q}}\rangle(t) &= \left(\frac{\rho-\nu J(t)}{\sqrt{1-2\nu J(t)\rho+\nu^2 J^2(t)}}\right)^2 \mathrm{d}\langle W_\alpha^{\mathbb{Q}}\rangle(t) + \left(\frac{\sqrt{1-\rho^2}}{\sqrt{1-2\nu J(t)\rho+\nu^2 J^2(t)}}\right)^2 \mathrm{d}\langle \tilde{W}^{\mathbb{Q}}\rangle(t) \\ &= \left(\frac{\rho-\nu J(t)}{\sqrt{1-2\nu J(t)\rho+\nu^2 J^2(t)}}\right)^2 \mathrm{d}t + \left(\frac{\sqrt{1-\rho^2}}{\sqrt{1-2\nu J(t)\rho+\nu^2 J^2(t)}}\right)^2 \mathrm{d}t \\ &= \mathrm{d}t, \end{aligned}$$

where the independence of $W_\alpha^{\mathbb{Q}}(t)$ and $\tilde{W}^{\mathbb{Q}}(t)$ was used to state that $\mathrm{d}\langle W_\alpha^{\hat{\mathbb{Q}}}, \tilde{W}^{\mathbb{Q}}\rangle(t) = 0$. It now holds that

$$\langle B^{\mathbb{Q}}\rangle(t) = \int_0^t \mathrm{d}s = t,$$

which means that by Lévy's characterization shown in Lemma A.3.1, $B^{\mathbb{Q}}(t)$ is indeed a Brownian motion under $\mathbb{Q}$. □

B.2. Application of Itô's lemma to $h(t)$

In preparation of the main lemma statement of Section 3.3.1, an application of Itô's lemma was deemed necessary. In this section, this application is discussed by the following corollary with associated proof.

Corollary B.2.0.1. *Let*

$$h(t) \coloneqq \int_0^{z(t)} \frac{1}{\nu\sqrt{1+2\rho x+x^2}}\,\mathrm{d}x,$$

where

$$\mathrm{d}z(t) = \nu\sqrt{1-2\rho z(t)+z^2(t)}\,\mathrm{d}W^T(t),$$

with $W^T(t)$ a standard Brownian motion under the T forward measure. An application of Itô's lemma to $h(t)$ then yields

$$\mathrm{d}h(t) = \mathrm{d}W^T(t) - \frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\,\mathrm{d}t,$$

where

$$\bar{h} \coloneqq \frac{1}{2\nu}\ln\left(\frac{1+\rho}{1-\rho}\right).$$

Proof. This proof is constructed by starting with the result stated in the lemma, and proving that the mentioned dynamics for $h(t)$ can be rewritten into the result directly following from an application of Itô's lemma to $h(t)$.

Recall from Theorem 3.2.2 that explicit evaluation of $h(t)$ yields

$$h(t) = \frac{\log\left(\sqrt{1+2\rho z(t)+z^2(t)}+\rho+z(t)\right)}{\nu} - \frac{\log(1+\rho)}{\nu}. \tag{B.19}$$

To capture the most interesting term in the dynamics of $h(t)$ more simply, let

$$y(t) \coloneqq -\frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\,\mathrm{d}t.$$

Continuing, it is known that the following equality holds

$$\tanh(x) = \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)},$$

which means that $y(t)$ can be rewritten to obtain

$$\begin{aligned}
y(t) &= -\frac{\nu}{2}\frac{\exp(\nu(h(t)+\bar{h}))-\exp(-\nu(h(t)+\bar{h}))}{\exp(\nu(h(t)+\bar{h}))+\exp(-\nu(h(t)+\bar{h}))} \\
&= -\frac{\nu}{2}\frac{\exp\left(\nu\left(h(t)+\frac{1}{2\nu}\ln\left(\frac{1+\rho}{1-\rho}\right)\right)\right)-\exp\left(-\nu\left(h(t)+\frac{1}{2\nu}\ln\left(\frac{1+\rho}{1-\rho}\right)\right)\right)}{\exp\left(\nu\left(h(t)+\frac{1}{2\nu}\ln\left(\frac{1+\rho}{1-\rho}\right)\right)\right)+\exp\left(-\nu\left(h(t)+\frac{1}{2\nu}\ln\left(\frac{1+\rho}{1-\rho}\right)\right)\right)} \\
&= -\frac{\nu}{2}\frac{\exp(\nu h(t))\left(\frac{1+\rho}{1-\rho}\right)^{\frac{1}{2}}-\exp(-\nu h(t))\left(\frac{1+\rho}{1-\rho}\right)^{-\frac{1}{2}}}{\exp(\nu h(t))\left(\frac{1+\rho}{1-\rho}\right)^{\frac{1}{2}}+\exp(-\nu h(t))\left(\frac{1+\rho}{1-\rho}\right)^{-\frac{1}{2}}} \\
&= -\frac{\nu}{2}\frac{\exp(\nu h(t))-\exp(-\nu h(t))\left(\frac{1+\rho}{1-\rho}\right)^{-1}}{\exp(\nu h(t))+\exp(-\nu h(t))\left(\frac{1+\rho}{1-\rho}\right)^{-1}}.
\end{aligned} \tag{B.20}$$

Additionally, from Equation (B.19) it follows that

$$\exp(\nu h(t)) = \exp\left(\log\left(\sqrt{1+2\rho z(t)+z^2(t)}+\rho+z(t)\right)-\log(1+\rho)\right) = \frac{\sqrt{1+2\rho z(t)+z^2(t)}+z(t)+\rho}{1+\rho}, \quad \text{(B.21)}$$

$$\exp(-\nu h(t)) = \exp\left(-\log\left(\sqrt{1+2\rho z(t)+z^2(t)}+\rho+z(t)\right)+\log(1+\rho)\right) = \frac{1+\rho}{\sqrt{1+2\rho z(t)+z^2(t)}+z(t)+\rho}. \quad \text{(B.22)}$$

Going back to Equation (B.20) with these new results, gives

$$y(t) = -\frac{\nu}{2}\frac{\frac{\sqrt{1+2\rho z(t)+z^2(t)}+z(t)+\rho}{1+\rho} - \frac{1+\rho}{\sqrt{1+2\rho z(t)+z^2(t)}+z(t)+\rho}\left(\frac{1+\rho}{1-\rho}\right)^{-1}}{\frac{\sqrt{1+2\rho z(t)+z^2(t)}+z(t)+\rho}{1+\rho} + \frac{1+\rho}{\sqrt{1+2\rho z(t)+z^2(t)}+z(t)+\rho}\left(\frac{1+\rho}{1-\rho}\right)^{-1}}. \quad \text{(B.23)}$$

The cumbersome simplification of Equation (B.23) is omitted, as it is a simple but lengthy rewriting problem arising from applying a least common multiplier to the terms in the numerator and denominator, adding the resulting fractions and finally applying a least common multiplier one more time. Performing these steps will give the following result

$$y(t) = -\frac{\nu}{2}\frac{\frac{2(z(t)+\rho)}{1+\rho}}{\frac{2\sqrt{1+2\rho z(t)+z^2(t)}}{1+\rho}} = -\frac{\nu}{2}\frac{z(t)+\rho}{\sqrt{1+2\rho z(t)+z^2(t)}}.$$

Therefore, the dynamics for $h(t)$ that should follow from application of Itô's lemma to $h(t)$, should be (according to the lemma at hand)

$$\mathrm{d}h(t) = \mathrm{d}W^T(t) - \frac{\nu}{2}\frac{z(t)+\rho}{\sqrt{1+2\rho z(t)+z^2(t)}}\,\mathrm{d}t. \quad \text{(B.24)}$$

Finally, direct application of Itô's lemma to Equation (B.19) results in

$$\begin{aligned}
\mathrm{d}h(t) &= \frac{\partial h}{\partial t}\mathrm{d}t + \frac{\partial h}{\partial z}\mathrm{d}z(t) + \frac{1}{2}\frac{\partial^2 h}{\partial z^2}(\mathrm{d}z(t))^2 \\
&= \frac{1}{\sqrt{1+2\rho z(t)+z^2(t)}}\mathrm{d}z(t) - \frac{1}{2}\frac{z(t)+\rho}{\nu\left(1+2\rho z(t)+z^2(t)\right)\sqrt{1+2\rho z(t)+z^2(t)}}(\mathrm{d}z(t))^2 \\
&= \mathrm{d}W^T(t) - \frac{1}{2}\frac{z(t)+\rho}{\nu\left(1+2\rho z(t)+z^2(t)\right)\sqrt{1+2\rho z(t)+z^2(t)}}\nu^2(1+2\rho z(t)+z^2(t))\,\mathrm{d}t \\
&= \mathrm{d}W^T(t) - \frac{\nu}{2}\frac{z(t)+\rho}{\sqrt{1+2\rho z(t)+z^2(t)}}\mathrm{d}t.
\end{aligned} \quad \text{(B.25)}$$

As it turns out, Equation (B.25) and Equation (B.24) are equal, which means that the lemma has been successfully proven. □

B.3. Lemma 3.3.1 clarifications

This section contains two lemmas used in the proof of Lemma 3.3.1. In order to prevent redundant notation, this section makes use of the notation and variables defined in Lemma 3.3.1.

Lemma B.3.1. *Let $\theta^*(t)$ be an exponential martingale [46] under the T forward measure described by*

$$\frac{\mathrm{d}\theta^*(t)}{\theta^*(t)} = \frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\,\mathrm{d}W_F^T(t),$$

such that

$$\theta^*(t) = \theta^*(0)\,\mathrm{e}^{-\frac{1}{2}\left(\frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\right)^2 t + \frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))W_F^T(t)}.$$

Solving $\mathrm{d}\left(\frac{f(h(t))}{\theta^*(t)}\right) = \mathcal{O}(\mathrm{d}t)$ *for f then yields*

$$f(h(t)) \coloneqq \sqrt{\cosh(\nu(h(t)+\bar{h}))}.$$

Proof. By the product rule

$$\mathrm{d}\left(\frac{f(h(t))}{\theta^*(t)}\right) = \frac{1}{\theta^*(t)}\,\mathrm{d}(f(h(t))) - \frac{f(h(t))}{\theta^*(t)}\frac{\mathrm{d}\theta^*(t)}{\theta^*(t)}. \tag{B.26}$$

The following has to be solved for f

$$\mathrm{d}\left(\frac{f(h(t))}{\theta^*(t)}\right) = \mathcal{O}(\mathrm{d}t).$$

Hence, filling in what is known into Equation (B.26), setting it equal to $\mathcal{O}(\mathrm{d}t)$ and multiplying the left and right hand side by $\theta^*(t)$, yields

$$\mathrm{d}f(h(t)) - f(h(t))\left(\frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\,\mathrm{d}W^T(t)\right) = \mathcal{O}(\mathrm{d}t). \tag{B.27}$$

Furhtermore, straightforward application of Itô's lemma to find an expression for $\mathrm{d}f(h(t))$, results in

$$\mathrm{d}f(h(t)) = \frac{\partial f(h(t))}{\partial t}\,\mathrm{d}t + \frac{\partial f(h(t))}{\partial h}\,\mathrm{d}h(t) + \frac{1}{2}\frac{\partial^2 f(h(t))}{\partial h^2}(\mathrm{d}h(t))^2, \tag{B.28}$$

where, from Section B.2, it is known that

$$\mathrm{d}h(t) = \mathrm{d}W^T(t) - \frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\,\mathrm{d}t. \tag{B.29}$$

Substituting Equation (B.29) into Equation (B.28) (using the convention that $(\mathrm{d}t)^2 = \mathrm{d}W^T(t)\,\mathrm{d}t = 0$ and $(\mathrm{d}W^T(t))^2 = \mathrm{d}t$), and then substituting the result into Equation (B.27), yields

$$\frac{\partial f(h(t))}{\partial t}\,\mathrm{d}t + \frac{\partial f(h(t))}{\partial h}\left(\mathrm{d}W^T(t) - \frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\,\mathrm{d}t\right) + \frac{1}{2}\frac{\partial^2 f(h(t))}{\partial h^2}\,\mathrm{d}t - f(h(t))\left(\frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\,\mathrm{d}W^T(t)\right) = \mathcal{O}(\mathrm{d}t). \tag{B.30}$$

It becomes clear that in order for Equation (B.30) to hold, it must be true that

$$\begin{aligned}\frac{\mathrm{d}f(h(t))}{\mathrm{d}h}\,\mathrm{d}W^T(t) - f(h(t))\left(\frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\,\mathrm{d}W^T(t)\right) &= 0,\\ \frac{\mathrm{d}f(h(t))}{\mathrm{d}h} - f(h(t))\left(\frac{\nu}{2}\tanh(\nu(h(t)+\bar{h}))\right) &= 0.\end{aligned} \tag{B.31}$$

Equation (B.31) is a simple ODE which can be solved to provide the following expression for f

$$f(h(t)) = \sqrt{\cosh(\nu(h(t)+\bar{h}))}. \tag{B.32}$$

□

Lemma B.3.2. *Given the function*

$$f(x) \coloneqq \sqrt{\cosh(\nu(x+\bar{h}))},$$

a second order Taylor expansion in $\mathrm{e}^{-\nu|x+\bar{h}|}$ *yields*

$$\frac{1}{f(x)} \approx \mathrm{e}^{-(\nu/2)|x+\bar{h}|}\frac{3-\mathrm{e}^{-\nu|x+\bar{h}|}}{2}.$$

Proof. Let $z = \nu(x+\bar{h})$, so that by definition of the variables in this expression it is safe to assume that $z \geq 0$. Using a standard trigonometric identity, the following can be written

$$\begin{aligned}(\cosh(z))^{-\frac{1}{2}} &= \left(\frac{\mathrm{e}^z + \mathrm{e}^{-z}}{2}\right)^{-\frac{1}{2}}\\ &= \mathrm{e}^{-(z/2)}\left(\frac{1+\xi^2}{2}\right)^{-\frac{1}{2}},\end{aligned} \tag{B.33}$$

where $\xi = e^{-z}$. Applying a second order Taylor expansion around $\xi = 1$ to Equation (B.33), yields

$$\left(\frac{1+\xi^2}{2}\right)^{-\frac{1}{2}} = 1 - \frac{1}{2}(\xi - 1) + \mathcal{O}((\xi-1)^2) \approx \frac{1}{2}(3-\xi). \tag{B.34}$$

Combining Equation (B.33) and Equation (B.34), gives the result

$$\begin{aligned}
\frac{1}{f(x)} &= (\cosh(z))^{-\frac{1}{2}} \\
&= e^{-(z/2)}\left(\frac{1+\xi^2}{2}\right)^{-\frac{1}{2}} \\
&\approx e^{-(z/2)}\frac{1}{2}(3-e^{-z}) \\
&= e^{-|z|/2}\frac{1}{2}(3-e^{-|z|}) \\
&= e^{-(\nu/2)|x+\tilde{h}|}\frac{3-e^{-\nu|x+\tilde{h}|}}{2}.
\end{aligned} \tag{B.35}$$

□

Bibliography

[1] Antonov, A., Konikov, M. & Spector, M. *Mixing SABR models for Negative Rates*. SSRN, 2015.

[2] Antonov, A., Konikov, M. & Spector, M. *SABR spreads its wings*. Risk, 2013.

[3] Antonov, A., Konikov, M. & Spector, M. *The Free Boundary SABR: Natural Extension to Negative Rates*. SSRN, 2015.

[4] Antonov, A., Konikov, M. & Spector, M. *Modern SABR Analytics*. Springer, 2019.

[5] Baadsgaard, M. *Estimation in stochastic differential equations. Master's thesis*. Denmark: Technical University of Denmark, 1996.

[6] Baadsgaard, M., Nygaard Nielsen, J., Spliid, H., Madsen, H. & Preisel, M. *Estimation in stochastic differential equations with a state dependent diffusion term*. IFAC Proceedings Volumes, Vol. 30, pp. 1369-1374, 1997.

[7] Baldeaux, J. *Exact Simulation of the 3/2 Model*. Australia: University of Technology Sydney, 2011.

[8] Balland, P. & Tran, Q. *SABR goes normal*. Risk, 2013.

[9] Bang, D. *Local stochastic volatility: shaken, not stirred*. Risk, 2018.

[10] Bang, D. *Local-Stochastic Volatility for Vanilla Modeling: A Tractable and Arbitrage Free Approach to Option Pricing*. SSRN, 2018.

[11] Behvand, F. *Forward Implied Volatility*. United Kingdom: University of Oxford, 2010.

[12] Berestycki, H., Busca, J. & Florent, I. *Computing the implied volatility in stochastic volatility models*. Communications on Pure and Applied Mathematics, Vol. 57, pp. 1352-1373, 2004.

[13] Black, F. & Scholes, M. *The Pricing of Options and Corporate Liabilities*. Journal of Political Economy, Vol. 81, pp. 637-654, 1973.

[14] Bliss, R.R. & Panigirtzoglou, N. *Option-implied risk aversion estimates*. The Journal of Finance, Vol. 59, pp. 407-446, 2004.

[15] Bomze, I.M. *Lotka-Volterra equation and replicator dynamics: A two-dimensional classification*. Biological Cybernetics, Vol. 48, pp. 201-211, 1983.

[16] Brenner, M. & Subrahmanyam, M.G. *A Simple Formula to Compute the Implied Standard Deviation*. Financial Analysts Journal, Vol. 44, pp. 80-83, 1988.

[17] Chan, K.C., Karolyi, G.A., Longstaff, F.A. & Sanders, A.B. (1992) An empirical investigation of alternative models of the short-term interest rate. Journal of Finance, Vol. 47, pp. 1209-1227, 1992.

[18] Chassagneux, J., Jacquier, A. & Mihaylov, I. *An explicit Euler scheme with strong rate of convergence for financial SDEs with non-Lipschitz coefficients*. SIAM Journal on Financial Mathematics, Vol. 7, pp. 993-1021, 2018.

[19] Cheridito, P., Kawaguchi, H. & Maejima, M. *Fractional Ornstein-Uhlenbeck processes*. Electronic Journal of Probability, Vol. 8, pp. 1-14, 2003.

[20] Comte, F. & Renault, E. *Long memory in continuous-time stochastic volatility models*. Mathematical Finance, Vol. 8, pp. 291-323, 1998.

[21] Cui, Y., del Baño Rollin, S. & Germano, G. *Full and fast calibration of the Heston stochastic volatility model*. European Journal of Operational Research, Vol. 263, pp. 625–638, 2017.

[22] Derman, E. & Kani, I. *Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility*. International Journal of Theoretical and Applied Finance, Vol. 1, pp. 61–110, 1998.

[23] Dirac, P. *The Principles of Quantum Mechanics (Fourth ed.)* Oxford University Press, 1958.

[24] Dupire, B. *Pricing with a smile*. Risk Magazine, Vol. 7, pp. 18-20, 1994.

[25] Forman, J. & Sørensen, M. *The Pearson Diffusion: A Class of Statistically Tractable Diffusion Processes*. Scandinavian Journal of Statics, Vol. 35, pp. 438-465, 2008.

[26] Fubini, G. *Opere scelte (Second ed.)*. Cremonese, pp. 243–249, 1958.

[27] Gatheral, J. *The Volatility Surface - A Practitioner's Guide*. Wiley Finance, 2006.

[28] Gatheral, J., Jaisson, T. & Rosenbaum, M. *Volatility is rough*. Quantitative Finance, Vol. 18, pp. 933-949, 2018.

[29] Gavin, H.P. *The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems*. United States: Duke University, 2019.

[30] Gyöngy, I. *Mimicking the One-Dimensional Marginal Distributions of Processes Having an Ito Differential*. Probability Theory and Related Fields 71, pp. 501-516, 1986.

[31] Hagan, P.S., Kumar, D., Lesniewski, A. & Woodward, D. *Managing Smile Risk*. Wilmott Magazine, pp. 84-108, 2002.

[32] Hambly, B.M. & Vaicenavicius, J. *The 3/2 model as a stochastic volatility approximation for a large-basket price-weighted index*. United Kingdom: University of Oxford, 2015.

[33] Iacus, S.M. *Simulation and Inference for Stochastic Differential Equations*. Springer, 2008.

[34] Jamshidian, F. *An exact bond option pricing formula*. Journal of Finance, Vol. 44, pp. 205-209, 1989.

[35] Kalman, R.E. & Bucy, R.S. *New results in linear filtering and prediction theory*. Journal of Basic Engineering. Vol. 83, pp. 95-108, 1961.

[36] Kotzé, A., Oosthuizen, R. & Pindza, E. *Implied and Local Volatility Surfaces for South African Index and Foreign Exchange Options*. Journal of Risk and Financial Management, Vol. 8, pp. 43-82, 2015.

[37] Kristensen, N.R., Madsen, H. & Jørgensen S.B. *Parameter estimation in stochastic grey-box models*. Automatica, Vol. 40, pp. 225-237, 2004.

[38] Lamperti, J. *Semi-stable stochastic processes*. Transactions of the American Mathematical Society, Vol. 104, pp. 62-78, 1962.

[39] Lay, S.R. *Analysis: With an Introduction to Proof (Fifth ed.)*. Pearson, 2014.

[40] Neuenkirch, A. & Szpruch, L. *First order strong approximations of scalar SDEs with values in a domain*. Numerische Mathematik, Vol. 128, pp. 103-136, 2014.

[41] Nielsen, J.N. & Madsen, H. *Applying the EKF to stochastic differential equations with level effects*. Automatica, Vol. 37, pp. 107-112, 2001.

[42] Massey, F.J. *The Kolmogorov-Smirnov Test for Goodness of Fit*. Journal of the American Statistical Association, Vol. 46, pp. 68-78, 2012.

[43] Møller, J.K. & Madsen, H. *From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform*. Denmark: Technical University of Denmark, 2010.

[44] Munkres, J.R. *Analysis on Manifolds*. United States: Massachusetts Institute of Technology. Westview Press, 1991.

[45] Nuzman, C.J. & Poor, V.H. *Linear estimation of self-similar processes via Lamperti's transformation*. Journal of Applied Probability, Vol. 37, pp. 429-452, 2000.

[46] Øksendal, B. *Stochastic Differential Equations. An Introduction with Applications (Fifth ed.)*. Springer, 2000.

[47] Oosterlee, C.W. & Grzelak, L.A. *Mathematical Modeling and Computation in Finance*. World Scientific, 2020.

[48] Powell, M.J.D. *An efficient method for finding the minimum of a function of several variables without calculating derivatives*. Computer Journal, Vol. 7, pp. 155-162, 1964.

[49] Protter, H.M. & Morrey, B.C. Jr. *Differentiation under the Integral Sign. Intermediate Calculus (Second ed.)*. Springer, pp. 421-426, 1985.

[50] van der Stoep, A.W., Grzelak, L.A. & Oosterlee, C.W. *The Heston Stochastic-Local Volatility Model: Efficient Monte Carlo Simulation*. International Journal of Theoretical and Applied Finance, Vol. 17, 2014.

[51] Wang, C. *Lévy's Characterization of Brownian Motion*. Taiwan: National Taiwan University, 2015.