MONASH University

Volatility modelling and calibration by optimal transport

Shiyi Wang

A thesis submitted for the degree of
Doctor of Philosophy
at Monash University in 2021
School of Mathematics

Copyright notice

Abstract

This thesis studies volatility modelling and calibration problems from the point of view of optimal transport. The manuscript is divided into three parts.

In the first part of the thesis, we examine the application of continuous-time martingale optimal transport to calibrate local volatility models from European options market prices. The calibration problem is formulated as a convex optimisation problem, for which we provide a numerical solution using the alternating direction method of multipliers (ADMM).

In the second part, we continue to extend the method for calibrating local-stochastic volatility (LSV) models. We study a semimartingale optimal transport problem with a finite number of discrete constraints motivated by fitting models to observable market option prices. This problem maximises a convex objective function on a convex set of probability measures. We prove that, if there exists a solution, we can find an optimal probability measure under which the semimartingale is a Markov process that fully matches the market option prices, whose drift and diffusion are functions of time and state variables. Focusing only on these Markov processes, we provide a partial differential equation formulation along with its dual counterpart. We propose a gradient descent algorithm for the numerical solution of the dual formulation, which involves solving a fully non-linear Hamilton–Jacobi–Bellman equation at each iteration. To demonstrate the numerical solution, we provide examples of calibrating a Heston-like LSV model with simulated data and foreign exchange market data.

Finally, in the third part, we apply the results developed in part two to solve the joint calibration problem of S&P 500 (SPX) and VIX options and futures, which has been known as a challenging problem for many years. To achieve this, we consider a two dimensional semimartingale whose first coordinate process is the logarithm of the SPX price and whose second coordinate process is the expected forward quadratic variation of the first coordinate process. Then the option prices and future prices of SPX and VIX can be formulated as the discrete constraints considered in part two. Therefore, the results developed in part two can be directly applied. In addition to the numerical solution for the dual formulation, we introduce a smoothing technique to smooth the model volatility surfaces and skews. Both numerical examples with simulated data and market data are presented.

Keywords: stochastic volatility, calibration, optimal transport, duality, SPX, VIX, HJB equation.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Shiyi Wang
25 September 2021

Acknowledgements

Firstly and foremost, I would like to express my heartfelt gratitude to my supervisors Prof. Grégoire Loeper and Dr. Ivan Guo, for their unwavering support and patience throughout my PhD studies. Without their guidance in the academic world, the research detailed in this dissertation would not have been possible.

I would also like to thank A/Prof. Andrea Collevecchio, Dr. Hasan Fallahgoul and Dr. Kihun Nam, members of my PhD committee, for their valuable time and feedback. My thanks also goes to my collaborator, Prof. Jan Obłój, for many enlightening discussions and the effort he has put into our manuscript.

I must also thank Dr. Oscar Yu Tian and John Chan, for their assistance in many aspects during my time at Monash University. I also thank my industry mentors, Steve McCarthy and Anna Liu for their guidance in the professional environment during my internship at NAB.

Many thanks go to my smart and lovely colleagues from the mathematics department, especially to my office mates, Abraham Nasrawi, Ben Amiet and Daniel Coelho. I always feel very lucky to be in the same office with you. I thank Wei Ning for her advice and suggestions in preparing the thesis manuscript. My thanks extend to my close friends Yifan Zhong, Siyuan Wu and Ziwen Zhong for their support and more, as well as many interesting discussions on random topics.

Last but not the least, I am deeply grateful for the love and support I have received from my family. For her selfless love and sacrifice, I dedicate this thesis to my mother. The last word is saved for my partner Huijun Zhan, for her emotional support and the countless delicious dinners she has prepared over the past few years.

The List of publications

The body of this thesis is based on three manuscripts that have been published already or submitted for publication.

- Chapter 2 is based on the manuscript titled "Local volatility calibration by optimal transport", which was co-authored by Ivan Guo and Grégoire Loeper. The manuscript has been published in the book, 2017 MATRIX Annals.
- Chapter 3 is based on the manuscript titled "Calibration of local-stochastic volatility models by optimal transport", which was co-authored by Ivan Guo and Grégoire Loeper. The manuscript has been publish by the journal, Mathematical Finance.
- Chapter 4 is based on the manuscript titled "Joint modelling and calibration of SPX and VIX by optimal transport", which was co-authored by Ivan Guo, Grégoire Loeper and Jan Obłój. The manuscript has been accepted by the journal, SIAM Journal on Financial Mathematics.

List of Notations

$\mathbb{R}^d$	a d-dimensional Euclidean space
$\mathbb{R}_+$	non-negative real numbers
$O^{m\times n}$	a null matrix of size $m \times n$
$I^{m\times n}$	an all-ones matrix of size $m \times n$
$\mathbb{S}^d$	the set of $d \times d$ symmetric matrices
$\mathbb{S}^d_+$	the set of $d \times d$ positive semidefinite symmetric matrices
Λ	the space $[0, T] \times \mathbb{R}^d$
$\mathcal{X}$	the space $\mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d$
$\mathrm{tr}(A)$	the trace of the square matrix A
$x^\intercal$	the transpose of the vector (or matrix) x
$A : B$	the sum of the scalar product of square matrices A and B, or $\mathrm{tr}(A^\intercal B)$
E	a Polish space equipped with its Borel σ-algebra
$C(E)$	the space of continuous functions on E
$C(E, \mathbb{R}^d)$	the space of continuous functions on E, valued in $\mathbb{R}^d$
$C_b(E)$	the space of bounded continuous functions on E
$C_b(E, \mathbb{R}^d)$	the space of bounded continuous functions on E, valued in $\mathbb{R}^d$
$C_b(E, \mathbb{R}^d)^*$	the topological dual of $C_b(E, \mathbb{R}^d)$
$C^2_b(E)$	the space of twice continuously differentiable functions on E with bounded partial derivatives up to order 2, and it is equipped with the norm given by the supremum of all partial derivatives up to order 2
$C_0(E)$	the space of continuous functions on E
$C^\infty_c(E)$	the space of smooth functions with compact support on E
$BV(E, \mathbb{R}^d)$	the space of functions of bounded variation on E, valued in $\mathbb{R}^d$
$L^1(d\mu, \mathbb{R}^d)$	the space of μ-integrable functions, valued in $\mathbb{R}^d$
$\mathcal{P}(E)$	the space of Borel probability measures on E
$\mathcal{M}(E, \mathbb{R}^d)$	the space of finite signed Borel measures on E, valued in $\mathbb{R}^d$ and endowed with the weak-$*$ topology
$\mathcal{M}_+(E, \mathbb{R}^d)$	the subset of nonnegative measures in $\mathcal{M}(E, \mathbb{R}^d)$
$\langle \cdot, \cdot \rangle$	the duality bracket between $C_b(\Lambda, \mathcal{X})$ and $C_b(\Lambda, \mathcal{X})^*$
$\mathbb{E}^{\mathbb{P}}(X \mid \mathcal{G})$	the conditional expectation of X under $\mathbb{P}$ given $\mathcal{G}$
$\mathbb{E}^{\mathbb{P}}_{t,x}$	the conditional expectation under $\mathbb{P}$ given $X_t = x$, or $\mathbb{E}^{\mathbb{P}}(\cdot \mid X_t = x)$
f^*	the convex conjugate or the Legendre–Fenchel transform of function f
f^{**}	the convex bi-conjugate of function f
$\partial_t \phi$	the first derivative of ϕ in time
$\partial_x \phi$ or $\nabla_x \phi$	the first derivative of ϕ in space
$\partial_{xx} \phi$ or $\nabla^2_x \phi$	the second derivative of ϕ in space

Contents

Chapter 1

Introduction

The main goal of this thesis is to investigate the problem of volatility modelling and calibration through the lens of optimal transport. The models we consider belong to a class of continuous semimartingales. To calibrate these models to observable market option prices, we formulate the calibration problems as convex optimisation problems, which are then solved by efficient numerical methods. When considering option price constraints instead of the terminal distribution constraint that appears in the classical optimal transport problem, our method is similar to the one studied by Avellaneda et al. [4] in which the local volatility model calibration problem was solved via entropy minimisation. Being nonparametric, if there exists a solution to the calibration problem, the method will lead us to an optimal probability measure under which the model fully matches the observable market option and future prices, i.e., an exact calibration. In addition, by using a dimension reduction technique, the calibrated model is Markovian in the state variables, which allows for hedging and pricing with conventional numerical methods.

This thesis is comprised of four chapters. The present chapter introduces the background of the problem of volatility modelling and calibration and the theory of optimal transport. Chapter 2 is based on our manuscript [51] entitled *Local volatility calibration by optimal transport*. In this work, we analyse a continuous-time martingale optimal transport problem in the spirit of the seminal work by Benamou and Brenier [8] for the classical optimal transport problem. The connections between optimal transport and the local volatility calibration are then established. Chapter 3 is based on our manuscript [54] entitled *Calibration of local-stochastic volatility models by optimal transport*. In this work, we study a semimartingale optimal transport problem, inspired by Tan and Touzi [103], and its application to the calibration of local-stochastic volatility models. In particular, motivated by fitting models to observable market option prices, we replace the terminal distribution constraint with a finite number of constraints (namely, options prices). A duality result for this modified problem and a numerical solution by gradient descent for the dual formulation are presented. Chapter 4 is based on our manuscript [52] entitled *Joint modelling and calibration of SPX and VIX by optimal transport*. In this work, we apply the method developed in Chapter 3 to construct a stochastic volatility model that can be jointly and exactly calibrated to the option and future prices of SPX and VIX.

1.1 Background

In the world of mathematical finance, the origin of option pricing — one of the most fundamental problems — traces back to the classical Black–Scholes framework introduced by Black and Scholes [14] and Merton [86] in 1973. A core assumption of their framework is that the underlying risky asset has constant volatility. However, this assumption can be easily dispelled, because the volatility surfaces implied by the observable market option prices are known to exhibit "smiles" or "skews". Ever since then, researchers and practitioners have put a lot of effort into developing sophisticated volatility models to explain this phenomenon.

Introduced as an extension of the Black–Scholes model, the local volatility model was proposed independently by Dupire [37] and Derman and Kani [35], in which the instantaneous volatility is a deterministic function of time and underlying asset price. The local volatility model is known as the simplest model that can capture the volatility smiles or skews, making it one of the most widely used models in the financial industry. Despite its simplicity and popularity, the local volatility model has been criticised for its unrealistic volatility dynamics, which means that the exotic option prices generated by the local volatility model are often inconsistent with those observed from the market. In contrast to the local volatility model, stochastic volatility models specify instantaneous volatility as a continuous stochastic process. Some classic instances of stochastic volatility models include the Heston model [64], the Hull–White model [67] and the SABR model [59]. These models tend to be more consistent with the market dynamics, but they struggle to fit short term market smiles and skews, and being parametric, they do not have enough degrees of freedom to match all vanilla market prices. The market practice is to use the so-called local-stochastic volatility (LSV) models which were first presented in Jex et al. [70]. The LSV models exploit the strengths of both the local volatility model and stochastic volatility models by incorporating a nonparametric local volatility component into stochastic volatility models. More details on LSV models will be given in Chapter 3.

In addition to the above models, there are other varieties that aim to accurately capture the market smiles and skews, including jump-diffusion models [76, 77, 87], pure jump models [28, 85], Lévy models [26, 99] and stochastic volatility models with jumps in the underlying or the volatility [25, 39]. In a series of papers [9, 10, 11, 12] and his book [13], Bergomi proposed and comprehensively reviewed a class of stochastic volatility models called the forward variance models. Instead of modelling the instantaneous volatility, these models specify the dynamics of forward variance, which provides a better fit to the future implied volatility surface and significantly reduces the pricing error of forward start options and cliquet options[1].

In recent years, the theory of optimal transport has been proven successful in applications to many fields, especially in mathematical finance. First, let us briefly introduce the mathematical field of optimal transport. The original optimal transport problem was addressed by Gaspard Monge [88] in the context of civil engineering in 1781. The goal of Monge's problem is to find a map to transport mass from one place to another with the minimum transportation cost. This problem was challenging to solve due to its nonlinear constraint. A breakthrough was made in 1940s by Leonid Kantorovich [72, 73] who presented two reformulations of Monge's problem based on linear programming methods, so the optimal transport problem is also known as the Monge–Kantorovich problem. Mathematically, given a cost function for measuring the cost of transporting a unit of mass from one point to another point, Kantorovich aims to minimise the integral of the cost function with respect to the set of all couplings (or so-called transport plans) that have given marginals. Kantorovich also provides a dual counterpart to the first formulation. These two formulations can be solved by a variety of modern mathematical techniques. Since then, optimal transport theory has attracted considerable attention with applications in

[1] A forward start option is an exotic option that is purchased at T_0, starts at T_1 and expires at T_2, where $T_0 < T_1 < T_2$. A cliquet option is an option that consists of a series of consecutive forward start options.

many areas such as cosmology [21, 82, 100], econometrics [43] and data science [92], etc. A comprehensive introduction to optimal transport can be found in the books of Rachev and Rüschendorf [93, 94] and the books of Villani [106, 107].

In a landmark paper published in 2000, Benamou and Brenier [8] proposed a continuous-time formulation of optimal transport. In this formulation, one aims to minimise the total transportation cost, in a fixed time interval, over all time-dependent densities and velocity fields, satisfying an initial and terminal density condition and a continuity equation. The continuity equation ensures that the mass is moving continuously in time and that the total mass is conserved during the transportation. In their paper, they also proposed a numerical solution based on an augmented Lagrangian method which is also widely known as the alternating direction method of multipliers (ADMM) in the recent literature[2]. In Brenier [20] (also see Loeper [82] for related works), the duality of the continuous-time optimal transport was formally established via the Fenchel–Rockafellar theorem (see e.g., Villani [106, Theorem 1.9]). This method was also applied in Huesmann and Trevisan [66] to investigate the duality of a continuous-time martingale optimal transport problem, that extends the optimal transport problem with an additional martingale constraint on the marginal distributions. Recently, Tan and Touzi [103] studied the problem of optimal transport with semimartingales with constraints on the marginals at initial and final times. More recently, in Guo and Loeper [50], the semimartingale optimal transport problem was further extended to a more general path-dependent setting. It is worth mentioning that two deep learning-based numerical methods were developed in Guo et al. [53] for solving the high-dimensional semimartingale optimal transport. Since both methods are mesh-free, they are not subject to the *curse of dimensionality*.

One of the most successful applications of optimal transport in mathematical finance is to solve the robust or model-free hedging problem. Instead of building sophisticated models based on market assumptions, the robust hedging problem prefers to have fewer beliefs in the financial market. This idea was formalised by Hobson in his famous paper [65], in which he derived lower and upper bounds for lookback options in continuous time by using call option prices and the no-arbitrage assumption. Following his work, martingale optimal transport was introduced to solve the robust hedging problem by Galichon et al. [44] in continuous time and by Beiglböck et al. [7] in discrete time. Their works have led to a growing literature on robust hedging and martingale optimal transport, see the recent book by Henry-Labordère [62] for an excellent review on these topics.

Another successful application of optimal transport is on the volatility model calibration problems. To the best of our knowledge, the first attempt of this application was made by Guo et al. [51], in which the local volatility calibration problem was formulated as a continuous-time martingale optimal transport problem and then numerically solved by the ADMM algorithm. Later in [50], Guo and Loeper studied a path-dependent model and expanded the available calibration instruments from European options to path-dependent options, such as Asian options, barrier options and lookback options. This was achieved by extending the semimartingale optimal transport problem of Tan and Touzi [103] to a path-dependent setting. Inspired by the so-called Schrödinger bridge problem, a problem closely related to optimal transport, Henry-Labordère [63] introduced a new class of stochastic volatility models. The calibration of these new models only requires to modify the drift, leaving the volatility of volatility unchanged. In a recent paper [56], using discrete-time martingale optimal transport, Guyon jointly and accurately reproduced the SPX and VIX smiles in discrete time. This joint calibration problem has been known to be highly chal-

[2]In the original paper by Benamou and Brenier [8] and some literature on optimal transport, the ADMM is also called ALG2 which is a name first used by the algorithm's inventors Fortin and Glowinski [41].

lenging. Guyon also proposed an efficient calibration method by extending the Sinkhorn matrix scaling algorithm [31], in the spirit of De March and Henry-Labordère [33]. More recently, the calibration framework developed in [50] was adapted to calibrate local-stochastic volatility models [54] and a joint model for SPX and VIX [52]. The highly connected works [51], [54] and [52] will be presented sequentially in Chapters 2, 3 and 4.

1.2 Contributions

The motivation of this thesis is to study volatility models and their calibration methods, and it comprises three connected research articles.

1.2.1 Local volatility model

In Chapter 2, we develop a calibration method for the local volatility model based on the theory of continuous-time martingale optimal transport. It is based on the published research article [51] entitled *Local volatility calibration by optimal transport*.

Introduced as an extension of the Black—Scholes model, the local volatility model has become one of the most popular models in the financial industry nowadays. In a local volatility model, the volatility function σ is a function of time t as well as the asset price S_t. The calibration of the local volatility function involves determining $\sigma(t,s)$ from available option prices. One of the most prominent calibration approaches is Dupire's formula [37] which is an analytical formula that allows to directly recover the local volatility function $\sigma(t,s)$ from available option prices. In particular, Dupire's formula is given by

$$\sigma^2(T,K)=\frac{\frac{\partial C(T,K)}{\partial T}+\mu_t K\frac{\partial C(T,K)}{\partial K}}{\frac{K^2}{2}\frac{\partial^2 C(T,K)}{\partial K^2}},$$

where μ_t is a deterministic function and $C(T,K)$ are European call option prices with maturity T and strike K. However, in practice, option prices are only available at discrete strikes and maturities. Hence interpolation is required in both variables to utilise this formula, leading to many inaccuracies and instabilities. Inspired by the seminal work of Benamou and Brenier [8] for the classical optimal transport, we introduce a variational approach for calibrating the local volatility function.

The first novelty of this work is that we adapt the variational formulation of the deterministic optimal transport and the augmented Lagrangian approach proposed by Benamou and Brenier [8] to the *martingale optimal transport* problem. More specifically, instead of considering a deterministic process whose density function is the solution of a continuity equation, we consider a martingale diffusion process whose density function solves a Fokker–Planck equation. We then formulate this optimisation problem in an augmented Lagrangian form which is then solved by an algorithm called the alternating direction method of multipliers (ADMM), in the spirit of Benamou and Brenier [8]. The second novelty is that, based on the developed theoretical and numerical outcomes of martingale optimal transport, we present a novel method to calibrate the local volatility function without the requirement of interpolating option prices in time. To the best of our knowledge, our work is the first attempt to formally establish the connection between optimal transport and volatility models calibration.

In the risk-neutral probability space, we suppose that the dynamic of an asset price X_t on $t\in[0,1]$ is given by a local volatility model

$$dX_t=\sigma(t,X_t)\,dW_t,\quad t\in[0,1],$$

where $\sigma(t,x)$ is a local volatility function and W_t is a one-dimensional Brownian motion. For the sake of simplicity, we assume the interest rates and dividends are zero. Denote by $\rho(t,x)$ the density function of X_t. As a direct consequence of Itô's formula, $\rho(t,x)$ solves the following Fokker-Planck equation:

$$\partial_t\rho(t,x) - \frac{1}{2}\partial_{xx}(\rho(t,x)\sigma^2(t,x)) = 0. \tag{1.1}$$

Suppose that the initial and the final densities are given by $\rho_0(x)$ and $\rho_1(x)$, which are recovered from European option prices via the Breeden–Litzenberger formula [19]. There exists a density function $\rho(t,x)$ that satisfies

$$\rho(0,x) = \rho_0(x), \quad \rho(1,x) = \rho_1(x) \tag{1.2}$$

and (1.1) if and only if ρ_0 and ρ_1 are in *convex order*[3]. This is known as Strassen's Theorem [102]. Let $D \subseteq \mathbb{R}$ be the support of $\{X_t, t \in [0,1]\}$. Given ρ_0 and ρ_1 that are in convex order and a convex cost function $F : \mathbb{R} \to \mathbb{R} \cup +\infty$, we are particularly interested in solving

$$\inf_{\rho,\sigma} \int_D \int_0^1 \rho(t,x) F\left(\frac{1}{2}\sigma^2(t,x)\right) dtdx, \tag{1.3}$$

subject to the constraints (1.1) and (1.2).

In Proposition 2.3.2, we restate two basic properties of the convex conjugate (also known as the *Legendre–Fenchel transform*). In Proposition 2.3.3, we derive the convex conjugate of $\rho F(\frac{m}{\rho})$ in ρ and $m := \frac{1}{2}\rho\sigma^2$ and an equality result. Introducing a Lagrange multiplier ϕ for the constraints (1.1) and (1.2) and using the results of Proposition 2.3.3, we reformulate (1.3) into a saddle point problem

$$\sup_{\mu} \inf_{\phi,q} L_r(\phi, q, \mu), \tag{1.4}$$

where $r > 0$ is a penalisation parameter and L_r is called an augmented Lagrangian. The optimal solution of (1.4) recovers the optimal solution of (1.3).

Next, we propose a numerical solution for solving (1.4) by adapting the ADMM algorithm used in [8]. The algorithm starts from an initial point (ϕ_0, q_0, μ_0), and it consists of three steps. In each step, it fixes two variables in (ϕ, q, μ) and solves (1.4) with respect to only the third (unfixed) variable. In the first step, by setting the functional derivative of $L_r(\phi, q, \mu)$ with respect to ϕ to zero, we show that the optimal ϕ can be obtained by solving a fourth-order linear PDE that has a bi-Laplacian operator. This PDE can be numerically solved by the standard finite difference method. In the second step, by setting the functional derivative of $L_r(\phi, q, \mu)$ with respect to q to zero, we show that the optimal q can be obtained by solving the point-wise minimisation problem (2.45). This problem can be solved analytically or numerically by Newton's method, depending on if a closed-form solution exists with the specific choice of the cost function F. In the last step, we update μ point-wise along the gradient of L_r with respect to μ. The algorithm repeats until the stopping criteria (2.48) is met with a preset tolerance.

In the numerical experiment, we provide a toy example with an initial distribution $N(0.5, 0.0025)$ at $t = 0$ and an final distribution $N(0.5, 0.01)$ at $t = 1$. The cost function F is chosen to be $F(x) = (x - \overline{x})^2$ if $x \geq 0$ or $F(x) = +\infty$ otherwise, where $\overline{x}$ is a chosen reference value for x. In this case, we set $\overline{x}$ to 0.00375, then the optimal variance

[3]The densities ρ_0 and ρ_1 are said to be in convex order if $\int_{\mathbb{R}} \varphi(x)\rho_0(x)dx \leq \int_{\mathbb{R}} \varphi(x)\rho_1(x)dx$ for all convex function $\varphi(x) : \mathbb{R} \to \mathbb{R}$.

is given by $\sigma^2 = 0.01 - 0.0025 = 0.0075$. The penalisation parameter r is set to 64. We present the results after 3000 iterations. In Figures 2.1 and 2.2, we plot the density function $\rho(t,x)$ and the variance function $\sigma^2(t,x)$ at six different times, respectively. Finally, the residuals across all iterations are shown in Figure 2.3. By comparing these figures with the analytical optimal solutions, we can see that the proposed numerical method effectively solves the problem.

1.2.2 Local-stochastic volatility models

In Chapter 3, we develop a calibration method for the local-stochastic volatility models based on the theory of semimartingale optimal transport. It is based on the published research manuscript [54] entitled *Calibration of local-stochastic volatility models by optimal transport*.

Local-Stochastic Volatility (LSV) models, introduced in Jex et al. [70], extend and take advantage of both the Local Volatility (LV) model and Stochastic Volatility (SV) models. The idea behind LSV models is to incorporate a nonparametric local factor (also called *leverage*) into the SV models. Thus, while keeping desirable properties of the SV model, the model can match all observed market prices (as long as one restricts to European claims). The calibration of local-stochastic volatility models involves determining the local factor from observable market option prices. Many of the existing calibration methods require a priori knowledge of the local volatility surface. This is usually obtained by using Dupire's formula [37]. However, only a finite number of options are available in practice. Thus, an interpolation of the implied volatility surface or option prices is often needed, which can lead to inaccuracies and instabilities. In the present work, inspired by the theory of semimartingale optimal transport [103], we introduce a variational approach for calibrating LSV models, which does not require any form of interpolation.

The first novelty of this work is that we study a semimartingale optimal transport problem with a finite number of, what we called, *discrete constraints*. These constraints are motivated by the fact that the expected discounted payoffs of a calibrated model should be equal to the observable market option prices. We introduce a dimension reduction technique that proves that there exists an optimal probability measure under which the semimartingale is also a Markov process. This allows us to deduce a PDE formulation of the problem along with its dual counterpart. In addition, we propose a algorithm by gradient descent to solve the dual formulation. The second novelty is that, inspired by the classical LSV model, we propose a generalised model we call the *OT-LSV model* whose instantaneous variance of the asset price and correlation are nonparametric functions of the asset price and a mean-reverting stochastic factor, which is then calibrated to observable market option prices by applying the developed theoretical results, without any form of interpolation.

In this work, we begin by considering a d-dimensional continuous semimartingale X that has the following representation under a probability measure $\mathbb{P}$:

$$dX_t = \alpha_t^{\mathbb{P}}\,dt + (\beta_t^{\mathbb{P}})^{\frac{1}{2}}\,dW_t^{\mathbb{P}}, \qquad t \in [0,T],$$

where $W^{\mathbb{P}}$ is a d-dimensional $\mathbb{P}$-Brownian motion, and $(\alpha^{\mathbb{P}},\beta^{\mathbb{P}})$ are adapted processes. We are particularly interested in a set of probability measures under which $(\alpha^{\mathbb{P}},\beta^{\mathbb{P}})$ are $\mathbb{P}$-integrable. Motivated by solving the volatility model calibration problem, we further restrict the probability measures to a subset $\mathcal{P}$ such that X has an initial value x_0 and satisfies a finite number of discrete constraints. Given a convex cost function F, we want to minimise $\mathbb{E}^{\mathbb{P}}\int_0^T F(\alpha_t^{\mathbb{P}},\beta_t^{\mathbb{P}})\,dt$ over all $\mathbb{P}\in\mathcal{P}$.

To find an optimal probability measure for this problem, we first apply a dimension reduction technique. As a direct consequence of Itô's formula and the superposition principle of Trevisan [105, Theorem 2.5] (or earlier Figalli [40, Theorem 2.6] for the bounded coefficients case), Lemma 3.3.1 restates that $\rho_t^{\mathbb{P}}$, the density of X_t under $\mathbb{P}$, is a weak solution to a Fokker–Planck equation, and there exists another probability measure $\mathbb{P}'$ under which $\rho^{\mathbb{P}} = \rho^{\mathbb{P}'}$ and X is a Markov process with $(\alpha^{\mathbb{P}'}, \beta^{\mathbb{P}'})$ that are functions of (t, X_t). This result can also be easily derived by the Markovian projection method of Brunick and Shreve [23]. In Lemma 3.3.3, we prove that if there exists a probability measure $\mathbb{P} \in \mathcal{P}$, then there exists another measure $\mathbb{P}'$ in the subset of $\mathcal{P}$ such that $\rho^{\mathbb{P}} = \rho^{\mathbb{P}'}$ and $(\alpha^{\mathbb{P}'}, \beta^{\mathbb{P}'})$ are functions of (t, X_t). Let us denote this subset by $\mathcal{P}_{loc}$. Using Lemma 3.3.1, Lemma 3.3.3 and Jensen's inequality, we prove in Proposition 3.3.4 that an optimal probability measure can be found in $\mathcal{P}_{loc}$. Focusing on finding an optimal measure in $\mathcal{P}_{loc}$, we introduce in Problem 2 a PDE formulation.

By closely following Brenier [20] and Loeper [82] and applying the Fenchel–Rockafellar theorem (see e.g., [106, Theorem 1.9] or [22, Chapter 1]), we establish the duality for the PDE formulation and hence introduce a dual counterpart in Theorem 3.3.6. In the dual formulation, given a vector $c \in \mathbb{R}^m$ that can be interpreted as the observable market option prices in the context of volatility model calibration, we maximise the objective function $\mathcal{V}(\lambda, \phi^\lambda) := \lambda \cdot c - \phi^\lambda(0, x_0)$ over (ϕ^λ, λ) where $\lambda \in \mathbb{R}^m$ are the Lagrange multipliers of the discrete constraints and ϕ^λ is a supersolution of a Hamilton–Jacobi–Bellman (HJB) equation. In Proposition 3.3.5, by using the shaken coefficients technique of Krylov [78], we show that the supremum of $\mathcal{V}(\lambda, \phi^\lambda)$ over ϕ^λ is achieved by the (unique) viscosity solution of the HJB equation. In addition, in Proposition 3.3.5, we show that the optimal $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ of the PDE formulation for $\mathbb{P} \in \mathcal{P}_{loc}$ can be found as by-products of solving the HJB equation in the dual formulation.

In Section 3.4, we fit the LSV model calibration problem into the developed semimartingale optimal transport framework. Let the first coordinate process of X be the logarithm of the underlying asset price, and let its second coordinate process be a mean-reverting stochastic factor. We want to calibrate X to the observable market option prices while retaining the desired properties in the model dynamics. To achieve this, in the first step, we define a convex function F that regularises our model X. In particular, we define our cost function F in Definition 3.4.3, which penalises deviations of the X from a reference model so that X has the LSV model dynamics as desired. In this case, the reference model is chosen to be the classical Heston model with parameters given by calibrating the Heston model to the observable market option prices. In the second step, we define the discrete constraints to ensure that the calibrated model X matches all the observable market option prices. We let G be a vector of discounted payoff functions of a set of European options, and we let τ and c be the corresponding vectors of their maturities and option prices, respectively. Then, the associated dual formulation of the reformulated LSV calibration problem is given in equation (3.24) along with the associated HJB equation in equation (3.25). Furthermore, in Lemma 3.4.5, we provide the gradient of the objective function with respect to λ. This allows us to use gradient-based numerical methods to solve the dual formulation. In addition, the gradients are formulated as the difference between the observable market option prices and the model prices with a given λ. As the objective function approaches its minimum, the gradient is approaching zero, so the dual formulation provides a natural interpretation in terms of fitting the model to observable market option prices.

In Section 3.5.1, we propose a numerical solution by gradient descent for solving the dual formulation. We start by setting an initial value (e.g., a null vector) to λ. Then, we calculate $\phi^\lambda(0, x_0)$ by numerically solving the HJB equation using an alternating direction implicit

(ADI) finite difference method. Next, we update λ through a gradient-based optimisation algorithm. In particular, we employed the L-BFGS algorithm [81] and obtained good convergence. The gradients can be computed by using Lemma 3.4.5, which requires one to compute the option prices by numerically solving one backward linear PDE per option. Since all these PDEs share the same linear operator, instead of solving them one by one, all these PDEs can be solved by inverting the linear operator only once at each time step, which highly reduces the computational cost. The algorithm will repeat until the maximum of absolute values of gradients (or the maximum absolute errors between market option prices and model option prices) is smaller than a certain threshold. The numerical solution is summarised in Algorithm 1. We also provide in Algorithm 2 another version enhanced by a technique called *policy iteration* for handling the nonlinearity of the HJB equation.

To illustrate the numerical method, we present two examples with simulated data in Section 3.5.2 and one example with the foreign exchange (FX) market data in Section 3.5.3. In the examples with simulated data, we calibrate the model X to a set of European call option prices generated by a Heston model with given parameters. To distinguish the Heston models between the one for generating option prices and the one taken as the reference, let us call the former the *Heston generating model* and call the later the *Heston reference model*. In the first example, we use the same set of parameters for both the Heston generating model and the Heston reference model. In this case, there exists a solution and the calibrated X recovers the Heston generating model. The instantaneous variance function is shown in Figure 3.2, which confirms that the algorithm works as expected. To further test the robustness of the method, in the second example, we choose different sets of parameters, as shown in Table 3.1, for the Heston generating model and the Heston reference model. As noted in the Remark 3.4.4, when the instantaneous variance function is independent of the stochastic factor, the model X reduces to a local volatility model that can be calibrated to any arbitrage-free option prices; thus, a solution exists in this example. Figure 3.3 shows the plot of the fraction between the instantaneous variance function and the second state variable, which serves as a reference in comparison with the so-called *leverage function* in the traditional LSV model. The correlation function is shown in Figure 3.4. The calibration results in terms of implied volatility are presented in Table 3.2 and Figure 3.5.

Finally, we provide an example with the FX market data used in Tian et al. [104] (also available in Section A.4). Table 3.3 presents the parameters for the Heston reference model used in this example. Figure 3.6 shows the implied volatility of market options, the calibrated and the uncalibrated model X for 1 month and 3 months maturities; Figure 3.7 shows the implied volatility of these models for 2 years and 5 years maturities.

1.2.3 Joint model for SPX and VIX

In Chapter 4, we further apply the semimartingale optimal transport framework developed in Chapter 3 to solve the problem of jointly calibrating models to SPX and VIX skews. It is based on the research manuscript [52] entitled *Joint modelling and calibration of SPX and VIX by optimal transport*.

The CBOE Volatility Index (VIX), also known as the stock market's "fear gauge", reflects the expectations of investors on the volatility of the S&P500 index (SPX) over the next 30 days. Although the index in itself is not a tradable asset, its derivatives such as futures and options are highly liquid. Since the VIX options started trading in 2006, researchers and practitioners have been putting a lot of effort in jointly calibrating models to the SPX and VIX options prices. This has been known as a long-standing puzzle. As commented by Guyon [56], inconsistencies might appear between the volatility-of-volatility inferred from SPX and VIX, making it an extremely difficult task to build a continuous model[4] that jointly captures the market SPX and VIX volatility skews. In Chapter 4, by fitting the problem into the semimartingale optimal transport framework developed in Chapter 3, we propose a novel method to calibrate a nonparametric continuous model jointly and exactly to the prices of SPX options, VIX options and VIX futures.

The main novelty of this work is that, instead of modelling the VIX index or the instantaneous variance of the SPX, we consider a semimartingale X whose first coordinate process X^1 is the logarithm of the SPX price, and whose second coordinate process X^2 is defined as the forward expected quadratic variation of X^1. By doing so, the calibration exercise only depends on the marginals of X at fixed times. Hence, the joint calibration problem immediately falls into the class of the semimartingale optimal transport problems studied in Chapter 3. Next, we propose a PDE formulation and a dual counterpart for the joint calibration problem. In addition, we provide a numerical method by gradient descent for solving the dual formulation along with some numerical treatments for eliminating instabilities.

The model we consider is a two-dimensional continuous semimartingale $X = (X^1, X^2)$ that has the following representation under a probability measure $\mathbb{P}$:

$$dX_t = \alpha_t^{\mathbb{P}}\, dt + (\beta_t^{\mathbb{P}})^{\frac{1}{2}}\, dW_t^{\mathbb{P}}, \qquad t \in [0, T],$$

where $W^{\mathbb{P}}$ is a two-dimensional $\mathbb{P}$-Brownian motion, and $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ are adapted processes. We want X^1 to be the logarithm of the SPX price, which solves

$$X_t^1 = X_0^1 - \frac{1}{2}\int_0^t \sigma_s^2\, ds + \int_0^t \sigma_s\, dW_s, \quad 0 \le t \le T,$$

where σ is an adapted process and W is a one-dimensional Brownian motion. For X^2, we model it as the half of the forward expected quadratic variation of X^1 on $[t, T]$ observed at t, that is

$$X_t^2 = \mathbb{E}^{\mathbb{P}}\left(\frac{1}{2}\int_t^T \sigma_s^2\, ds \,\middle|\, \mathcal{F}_t\right), \quad 0 \le t \le T.$$

The desired model dynamics can be captured by probability measures $\mathbb{P}$ under which

$$\alpha_t^{\mathbb{P}} = \begin{bmatrix} -\frac{1}{2}\sigma_t^2 \\ -\frac{1}{2}\sigma_t^2 \end{bmatrix} \quad \text{and} \quad \beta_t^{\mathbb{P}} = \begin{bmatrix} \sigma_t^2 & (\beta_t^{\mathbb{P}})_{12} \\ (\beta_t^{\mathbb{P}})_{12} & (\beta_t^{\mathbb{P}})_{22} \end{bmatrix}, \quad 0 \le t \le T, \tag{1.5}$$

where $(\beta_t^{\mathbb{P}})_{12} = d\langle X^1, X^2\rangle_t \,/\, dt$ and $(\beta_t^{\mathbb{P}})_{22} = d\langle X^2\rangle_t \,/\, dt$, and with the additional property that $X_T^2 = 0$ $\mathbb{P}$-a.s.

In the semimartingale optimal transport framework, the above settings can be obtained by choosing a suitable cost function and proper discrete constraints. In particular, we choose the cost function F that takes the form of equation (4.4). Although both the calibration method and the model are nonparametric, the cost function requires some reference values for the covariance of the model X. Later in Section 4.2.3, we will derive the expression of the Heston model in terms of X^1 and X^2, and then we take the covariance from the derived expression as the required reference values.

To ensure that the calibrated model X matches the observable market prices of SPX and VIX options and futures, we impose on X a finite number of discrete constraints that take the form of $\mathbb{E}^{\mathbb{P}} G(X_\tau) = c$, where G is the discounted payoff function of a product in terms of X^1 and X^2, τ is the maturity and c is the observable market price of the product.

[4] By continuous model, we mean a continuous-time model with continuous paths.

The additional property $X_T^2 = 0$ $\mathbb{P}$-a.s. is also implemented as a discrete constraint. Let $\mathcal{P}_{joint}$ be the set of probability measures such that, under any $\mathbb{P} \in \mathcal{P}_{joint}$, $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ are $\mathbb{P}$-integrable, X has an initial value x_0 and all discrete constraints are satisfied. We are particularly interested in solving the following reformulated joint calibration problem:

$$V = \inf_{\mathbb{P} \in \mathcal{P}_{joint}} \mathbb{E}^{\mathbb{P}} \int_0^T F(\alpha_s^{\mathbb{P}}, \beta_s^{\mathbb{P}})\, ds. \tag{1.6}$$

In Chapter 4.3, we introduce some important results by utilising the theoretical developments of Chapter 3. Let $\mathcal{P}_{joint}^{loc}$ be a subset of $\mathcal{P}_{joint}$ such that, under any $\mathbb{P} \in \mathcal{P}_{joint}^{loc}$, the semimartingale X is a Markov process, whose drift $\alpha_t^{\mathbb{P}}$ and diffusion $\beta_t^{\mathbb{P}}$ are functions of t and X_t. Proposition 4.3.2 is immediately followed by the dimension reduction method developed in Chapter 3, which shows that an optimal probability measure that achieves the infimum in (1.6) can be found in $\mathcal{P}_{joint}^{loc}$. Based on this result, we introduce in Proposition 4.3.3 a PDE formulation in which we seek a triple of functions subject to a Fokker–Planck equation and the discrete constraints. Further in Theorem 4.3.4, we introduce a dual formulation in which we solve

$$V = \sup_{\lambda := (\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^{\xi}) \in \mathbb{R}^{m+n+2}} \lambda^{SPX} \cdot u^{SPX} + \lambda^{VIX,f} u^{VIX,f} + \lambda^{VIX} \cdot u^{VIX} - \phi^{\lambda}(0, X_0), \tag{1.7}$$

where ϕ^{λ} is the viscosity solution to an HJB equation; $u^{SPX} \in \mathbb{R}^m, u^{VIX,f} \in \mathbb{R}, u^{VIX} \in \mathbb{R}^n$ are the prices of m SPX options, VIX futures, n VIX options, respectively; $\lambda^{SPX} \in \mathbb{R}^m, \lambda^{VIX,f} \in \mathbb{R}, \lambda^{VIX} \in \mathbb{R}^n$ are the Lagrange multipliers of the discrete constraints for matching the model prices to the market prices of SPX options, VIX futures, VIX options, respectively; and $\lambda^{\xi} \in \mathbb{R}$ is the Lagrange multiplier of the constraint $X_T^2 = 0$ $\mathbb{P}$-a.s.[5] Additionally, Theorem 4.3.4 states that the optimal $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ can be obtained as by-products of solving the HJB equation in the dual formulation. Next, in Lemma 4.3.6, we derive the gradients of the objective function with respect to each element of λ. Note that ϕ^{λ} depends on λ implicitly, since both ϕ^{λ} and the vector λ appear in the HJB equation. Interestingly, the gradients can be formulated as the difference between the observable market option and future prices and the model prices at the current optimisation iteration. This provides a natural interpretation for the dual formulation in terms of matching the model to observable market option prices.

In Section 4.4.1, by adapting the algorithm developed in Chapter 3, we present a gradient descent numerical method for solving the dual formulation. At each iteration of the optimisation process, we solve the HJB equations (4.24) and (4.25) by an alternating direction implicit (ADI) finite difference method. Due to the presence of Dirac delta functions in the HJB equations, the solution might have discontinuities in time. To handle this issue, some numerical treatments are provided in Section 4.4.1. When the chosen reference model is far away from the one that describes the market dynamics, the calibrated volatility surfaces might be spiky, and the volatility skews might be hump-shaped. To overcome this issue, in Section 4.4.3, we introduce a *reference measure iteration* method to smooth the volatility surfaces and skews. This smoothing method will be later applied in all numerical examples.

We illustrate the method with two numerical examples. In the first example, we use simulated data to test the significance of the reference values after applying the reference measure iteration method. We generate SPX and VIX option and future prices by a Heston

[5]Please note that although λ^{ξ} does not appear in (1.7), it has an implicit dependence with ϕ^{λ} via the HJB equation.

model with a given set of parameters. Then, we calibrate the model X to these prices by solving the dual formulation with reference values from two different models: a Heston model with a different set of parameters and a model with constant covariance. All parameters used in this example are shown in Table 4.1. In Figures 4.1 and 4.2, we show the volatility skews of the both cases. Figures 4.3 and 4.3 visualise the simulated dynamics of X^1 and X^2. These figures show that the dynamics of the model can be affected by the chosen reference values. In Table 4.2, we provide a complete set of numerical results in option prices and implied volatility. We further display the volatility behaviour of the three models in Appendix B.3. These results confirm that the model accurately captures the simulated option and future prices while keeping the desired model dynamics.

Finally, we test the robustness of the method by using the market data from September 1st, 2020, including monthly SPX options maturing at 17 days and 45 days and monthly VIX futures and options maturing at 15 days. The reference values are chosen from a pure Heston model that has been calibrated to the market data. The calibrated parameters and the initial values of X are given in Table 4.3. Even with these parameters, the VIX skew generated by this Heston model is highly unrealistic. Therefore, we apply the reference measure iteration method to improve the reference values iteratively. The calibrated model volatility skews are plotted in Figure 4.5, and the simulation of X is given in Figure 4.6. The volatility functions are displayed in Appendix B.4. The results verify that the proposed method is effective for the market data as well.

Chapter 2

Local Volatility Calibration by Optimal Transport

The objective of this chapter is to study the connection between optimal transport and the local volatility model calibration problem. The most common approach of calibrating local volatility among industry practitioners is based on the celebrated Dupire's formula [37], which requires the knowledge of vanilla option prices for a continuum of strikes and maturities that can only be obtained via some form of price interpolation. In this chapter, we formulate the calibration problem as a time continuous martingale optimal transport problem, which seeks a martingale diffusion process that matches the known densities of an asset price at two different dates, while minimising a chosen cost function. Inspired by the seminal work of Benamou and Brenier [8], we formulate the problem as a convex optimisation problem, derive its dual formulation, and solve it numerically via an augmented Lagrangian method and the alternative direction method of multipliers (ADMM) algorithm. Numerical experiment with simulated data shows that the proposed solution effectively reconstructs the dynamic of the asset price between the two dates by recovering the optimal local volatility function, without requiring any time interpolation of the option prices.

2.1 Introduction

A fundamental assumption of the classical Black-Scholes option pricing framework is that the underlying risky asset has a constant volatility. However, this assumption can be easily dispelled by the option prices observed in the market, where the implied volatility surfaces are known to exhibit "skews" or "smiles". Over the years, many sophisticated volatility models have been introduced to explain this phenomenon. One popular class of model is the local volatility models. In a local volatility model, the volatility function $\sigma(t, S_t)$ is a function of time t as well as the asset price S_t. The calibration of the local volatility function involves determining σ from available option prices.

One of the most prominent approaches for calibrating local volatility is introduced by the path-breaking work of Dupire [37], which provides a method to recover the local volatility function $\sigma(t, s)$ if the prices of European call options $C(T, K)$ are known for a continuum of maturities T and strikes K. In particular, the famous Dupire's formula is given by

$$\sigma^2(T, K) = \frac{\frac{\partial C(T,K)}{\partial T} + \mu_t K \frac{\partial C(T,K)}{\partial K}}{\frac{K^2}{2} \frac{\partial^2 C(T,K)}{\partial K^2}}, \tag{2.1}$$

where μ_t is a deterministic function. However, in practice, option prices are only available at discrete strikes and maturities, hence interpolation is required in both variables to utilise this formula, leading to many inaccuracies. Furthermore, the numerical evaluation of the second derivative in the denominator can potentially cause instabilities in the volatility surface as well as singularities. Despite these drawbacks, Dupire's formula and its variants are still used prevalently in the financial industry today.

In this chapter, we introduce a new technique for the calibration of local volatility functions that adopts a variational approach inspired by optimal transport. We begin by recovering the probability density of the underlying asset at times t_0 and t_1 from the prices of European options expiring at t_0 and t_1. Then, instead of interpolating between different maturities, we seek a martingale diffusion process which transports the density from t_0 to t_1, while minimising a particular cost function. This is similar to the classical optimal transport problem, with the additional constraint that the diffusion process must be a martingale driven by a local volatility function. In the case where the cost function is convex, we find that the problem can be reformulated as a convex optimisation problem under linear constraints. Theoretically, the stochastic control problem can be reformulated as an optimisation problem which involves solving a non-linear PDE at each step, and the PDE is closely connected with the ones studied in Bouchard et al. [16, 17] and Loeper [83] in the context of option pricing with market impact. In this chapter, we approach the problem via the augmented Lagrangian method and the alternative direction method of multipliers (ADMM) algorithm, which was also used in Benamou and Brenier [8] for classical optimal transport problems.

This chapter is organised as follows. In Section 2.2, we introduce the classical optimal transport problem as formulated by Benamou and Brenier [8]. In Section 2.3, we introduce the martingale optimal transport problem and its augmented Lagrangian. The numerical method is detailed in Section 2.4 and numerical results are given in Section 2.5.

2.2 Optimal Transport

In this section, we briefly outline the optimal transport problem as formulated by Benamou and Brenier [8]. Given density functions $\rho_0, \rho_1 : \mathbb{R}^d \to [0, \infty)$ with equal total mass $\int_{\mathbb{R}^d} \rho_0(x)dx = \int_{\mathbb{R}^d} \rho_1(x)dx$. We say that a map $s : \mathbb{R}^d \to \mathbb{R}^d$ is an admissible transport plan if it satisfies

$$\int_{x\in A} \rho_1(x)dx = \int_{s(x)\in A} \rho_0(x)dx, \tag{2.2}$$

for all bounded subset $A \subset \mathbb{R}^d$. Let $\mathcal{T}$ denote the collection of all admissible maps. Given a cost function $c(x, y)$, which represents the transportation cost of moving one unit of mass from x to y, the optimal transport problem is to find an optimal map $s^* \in \mathcal{T}$ that minimises the total cost

$$\inf_{s\in\mathcal{T}} \int_{\mathbb{R}^d} c(x, s(x))\rho_0(x)dx. \tag{2.3}$$

In particular, when $c(x, y) = |y - x|^2$ where $|\cdot|$ denotes the Euclidean norm, this problem is known as the L^2 Monge-Kantorovich problem (MKP).

The L^2 MKP is reformulated in [8] in a fluid mechanic framework. In the time interval $t \in [0, 1]$, consider all possible smooth, time-dependent, densities $\rho(t, x) \geq 0$ and velocity fields $v(t, x) \in \mathbb{R}^d$, that satisfy the continuity equation

$$\partial_t\rho(t, x) + \nabla \cdot (\rho(t, x)v(t, x)) = 0, \qquad \forall t \in [0, 1],\ \forall x \in \mathbb{R}^d, \tag{2.4}$$

and the initial and final conditions

$$\rho(0, x) = \rho_0, \quad \rho(1, x) = \rho_1. \tag{2.5}$$

In [8], it is proven that the L^2 MKP is equivalent to finding an optimal pair (ρ^*, v^*) that minimises

$$\inf_{\rho,v} \int_{\mathbb{R}^d} \int_0^1 \rho(t,x)|v(t,x)|^2 dt dx, \tag{2.6}$$

subject to the constraints (2.4) and (2.5). This problem is then solved numerically in [8] via an augmented Lagrangian approach. The specific numerical algorithm used is known as the alternative direction method of multipliers (ADMM), which has applications in statistical learning and distributed optimisation.

2.3 Definition of the martingale problem

Let $(\Omega, \mathbb{F}, \mathbb{Q})$ be a probability space, where $\mathbb{Q}$ is the risk-neutral measure. Suppose the dynamic of an asset price X_t on $t \in [0,1]$ is given by the local volatility model

$$dX_t = \sigma(t, X_t) dW_t, \quad t \in [0,1], \tag{2.7}$$

where $\sigma(t,x)$ is a local volatility function and W_t is a one-dimensional Brownian motion. For the sake of simplicity, suppose the interest and dividend rates are zero. Denote by $\rho(t,x)$ the density function of X_t and $\gamma(t,x) = \sigma(t,x)^2/2$ the diffusion coefficient. It is well known that $\rho(t,x)$ follows the Fokker-Planck equation

$$\partial_t \rho(t,x) - \partial_{xx}(\rho(t,x)\gamma(t,x)) = 0. \tag{2.8}$$

Suppose that the initial and the final densities are given by $\rho_0(x)$ and $\rho_1(x)$, which are recovered from European option prices via the Breeden-Litzenberger [19] formula,

$$\rho_T(K) = \frac{\partial^2 C(T,K)}{\partial K^2}.$$

Let $F : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a convex cost function. We are interested in minimising the quantity

$$\mathbb{E}\left(\int_0^1 F(\gamma(t, X_t))\, dt \right) = \int_D \int_0^1 \rho(t,x) F(\gamma(t,X_t))\, dt dx,$$

where $F(x) = +\infty$ if $x < 0$, and $D \subseteq \mathbb{R}$ is the support of $\{X_t, t \in [0,1]\}$. Unlike the classical optimal transport problem, the existence of a solution here requires an additional condition: there exists a martingale transport plan if and only if ρ_0 and ρ_1 satisfy:

$$\int_{\mathbb{R}} \varphi(x)\rho_0(x) dx \le \int_{\mathbb{R}} \varphi(x)\rho_1(x) dx,$$

for all convex function $\varphi(x) : \mathbb{R} \to \mathbb{R}$. This is known as Strassen's Theorem [102]. This condition is naturally satisfied by financial models in which the asset price follows a martingale diffusion process.

Remark 2.3.1. The formulation here is actually quite general and it can be easily adapted to a large family of models. For example, the case of a geometric Brownian motion with local volatility can be recovered by substituting $\tilde{\sigma}(t, X_t)X_t = \sigma(t, X_t)$ everywhere, including in the Fokker-Planck equation. The cost function F would then also be dependent on x. The later arguments involving convex conjugates still hold since F remains a convex function of $\tilde{\sigma}$.

Since $\rho F(\gamma)$ is not convex in (ρ,γ) (which is crucial for our method), the substitution $m(t,x) := \rho(t,x)\gamma(t,x)$ is applied. So we obtain the following *martingale optimal transport problem*:

$$\inf_{\rho,m} \int_D \int_0^1 \rho(t,x) F\left(\frac{m(t,x)}{\rho(t,x)}\right) dtdx, \tag{2.9}$$

subject to the constraints:

$$\rho(0,x) = \rho_0(x), \quad \rho(1,x) = \rho_1(x), \tag{2.10}$$
$$\partial_t \rho(t,x) - \partial_{xx} m(t,x) = 0. \tag{2.11}$$

Using the convexity of F, the term $\rho F(m/\rho)$ can be easily verified to be convex in (ρ, m). Also note that we have the natural restrictions of $\rho > 0$ and $m \geq 0$. Note that $m \geq 0$ is enforced by penalising the cost function F, and $\rho > 0$ will be encoded in the convex conjugate formulation. (see Proposition 2.3.2)

Next, introduce a time-space dependent Lagrange multiplier $\phi(t,x)$ for the constraints (2.10) and (2.11) . Hence the associated Lagrangian is

$$L(\phi,\rho,m) = \int_{\mathbb{R}} \int_0^1 \rho(t,x) F\left(\frac{m(t,x)}{\rho(t,x)}\right) + \phi(t,x)\big(\partial_t \rho(x) - \partial_{xx}(m(t,x))\big) dtdx. \tag{2.12}$$

Integrating (2.12) by parts and letting $m = \rho\gamma$ vanish on the boundaries of D, the martingale optimal transport problem can be reformulated as the following saddle point problem:

$$\begin{aligned}\inf_{\rho,m} \sup_{\phi} L(\phi,\rho,m) = \inf_{\rho,m} \sup_{\phi} \int_D \int_0^1 \left(\rho F\left(\frac{m}{\rho}\right) - \rho \partial_t \phi - m \partial_{xx}\phi\right) dtdx \\ - \int_D \left(\phi(0,x)\rho_0 - \phi(1,x)\rho_1\right) dx. \end{aligned} \tag{2.13}$$

As shown by Theorem 3.6 in [103], (2.13) has an equivalent dual formulation which leads to the following representation:

$$\begin{aligned}\sup_{\phi} \inf_{\rho,m} L(\phi,\rho,m) = \sup_{\phi} \inf_{\rho} \int_D \int_0^1 -\rho \left(\partial_t \phi + F^*(\partial_{xx}\phi)\right) dtdx \\ - \int_D \left(\phi(0,x)\rho_0 - \phi(1,x)\rho_1\right) dx. \end{aligned} \tag{2.14}$$

In particular, the optimal ϕ must satisfy the condition

$$\partial_t \phi + F^*(\partial_{xx}\phi) = 0, \tag{2.15}$$

where F^* is the convex conjugate of F (see (2.16) and Proposition 2.3.2). We will later use (2.15) to check the optimality of our algorithm.

Augmented Lagrangian Approach

Similar to [8], we solve the martingale optimal transport problem using the augmented Lagrangian approach. Let us begin by briefly recalling the well-known definition and properties of the convex conjugate. For more details, the readers are referred to Section 12 of Rockafellar [96].

Fix $D \subseteq \mathbb{R}^d$, let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a proper convex and lower semi-continuous function. Then the *convex conjugate* of f is the function $f^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(y) := \sup_{x \in \mathbb{R}^d} (x \cdot y - f(x)). \tag{2.16}$$

The convex conjugate is also often known as the *Legendre-Fenchel transform*.

Proposition 2.3.2. *We have the following properties:*
(i) f^* is a proper convex and lower semi-continuous function with $f^{**} \equiv f$;
(ii) if f is differentiable, then $f(x) + f^*(f'(x)) = xf'(x)$.

Returning to the problem at hand, recall that $G(x,y) := xF(y/x), x > 0$ is convex in (x,y). By adopting the convention of $G(x,y) = \infty$ whenever $x \leq 0$, it can be expressed in terms of the convex conjugate, as shown in the following proposition.

Proposition 2.3.3. *Denote by F^* the convex conjugate of F.*
(i) Let $G(x,y) = xF(y/x)$, the convex conjugate of G is given by:

$$G^*(a,b) = \begin{cases} 0, & \textit{if } a + F^*(b) \leq 0, \\ \infty, & \textit{otherwise.} \end{cases} \tag{2.17}$$

(ii) For $x > 0$, We have the following equality,

$$xF\left(\frac{y}{x}\right) = \sup_{(a,b)\in\mathbb{R}^2} \{ax + by : a + F^*(b) \leq 0\}. \tag{2.18}$$

Proof. (i) By definition, the convex conjugate of G is given by

$$G^*(a,b) = \sup_{(x,y)\in\mathbb{R}^2} \left\{ax + by - xF\left(\frac{y}{x}\right) : x > 0\right\} \tag{2.19}$$

$$= \sup_{(x,y)\in\mathbb{R}^2} \left\{ax + x\left(b\frac{y}{x} - F\left(\frac{y}{x}\right)\right) : x > 0\right\} \tag{2.20}$$

$$= \sup_{x>0} \left\{x(a + F^*(b))\right\}, \tag{2.21}$$

If $a + F^*(b) \leq 0$, the supremum is achieved by limit $x \to 0$, otherwise, G^* becomes unbounded as x increases. This establishes part (i).

(ii) The required equality follows immediately from part (i) and the fact that

$$xF\left(\frac{y}{x}\right) = \sup_{(a,b)\in\mathbb{R}^2} \{ax + by - G^*(a,b) : a + F^*(b) \leq 0\}.$$

□ □

Now we are in a position to present the augmented Lagrangian. First, let us introduce the following notations:

$$K = \left\{(a,b) : \mathbb{R}\times\mathbb{R} \to \mathbb{R}\times\mathbb{R} \;\middle|\; a + F^*(b) \leq 0\right\}, \tag{2.22}$$

$$\mu = (\rho, m) = (\rho, \rho\gamma), \quad q = (a,b), \quad \langle \mu, q\rangle = \int_D \int_0^1 \mu \cdot q, \tag{2.23}$$

$$H(q) = G^*(a,b) = \begin{cases} 0, & \text{if } q \in K, \\ \infty, & \text{otherwise,} \end{cases} \tag{2.24}$$

$$J(\phi) = \int_D [\phi(0,x)\rho_0 - \phi(1,x)\rho_1], \tag{2.25}$$

$$\nabla_{t,xx} = (\partial_t, \partial_{xx}). \tag{2.26}$$

By using the above notations, we can express the equality from Proposition 2.3.3 (ii) in the following way,

$$\rho F\left(\frac{m}{\rho}\right) = \sup_{\{a,b\}\in K} \{a\rho + bm\} = \sup_{q\in K}\{\mu \cdot q\}. \tag{2.27}$$

Since the restriction $q \in K$ is checked point-wise for every (t, x), we can exchange the supremum with the integrals in the following equality

$$\int_D \int_0^1 \sup_{q \in K} \{\mu \cdot q\} = \sup_q \Big\{ -H(q) + \int_D \int_0^1 \mu \cdot q \Big\} = \sup_q \Big\{ -H(q) + \langle \mu, q \rangle \Big\}. \tag{2.28}$$

Therefore, the saddle point problem specified by (2.13) can be rewritten as

$$\sup_\mu \inf_{\phi, q} \Big\{ H(q) + J(\phi) + \langle \mu, \nabla_{t,xx}\phi - q \rangle \Big\}. \tag{2.29}$$

Note that in the new saddle point problem (2.29), μ is the Lagrange multiplier of the new constraint $\nabla_{t,xx}\phi = q$. In order to turn this into a convex problem, we define the augmented Lagrangian as follows:

$$L_r(\phi, q, \mu) = H(q) + J(\phi) + \langle \mu, \nabla_{t,xx}\phi - q \rangle + \frac{r}{2} \langle \nabla_{t,xx}\phi - q, \nabla_{t,xx}\phi - q \rangle, \tag{2.30}$$

where $r > 0$ is a penalisation parameter. The saddle point problem then becomes

$$\sup_\mu \inf_{\phi, q} L_r(\phi, q, \mu), \tag{2.31}$$

which has the same solution as (2.13).

2.4 Numerical Method

In this section, we describe in detail the alternative direction method of multipliers (ADMM) algorithm for solving the saddle point problem given by (2.30) and (2.31). In each iteration, using $(\phi^{n-1}, q^{n-1}, \mu^{n-1})$ as a starting point, the ADMM algorithm performs the following three steps:

$$\text{Step A:} \quad \phi^n = \arg\min_\phi L_r(\phi, q^{n-1}, \mu^{n-1}), \tag{2.32}$$

$$\text{Step B:} \quad q^n = \arg\min_q L_r(\phi^n, q, \mu^{n-1}), \tag{2.33}$$

$$\text{Step C:} \quad \mu^n = \arg\max_\mu L_r(\phi^n, q^n, \mu). \tag{2.34}$$

Step A: $\phi^n = \arg\min_\phi L_r(\phi, q^{n-1}, \mu^{n-1})$

To find the function ϕ^n that minimises $L_r(\phi, q^{n-1}, \mu^{n-1})$, we set the functional derivative of L_r with respect to ϕ to zero:

$$J(\phi) + \langle \mu^{n-1}, \nabla_{t,xx}\phi \rangle + r \langle \nabla_{t,xx}\phi^n - q^{n-1}, \nabla_{t,xx}\phi \rangle = 0. \tag{2.35}$$

By integrating by parts, we arrive at the following variational equation

$$-r(\partial_{tt}\phi^n - \partial_{xxxx}\phi^n) = \partial_t(\rho^{n-1} - ra^{n-1}) - \partial_{xx}(m^{n-1} - rb^{n-1}), \tag{2.36}$$

with Neumann boundary conditions in time $\forall x \in D$:

$$r\partial_t \phi^n(0, x) = \rho_0 - \rho^{n-1}(0, x) + ra^{n-1}(0, x), \tag{2.37}$$

$$r\partial_t \phi^n(1, x) = \rho_1 - \rho^{n-1}(1, x) + ra^{n-1}(1, x). \tag{2.38}$$

For the boundary conditions in space, let $D = [\underline{D}, \overline{D}]$. We give the following boundary condition to the diffusion coefficient:

$$\gamma(t, \underline{D}) = \gamma(t, \overline{D}) = \overline{\gamma} := \arg\min_{\gamma\in\mathbb{R}} F(\gamma).$$

From (2.13) and (2.15), we know $\partial_{xx}\phi$ is the dual variable of γ. Since $\overline{\gamma}$ minimises F, the corresponding $\partial_{xx}\phi$ must be zero. Therefore, we have the following boundary conditions:

$$\partial_{xx}\phi(t, \underline{D}) = \partial_{xx}\phi(t, \overline{D}) = 0, \quad \forall t \in [0, 1]. \tag{2.39}$$

In [8], periodic boundary conditions were used in the spatial dimension and a perturbed equation was used to yield a unique solution. Since periodic boundary conditions are inappropriate for martingale diffusion and we are dealing with a bi-Laplacian term in space, we impose the following additional boundary conditions in order to enforce a unique solution:

$$\phi(t, \underline{D}) = \phi(t, \overline{D}) = 0, \quad \forall t \in [0, 1]. \tag{2.40}$$

Now, the 4th order linear PDE (2.36) can be numerically solved by the finite difference method or the finite element method.

Step B: $q^n = \arg\min_q L_r(\phi^n, q, \mu^{n-1})$

Since $H(q)$ is not differentiable, we cannot differentiate L_r with respect to q. Nevertheless, we can simply obtain q^n by solving the minimisation problem

$$\inf_q L_r(\phi^n, q, \mu^{n-1}). \tag{2.41}$$

This is equivalent to solving

$$\inf_{q\in K} \left\langle \nabla_{t,xx}\phi^n + \frac{\mu^{n-1}}{r} - q, \nabla_{t,xx}\phi^n + \frac{\mu^{n-1}}{r} - q \right\rangle. \tag{2.42}$$

Now, let us define

$$p^n(t,x) = \{\alpha^n(t,x), \beta^n(t,x)\} = \nabla_{t,xx}\phi^n(t,x) + \frac{\mu^{n-1}(t,x)}{r}, \tag{2.43}$$

then we can find $q^n(t,x) = \{a^n(t,x), b^n(t,x)\}$ by solving

$$\inf_{\{a,b\}\in\mathbb{R}\times\mathbb{R}} \Big\{ (a(t,x) - \alpha^n(t,x))^2 + (b(t,x) - \beta^n(t,x))^2 : a + F^*(b) \le 0 \Big\} \tag{2.44}$$

point-wise in space and time. This is a simple one-dimensional projection problem. If $\{\alpha^n, \beta^n\}$ satisfies the constraint $\alpha^n + F^*(\beta^n) \le 0$, then it is also the minimum. Otherwise, the minimum must occur on the boundary $a + F^*(b) = 0$. In this case we substitute the condition into (2.44) to obtain

$$\inf_{b\in\mathbb{R}} \Big\{ (F^*(b(t,x)) + \alpha(t,x))^2 + (b(t,x) - \beta(t,x))^2 \Big\}, \tag{2.45}$$

which must be solved point-wise. The minimum of (2.45) can be found using standard root finding methods such as Newton's method. In some simple cases it is even possible to compute the solution analytically.

Step C: $\mu^n = \arg\max_\mu L_r(\phi^n, q^n, \mu)$

Begin by computing the gradient by differentiating the augmented Lagrangian L_r respect to μ. Then, simply update μ by moving it point-wise along the gradient as follows,

$$\mu^n(t,x) = \mu^{n-1}(t,x) + r(\nabla_{t,xx}\phi^n(t,x) - q^n(t,x)). \tag{2.46}$$

Stopping criteria:

Recall the HJB equation (2.15):

$$\partial_t \phi + F^*(\partial_{xx}\phi) = 0. \tag{2.47}$$

We use (2.47) to check for optimality. Define the residual:

$$res^n = \max_{t\in[0,1], x\in D} \rho \left| \partial_t \phi + F^*(\partial_{xx}\phi) \right|. \tag{2.48}$$

This quantity converges to 0 when it approaches the optimal solution of the problem. The residual is weighted by the density ρ to alleviate any potential issues caused by small values of ρ.

2.5 Numerical Results

The algorithm was implemented and tested on the following simple example. Consider the computational domain $x \in [0,1]$ and the time interval $t \in [0,1]$. We set the initial and final distributions to be $X_0 \sim N(0.5, 0.05^2)$ and $X_1 \sim N(0.5, 0.1^2)$ respectively, where $N(\mu, \sigma^2)$ denotes the normal distribution. The following cost function was chosen:

$$F(\gamma) = \begin{cases} (\gamma - \overline{\gamma})^2, & \gamma \geq 0, \\ +\infty, & \text{otherwise}, \end{cases} \tag{2.49}$$

where $\overline{\gamma}$ was set to 0.00375 so that the optimal value of variance is constant $\sigma^2 = 0.1^2 - 0.05^2 = 0.0075$. Then we discretised the space-time domain as a 128×128 lattice. The penalisation parameter is set to $r = 64$. The results after 3000 iterations are shown in Figures 2.1 and 2.2, and the convergence of the residuals is shown in Figure 2.3. The convergence speed decays quickly, but we reach a good approximation after about 500 iterations. The noisy tails in Figure 2.2 correspond to regions where the density ρ is close to zero. The diffusion process has a very low probability of reaching these regions, so the value of σ^2 has little impact. In areas where ρ is not close to zero, σ^2 remains constant which matches the analytical solution.

2.6 Summary

this chapter focuses on a new approach for the calibration of local volatility models. Given the distributions of the asset price at two fixed dates, the technique of optimal transport is applied to interpolate the distributions and recover the local volatility function, while maintaining the martingale property of the underlying process. Inspired by [8], the problem is first converted into a saddle point problem, and then solved numerically by an augmented

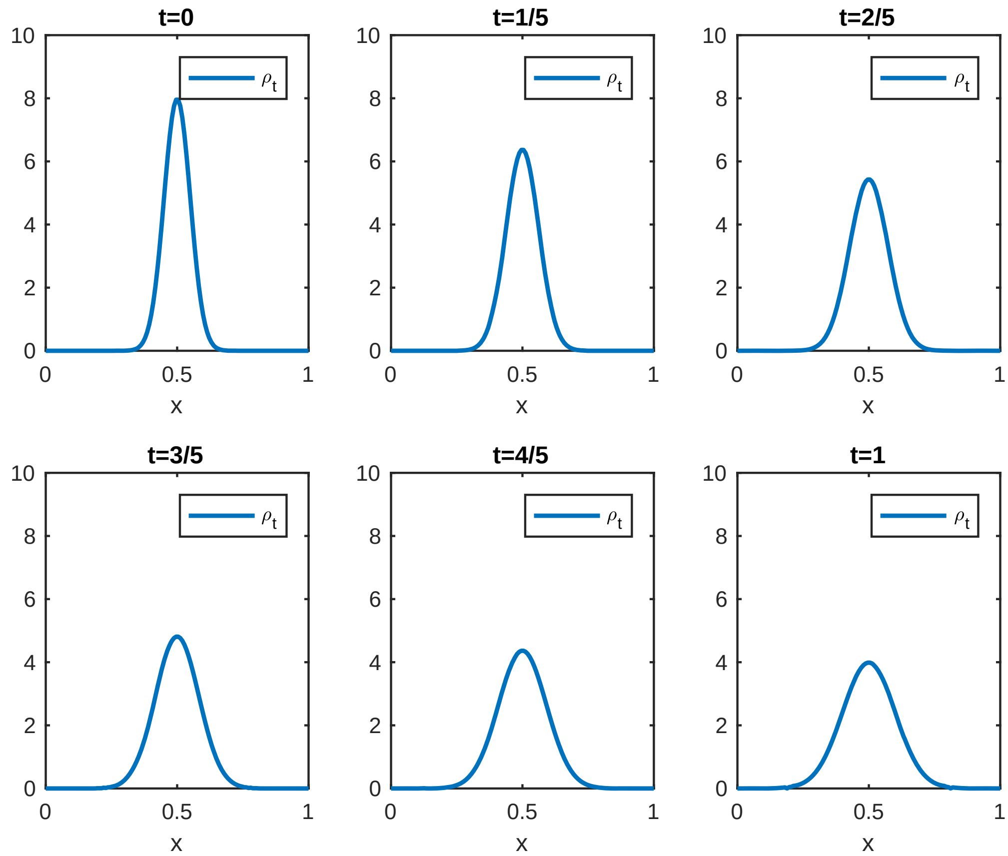

Figure 2.1: The density function $\rho(t, x)$.

Lagrangian approach and the alternative direction method of multipliers (ADMM) algorithm. The algorithm performs well on a simple case in which the numerical solution matches its analytical counterpart. The main drawback of this method is due to the slow convergence rate of the ADMM algorithm. We observed that a higher penalisation parameter may lead to faster convergence. Further research is required to conduct more numerical experiment, improve the efficiency of the algorithm and apply it to more complex cases.

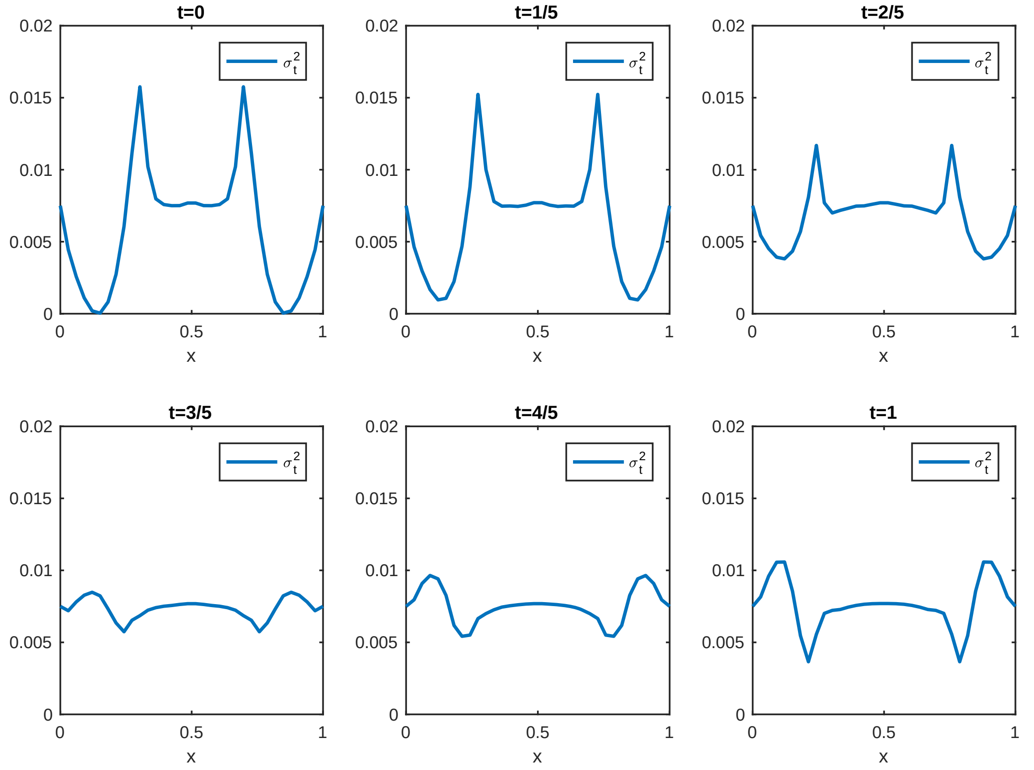

Figure 2.2: The variance $\sigma^2(t,x)$.

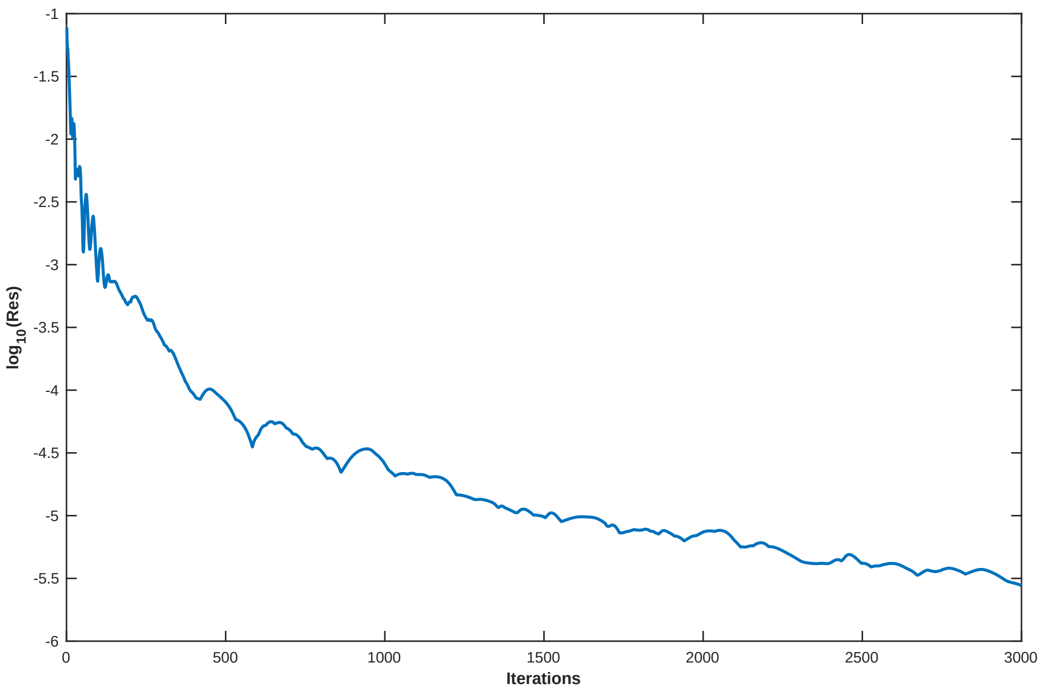

Figure 2.3: The residual res^n.

Chapter 3

Calibration of local-stochastic volatility models by optimal transport

The objective of this chapter is to extend the approach developed in Chapter 2 to the calibration of the so-called local-stochastic volatility (LSV) models. Although the local volatility model can be exactly calibrated to any arbitrage-free implied volatility surface, it has been criticised for its unrealistic volatility dynamics. In this chapter, we consider a class of LSV models whose volatility is in the form of a function of time, underlying asset price and a mean-reverting stochastic factor. We formulate the calibration problem as a semimartingale optimal transport problem. Rather than considering the classical constraints on marginal distributions at initial and final time as in Chapter 2, we optimise our cost function given the spot price and the prices of a finite number of European options. We further formulate the problem as a convex optimisation problem, for which we provide a PDE formulation along with its dual counterpart. Then we develop a gradient descent method to numerically solve the dual formulation, which involves solving a fully non-linear Hamilton–Jacobi–Bellman equation at each optimisation iteration. The method is tested by calibrating a LSV model with simulated data and foreign exchange market data. Numerical results show that the method effectively calibrates the model fully to the European option prices with both simulated data and market data.

3.1 Introduction

Since the introduction of the Black–Scholes model, a lot of effort has been put on developing sophisticated volatility models that properly capture the market dynamics. In the space of equities and currencies, the most widely used models are the Local Volatility (LV) model by Dupire [37] and the Stochastic Volatility (SV) models [see e.g., 46, 64]. Introduced as an extension of the Black–Scholes model, the LV model can be exactly calibrated to any arbitrage-free implied volatility surface. Despite this feature, the LV model has often been criticised for its unrealistic volatility dynamics. The SV models tend to be more consistent with the market dynamics, but they struggle to fit short term market smiles and skews, and being parametric, they do not have enough degrees of freedom to match all vanilla market prices. A better fit can be obtained by increasing the number of stochastic factors in the SV models; however, this also increases the complexity of calibration and pricing.

Local-Stochastic Volatility (LSV) models, introduced in Jex et al. [70], naturally extend and take advantage of both approaches. The idea behind LSV models is to incorporate a local, non-parametric, factor into the SV models. Thus, while keeping consistent dynamics, the model can match all observed market prices (as long as one restricts to European claims). The determination of this local factor (also called *leverage*) is based on the mimicking theorem by Gyöngy [58]. Research into the numerical calibration of LSV models has been developed in two different directions. One is based on a Monte Carlo approach, with Henry-Labordère [61], followed by Guyon and Henry-Labordère [57] using a so-called McKean's particle method. Another approach relies on solving the Fokker–Planck equation as in Ren et al. [95]. Engelmann et al. [38] used the finite volume method (FVM) to solve the partial differential equation (PDE), while Tian et al. [104] considered time-dependent parameters. In a more recent study, Wyns and Du Toit [108] considered a method that combines the FVM with alternating direction implicit (ADI) schemes.

All of the calibration methods mentioned above require a priori knowledge of the Local Volatility surface. This is usually obtained by using Dupire's formula [37] assuming the knowledge of vanilla options for all strikes and maturities. However, only a finite number of options are available in practice. Thus, an interpolation of the implied volatility surface or option prices is often needed, which can lead to inaccuracies and instabilities. Inaccuracies can come from the usage of a parametric model for the volatility surface that will not match perfectly market prices by definition. Instabilities can come form the interpolating model being not arbitrage-free. It also raises the question of what arbitrary shape of extrapolation one is going to take for very out of the money strikes. Moreover, there is no a priori control on the regularity of the leverage function, and even its very existence remains an open problem, although some results for small times have been obtained in Abergel and Tachet [1] [cf. 98, for an application of Tikhonov regularisation technique to the LSV calibration problem]. Other related works include Jourdain and Zhou [71], Lacker et al. [80]. In a recent work of Cuchiero et al. [30], the LSV calibration problem was addressed from a deep learning point of view. In particular, the leverage function is parameterised by a class of feed-forward neural networks, and the model is calibrated by a generative adversarial network approach. In the present work, inspired by the theory of optimal transport, we introduce a variational approach for calibrating LSV models that does not require any form of interpolation.

In this chapter, we further extend the approach of Guo et al. [51] and Guo and Loeper [50] to the calibration of LSV models. The calibration problem is formulated as a semi-martingale optimal transport problem. Unlike Tan and Touzi [103], we consider a finite number of discrete constraints given by the prices of European claims. As a consequence of Jensen's inequality, we show that an optimal diffusion process can be chosen to be Markovian in the state variables given by the initial SV model. This result leads to a PDE formulation. By following the duality theory of optimal transport introduced in Brenier [20] and a smoothing argument used in Bouchard et al. [18], we establish a dual formulation. We also provide a numerical method to solve a fully non-linear Hamilton–Jacobi–Bellman (HJB) equation arising in the dual formulation. Finally, numerical examples show that the model can be fully calibrated to the European options with both simulated data and FX market data.

Despite its accuracy, our method is quite demanding in terms of computational power. The gradient descent demands at each step to solve one non-linear $2d$ PDE, and the computation of the gradient requires one (linear) $2d$ PDE per instrument. The most costly operation of numerically solving a linear PDE is inverting a large sparse matrix. However, this operation only needs to be carried once per time step because the computations of all components of the gradient are computed by solving the same linear PDE but with different terminal conditions. Alternatively, all gradients can be efficiently computed in one Monte Carlo simulation, which is a choice we did not make here for the sake of accuracy. Going into higher dimensions (a multi-factor stochastic volatility model for example) would require to increase the dimension of the PDEs, which is problematic as soon as $d \geq 3$. When the goal is only to solve the usual LSV calibration problem, i.e., to find the leverage function, other methods achieve the result faster. For example, for a one-factor model, the PDE method of Ren et al. [95] only requires solving a two-dimensional non-linear PDE once, and the particle method of Guyon and Henry-Labordère [57] is even faster and can be applied to high dimensional cases (e.g., calibrating an LSV model with multiple stochastic factors). On the other hand, with the technique developed in this chapter, one can fit path-dependent products [50], SPX and VIX options [52] and here LSV models. Therefore the interest of our method is clearly its broad range of applications, at the cost of a relatively heavy computational cost. We also believe that with the recent developments of numerical methods for solving non-linear PDEs in high dimensions [see e.g., 60], our method can be greatly improved in terms of computational speed, and become applicable in high dimensions. Also notice that, being based on gradient descent, for a slight update of the market data, only a few gradient iterations should be needed to update the model. Finally, our method provides a rigorous existence result of an LSV type model. Previous works by Abergel and Tachet [1] only provide an existence result for small times (see also Jourdain and Zhou [71], Lacker et al. [80]).

This chapter is organised as follows: In Section 2, we introduce some preliminary definitions. In Section 3, we show the connection between the semi-martingale optimal transport problem and a PDE formulation. Duality results are then established for the PDE formulation. In Section 4, we demonstrate the calibration method using a Heston-like LSV model. Numerical method and results with both simulated data and FX market data are provided in Section 5.

3.2 Preliminaries

Given a Polish space E equipped with its Borel σ-algebra, let $C(E)$ be the space of continuous functions on E and $C_b(E)$ be the space of bounded continuous functions. Denote by $\mathcal{M}(E)$ the space of finite signed Borel measures endowed with the weak-$*$ topology. Let $\mathcal{M}_+(E) \subset \mathcal{M}(E)$ denote the subset of nonnegative measures. If E is compact, the topological dual of $C_b(E)$ is given by $C_b(E)^* = \mathcal{M}(E)$. More generally, if E is non-compact, $C_b(E)^*$ is larger than $\mathcal{M}(E)$. Let $\mathcal{P}(E)$ be the space of Borel probability measures, $BV(E)$ be the space of functions of bounded variation and $L^1(d\mu)$ be the space of μ-integrable functions. We also write $C_b(E, \mathbb{R}^d)$, $\mathcal{M}(E, \mathbb{R}^d)$, $BV(E, \mathbb{R}^d)$ and $L^1(d\mu, \mathbb{R}^d)$ as the vector-valued versions of their corresponding spaces. If $\mu_t(x) = \mu(t, x)$ is a measure defined on $[0, T] \times \mathbb{R}^d$, we will write $d\mu$ or $d\mu_t dt$ in short for $\mu(t, dx)dt$. Denote by $\mathbb{S}^d$ the set of $d \times d$ symmetric matrices and $\mathbb{S}^d_+ \subset \mathbb{S}^d$ the set of positive semidefinite matrices. For any matrices $A, B \in \mathbb{S}^d$, we write $A : B := \operatorname{tr}(A^\intercal B)$ for their scalar product. For convenience, let $\Lambda = [0, T] \times \mathbb{R}^d$ and $\mathcal{X} = \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d$. We use the notation $\langle \cdot, \cdot \rangle$ to denote the duality bracket between $C_b(\Lambda, \mathcal{X})$ and $C_b(\Lambda, \mathcal{X})^*$.

Let $\Omega := C([0, T], \mathbb{R}^d), T > 0$ be the canonical space with the canonical process X and the canonical filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ generated by X. We denote by $\mathcal{P}$ the collection of all probability measures $\mathbb{P}$ on $(\Omega, \mathcal{F}_T)$ under which $X \in \Omega$ is an $(\mathbb{F}, \mathbb{P})$-semi-martingale given by

$$X_t = X_0 + A_t + M_t, \quad t \in [0, T], \quad \mathbb{P}\text{-a.s.},$$

where M is an $(\mathbb{F}, \mathbb{P})$-martingale with quadratic variation $\langle X_t \rangle = \langle M_t \rangle = B_t$, and the processes A and B are $\mathbb{P}$-a.s. absolutely continuous with respect to t. We say $\mathbb{P}$ is *characterised* by $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ if

$$\alpha_t^{\mathbb{P}} = \frac{dA_t^{\mathbb{P}}}{dt}, \quad \beta_t^{\mathbb{P}} = \frac{dB_t^{\mathbb{P}}}{dt},$$

where $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ take values in the space $\mathbb{R}^d \times \mathbb{S}_+^d$. Note that $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ is $\mathbb{F}$-adapted and determined up to $d\mathbb{P} \times dt$, almost everywhere. Let $\mathcal{P}^1 \subset \mathcal{P}$ be the subset of probability measures $\mathbb{P}$ under which the characteristics $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ are $\mathbb{P}$-integrable on the interval $[0, T]$. In other words,

$$\mathbb{E}^{\mathbb{P}} \left(\int_0^T |\alpha_t^{\mathbb{P}}| + |\beta_t^{\mathbb{P}}| \, dt \right) < +\infty,$$

where $|\cdot|$ is the L^1-norm.

Given a vector $\tau := (\tau_1, \ldots, \tau_m) \in (0, T]^m$, denote by G a vector of m functions such that each function $G_i \in C_b(\mathbb{R}^d)$ for $i = 1, \ldots, m$. Given a Dirac measure $\mu_0 = \delta_{x_0}$ and a vector $c \in \mathbb{R}^m$, we define $\mathcal{P}(\mu_0, \tau, c, G) \subset \mathcal{P}^1$ as follows:

$$\mathcal{P}(\mu_0, \tau, c, G) := \{\mathbb{P} : \mathbb{P} \in \mathcal{P}^1,\, \mathbb{P} \circ X_0^{-1} = \mu_0 \text{ and } \mathbb{E}^{\mathbb{P}}[G_i(X_{\tau_i})] = c_i,\, i = 1, \ldots, m\}.$$

Assumption 3.2.1. The final time T coincides with the longest maturity, i.e., $T = \max_k \tau_k$.

For technical reasons, we restrict ourselves to functions G_i in $C_b(\mathbb{R}^d)$. In the context of volatility models calibration, G_i are discounted European payoffs. Although the call option payoff functions are not technically in $C_b(\mathbb{R}^d)$, we only work with them in a truncated (compact) space in practice. Alternatively, one may consider only put options using put-call parity. It is possible to relax the assumption $G_i \in C_b(\mathbb{R}^d)$, but it would require a different set up in topological spaces.

3.3 Main results

3.3.1 Formulations

In this section, we first formulate the semi-martingale optimal transport problem under discrete constraints. Then a PDE formulation is introduced along with its dual counterpart.

Define the cost function $F : \Lambda \times \mathbb{R}^d \times \mathbb{S}^d \to \mathbb{R} \cup \{+\infty\}$ where $F(t, x, \alpha, \beta) = +\infty$ if $\beta \notin \mathbb{S}_+^d$, and $F(t, x, \alpha, \beta)$ is nonnegative, proper, lower semi-continuous, strongly convex and coercive in (α, β) and uniformly in (t, x). By F being strongly convex in (α, β) we mean that there exists a constant $C > 0$ such that for all $t, x, \alpha, \beta, \alpha', \beta'$ and any subderivative ∇F, where ∇ is performed over (α, β), if $F(t, x, \alpha, \beta)$ is finite then

$$F(t, x, \alpha', \beta') \geq F(t, x, \alpha, \beta) + \langle \nabla F(t, x, \alpha, \beta), (\alpha' - \alpha, \beta' - \beta) \rangle + C(\|\alpha' - \alpha\|^2 + \|\beta' - \beta\|^2),$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$ and $\mathbb{S}^d$. By F being coercive in (α, β) we mean that there exist constants $p > 1$ and $C > 0$ for all t, x, α, β such that

$$|\alpha|^p + |\beta|^p \leq C(1 + F(t, x, \alpha, \beta)).$$

The convex conjugate of F with respect to (α, β) is denoted by $F^* : \Lambda \times \mathbb{R}^d \times \mathbb{S}^d \to \mathbb{R} \cup \{+\infty\}$ and is given by

$$F^*(t, x, a, b) := \sup_{\alpha \in \mathbb{R}^d, \beta \in \mathbb{S}^d} \{\alpha \cdot a + \beta : b - F(t, x, \alpha, \beta)\}. \tag{3.1}$$

We remark that $\mathbb{S}^d$ in (3.1) can be replaced by $\mathbb{S}^d_+$ due to the assumption that $F(t,x,\alpha,\beta)$ is finite only if $\beta \in \mathbb{S}^d_+$. For simplicity, we write $F(\alpha,\beta) := F(t,x,\alpha,\beta)$ and $F^*(a,b) := F^*(t,x,a,b)$ if there is no ambiguity. Note that our definition of strongly convex does not require F to be differentiable, since only subderivatives are used. Nevertheless, it implies that F is strictly convex and thus F^* is differentiable. In addition, the coercivity of F implies that F^* is finite.

Adopting the convention $\inf \emptyset = +\infty$, we are interested in the following minimisation problem:

Problem 1. Given μ_0, τ, c and G, we want to find

$$\mathcal{V} = \inf_{\mathbb{P} \in \mathcal{P}(\mu_0,\tau,c,G)} \mathbb{E}^{\mathbb{P}} \int_0^T F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})\, dt.$$

The problem is said to be *admissible* if $\mathcal{P}(\mu_0,\tau,c,G)$ is nonempty and the infimum above is finite.

It is well known that the marginal distributions of diffusion processes at fixed times solve the Fokker–Planck equation in the weak sense. The converse result was given by Figalli [40] and Trevisan [105]. For brevity, we write $\mathbb{E}^{\mathbb{P}}_{t,x} := \mathbb{E}^{\mathbb{P}}[\,\cdot \mid X_t = x]$. As an immediate consequence of Itô's formula and Theorem 2.5 in Trevisan [105], we introduce the following lemma.

Lemma 3.3.1. *Let $\mathbb{P} \in \mathcal{P}^1$ and $\rho_t^{\mathbb{P}} = \rho^{\mathbb{P}}(t,\cdot) = \mathbb{P} \circ X_t^{-1}$ be the marginal distribution of X_t under $\mathbb{P}$, $t \le T$. Then $\rho^{\mathbb{P}}$ is a weak solution to the Fokker–Planck equation:*

$$\begin{cases} \partial_t \rho_t^{\mathbb{P}} + \nabla_x \cdot (\rho_t^{\mathbb{P}} \mathbb{E}^{\mathbb{P}}_{t,x} \alpha_t^{\mathbb{P}}) - \dfrac{1}{2} \displaystyle\sum_{i,j} \partial_{ij} (\rho_t^{\mathbb{P}} (\mathbb{E}^{\mathbb{P}}_{t,x} \beta_t^{\mathbb{P}})_{ij}) = 0 & \text{in } [0,T] \times \mathbb{R}^d, \\ \rho_0^{\mathbb{P}} = \delta_{X_0} & \text{in } \mathbb{R}^d. \end{cases} \tag{3.2}$$

Moreover, there exists another probability measure $\mathbb{P}' \in \mathcal{P}^1$, characterised by $(\alpha^{\mathbb{P}'}, \beta^{\mathbb{P}'})$, under which X has the same marginals, $\rho^{\mathbb{P}'} = \rho^{\mathbb{P}}$, and is a Markov process solving

$$\begin{cases} dX_t = \alpha^{\mathbb{P}'}(t,X_t)dt + (\beta^{\mathbb{P}'}(t,X_t))^{\frac{1}{2}}\, dW_t^{\mathbb{P}'}, \quad 0 \le t \le T, \\ X_0 = x_0, \end{cases} \tag{3.3}$$

where $W^{\mathbb{P}'}$ is a $\mathbb{P}'$-Brownian motion, $\alpha^{\mathbb{P}'}(t,X_t) = \mathbb{E}^{\mathbb{P}}_{t,X_t} \alpha_t^{\mathbb{P}}$ and $\beta^{\mathbb{P}'}(t,X_t) = \mathbb{E}^{\mathbb{P}}_{t,X_t} \beta_t^{\mathbb{P}}$.

The above lemma provides a solution to study semi-martingales via Markov processes in the form of (3.3). It is worth noting that the idea of using diffusion processes to mimic an Itô process by matching their marginals at fixed times traces back to the classical mimicking theorem of Gyöngy [58]. The uniform ellipticity condition of Gyöngy's mimicking theorem was later relaxed by Brunick and Shreve [23]. In fact, if X is an Itô process under $\mathbb{P}$, Lemma 3.3.1 can be seen as a reformulation of Brunick and Shreve [23, Corollary 3.7] which was constructed by a completely different approach. The Markov processes X in (3.3) are also called Markovian projections in the literature. Note that, in Brunick and Shreve [23], even though the main results are given for Itô processes, the authors first provide more general results for semi-martingales (see Brunick and Shreve [23, Theorem 7.1]) and then prove the main results for Itô processes by the Itô representation theorem. Therefore, Lemma 3.3.1 can also be proved by the results of Brunick and Shreve [23].

Definition 3.3.2. Define $\mathcal{P}_{loc}(\mu_0,\tau,c,G)$ to be the subset of $\mathcal{P}(\mu_0,\tau,c,G)$ such that, under any $\mathbb{P} \in \mathcal{P}_{loc}(\mu_0,\tau,c,G)$, X is a Markov process that takes the form of (3.3).

Lemma 3.3.3. *If $\mathcal{P}(\mu_0,\tau,c,G)$ is not empty, then $\mathcal{P}_{loc}(\mu_0,\tau,c,G)$ is not empty. Moreover, for any $\mathbb{P}\in\mathcal{P}(\mu_0,\tau,c,G)$, there exists a $\mathbb{P}'\in\mathcal{P}_{loc}(\mu_0,\tau,c,G)$ such that X has the same marginals under $\mathbb{P}$ and $\mathbb{P}'$.*

Proof. Assume that $\mathcal{P}(\mu_0,\tau,c,G)$ is not empty, for any $\mathbb{P}\in\mathcal{P}(\mu_0,\tau,c,G)$, by Lemma 3.3.1, there exists $\mathbb{P}'\in\mathcal{P}^1$ such that X is a Markov process that has the same marginals $\rho^{\mathbb{P}'}=\rho^{\mathbb{P}}$ and takes the form of (3.3) with coefficients $(\alpha^{\mathbb{P}'}(t,X_t),\beta^{\mathbb{P}'}(t,X_t))=(\mathbb{E}^{\mathbb{P}}_{t,X_t}\alpha^{\mathbb{P}}_t,\mathbb{E}^{\mathbb{P}}_{t,X_t}\beta^{\mathbb{P}}_t)$. Since $\rho^{\mathbb{P}'}=\rho^{\mathbb{P}}$, X has the initial marginal μ_0 and satisfies $\mathbb{E}^{\mathbb{P}'}[G_i(X_{\tau_i})]=c_i$ for all $i=1,\dots,m$ under both $\mathbb{P}$ and $\mathbb{P}'$. Thus, $\mathbb{P}'\in\mathcal{P}_{loc}(\mu_0,\tau,c,G)$. □

Applying Lemma 3.3.3 and taking advantage of the convexity of the cost function, we establish the following result:

Proposition 3.3.4. *Given μ_0,τ,c and G, then*

$$\mathcal{V}=\inf_{\mathbb{P}\in\mathcal{P}(\mu_0,\tau,c,G)}\mathbb{E}^{\mathbb{P}}\int_0^T F(\alpha^{\mathbb{P}}_t,\beta^{\mathbb{P}}_t)\,dt=\inf_{\mathbb{P}\in\mathcal{P}_{loc}(\mu_0,\tau,c,G)}\mathbb{E}^{\mathbb{P}}\int_0^T F(\alpha^{\mathbb{P}}(t,X_t),\beta^{\mathbb{P}}(t,X_t))\,dt. \tag{3.4}$$

Proof. If $\mathcal{P}(\mu_0,\tau,c,G)$ is empty, then $\mathcal{P}_{loc}(\mu_0,\tau,c,G)$ is empty since $\mathcal{P}_{loc}(\mu_0,\tau,c,G)\subset\mathcal{P}(\mu_0,\tau,c,G)$. Thus, (3.4) holds and $\mathcal{V}=+\infty$.

If $\mathcal{P}(\mu_0,\tau,c,G)$ is not empty, by Lemma 3.3.3, $\mathcal{P}_{loc}(\mu_0,\tau,c,G)$ is not empty. For any $\mathbb{P}\in\mathcal{P}(\mu_0,\tau,c,G)$, let $\mathbb{P}'\in\mathcal{P}_{loc}(\mu_0,\tau,c,G)$ be a probability measure such that X has the same marginals under $\mathbb{P}$ and $\mathbb{P}'$. Applying Jensen's inequality together with the tower property of conditional expectation, we have

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\int_0^T F(\alpha^{\mathbb{P}}_t,\beta^{\mathbb{P}}_t)\,dt&=\mathbb{E}^{\mathbb{P}}\int_0^T \mathbb{E}^{\mathbb{P}}_{t,X_t}F(\alpha^{\mathbb{P}}_t,\beta^{\mathbb{P}}_t)\,dt\\&\geq\mathbb{E}^{\mathbb{P}}\int_0^T F(\mathbb{E}^{\mathbb{P}}_{t,X_t}\alpha^{\mathbb{P}}_t,\mathbb{E}^{\mathbb{P}}_{t,X_t}\beta^{\mathbb{P}}_t)\,dt\\&=\mathbb{E}^{\mathbb{P}'}\int_0^T F(\alpha^{\mathbb{P}'}(t,X_t),\beta^{\mathbb{P}'}(t,X_t))\,dt.\end{aligned}\tag{3.5}$$

The last $\mathbb{E}^{\mathbb{P}}$ is replaced by $\mathbb{E}^{\mathbb{P}'}$ because the marginal of X is the same under $\mathbb{P}$ and $\mathbb{P}'$. Since $\mathcal{P}_{loc}(\mu_0,\tau,c,G)\subset\mathcal{P}(\mu_0,\tau,c,G)$, taking infimum over all $\mathbb{P}\in\mathcal{P}(\mu_0,\tau,c,G)$ on the left-hand side and over all $\mathbb{P}'\in\mathcal{P}_{loc}(\mu_0,\tau,c,G)$ on the right-hand side of (3.5), we obtain the required result. □

Proposition 3.3.4 shows that it suffices to consider only the probability measures in $\mathcal{P}_{loc}(\mu_0,\tau,c,G)$. Thus, by the connections established in Lemma 3.3.1, Problem 1 can be studied via PDE methods. Following the Benamou–Brenier formulation of the classical optimal transport from Benamou and Brenier [8], we introduce the following problem:

Problem 2 (PDE formulation)**.** Given μ_0,τ,c and G, we want to solve

$$\mathcal{V}=\inf_{\rho,\alpha,\beta}\int_0^T\int_{\mathbb{R}^d}F(\alpha(t,x),\beta(t,x))\,\rho(t,dx)dt, \tag{3.6}$$

among all $(\rho,\alpha,\beta)\in C([0,T],\mathcal{P}(\mathbb{R}^d)-w*)\times L^1(d\rho_t dt,\mathbb{R}^d)\times L^1(d\rho_t dt,\mathbb{S}^d)$ satisfying (in the distributional sense)

$$\partial_t\rho(t,x)+\nabla_x\cdot(\rho(t,x)\alpha(t,x))-\frac{1}{2}\sum_{i,j}\partial_{ij}(\rho(t,x)\beta_{ij}(t,x))=0, \tag{3.7}$$

$$\int_{\mathbb{R}^d}G_i(x)\,\rho(\tau_i,dx)=c_i,\ \forall i=1,\dots,m,\qquad\text{and}\qquad\rho(0,\cdot)=\mu_0. \tag{3.8}$$

The interchange of integrals in (3.6) is justified by Fubini's theorem as F is nonnegative. For the weak continuity of measure ρ in time, the reader can refer to Loeper [82, Theorem 3].

Based on the results of Sections 3.3.2 and 3.3.3 below, we shall introduce a dual formulation of Problem 2. In the proposition below, $C_b^2(\mathbb{R}^d)$ is the space of twice continuously differentiable functions with bounded partial derivatives up to order 2, and it is equipped with the norm given by the supremum of all partial derivatives up to order 2. The subscript of ϕ_λ indicates the implicit dependence of ϕ on λ via the HJB equation. The definition of the viscosity solution to (3.10) and the proof will be given in Section 3.3.3.

Proposition 3.3.5 (Dual formulation)**.** *If Problem 1 is admissible, then*

$$\mathcal{V} = \sup_{\lambda \in \mathbb{R}^m} \left\{ \sum_{i=1}^{m} \lambda_i c_i - \int_{\mathbb{R}^d} \phi_\lambda(0,x)\, d\mu_0 \right\}, \tag{3.9}$$

where ϕ is the viscosity solution to the HJB equation

$$\partial_t \phi_\lambda + \sum_{i=1}^{m} \lambda_i G_i \delta_{\tau_i} + F^*(\nabla_x \phi_\lambda, \frac{1}{2}\nabla_x^2 \phi_\lambda) = 0, \quad \text{in } [0,T) \times \mathbb{R}^d, \tag{3.10}$$

with the terminal condition $\phi_\lambda(T,\cdot) = 0$. Moreover, if there exists $(\rho,\alpha,\beta) \in C([0,T],\mathcal{P}(\mathbb{R}^d)-w*) \times L^1(d\rho_t dt, \mathbb{R}^d) \times L^1(d\rho_t dt, \mathbb{S}^d)$ satisfying (3.7) and (3.8) (in the distributional sense), then the infimum of Problem 2 is attained. If the supremum is attained by some $\lambda^* \in \mathbb{R}^m$ and (ρ,α,β) is an optimal solution of Problem 2, then (α,β) is given by

$$(\alpha,\beta) = \nabla F^*(\nabla_x \phi_{\lambda^*}, \frac{1}{2}\nabla_x^2 \phi_{\lambda^*}), \quad d\rho_t dt - \textit{almost everywhere}. \tag{3.11}$$

Before ending this section, it is worth commenting on the admissibility of Problem 1. We have chosen to impose the admissibility assumption in order to simplify our presentation and arguments. With some modifications, it is possible to remove this assumption from the primal problem and still obtain duality. In particular, both sides of the duality would be infinite if the problem is not admissible. Then, characterising the admissibility of Problem 1 corresponds to checking the finiteness of the dual problem, and can be seen as a more elaborate analogue of Strassen's theorem for the classical optimal transport problem.

3.3.2 Duality

This section is devoted to establishing the duality by closely following Loeper [82, Section 3.2] [see also 20, 66].

Theorem 3.3.6. *If Problem 1 is admissible, then*

$$\mathcal{V} = \sup_{\phi,\lambda} \left\{ \sum_{i=1}^{m} \lambda_i c_i - \int_{\mathbb{R}^d} \phi(0,x)\, d\mu_0 \right\}, \tag{3.12}$$

where the supremum is taken over all $(\phi,\lambda) \in BV([0,T], C_b^2(\mathbb{R}^d)) \times \mathbb{R}^m$ satisfying

$$\partial_t \phi + \sum_{i=1}^{m} \lambda_i G_i \delta_{\tau_i} + F^*(\nabla_x \phi, \frac{1}{2}\nabla_x^2 \phi) \le 0 \qquad \text{in } [0,T) \times \mathbb{R}^d, \tag{3.13}$$

and $\phi(T,\cdot) = 0$. Moreover, if there exists $(\rho,\alpha,\beta) \in C([0,T],\mathcal{P}(\mathbb{R}^d)-w*) \times L^1(d\rho_t dt, \mathbb{R}^d) \times L^1(d\rho_t dt, \mathbb{S}^d)$ satisfying (3.7) and (3.8) (in the distributional sense), then the infimum of Problem 2 is attained.

Proof. The proof relies on the Fenchel–Rockafellar theorem which plays a key role in the applications of convex analysis. One may note that the objective function (3.6) is not convex in (ρ, α, β) since $F(\alpha, \beta)\rho$ is not convex in (ρ, α, β). As we will see below, (3.6) can be written as the convex conjugate (which is always convex) of another function with respect to $(\rho, \mathcal{A} := \alpha\rho, \mathcal{B} := \beta\rho)$ and $(\mathcal{A}, \mathcal{B})$ are absolutely continuous with respect to ρ. In addition, the constraints (3.7) and (3.8) are linear in $(\rho, \mathcal{A}, \mathcal{B})$. Therefore, throughout the proof, we will work on $(\rho, \mathcal{A}, \mathcal{B})$ instead. For simplicity, we will write $d\mathcal{A}$ and $d\mathcal{B}$ in short for $\alpha(t,x)\rho(t,dx)dt$ and $\beta(t,x)\rho(t,dx)dt$, respectively.

Formulate the constraints (3.7) and (3.8) in the following weak form:

$$\forall \phi \in C_c^\infty(\Lambda), \qquad \int_\Lambda \partial_t\phi \, d\rho + \nabla_x\phi \cdot d\mathcal{A} + \frac{1}{2}\nabla_x^2\phi : d\mathcal{B} + \int_{\mathbb{R}^d} \phi(0,\cdot)\, d\mu_0 = 0, \quad \phi(T,\cdot) = 0 \tag{3.14}$$

$$\forall \lambda \in \mathbb{R}^m, \qquad \int_\Lambda \sum_{i=1}^m \lambda_i G_i \delta_{\tau_i} d\rho - \sum_{i=1}^m \lambda_i c_i = 0. \tag{3.15}$$

where $C_c^\infty(\Lambda)$ is the space of smooth functions with compact support on Λ. Thus Problem 2 can be reformulated as the following saddle point problem:

$$\begin{aligned}\mathcal{V} = \inf_{\rho,\mathcal{A},\mathcal{B}} \sup_{\phi,\lambda} \Bigg\{ \int_\Lambda F\left(\frac{d\mathcal{A}}{d\rho}, \frac{d\mathcal{B}}{d\rho}\right) d\rho - \partial_t\phi\, d\rho - \nabla_x\phi \cdot d\mathcal{A} - \frac{1}{2}\nabla_x^2\phi : d\mathcal{B} - \int_{\mathbb{R}^d} \phi(0,\cdot)\, d\mu_0 \\ - \int_\Lambda \sum_{i=1}^m \lambda_i G_i \delta_{\tau_i}\, d\rho + \sum_{i=1}^m \lambda_i c_i \Bigg\}.\end{aligned} \tag{3.16}$$

The strategy of the proof is to first construct a function Φ whose convex conjugate Φ^* is equal to the objective function of Problem 2, and construct another function Ψ whose convex conjugate Ψ^* is equal to the rest part inside the infimum of (3.16) so that $\mathcal{V} = \inf_{\rho,\mathcal{A},\mathcal{B}}(\Phi^* + \Psi^*)(\rho, \mathcal{A}, \mathcal{B})$. Then, the duality is established by applying the Fenchel–Rockafellar theorem.

Adopting the terminology of Huesmann and Trevisan [66], we say the triple (r, a, b) is *represented* by (ϕ, λ) if it satisfies

$$\begin{aligned} r + \partial_t\phi + \sum_{i=1}^m \lambda_i G_i \delta_{\tau_i} &= 0, \\ a + \nabla_x\phi &= 0, \\ b + \frac{1}{2}\nabla_x^2\phi &= 0. \end{aligned}$$

If we choose (r, a, b) from $C_b(\Lambda, \mathcal{X})$, by the first equation above, $\partial_t\phi$ is a measure because of the presence of the Dirac delta functions. Thus, ϕ has bounded variation with respect to t on $[0,T]$ and has possible jump discontinuities at $t = \tau_i$. Now, define functionals $\Phi : C_b(\Lambda, \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ and $\Psi : C_b(\Lambda, \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ as follows:

$$\Phi(r,a,b) = \begin{cases} 0 & \text{if } r + F^*(a,b) \le 0, \\ +\infty & \text{otherwise,} \end{cases}$$

$$\Psi(r,a,b) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} \phi(0,x)\, d\mu_0 - \sum_{i=1}^m \lambda_i c_i & \text{if } (r,a,b) \text{ is represented by } (\phi,\lambda) \\ & \text{in } BV([0,T], C_b^2(\mathbb{R}^d)) \times \mathbb{R}^m \text{ with } \phi(T,\cdot) = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Note that Ψ is well-defined. If $\Psi(r,a,b) < +\infty$ for some (r,a,b) that is represented by some (ϕ,λ), then (ϕ,λ) satisfies the constraints (3.14) and (3.15), otherwise we can arbitrarily scale (ϕ,λ) in (3.16) then $\mathcal{V}$ becomes unbounded. Assume that (r,a,b) can be represented by both $(\hat{\phi},\hat{\lambda})$ and $(\tilde{\phi},\tilde{\lambda})$, then we have

$$\partial_t(\hat{\phi}-\tilde{\phi}) + \sum_{i=1}^{m}(\hat{\lambda}_i - \tilde{\lambda}_i)G_i\delta_{\tau_i} = 0. \tag{3.17}$$

Integrating (3.17) with any ρ that satisfies (3.15) and $\rho(0,\cdot) = \mu_0$, we have $\int_{\mathbb{R}^d}\hat{\phi}(0,x)\,d\mu_0 - \sum_{i=1}^m \hat{\lambda}_i c_i = \int_{\mathbb{R}^d}\tilde{\phi}(0,x)\,d\mu_0 - \sum_{i=1}^m \tilde{\lambda}_i c_i$, so the value of Ψ does not depend on the choice of (ϕ,λ) and hence Ψ is well-defined.

Denote by Φ^* and Ψ^* the convex conjugates of Φ and Ψ, respectively. For Φ, its convex conjugate $\Phi^* : C_b(\Lambda,\mathcal{X})^* \to \mathbb{R}\cup\{+\infty\}$ is given by

$$\Phi^*(\rho,\mathcal{A},\mathcal{B}) = \sup_{(r,a,b)\in C_b(\Lambda,\mathcal{X})}\{\langle (r,a,b),(\rho,\mathcal{A},\mathcal{B})\rangle \ ; \ r + F^*(a,b) \le 0\}.$$

As shown in Lemma A.1.1, if we restrict Φ^* to $\mathcal{M}(\Lambda,\mathcal{X})$, then

$$\Phi^*(\rho,\mathcal{A},\mathcal{B}) = \begin{cases} \displaystyle\int_\Lambda F\left(\frac{d\mathcal{A}}{d\rho},\frac{d\mathcal{B}}{d\rho}\right)d\rho & \text{if } \rho\in\mathcal{M}_+(\Lambda,\mathcal{X}) \text{ and } (\mathcal{A},\mathcal{B}) \ll \rho, \\ +\infty & \text{otherwise.} \end{cases}$$

Next, $\Psi^* : C_b(\Lambda,\mathcal{X})^* \to \mathbb{R}\cup\{+\infty\}$ is given by

$$\Psi^*(\rho,\mathcal{A},\mathcal{B}) = \sup_{(r,a,b)}\left\{\langle (r,a,b),(\rho,\mathcal{A},\mathcal{B})\rangle - \int_{\mathbb{R}^d}\phi(0,x)\,d\mu_0 + \sum_{i=1}^m \lambda_i c_i\right\},$$

where the supremum is taken over all triples $(r,a,b)\in C_b(\Lambda,\mathcal{X})$ represented by (ϕ,λ) in $BV([0,T],C_b^2(\mathbb{R}^d))\times\mathbb{R}^m$. In terms of (ϕ,λ),

$$\Psi^*(\rho,\mathcal{A},\mathcal{B}) = \sup_{\phi,\lambda}\left\{\langle(-\partial_t\phi - \sum_{i=1}^m \lambda_i G_i\delta_{\tau_i}, -\nabla_x\phi, -\frac{1}{2}\nabla_x^2\phi),(\rho,\mathcal{A},\mathcal{B})\rangle - \int_{\mathbb{R}^d}\phi(0,x)\,d\mu_0 + \sum_{i=1}^m\lambda_i c_i\right\}.$$

As proved in Lemma A.2.1, the objective $\mathcal{V}$ can be expressed as

$$\mathcal{V} = \inf_{(\rho,\mathcal{A},\mathcal{B})\in\mathcal{M}(\Lambda,\mathcal{X})}(\Phi^*+\Psi^*)(\rho,\mathcal{A},\mathcal{B}) = \inf_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda,\mathcal{X})^*}(\Phi^*+\Psi^*)(\rho,\mathcal{A},\mathcal{B}).$$

Let $O^{m\times n}$ denote a null matrix of size $m\times n$. Consider the point $(r,a,b) = (-1,O^{d\times 1},O^{d\times d})$ which can be represented by $(\phi,\lambda) = (T-t,O^{m\times 1})$. As F is nonnegative, at $(-1,O^{d\times 1},O^{d\times d})$ we have

$$-1 + F^*(O^{d\times 1},O^{d\times d}) = -1 - \inf_{\alpha\in\mathbb{R}^d,\beta\in\mathbb{S}^d}F(\alpha,\beta) < 0.$$

This shows that

$$\Phi(-1,O^{d\times 1},O^{d\times d}) = 0, \quad \Psi(-1,O^{d\times 1},O^{d\times d}) = 0.$$

Thus, at $(-1,O^{d\times 1},O^{d\times d})$, Φ is continuous with respect to the uniform norm (since F^* is continuous in $\mathrm{dom}(F^*)$), and Ψ is finite. Furthermore, as the convex functionals Φ and Ψ

take values in $(-\infty, +\infty]$, all of the required conditions are fulfilled to apply the Fenchel–Rockafellar duality theorem [see e.g., 22, Chapter 1]. We then obtain

$$\mathcal{V} = \inf_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda,\mathcal{X})^*} \{\Phi^*(\rho,\mathcal{A},\mathcal{B}) + \Psi^*(\rho,\mathcal{A},\mathcal{B})\} = \sup_{(r,a,b)\in C_b(\Lambda,\mathcal{X})} \{-\Phi(-r,-a,-b) - \Psi(r,a,b)\},$$

and the infimum is in fact attained. Consequently,

$$\mathcal{V} = \sup_{(r,a,b)} \left\{ -\int_{\mathbb{R}^d} \phi(0,x)\, d\mu_0 + \sum_{i=1}^m \lambda_i c_i \; ; \; -r + F^*(-a,-b) \le 0 \right\},$$

where the supremum is restricted to all (r, a, b) represented by $(\phi, \lambda) \in BV([0,T], C_b^2(\mathbb{R}^d)) \times \mathbb{R}^m$. Writing (r, a, b) in terms of (ϕ, λ) with $\phi(T, \cdot) = 0$, we obtain the required result. $\square$

3.3.3 Viscosity solutions

Adopting the concept of viscosity solutions, it can be shown that the supremum of the objective with respect to ϕ is achieved by the viscosity solution of the HJB equation (3.10). Due to presence of the Dirac delta functions in (3.10), we shall introduce a suitable definition of the viscosity solution that allows to have jump discontinuities in time.

Definition 3.3.7. Denote by $\operatorname{set}(\tau)$ the set of entries of vector τ and by K the cardinality of $\operatorname{set}(\tau)$. Let $t_0 = 0$, we define disjoint intervals $I_k := [t_{k-1}, t_k)$ such that

$$\bigcup_{k=1}^{K} I_k = [0, T),$$

where $t_{k-1} < t_k$ and $t_k \in \operatorname{set}(\tau)$ for all $k = 1, \ldots, K$.

Definition 3.3.8 (Viscosity solution)**.** For any $\lambda \in \mathbb{R}^m$, we say ϕ is a viscosity subsolution (resp., supersolution) of (3.10) if ϕ is a classical (continuous) viscosity subsolution (resp., supersolution) of (3.10) in $I_k \times \mathbb{R}^d$ for all $k = 1, \ldots, K$, and has jump discontinuities:

$$\phi(t,x) = \phi(t^-,x) - \sum_{i=1}^m \lambda_i G_i(x) \mathbb{1}(t = \tau_i) \qquad \forall (t,x) \in \tau \times \mathbb{R}^d.$$

Then, ϕ is called a viscosity solution of (3.10) if ϕ is both a viscosity subsolution and a viscosity supersolution of (3.10).

Remark 3.3.9 (Comparison principle)**.** The comparison principle still holds for viscosity solutions of (3.10). Let u and v be a viscosity subsolution and a viscosity supersolution of the equation (3.10), respectively. At the terminal time T, $u(T,\cdot) \le v(T,\cdot)$. Since $t_K = T$ is in $\operatorname{set}(\tau)$ and u, v have the same jump size at $\{T\} \times \mathbb{R}^d$, we get $u(T^-,\cdot) \le v(T^-,\cdot)$. Next, in the interval $I_K = [t_{K-1}, t_K)$, by the classical comparison principle, we get $u \le v$ on I_K. Applying this argument for all intervals I_k for $k = 1, \ldots, K$, we conclude that

$$u(t,x) \le v(t,x), \quad \forall (t,x) \in [0,T] \times \mathbb{R}^d.$$

Also, $u(0,\cdot) \le v(0,\cdot)$.

Remark 3.3.10 (Existence and uniqueness)**.** As a consequence of the comparison principle, there exists a unique viscosity solution of (3.10). The uniqueness is a direct consequence of the comparison principle. The existence can be obtained by Perron's method [see 29] under which the comparison principle is a key argument.

Now we shall prove Proposition 3.3.5. The proof relies on a smoothing argument used in Bouchard et al. [18], which is based on the shaken coefficients technique of Krylov [78]. The proof is similar to Theorem 2.4 in Bouchard et al. [18], which we sketch here for completeness.

Proof of Proposition 3.3.5. Denote by φ a viscosity solution of the equation (3.10) with any $\lambda \in \mathbb{R}^m$. From Remark 3.3.10, we know that such φ exists and is unique. The first part of the proposition is proved in two steps:
Step 1. Assuming that there exists a sequence of supersolutions of (3.10) in $BV([0,T], C_b^2(\mathbb{R}^d))$ converging to φ pointwise, we can show that φ achieves the supremum with respect to ϕ in the objective of the dual (3.12). Let $\phi \in BV([0,T], C_b^2(\mathbb{R}^d))$ be any solution that satisfies (3.13), and ϕ is also a (viscosity) supersolution of (3.10). By Remark 3.3.9, we have $\varphi(0,x) \leq \phi(0,x)$ for all $x \in \mathbb{R}^d$, hence

$$\sum_{i=1}^m \lambda_i c_i - \int_{\mathbb{R}^d} \phi(0,x)\, d\mu_0 \leq \sum_{i=1}^m \lambda_i c_i - \int_{\mathbb{R}^d} \varphi(0,x)\, d\mu_0. \tag{3.18}$$

The equality can be achieved in (3.18) by taking the supremum with respect to ϕ on the left-hand side of (3.18).
Step 2. Now, we shall construct the sequence of supersolutions required in Step 1. Let us introduce the regularising kernel $r_\varepsilon : \mathbb{R}^d \to \mathbb{R}$ such that $r_\varepsilon(x) = \frac{1}{\varepsilon^d} r'\left(\frac{x}{\varepsilon}\right)$ where r' is some compactly supported function that satisfies $\int_{\mathbb{R}^d} r'(x)\, dx = 1$. Then we define $\varphi_\varepsilon = \varphi * r_\varepsilon$ where the convolution acts only on the variable x. By applying the result of Bouchard et al. [18] which relies critically on the fact that $F^*(a,b)$ is convex in (a,b), it can be shown that φ_ε are supersolutions of equation (3.10). If we send ε to 0, the supersolutions φ_ε converge to the viscosity solution φ pointwise. The desired sequence is then constructed.

Now we prove the second part of the proposition. Let $(\rho^*, \alpha^*, \beta^*)$ be the optimal solution of Problem 2, then $(\rho^*, \rho^*\alpha^*, \rho^*\beta^*)$ also achieves the infimum (3.16). Assume that there exists an optimal solution $\lambda^* \in \mathbb{R}^m$ that solves (3.9), then $(\phi_{\lambda^*}, \lambda^*)$ also achieve the supremum in (3.16). With the optimal solutions defined above, we can reformulate (3.16) as

$$\begin{aligned}
0 &= \int_0^T \int_{\mathbb{R}^d} \left(F(\alpha^*, \beta^*) - \partial_t \phi_{\lambda^*} - \nabla_x \phi_{\lambda^*} \cdot \alpha^* - \frac{1}{2} \nabla_x^2 \phi_{\lambda^*} : \beta^* - \sum_{i=1}^m \lambda_i^* G_i \delta_{\tau_i} \right) d\rho_t^* dt \\
&= \int_0^T \int_{\mathbb{R}^d} \left(F(\alpha^*, \beta^*) + F^* \left(\nabla_x \phi_{\lambda^*}, \frac{1}{2} \nabla_x^2 \phi_{\lambda^*} \right) - \nabla_x \phi_{\lambda^*} \cdot \alpha^* - \frac{1}{2} \nabla_x^2 \phi_{\lambda^*} : \beta^* \right) d\rho_t^* dt.
\end{aligned}$$

Let $(\tilde{\alpha}, \tilde{\beta})$ be defined by

$$(\tilde{\alpha}, \tilde{\beta}) = \nabla F^* \left(\nabla_x \phi_{\lambda^*}, \frac{1}{2} \nabla_x^2 \phi_{\lambda^*} \right), \quad \left(\nabla_x \phi_{\lambda^*}, \frac{1}{2} \nabla_x^2 \phi_{\lambda^*} \right) = \nabla F(\tilde{\alpha}, \tilde{\beta}).$$

Hence, by the definition of convex conjugate and the strong convexity of F,

$$\begin{aligned}
0 &= \int_0^T \int_{\mathbb{R}^d} \left(F(\alpha^*, \beta^*) - F(\tilde{\alpha}, \tilde{\beta}) - \nabla_x \phi_{\lambda^*} \cdot (\alpha^* - \tilde{\alpha}) - \frac{1}{2} \nabla_x^2 \phi_{\lambda^*} : (\beta^* - \tilde{\beta}) \right) d\rho_t^* dt \\
&\geq \int_0^T \int_{\mathbb{R}^d} C \left(\|\alpha^* - \tilde{\alpha}\|^2 + \|\beta^* - \tilde{\beta}\|^2 \right) d\rho_t^* dt \geq 0,
\end{aligned}$$

where $C > 0$ is a constant. Therefore, $(\alpha^*, \beta^*) = (\tilde{\alpha}, \tilde{\beta})$, $d\rho_t^* dt$-almost everywhere. The proof is completed. □

3.4 LSV Calibration

In this section, we illustrate our method by calibrating a Heston-like LSV model. This method could also be easily extended to other LSV models. We consider the LSV model with following dynamics under the risk-neutral measure:

$$\begin{cases} dZ_t = (r(t) - q(t) - \frac{1}{2}\sigma^2(t, Z_t, V_t))\, dt + \sigma(t, Z_t, V_t)\, dW_t^Z, \\ dV_t = \kappa(\theta - V_t)\, dt + \xi\sqrt{V_t}\, dW_t^V, \\ dW_t^Z dW_t^V = \eta(t, Z_t, V_t)\, dt, \end{cases} \tag{3.19}$$

where Z_t is the logarithm of the stock price at time t. The interpretations of r and q differ between financial markets. In the equity market, r is the risk-free rate and q is the dividend yield. In the FX market, r is the domestic interest rate and q is the foreign interest rate. The parameters κ, θ, ξ have the same interpretation as in the Heston model. In our method, we assume these parameters are given and obtained by calibrating a pure Heston model. Note that in the literature, the widely considered LSV model has a volatility function $\sigma(t, Z_t, V_t) = L(t, Z_t)\sqrt{V_t}$ and a constant correlation η, where $L(t, Z_t)$ is known as the leverage function. By contrast, we consider a local-stochastic volatility $\sigma > 0$ and a local-stochastic correlation $\eta \in [-1, 1]$ whose values depend on (t, Z_t, V_t). Our objective is to calibrate $\sigma(t, Z, V)$ and $\eta(t, Z, V)$ so that model prices exactly match market prices.

Remark 3.4.1. If the volatility $\sigma(t, Z, V) \equiv \sqrt{V}$ and the correlation $\eta(t, Z, V)$ is a constant, the LSV model reduces to a pure Heston model. Furthermore, if $\sigma(t, Z, V)$ is independent of the variable V, the model is equivalent to a local volatility model.

Consider a probability measure $\mathbb{P} \in \mathcal{P}^1$ and a two-dimensional $\mathbb{P}$-semi-martingale X_t. The process X_t has dynamics (3.19), i.e., $X_t = (Z_t, V_t)$, if $\mathbb{P}$ is characterised by $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ such that

$$(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}}) = \left(\begin{bmatrix} r_t - q_t - \frac{1}{2}\sigma_t^2 \\ \kappa(\theta - V_t) \end{bmatrix}, \begin{bmatrix} \sigma_t^2 & \eta_t \xi \sqrt{V_t}\sigma_t \\ \eta_t \xi \sqrt{V_t}\sigma_t & \xi^2 V_t \end{bmatrix} \right), \quad t \in [0, T], \tag{3.20}$$

with functions $\sigma_t = \sigma(t, Z_t, V_t)$ and $\eta_t = \eta(t, Z_t, V_t)$. Recall that the parameters (κ, θ, ξ) are assumed to be given. Also, r_t and q_t are known and V_t is a state variable. Hence, the only unknown variables in (3.20) are σ_t and η_t. As we will see below, σ_t will be the only free variable in the calibration. Given m European options with prices $c \in \mathbb{R}_+^m$, maturities $\tau = (\tau_1, \ldots, \tau_m) \in (0, T]^m$ and discounted payoffs $G = (G_1, \ldots, G_m)$ where $G_i : \mathbb{R}^2 \to \mathbb{R}_+$ (e.g., $G_i(x) = e^{-\int_0^{\tau_i} r(s)ds}(e^{x_1} - K)^+$ if the i-th option is a European call with strike K and maturity τ_i, where x_1 stands for the first element of x). If X_t has an initial distribution $\mu_0 = \delta_{(Z_0, V_0)}$ and is exactly calibrated to these European options, then $\mathbb{P} \in \mathcal{P}(\mu_0, \tau, c, G)$. One way to build a calibrated LSV model is to solve

$$\mathcal{V} = \inf_{\mathbb{P} \in \mathcal{P}(\mu_0, \tau, c, G)} \mathbb{E}^{\mathbb{P}} \int_0^T F(t, X_t, \alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})\, dt, \tag{3.21}$$

where F is a suitable convex cost function that forces $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ to take the form of (3.20).

One possible way to choose the cost function F is based on the idea of minimising the difference between each element of $\beta^{\mathbb{P}}$ and a reference value while keeping $\beta^{\mathbb{P}}$ in $\mathbb{S}_+^2$. However, it is often impossible to find an explicit formula to approximate F^*. Thus numerical optimisation is needed, which makes the method computationally expensive. To overcome this issue, we choose the correlation

$$\eta_t = \frac{\sqrt{V_t}}{\sigma_t}\overline{\eta}, \quad t \in [0, T], \tag{3.22}$$

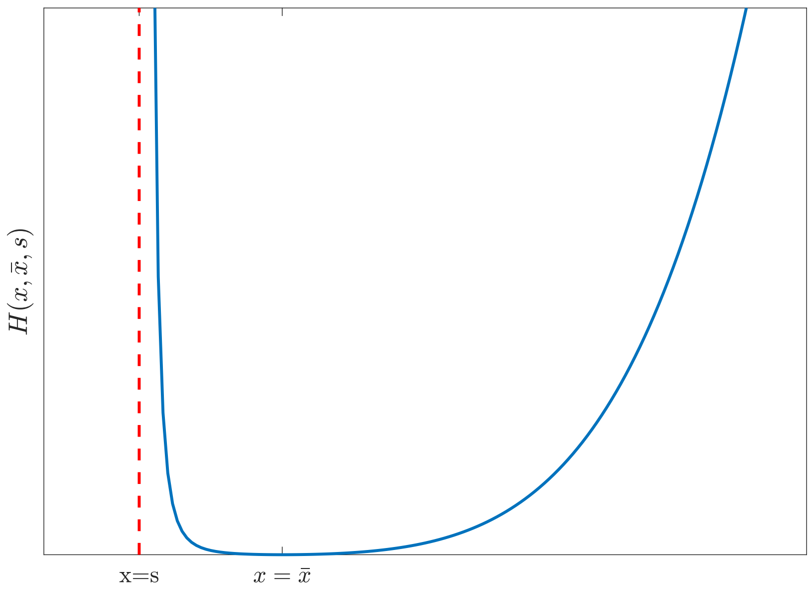

Figure 3.1: The function $H(x, \overline{x}, s)$ for a given $\overline{x}$ and a given $s < \overline{x}$.

where $\overline{\eta}$ is a constant correlation obtained (along with κ, θ, ξ) by calibrating a pure Heston model. In this case, $\beta_t^{\mathbb{P}}$ is positive semidefinite if and only if $\sigma_t^2 \geq \overline{\eta}^2 V_t$ for $t \leq T$.

Definition 3.4.2. Define function $H : \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ such that

$$H(x, \overline{x}, s) := \begin{cases} a(\dfrac{x-s}{\overline{x}-s})^{1+p} + b(\dfrac{x-s}{\overline{x}-s})^{1-p} + c & \text{if } x > s \text{ and } \overline{x} > s, \\ +\infty & \text{otherwise.} \end{cases}$$

The parameter p is a constant greater than 1, and a, b, c are constants determined to minimise the function at $x = \overline{x}$ with $\min H = 0$.

Given $\overline{x}$ and s satisfying $\overline{x} > s$, the function H is convex in x and minimised at $\overline{x}$. It is finite only when $x > s$. In the numerical examples (see Section 3.5.2 and 3.5.3 below), the parameter p is set to 4. A plot of H is given in Figure 3.1. Then, we define the cost function as follows.

Definition 3.4.3. The cost function $F : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^2 \times \mathbb{S}^2 \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$F(t, Z, V, \alpha, \beta) := \begin{cases} H(\beta_{11}, V, \overline{\eta}^2 V) & \text{if } (\alpha, \beta) \in \Gamma(t, V), \\ +\infty & \text{otherwise,} \end{cases} \tag{3.23}$$

where the convex set Γ is defined as

$$\Gamma(t, V) := \{(\alpha, \beta) \in \mathbb{R}^2 \times \mathbb{S}^2 \mid \alpha_1 = r(t) - q(t) - \beta_{11}/2,\ \alpha_2 = \kappa(\theta - V), \\ \beta_{12} = \beta_{21} = \overline{\eta}\xi V,\ \beta_{22} = \xi^2 V\}.$$

Remark 3.4.4. The function H penalises deviations of the LSV model from a pure Heston model by choosing $\overline{x} = V$ (see Remark 3.4.1). This approach seeks to retain the attractive features of the Heston model while still matching all the market prices. We also set $s = \overline{\eta}^2 V$ to ensure that $\sigma^2 > \overline{\eta}^2 V$, hence β remains positive definite and the correlation η is in $[-1, 1]$. The set Γ forces X_t to have dynamics of the form (3.19) with η defined in (3.22) by restricting the characteristics in Γ. In particular, it remains risk neutral.

By applying Proposition 3.3.5, the dual formulation of (3.21) with the cost function (3.23) is as follows:

$$\mathcal{V} = \sup_{\lambda \in \mathbb{R}^m} \left\{ \sum_{i=1}^m \lambda_i c_i - \phi_\lambda(0, Z_0, V_0) \right\}, \tag{3.24}$$

where ϕ_λ is the viscosity solution to the HJB equation

$$\begin{aligned} \partial_t \phi_\lambda + \sum_{i=1}^m \lambda_i G_i \delta_{\tau_i} + \sup_{\beta_{11}} \Big\{ & (r - q - \frac{1}{2}\beta_{11})\partial_Z \phi_\lambda + \kappa(\theta - V)\partial_V \phi_\lambda + \overline{\eta}\xi V \partial_{ZV}\phi_\lambda \\ & + \frac{1}{2}\beta_{11}\partial_{ZZ}\phi_\lambda + \frac{1}{2}\xi^2 V \partial_{VV}\phi_\lambda - H(\beta_{11}, V, \overline{\eta}^2 V) \Big\} = 0, \end{aligned} \tag{3.25}$$

with a terminal condition $\phi_\lambda(T, \cdot) = 0$.

Given any $\lambda \in \mathbb{R}^m$, we can calculate $\phi_\lambda(0, Z_0, V_0)$ by numerically solving the HJB equation (3.25). The optimal λ can be found through a standard optimisation algorithm (see Section 3.5.1 below). The convergence of the algorithm can be improved by providing the gradient of the objective. Let $\overline{\beta}_\lambda$ denote the optimal β_{11} that solves the supremum in (3.25), which also implicitly depends on λ. In fact, solving the supremum in (3.25) is equivalent to solving the following equation for σ^2:

$$(\partial_{ZZ}\phi_\lambda - \partial_Z \phi_\lambda)/2 = \partial_{\sigma^2} H(\sigma^2, V, \overline{\eta}^2 V), \tag{3.26}$$

for which a closed-form solution is available. We also denote by $\mathbb{P}_\lambda \in \mathcal{P}^1$ a probability measure characterised by $(\alpha^{\mathbb{P}_\lambda}, \beta^{\mathbb{P}_\lambda})$ defined in (3.20) with $(\sigma_t, \eta_t) = (\sqrt{(\overline{\beta}_\lambda)_t}, \overline{\eta}\sqrt{V_t/(\overline{\beta}_\lambda)_t})$, $t \leq T$.

Lemma 3.4.5. *Define $J(\lambda) = \sum_{i=1}^m \lambda_i c_i - \phi_\lambda(0, Z_0, V_0)$. The gradient of $J(\lambda)$ with respect to λ_i can be formulated as:*

$$\partial_{\lambda_i} J(\lambda) = c_i - \mathbb{E}^{\mathbb{P}_\lambda} G_i(X_{\tau_i}), \quad \forall i = 1, \ldots, m. \tag{3.27}$$

In addition, $\mathbb{E}^{\mathbb{P}_\lambda} G_i(X_{\tau_i}) = \phi'(0, Z_0, V_0)$ where ϕ' solves

$$\partial_t \phi' + (r - q - \frac{1}{2}\overline{\beta}_\lambda)\partial_Z \phi' + \kappa(\theta - V)\partial_V \phi' + \overline{\eta}\xi V \partial_{ZV}\phi' + \frac{1}{2}\overline{\beta}_\lambda \partial_{ZZ}\phi' + \frac{1}{2}\xi^2 V \partial_{VV}\phi' = 0, \tag{3.28}$$

with the terminal condition $\phi'(\tau_i, \cdot) = G_i$.

Proof. Given a λ and the associated $\overline{\beta}_\lambda$, the HJB equation (3.25) reduces to

$$\begin{aligned} \partial_t \phi_\lambda + \sum_{i=1}^m \lambda_i G_i \delta_{\tau_i} + (r - q - \frac{1}{2}\overline{\beta}_\lambda)\partial_Z \phi_\lambda + \kappa(\theta - V)\partial_V \phi_\lambda + \overline{\eta}\xi V \partial_{ZV}\phi_\lambda & \\ + \frac{1}{2}\overline{\beta}_\lambda \partial_{ZZ}\phi_\lambda + \frac{1}{2}\xi^2 V \partial_{VV}\phi_\lambda - H(\overline{\beta}_\lambda, V, \overline{\eta}^2 V) = 0. & \end{aligned} \tag{3.29}$$

Since λ, ϕ_λ and $\overline{\beta}_\lambda$ are related implicitly, by taking implicit partial differentiation of (3.29) to compute $\phi' := \partial_{\lambda_i}\phi_\lambda$ for any $i = 1, \ldots, m$, we obtain the following PDE

$$\begin{aligned} \partial_t \phi' + (r - q - \frac{1}{2}\overline{\beta}_\lambda)\partial_Z \phi' + \kappa(\theta - V)\partial_V \phi' + \overline{\eta}\xi V \partial_{ZV}\phi' & \\ + \frac{1}{2}\overline{\beta}_\lambda \partial_{ZZ}\phi' + \frac{1}{2}\xi^2 V \partial_{VV}\phi' = -G_i \delta_{\tau_i}. & \end{aligned} \tag{3.30}$$

With the terminal condition $\phi'(T, \cdot) = 0$, (3.30) can be solved by the Feynman–Kac formula [see e.g., 74, Theorem 7.6]. Thus,

$$\phi'(0, Z_0, V_0) = \mathbb{E}^{\mathbb{P}_\lambda} G_i(X_{\tau_i}).$$

Moreover, solving (3.30) with $\phi'(T, \cdot) = 0$ is equivalent to solving (3.28) with $\phi'(\tau_i, \cdot) = G_i$. The proof is completed. □

Remark 3.4.6. Note that $\mathbb{E}^{\mathbb{P}_\lambda} G_i(X_{\tau_i})$ is the price of the i-th European option calculated by X_t under $\mathbb{P}_\lambda$, which we refer to as the model price, and c_i is the market price. Instead of solving (3.30) once for each option, we can perform a Monte Carlo simulation to efficiently calculate the model prices for all options. However, for the sake of accuracy, we still choose to solve (3.30) in the numerical examples below. Moreover, as the gradient is decreasing to zero while the solution is moving towards the optimal solution, the optimisation process can be interpreted as matching the model X_t to market prices.

3.5 Numerical aspects

3.5.1 Numerical method

In this section, we present a numerical method for solving the dual formulation. To shorten notations, we will simply write ϕ for ϕ_λ from now on. Starting with an initial $\lambda = \lambda^0$ (e.g., setting it to a null vector), we solve the HJB equation (3.25) to calculate $\phi(0, Z_0, V_0)$ and hence calculate $J(\lambda^0)$. Then, J is maximised over $\lambda \in \mathbb{R}^m$ through an optimisation algorithm. In particular, we employed the L-BFGS algorithm [81] and obtained good convergence. The optimisation process can be accelerated by providing the gradient $\nabla J(\lambda)$ which can be numerically computed by (3.27). We measure the optimality by the maximum absolute value on the gradient. In other words, by setting a threshold ϵ_1, the algorithm terminates when the following stopping criterion is reached:

$$\|\nabla J(\lambda)\|_\infty \leq \epsilon_1.$$

For solving the HJB equation (3.25), we use an alternating direction implicit (ADI) method together with the central finite difference scheme. In the numerical examples below, we employ the Douglas scheme from In 't Hout and Foulon [68]. Given a λ, we solve the HJB equation backward. Consider a discretisation $\{t_k\}$ of the time interval $[0, T]$ such that $0 = t_0 < t_1 < \cdots < t_{N_T} = T, N_T \in \mathbb{N}$. Without loss of generality, we assume that $\mathrm{set}(\tau) \subset \{t_k\}$. At each time step t_k, we approximate $\sigma^2_{t_k}$ by solving (3.26) with $\phi = \phi_{t_{k+1}}$ for which an analytical solution can be found. At $t = t_k$, with the approximated $\sigma^2_{t_k}$, the HJB equation (3.25) is solved by the ADI finite difference method. Note that this approximation scheme of σ is similar to the one used in Ren et al. [95] for approximating the leverage function.

Let $\hat{\tau}_i$ be an element in $\mathrm{set}(\tau)$ such that $\cup_{i=1}^K \{\hat{\tau}_i\} = \mathrm{set}(\tau)$ and $0 =: \hat{\tau}_0 < \hat{\tau}_1 < \ldots < \hat{\tau}_K = T$ (see Definition 3.3.7 for the definitions of $\mathrm{set}(\tau)$ and K). Denote by D the spatial computational domain and by ∂D the boundary of D. When numerically solving the HJB equation (3.25), we impose the following boundary conditions for the spatial dimensions:

$$\forall i = 1, \ldots, K \qquad \nabla^2_x \phi(t, x) = \nabla^2_x \phi(\hat{\tau}_i^-, x), \qquad (t, x) \in [\hat{\tau}_{i-1}, \hat{\tau}_i) \times \partial D,$$

In addition, we set a sufficiently large D to reduce the impact of the boundary conditions.

To handle the jump discontinuities caused by the presence of the Dirac delta terms, we can solve the HJB equation interval-wise in the intervals separated by the maturities,

Algorithm 1: LSV calibration

Data: Market prices of European option

Result: A calibrated OT-LSV model that matches all market prices

Set an initial λ

do

  `/* Solving the HJB equation */`

  for $k = N_T - 1, \dots, 0$ **do**

    if t_{k+1} *is equal to the maturity of any calibrating options.* **then**

      $\phi_{t_{k+1}} \leftarrow \phi_{t_{k+1}} + \sum_{i=1}^m \lambda_i G_i \mathbb{1}(t_{k+1} = \tau_i)$

    end

    Approximate $\sigma_{t_k}^2$ by solving (3.26) with $\phi = \phi_{t_{k+1}}$

    Solve the HJB equation (3.25) by the ADI method at $t = t_k$

  end

  `/* Calculating model prices and gradient */`

  Solve (3.28) to calculate the model prices by the ADI method

  Calculate the gradient $\nabla J(\lambda)$ by (3.27)

  Update λ by the L-BFGS algorithm

while $\|\nabla J(\lambda)\|_\infty > \epsilon_1$

and the jump discontinuity can be incorporated into the terminal condition of the HJB equation in each interval. More precisely, if t_{k+1} is equal to the maturity of any calibrating options, we incorporate the jump discontinuity by adding $\sum_{i=1}^m \lambda_i G_i \mathbb{1}(t_{k+1} = \tau_i)$ to $\phi_{t_{k+1}}$. The numerical method is summarised in Algorithm 1.

Due to the non-linearity of the HJB equation, when the time step sizes are too large, it might not be accurate to simply approximate $\sigma_{t_k}^2$ by solving (3.26) with $\phi = \phi_{t_{k+1}}$ once per time step. Therefore, we slightly modify the algorithm by including an iterative step to improve the accuracy of the approximation of $\sigma_{t_k}^2$. In the literature, this iterative step is known as *policy iteration*, see e.g., Ma and Forsyth [84]. Specifically, at each time step t_k, we first approximate $\sigma_{t_k}^2$ by solving (3.26) with $\phi = \phi_{t_{k+1}}$. Next, we obtain ϕ_{t_k} by solving the HJB equation (3.25) with $\sigma_{t_k}^2$, and then approximate $\sigma_{t_k}^2$ again by solving (3.26) with $\phi = \phi_{t_k}$. This process is repeated until ϕ_{t_k} converges. For completeness, the modified numerical method with policy iteration is summarised in Algorithm 2. For the sake of accuracy, we use Algorithm 2 in both Section 3.5.2 and Section 3.5.3.

In our experiments, we notice that the algorithm can provide satisfactory calibration results even with coarse grids. However, it is crucial to ensure that the grids are fine enough, because we do not want to calibrate the wrong model prices to the calibrating option prices. In fact, we observe that the algorithm converges faster with finder grids, because the numerical approximations of the gradients are more accurate with finer grids.

3.5.2 Numerical example: simulated data

In this section, we provide two numerical examples with simulated data to demonstrate the calibration method. In both examples, the risk-free rate is set to a constant $r = 0.05$ and the dividend yield is set to $q = 0$. Let $Z_0 = \ln 100$ and $V_0 = 0.04$ for both models. We consider a uniform mesh over the spatial computational domain $D = [Z_0 - 4\sqrt{V_0}, Z_0 + 4\sqrt{V_0}] \times [0, 0.5]$ and use 101 points for each dimension. We also consider a uniform mesh over the time interval $[0, 1]$ with $N_T = 100$. The LSV model is calibrated to a set of European call options generated by a Heston model with given parameters. For clarity, we will refer

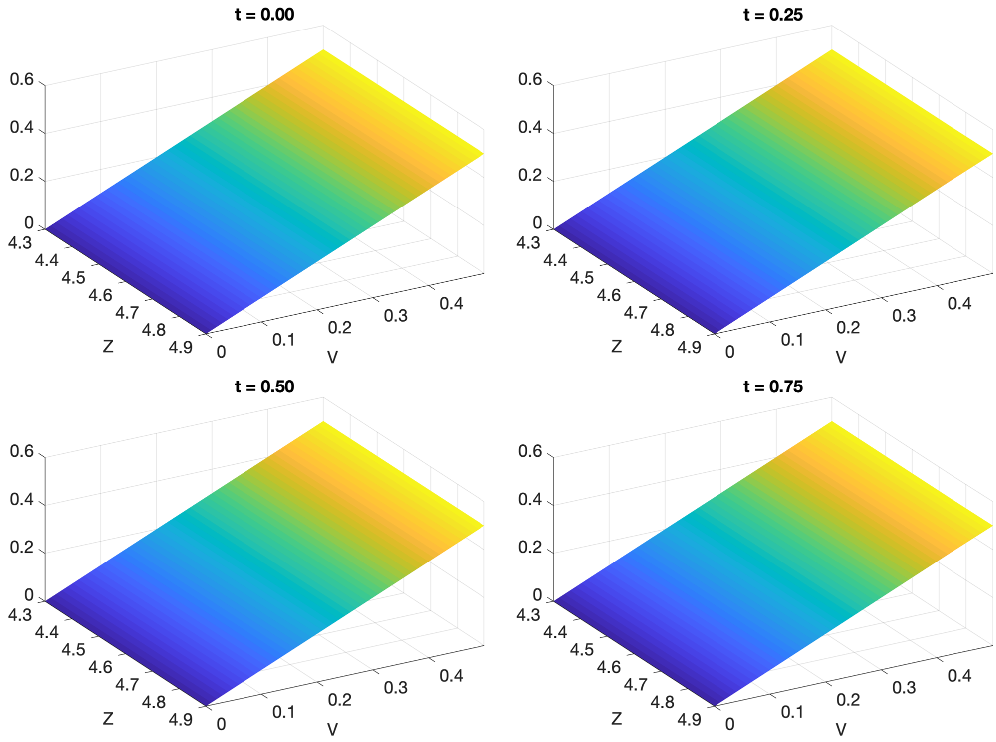

Figure 3.2: The volatility function $\sigma^2(t, Z, V)$ in Example 1

to the LSV model as the *OT-LSV model* and refer to the Heston model as the *Heston generating model*. The option prices are calculated at maturities in $\{0.2, 0.4, 0.6, 0.8, 1.0\}$ and at 18 different strikes in $[Z_0 - 1.4\sqrt{V_0}, Z_0 + 1.4\sqrt{V_0}]$.

Example 1

In the first example, we use parameters $(\kappa, \theta, \xi, \overline{\eta}) = (0.5, 0.04, 0.16, -0.4)$ for both the OT-LSV model and the Heston generating model. This example represents a trivial case, since if we use the same set of parameters for both models, the optimal solution of the dual formulation is a null vector $\lambda = \mathbf{0} \in \mathbb{R}^m$, and hence $\mathcal{V} = 0$. In this case, under the optimal measure of Problem 1, $\sigma^2(t, Z, V) = V$ and $\eta(t, Z, V) = \overline{\eta}$. Setting a threshold $\epsilon_1 = 10^{-6}$, we obtain the expected results. The plot of $\sigma^2(t, Z, V)$ is provided in Figure 3.2.

Example 2

In the second example, we give different parameters to the OT-LSV model and the Heston generating model (see Table 3.1). As noted in Remark 3.4.1, the OT-LSV model reduces to a LV model if $\sigma^2(t, Z, V)$ is independent of V. Also, it is well known that an LV model can be calibrated to any arbitrage-free option prices. In this example, the Heston generating model has characteristics that are outside of Γ in the cost function F, so the Heston generating model would lead to an infinite cost. However, since the generated option prices are arbitrage free, a finite cost is still achievable by the OT-LSV model and the problem is admissible, i.e., $\mathcal{P}(\mu_0, \tau, c, G) \neq \emptyset$ and $\mathcal{V} < +\infty$.

By setting the threshold $\epsilon_1 = 0.0005$, we obtained accurate calibration results. The calibration results for a subset of options are given in Table 3.2. If σ^2 is in the form of $\sigma^2(t, Z, V) = L^2(t, Z)V$ for some function L, then $L(t, Z)$ is called the leverage function and the OT-LSV model recovers the traditional LSV model considered in most of the literature.

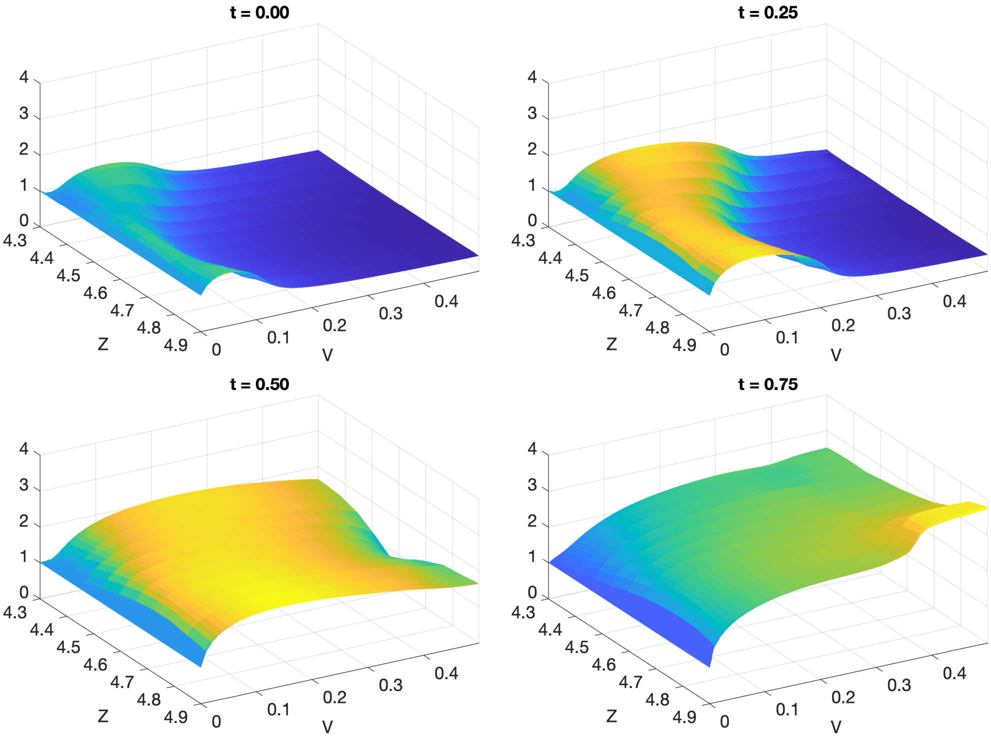

Figure 3.3: The function $\sigma^2(t, Z, V)/V$ in Example 2

	κ	θ	ξ	$\overline{\eta}$
Heston generating model	2.0	0.09	0.10	-0.6
OT-LSV model	0.5	0.04	0.16	-0.4

Table 3.1: The parameters of the Heston generating model and the OT-LSV model in Example 2

Thus, we plot the function $\sigma^2(t, Z, V)/V$ in Figure 3.3 for comparison with $L^2(t, Z)$. The plot of the correlation function $\eta(t, Z, V)$ is also provided in Figure 3.4. Finally, we show the implied volatility of the Heston generated option prices and the OT-LSV generated option prices in Figure 3.5. We can see that the OT-LSV model is well-calibrated to the Heston generated option prices.

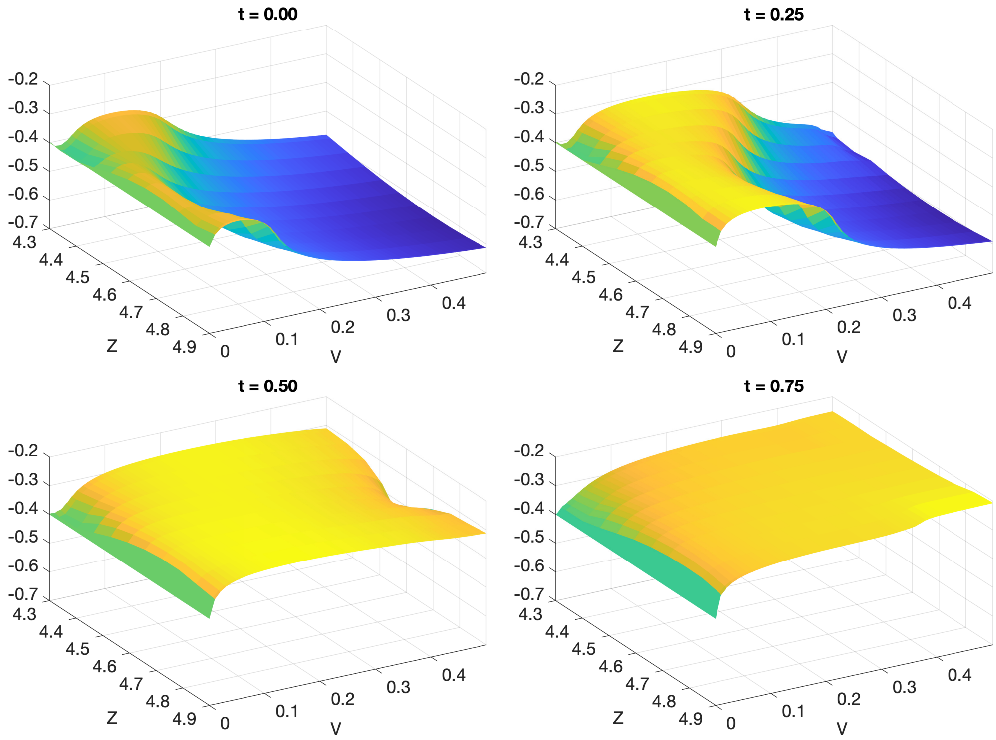

Figure 3.4: The correlation function $\eta(t, Z, V)$ in Example 2

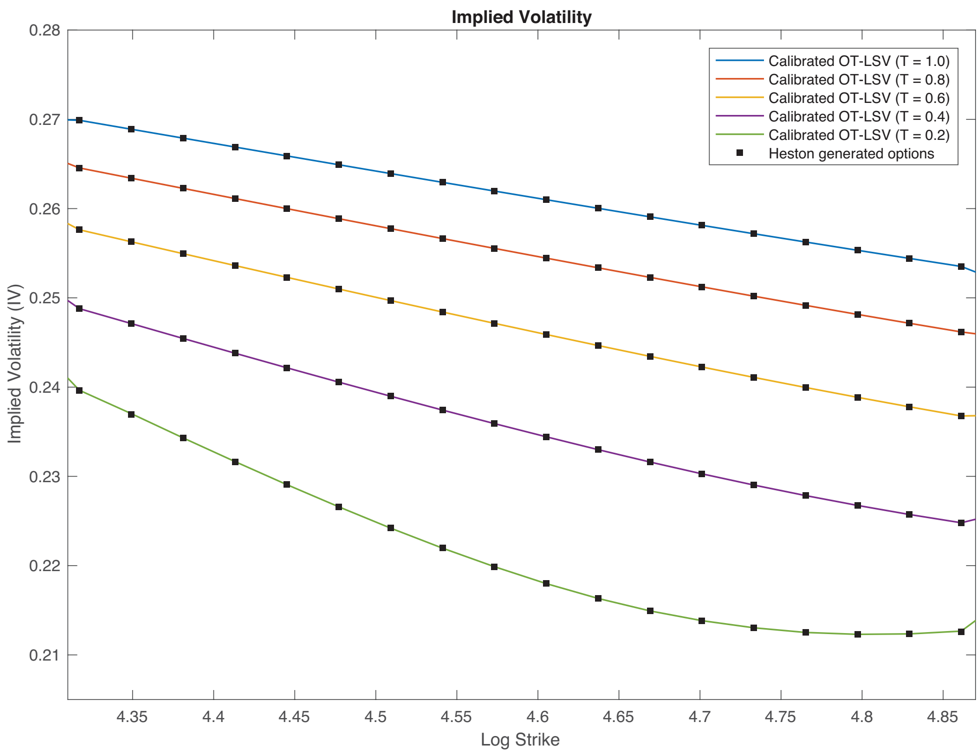

Figure 3.5: The implied volatility of the Heston generated options and the calibrated OT-LSV model in Example 2

Maturity	Log-strike	Implied vol (Heston)	Implied vol (OT-LSV)	Error
T = 0.2	4.3492	0.2396	0.2396	1.55E-05
	4.4452	0.2291	0.2291	1.09E-06
	4.5732	0.2199	0.2199	8.89E-06
	4.7012	0.2138	0.2138	8.56E-06
	4.8292	0.2123	0.2124	2.99E-06
T = 0.4	4.3492	0.2488	0.2488	1.82E-07
	4.4452	0.2422	0.2422	3.93E-06
	4.5732	0.2359	0.2359	2.03E-06
	4.7012	0.2303	0.2303	2.69E-06
	4.8292	0.2257	0.2257	5.20E-07
T = 0.6	4.3492	0.2576	0.2576	8.15E-06
	4.4452	0.2523	0.2523	2.14E-07
	4.5732	0.2471	0.2471	2.42E-06
	4.7012	0.2423	0.2423	6.52E-07
	4.8292	0.2378	0.2378	3.55E-06
T = 0.8	4.3492	0.2646	0.2646	1.97E-05
	4.4452	0.2600	0.2600	1.82E-06
	4.5732	0.2555	0.2555	2.72E-06
	4.7012	0.2512	0.2512	1.81E-06
	4.8292	0.2472	0.2472	2.13E-06
T = 1.0	4.3492	0.2699	0.2699	4.08E-06
	4.4452	0.2659	0.2659	6.81E-07
	4.5732	0.2620	0.2620	1.44E-06
	4.7012	0.2581	0.2581	1.54E-06
	4.8292	0.2544	0.2544	7.30E-07

Table 3.2: A subset of the implied volatility of the options generated by the Heston generating model and the calibrated OT-LSV model in Example 2

3.5.3 Numerical example: FX market data

In this example, we calibrate the OT-LSV model to the FX options data provided in Tian et al. [104]. The options data and the domestic and foreign yields are listed in Table A.1 and Table A.2. The parameters $(\kappa, \theta, \xi, \overline{\eta})$ are shown in Table 3.3, which are obtained by (roughly) calibrating a standard Heston model to the market option prices. In this case, $2\kappa\theta/\xi^2 = 0.169 \ll 1$ and the Feller condition is strongly violated.

Parameter	κ	θ	ξ	$\overline{\eta}$	Z_0	V_0
Value	0.8721	0.0276	0.5338	-0.3566	0.2287	0.012

Table 3.3: The parameters of the OT-LSV model in the FX market data example.

For the numerical settings, the spatial computational domain is set to $D = [-0.6, 1.0] \times [0, 2]$ with 101 points in each dimension. In order to improve the accuracy while still keeping a reasonable computation time, we employ a non-uniform mesh over D and place more points around (Z_0, V_0) [see e.g., 68, Section 2.2.]. For the time interval $[0, 5]$, we use 30 time steps with an equal step size between any two consecutive maturities, e.g., 30 time steps in $(0, 1/12]$ and 30 time steps in $(1/12, 1/6]$, and so on. Since there are 10 maturities (see Table A.1), we have 300 time steps for 5 years in total.

Setting a threshold of $\epsilon_1 = 6 \times 10^{-6}$, we obtain an exact calibration. The maximum difference between the model implied volatility and the market implied volatility is less than 1 basis point. Figure 3.6 shows the implied volatility of the short-maturity options (1 month and 3 months) for the market data, the uncalibrated LSV model and the OT-calibrated LSV model. Figure 3.7 shows the implied volatility of the long-maturity options (2 years and 5 years).

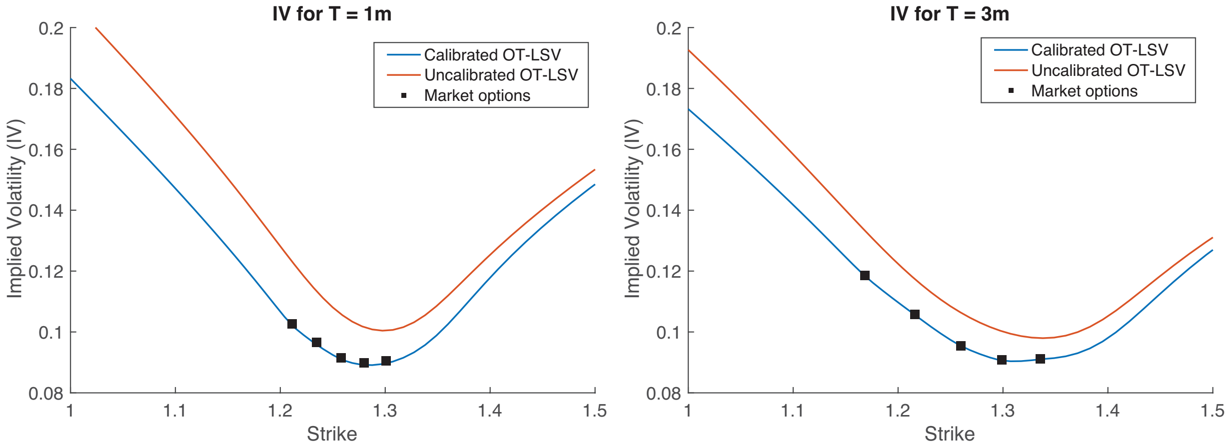

Figure 3.6: The implied volatility (IV) skews generated by both the uncalibrated and the calibrated OT-LSV model for 1 month and 3 months maturities in the FX market data example.

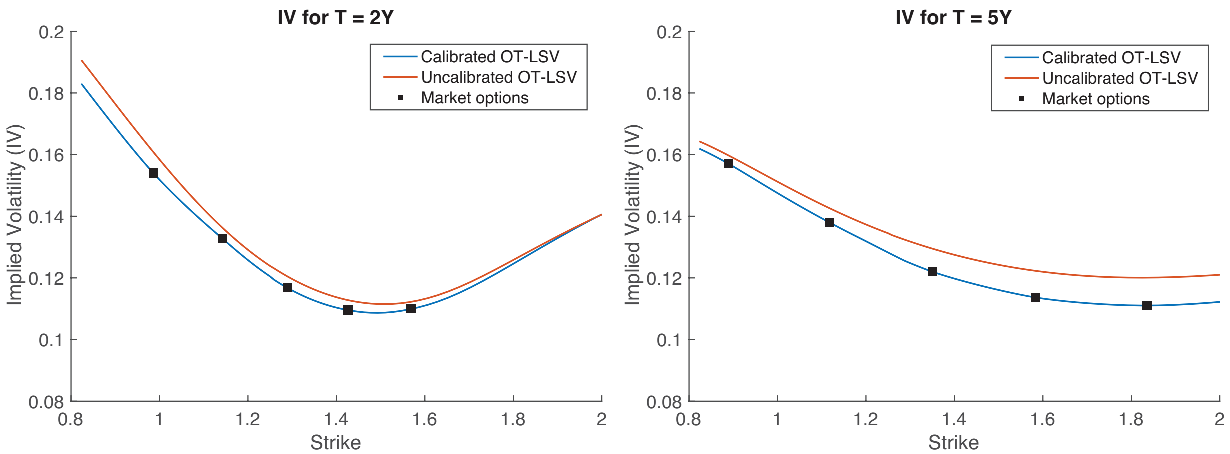

Figure 3.7: The implied volatility (IV) skews generated by both the uncalibrated and the calibrated OT-LSV model for 2 years and 5 years maturities in the FX market data example.

Chapter 4

Joint modelling and calibration of SPX and VIX by optimal transport

The objective of this chapter is to further extend the approach developed in Chapter 3 to address the joint calibration problem of options and futures of the S&P 500 index (SPX) and its volatility index (VIX). This problem has attracted the attention of many researchers and practitioners and has proven challenging. In this chapter, we introduce a time continuous formulation of the joint calibration problem. We consider a semimartingale X whose first coordinate process X^1 is the logarithm of the SPX price and whose second coordinate process X^2 is defined as the expectation of the forward quadratic variation of X^1. The reformulated joint calibration problem falls into the class of the semimartingale optimal transport problem studied in Chapter 3. Then, by following results developed in Chapter 3, we introduce a PDE formulation along with its dual counterpart. The solution, a calibrated diffusion process, can be represented via the solutions of the Hamilton–Jacobi–Bellman equations arising from the dual formulation. The method requires one to choose a reference measure for regularising the nonparametric model. When the chosen reference is very different from the one that describes the observable market option and future prices, the calibrated model might have hump-shaped volatility surfaces and volatility skews, which is not realistic. To address this issue, we introduce a reference measure iteration method that iteratively updates the reference measure while solving the dual formulation. Finally, the proposed calibration method is tested on both simulated data and market data. Numerical examples show that the model can be accurately calibrated to SPX options, VIX options and VIX futures simultaneously.

4.1 Introduction

The CBOE Volatility Index (VIX), also known as the stock market's "fear gauge", reflects the expectations of investors on the volatility of the S&P500 index (SPX) over the next 30 days. Although the index in itself is not a tradable asset, its derivatives such as futures and options are highly liquid. Since the VIX options started trading in 2006, researchers and practitioners have been putting a lot of effort in jointly calibrating models to the SPX and VIX options prices. It has proven to be a challenging problem. As noted by many authors (e.g., [69, 101]), inconsistencies might appear between the volatility-of-volatility inferred from SPX and VIX.

In the literature, the first attempt at jointly calibrating with continuous models[1] was made by Gatheral [45], who considered a two-factor stochastic volatility model. Other

[1]Continuous models refer to continuous-time models with continuous SPX paths.

attempts include a Heston model with stochastic volatility-of-volatility by Fouque and Saporito [42] and a regime-switching stochastic volatility model by Goutte et al. [48]. In addition, many authors have tried incorporating jumps into the SPX dynamics, see, e.g., [5, 27, 75, 90, 91]. However, even with jumps, these models have yet to achieve satisfactory accuracy, particularly for short maturities. This leads to a natural question of whether there exists a continuous model which can capture the SPX and VIX smiles simultaneously. In [2, 55], Acciaio and Guyon provide a necessary condition for the existence of such continuous models. Their work was followed by the contribution of Gatheral et al. [47] who introduced the so-called quadratic rough Heston model that aims to provide a good approximation for both SPX and VIX smiles with only six parameters. Notably, apart from continuous models, a remarkable result was obtained by Guyon [56] recently, who accurately reproduced the SPX and VIX smiles by modelling the distributions of SPX in discrete time.

In this chapter, we introduce a time continuous formulation of the joint calibration problem. Instead of directly modelling the instantaneous volatility of the SPX or the VIX index, we consider a semimartingale X whose first element X^1 is the logarithm of the SPX price and whose second element X^2 is defined as the expectation of the forward quadratic variation of X^1. By doing so, the calibration exercise only depends on the marginals of X at fixed times, and the joint calibration problem falls into the class of the semimartingale optimal transport problem studied in [54]. As a corollary of the superposition principle of Trevisan [105] (or earlier Figalli [40] for the bounded coefficients case), for any probability measure such that the drift and diffusion of X are adapted processes, there exists another measure under which the semimartingale X reduces to a time-inhomogeneous diffusion and has the same marginals at fixed times under both measures. It is worth noting that the idea of using diffusion processes to mimic an Itô process by matching their marginals at fixed times traces back to the classical mimicking theorem of Gyöngy [58], which was later extended by Brunick and Shreve [23] to remove the conditions of nondegeneracy and boundedness on the covariance of the Itô process. Based on this result, as shown in [54], it is sufficient to look for solutions among such diffusion processes. This allows us to deduce a PDE formulation of the problem along with its dual counterpart. The latter naturally gives rise to Hamilton–Jacobi–Bellman (HJB) equations which can be used to represent the solutions to the original problem. Importantly, being Markovian in the state variables, our calibrated model allows us to easily derive hedging strategies for any other options. Indeed, as long as the covariance matrix is invertible, the model is complete (see [32]) and all derivatives based on X can be fully delta hedged through dynamical trading in the SPX index and variance swaps on it.

In terms of numerical aspects, pricing of VIX derivatives involves evaluating the square root of a conditional expectation. This requires nested Monte Carlo or least square Monte Carlo methods. Nested Monte Carlo has good accuracy, but is computationally expensive. Least square Monte Carlo is efficient, but it is difficult to determine the sign of the error, which can be a useful piece of information in risk management. In the previous work of two authors of this chapter [49], the least square Monte Carlo approach was adapted for computing the duality bounds of VIX derivatives. In this chapter, by taking X^2 as the forward quadratic variation of X^1, we can use conventional Monte Carlo methods or PDE methods to calculate the prices of VIX options and futures. Then, X is calibrated by a gradient descent method proposed in [54], in which an HJB equation is numerically solved by a fully implicit finite difference method at each iteration. It should be mentioned that a similar numerical algorithm was studied much earlier in [4] in the context of entropy minimisation. Let us also point out that, by defining suitable state variables, our results are applicable to any calibration problem in which the calibration instruments have payoffs in the form of a function of a conditional expectation.

In fact, the calibration method presented in this chapter shares many common features with Guyon's approach [56]. For example, both methods are nonparametric and based on the theory of optimal transport, and both methods suffer from the curse of dimensionality when considering multiple maturities of VIX futures and options. Despite these similarities, there are many important differences as well. On one hand, Guyon's model is fitted to the distributions implied from market SPX and VIX options and futures, and our model is directly calibrated to the market prices of these products. On the other hand, Guyon's method seeks a three-dimensional joint probability measures on SPX and VIX at the start date of VIX and on SPX only on the end date of VIX. Our method recovers the whole trajectory distributions of SPX in a given time interval. We must acknowledge that, compared to Guyon's method, our method is more computationally expensive. We leave the study of reducing the computational complexity for future research.

This chapter is organised as follows. Section 4.2 introduces some basic notations and the formulation of the problem. Section 4.3 presents the main results including a dimension reduction result, the PDE formulation and the dual formulation. Section 4.4 describes the numerical method in detail. Finally, in Section 4.5, we provide numerical examples with both simulated data and market data.

4.2 Problem formulation

4.2.1 Preliminaries

Let E be a Polish space equipped with its Borel σ-algebra. We denote $C(E)$ the set of continuous functions on E and $C_b(E)$ the set of bounded continuous functions on E. Denote by $\mathcal{P}(E)$ the set of Borel probability measures endowed with the weak-$*$ topology. Let $BV(E)$ be the set of functions of bounded variation and $L^1(d\mu)$ be the set of μ-integrable functions. We also write $C(E,\mathbb{R}^d), C_b(E,\mathbb{R}^d), BV(E,\mathbb{R}^d)$ and $L^1(d\mu,\mathbb{R}^d)$ for the vector-valued versions of their corresponding sets.

Let $\Omega := C([0,T],\mathbb{R}^2)$ be the two-dimensional canonical space with the canonical process $X = (X^1, X^2)$, and let $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ be the canonical filtration generated by X. Denote by $\mathcal{P}$ the set of Borel probability measures on $(\Omega, \mathcal{F}_T), T > 0$. Let $\mathcal{P}^0 \subset \mathcal{P}$ denote the subset of measures such that, for each $\mathbb{P} \in \mathcal{P}^0$, $X \in \Omega$ is an $(\mathbb{F},\mathbb{P})$-semimartingale given by

$$X_t = X_0 + A_t + M_t, \qquad \langle X\rangle_t = \langle M\rangle_t = B_t, \quad \mathbb{P}\text{-a.s.}, \tag{4.1}$$

where M is an $(\mathbb{F},\mathbb{P})$-martingale and (A, B) is $\mathbb{P}$-a.s. absolutely continuous with respect to t. In particular, $\mathbb{P}$ is said to be characterised by $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$, which is defined in the following way,

$$\alpha_t^{\mathbb{P}} = \frac{dA_t}{dt}, \quad \beta_t^{\mathbb{P}} = \frac{dB_t}{dt}.$$

Note that $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ is $\mathbb{F}$-adapted and determined up to $d\mathbb{P} \times dt$, almost everywhere. In general, $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ takes values in the space $\mathbb{R}^2 \times \mathbb{S}^2_+$, where $\mathbb{S}^2$ is the set of symmetric matrices and $\mathbb{S}^2_+$ is the set of positive semidefinite matrices of order two. For any $A, B \in \mathbb{S}^2$, we write $A : B = \mathrm{tr}(A^\intercal B)$. Denote by $\mathcal{P}^1 \subset \mathcal{P}^0$ a set of probability measures $\mathbb{P}$ whose characteristics $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ are $\mathbb{P}$-integrable. In other words,

$$\mathbb{E}^{\mathbb{P}}\left(\int_0^T |\alpha_t^{\mathbb{P}}| + |\beta_t^{\mathbb{P}}|\, dt\right) < +\infty,$$

where $|\cdot|$ is the L^1-norm.

Denote by $F : [0,T] \times \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{S}^2 \to \mathbb{R} \cup \{+\infty\}$ a cost function, and denote by $F^* : [0,T] \times \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{S}^2 \to \mathbb{R} \cup \{+\infty\}$ the convex conjugate of F with respect to (α, β):

$$F^*(t,x,a,b) := \sup_{\alpha \in \mathbb{R}^2, \beta \in \mathbb{S}^2} \{\alpha \cdot a + \beta : b - F(t,x,\alpha,\beta)\}.$$

When there is no ambiguity, we will simply write $F(\alpha,\beta) := F(t,x,\alpha,\beta)$ and $F^*(a,b) := F^*(t,x,a,b)$.

4.2.2 The joint calibration problem

We are interested in risk-neutral measures under which the SPX price is a continuous martingale, as we assume for simplicity that both dividends and interests rates are null. Let S_t be the SPX price of the form

$$S_t = S_0 + \int_0^t \sigma_s S_s \, dW_s,$$

where σ is some adapted process and W is a one-dimensional Brownian motion. It then follows that X_t^1, the logarithm of S_t, is a semimartingale with dynamics

$$X_t^1 = X_0^1 - \frac{1}{2}\int_0^t \sigma_s^2 \, ds + \int_0^t \sigma_s \, dW_s, \quad 0 \le t \le T.$$

For such X^1, we then use X^2 to represent a half of the expectation of the forward quadratic variation of X^1 on $[t,T]$ observed at time t, that is

$$X_t^2 = X_{t,T}^2 := \mathbb{E}^{\mathbb{P}}\left(\frac{1}{2}\int_t^T \sigma_s^2 \, ds \,\middle|\, \mathcal{F}_t\right) = X_t^1 - \mathbb{E}^{\mathbb{P}}(X_T^1 \mid \mathcal{F}_t), \quad 0 \le t \le T. \tag{4.2}$$

From now on, we will interchangeably use X_t^2 for $X_{t,T}^2$ and vice versa, $X_{t,T}^2$ being used to emphasise the dependence of X^2 on T. Note that the second term on the right-hand side of (4.2) is the T-futures price on X^1 at time t and hence is a martingale. It follows that the modelling setting we just described is captured by probability measures $\mathbb{P} \in \mathcal{P}^1$ characterised by (α, β) such that

$$\alpha_t = \begin{bmatrix} -\frac{1}{2}\sigma_t^2 \\ -\frac{1}{2}\sigma_t^2 \end{bmatrix} \quad \text{and} \quad \beta_t = \begin{bmatrix} \sigma_t^2 & (\beta_t)_{12} \\ (\beta_t)_{12} & (\beta_t)_{22} \end{bmatrix}, \quad 0 \le t \le T, \tag{4.3}$$

where $(\beta_t)_{12} = d\langle X^1, X^2\rangle_t \,/\, dt$ and $(\beta_t)_{22} = d\langle X^2\rangle_t \,/\, dt$ and with the additional property that $X_{T,T}^2 = 0$ $\mathbb{P}$-a.s.

Remark 4.2.1. We note that this is a fully nonparametric description of all the models in $\mathcal{P}^1$ compatible with the market setting described above. In particular, we do not specify the dynamics of the volatility $(\sigma_t)_{t \le T}$. In Section 4.2.3, we show that X may reproduce Heston's stochastic volatility market dynamics. More generally, we believe X may capture the SPX and VIX smiles of a wide range of one-factor stochastic volatility models. However, to capture full model dynamics for other models including multi-factor stochastic volatility models, one would need to add some additional state variables so they can explicitly express $\mathbb{E}^{\mathbb{P}}(X_T^1 \mid \mathcal{F}_t)$ in terms of all state variables, which also increases the dimension of the problem.

In order to restrict the probability measures to those characterised by (α, β) of the form (4.3), we can define a cost function that penalises characteristics that are not in the following convex set:

$$\Gamma := \left\{ (\alpha, \beta) \in \mathbb{R}^2 \times \mathbb{S}^2_+ : \alpha_1 = \alpha_2 = -\frac{1}{2}\beta_{11} \right\}.$$

Define the convex cost function F as follows:

$$F(\alpha, \beta) = \begin{cases} \sum_{i,j=1}^{2} (\beta_{ij} - \overline{\beta}_{ij})^2 & \text{if } (\alpha, \beta) \in \Gamma, \\ +\infty & \text{otherwise}, \end{cases} \tag{4.4}$$

where $\overline{\beta}$ is a matrix of some reference values for β. Note that $\overline{\beta}$ may depend on (t, X_t). Then, F is finite if and only if (α, β) is in the form of (4.3). Furthermore, F allows for stability across calibration exercises through specification of a reference model $\overline{\beta}$. Employing F as the cost function, our aim will be to find a model which is the closest to $\overline{\beta}$ among the ones which calibrate fully to the given market data. We comment further on the significance of $\overline{\beta}$ below in Section 4.5.

The calibration instruments we consider are SPX European options, VIX options and VIX futures. The market prices of these derivatives can be imposed as constraints on X. Let G be a vector of m number of SPX option payoff functions[2]. For example, if the i-th option is a put option with a strike K_i, then the payoff function $G_i : \mathbb{R}^2 \to \mathbb{R}_+$ is given by $G_i(x) = \max(K_i - \exp(x_1), 0)$. Let $u^{SPX} \in \mathbb{R}^m$ be the SPX option prices and $\tau \in [0, T]^m$ be the vector of their maturities. The prices u^{SPX} can be imposed on X by restricting $\mathbb{P}$ to probability measures that satisfy

$$\mathbb{E}^{\mathbb{P}} G_i(X_{\tau_i}) = u_i^{SPX}, \qquad \forall i = 1, \dots, m.$$

Let $0 \le t_0 \le T$. The annualised realised variance of $S_t = \exp(X_t^1)$ over a time grid $t_0 < t_1 < \cdots < t_n = T$ is defined to be

$$AF \sum_{i=1}^{n} \left(\log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2,$$

where AF is an annualisation factor. For example, if t_i corresponds to the daily observation dates, then $AF = 100^2 \times 252/n$, and the realised variance is expressed in basis points per annum. As $\sup_{i=1,\dots,n} |t_i - t_{i-1}| \to 0$, the realised variance can be approximated by the quadratic variation of X_t^1, given by

$$AF \sum_{i=1}^{n} \left(\log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2 \xrightarrow{\mathbb{P}} \frac{100^2}{T - t_0} \int_{t_0}^{T} \sigma_t^2 \, dt.$$

The CBOE VIX index at t_0 is defined as the square root of a weighted average of out-of-money SPX call and put option prices with maturity $T = t_0 + 30$ days, which is an approximation of the implied volatility of a 30-day log-contract on the SPX. For models with continuous paths, the VIX index at t_0 can be expressed as the square root of the expected realised variance over the next 30 days (see [36] and [89]), that is

$$VIX_{t_0} = 100 \sqrt{\frac{2}{T - t_0} \mathbb{E}^{\mathbb{P}} \left(\frac{1}{2} \int_{t_0}^{T} \sigma_t^2 \, dt \,\middle|\, \mathcal{F}_{t_0} \right)} = 100 \sqrt{\frac{2}{T - t_0} X^2_{t_0, T}}.$$

[2] In the case of non-zero interest rate, the payoff functions in G should be discounted.

Consider VIX options and futures both with maturity t_0. Let H be a vector of n number of VIX option payoff functions. Similarly to G, if the i-th VIX option is a put option with a strike K_i, then the payoff function $H_i : \mathbb{R} \to \mathbb{R}_+$ is given by $H_i(x) = \max(K_i - x, 0)$. Let $J : \mathbb{R}^2 \to \mathbb{R}_+$ be given by $J(x) := 100\sqrt{2x_2/(T - t_0)}$. Let $u^{VIX,f} \in \mathbb{R}$ be the VIX futures price and let $u^{VIX} \in \mathbb{R}^n$ be the VIX option prices. Then, we want to further restrict $\mathbb{P}$ to those under which X also satisfies the following constraints:

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} J(X_{t_0}) &= u^{VIX,f}, \\ \mathbb{E}^{\mathbb{P}}(H_i \circ J)(X_{t_0}) &= u_i^{VIX}, \qquad \forall i = 1, \dots, n. \end{aligned}$$

Finally, to ensure that $X^2_{T,T} = 0$, one additional constraint is imposed on the model. Let $\xi : \mathbb{R}^2 \to \mathbb{R}_+$ be a function such that $\xi(x) = 0$ if and only if $x_2 = 0$. Here, we choose $\xi(x) := 1 - \exp(-(x_2)^2)$ and add constraint $\mathbb{E}^{\mathbb{P}}\xi(X_T) = 0$. This constraint can be interpreted as a contract that has payoff $\xi(X_T)$ at time T, and its price is always null. From now on, we call it the *singular contract*.

We assume that $X_0 = (X_0^1, X_{0,T}^2) \in \mathbb{R}^2$ is known, and the initial marginal of X is a Dirac measure on X_0. The value of X_0^1 is the logarithm of the current SPX price. In practice, $X^2_{0,T}$ can be inferred if the market prices of SPX call and put options maturing at T are available over a continuous spectrum of strikes:

$$X^2_{0,T} = \mathbb{E}^{\mathbb{P}}\left(\frac{1}{2}\int_0^T \sigma_s^2\, ds\right) = \int_0^{\hat{f}_T} \frac{\mathbb{E}^{\mathbb{P}}(k - S_T)^+}{k^2}\, dk + \int_{\hat{f}_T}^{\infty} \frac{\mathbb{E}^{\mathbb{P}}(S_T - k)^+}{k^2}\, dk,$$

where $\hat{f}_T = \mathbb{E}^{\mathbb{P}}(S_T)$ is the T-forward price of the SPX index (e.g., see [24]). If $X^2_{0,T}$ is not observable from the market, we can treat it as a parameter. Now, putting all the constraints together, we define a set of probability measures $\mathcal{P}(X_0, G, H, \tau, t_0, T, u^{SPX}, u^{VIX,f}, u^{VIX}) \subset \mathcal{P}^1$ as follows:

$$\begin{aligned} \mathcal{P}(X_0, G, H, \tau, t_0, T, u^{SPX}, u^{VIX,f}, u^{VIX}) := \{\mathbb{P} \in \mathcal{P}^1 : \mathbb{P} \circ X_0^{-1} &= \delta_{X_0}, \\ \mathbb{E}^{\mathbb{P}} G_i(X_{\tau_i}) &= u_i^{SPX},\ i = 1, \dots, m, \\ \mathbb{E}^{\mathbb{P}} J(X_{t_0}) &= u^{VIX,f}, \\ \mathbb{E}^{\mathbb{P}}(H_i \circ J)(X_{t_0}) &= u_i^{VIX},\ i = 1, \dots, n, \\ \mathbb{E}^{\mathbb{P}}\xi(X_T) &= 0\}. \end{aligned}$$

For simplicity, we write $\mathcal{P}_{joint}$ as a shorthand for $\mathcal{P}(X_0, G, H, \tau, t_0, T, u^{SPX}, u^{VIX,f}, u^{VIX})$. Any $\mathbb{P} \in \mathcal{P}_{joint}$ is a feasible risk-neutral measure under which the semimartingale X reproduces the market prices. If $\mathcal{P}_{joint}$ is empty, it means that the market data is not compatible with a continuous-time semimartingale model with continuous paths. Adopting the convention $\inf \emptyset = +\infty$, we formulate the joint calibration problem as a *semimartingale optimal transport problem under a finite number of discrete constraints*, as studied in [54]:

Problem 3. Given $X_0, G, H, \tau, t_0, , T, u^{SPX}, u^{VIX,f}$ and u^{VIX}, solve

$$V := \inf_{\mathbb{P} \in \mathcal{P}_{joint}} \mathbb{E}^{\mathbb{P}} \int_0^T F(\alpha_s^{\mathbb{P}}, \beta_s^{\mathbb{P}})\, ds. \tag{4.5}$$

The problem is said to be *admissible* if the infimum is finite and, in particular, $\mathcal{P}_{joint}$ is nonempty.

Remark 4.2.2. Let Y be an $\mathcal{F}_T$-measurable random variable. By identifying X_t^2 as a function of X_t^1 and $\mathbb{E}^{\mathbb{P}}(Y \mid \mathcal{F}_t)$, our results apply to any model calibration problem where the payoffs of the calibration instruments can be expressed as functions of X_t^1 and X_t^2.

Remark 4.2.3. When considering multiple maturities for VIX futures and options, we need to have one X^2 for each maturity, e.g., X^2_{t,T_1}, X^2_{t,T_2}, etc. Although there is no theoretical limitation for considering multiple maturities, from numerical and practical standpoints this is challenging as each additional maturity increases the PDE's dimension.

4.2.3 An example: the Heston model

The Heston model [64] is a one-factor stochastic volatility model which directly models the spot price S_t and the instantaneous variance ν_t under the risk-neutral measure. The model dynamics are given by

$$\begin{aligned} dS_t &= \sqrt{\nu_t} S_t \, dW_t^1, \\ d\nu_t &= -\kappa(\nu_t - \theta)\, dt + \omega \sqrt{\nu_t}\, dW_t^2, \\ \langle dW^1, dW^2 \rangle_t &= \eta \, dt, \end{aligned}$$

where W_t^1 and W_t^2 are standard Brownian motions with correlation η and $\kappa, \theta > 0$ with $2\kappa\theta > \omega^2$ so that $\nu_t > 0$ a.s. In this section, we rewrite the Heston dynamics in terms of X_t^1 and $X^2_{t,T}$ and hence specify the probability measure $\mathbb{P} \in \mathcal{P}^1$ which captures the Heston dynamics.

For X^1, it is obvious that $dX_t^1 = d\log(S_t) = -\frac{1}{2}\nu_t\, dt + \sqrt{\nu_t}\, dW_t^1$. For X^2, by applying Itô's formula, we have

$$X^2_{t,T} = \mathbb{E}^{\mathbb{P}}\left(\frac{1}{2}\int_t^T \nu_s \, ds \middle| \mathcal{F}_t \right) = \frac{1 - e^{-\kappa(T-t)}}{2\kappa}(\nu_t - \theta) + \frac{1}{2}\theta(T-t). \tag{4.6}$$

Define $A(t,\kappa) := (1 - e^{-\kappa(T-t)})/\kappa$, then a simple rearrangement of (4.6) gives that

$$\nu_t = A(t,\kappa)^{-1}(2X^2_{t,T} - \theta(T-t)) + \theta =: \nu(t, X^2_{t,T}, \kappa, \theta).$$

The above equation establishes a one-to-one relation between ν_t and $X^2_{t,T}$ at time t. Applying Itô's formula to $X^2_{t,T}$, we have

$$\begin{aligned} dX^2_{t,T} &= d\left(\frac{1}{2}A(t,\kappa)(\nu_t - \theta) + \frac{1}{2}\theta(T-t) \right) \\ &= \frac{1}{2}(\nu_t - \theta)\, dA(t,\kappa) + \frac{1}{2}A(t,\kappa)\, d\nu_t - \frac{1}{2}\theta\, dt \\ &= \left(\frac{1}{2}(\nu_t - \theta)(\kappa A(t,\kappa) - 1) - \frac{1}{2}\kappa A(t,\kappa)(\nu_t - \theta) - \frac{1}{2}\theta \right) dt + \frac{1}{2}A(t,\kappa)\omega\sqrt{\nu_t}\, dW_t^2 \\ &= -\frac{1}{2}\nu_t\, dt + \frac{1}{2}A(t,\kappa)\omega\sqrt{\nu_t}\, dW_t^2. \end{aligned}$$

Therefore, the Heston model can be reformulated as

$$\begin{aligned} dX_t^1 &= -\frac{1}{2}\nu(t, X^2_{t,T}, \kappa, \theta)\, dt + \sqrt{\nu(t, X^2_{t,T}, \kappa, \theta)}\, dW_t^1, \\ dX^2_{t,T} &= -\frac{1}{2}\nu(t, X^2_{t,T}, \kappa, \theta)\, dt + \frac{1}{2}A(t,\kappa)\omega\sqrt{\nu(t, X^2_{t,T}, \kappa, \theta)}\, dW_t^2, \\ \langle dW_t^1, dW_t^2 \rangle &= \eta\, dt. \end{aligned}$$

This dynamics can be captured by the probability measure $\mathbb{P} \in \mathcal{P}^0$ characterised by (α, β) such that, for $t \in [0,T]$,

$$(\alpha_t, \beta_t) = \left(\begin{bmatrix} -\frac{1}{2}\nu(t, X^2_{t,T}, \kappa, \theta) \\ -\frac{1}{2}\nu(t, X^2_{t,T}, \kappa, \theta) \end{bmatrix}, \begin{bmatrix} \nu(t, X^2_{t,T}, \kappa, \theta) & \frac{1}{2}\eta\omega A(t,\kappa)\nu(t, X^2_{t,T}, \kappa, \theta) \\ \frac{1}{2}\eta\omega A(t,\kappa)\nu(t, X^2_{t,T}, \kappa, \theta) & \frac{1}{4}\omega^2 A(t,\kappa)^2 \nu(t, X^2_{t,T}, \kappa, \theta) \end{bmatrix} \right). \tag{4.7}$$

Further, it is easy to check that $\mathbb{E}^{\mathbb{P}} \int_0^T \nu(t, X_{t,T}^2, \kappa, \theta)\, dt < \infty$ and hence $\mathbb{P} \in \mathcal{P}^1$. The characteristics (4.7) will be used in the numerical example provided in Section 4.5 for generating simulated option prices and will also be used as a reference model.

4.3 Main results

This section is devoted to presenting our main results. By following [54], we first present a dimension reduction result which shows that the optimal transportation cost can be achieved by a set of Markov processes. Focusing only on these Markov processes, we introduce a PDE formulation. Furthermore, we deduce a dual formulation and find the optimal characteristics as by-product of solving the dual formulation.

4.3.1 Dimension reduction

In this section, we show that if Problem 3 is admissible then the optimal transportation cost V can be found by minimising (4.5) over a subset of probability measures under which X is a (time inhomogeneous) Markov processes. Before proceeding, we introduce some notations for brevity. Denote by $\mathbb{E}^{\mathbb{P}}_{t,x}$ the conditional expectation $\mathbb{E}^{\mathbb{P}}(\,\cdot \mid X_t = x)$. For any square matrix $\beta \in \mathbb{S}^2_+$, we write $\beta^{\frac{1}{2}}$ such that $\beta = \beta^{\frac{1}{2}}(\beta^{\frac{1}{2}})^\intercal$. Now, let us restate Lemma 3.1 of [54].

Lemma 4.3.1. *Let $\mathbb{P} \in \mathcal{P}^1$ and $\rho^{\mathbb{P}}_t = \rho^{\mathbb{P}}(t, \cdot) = \mathbb{P} \circ X_t^{-1}$ be the marginal distribution of X_t under $\mathbb{P}$, $t \le T$. Then $\rho^{\mathbb{P}}$ is a weak solution to the Fokker–Planck equation:*

$$\begin{cases} \partial_t \rho^{\mathbb{P}}_t + \nabla_x \cdot (\rho^{\mathbb{P}}_t \mathbb{E}^{\mathbb{P}}_{t,x} \alpha^{\mathbb{P}}_t) - \dfrac{1}{2} \displaystyle\sum_{i,j} \partial_{ij} (\rho^{\mathbb{P}}_t (\mathbb{E}^{\mathbb{P}}_{t,x} \beta^{\mathbb{P}}_t)_{ij}) = 0 & \text{in } [0,T] \times \mathbb{R}^2, \\ \rho^{\mathbb{P}}_0 = \delta_{X_0} & \text{in } \mathbb{R}^2. \end{cases} \tag{4.8}$$

Moreover, there exists another probability measure $\mathbb{P}' \in \mathcal{P}^1$ under which X has the same marginals, $\rho^{\mathbb{P}'} = \rho^{\mathbb{P}}$, and is a Markov process solving

$$dX_t = \alpha^{\mathbb{P}'}(t, X_t) dt + (\beta^{\mathbb{P}'}(t, X_t))^{\frac{1}{2}}\, dW^{\mathbb{P}'}_t, \quad 0 \le t \le T, \tag{4.9}$$

where $W^{\mathbb{P}'}$ is a $\mathbb{P}'$-Brownian motion, $\alpha^{\mathbb{P}'}(t, X_t) = \mathbb{E}^{\mathbb{P}}_{t,X_t} \alpha^{\mathbb{P}}_t$ and $\beta^{\mathbb{P}'}(t, X_t) = \mathbb{E}^{\mathbb{P}}_{t,X_t} \beta^{\mathbb{P}}_t$.

Lemma 4.3.1 is a corollary of the superposition principle of Trevisan [105] and Figalli [40]. It is worth noting that the idea of using diffusion processes to mimic an Itô process by matching their marginals at fixed times (also called Markovian projection in the literature) traces back to the classical mimicking theorem of Gyöngy [58], which was later extended by Brunick and Shreve [23] to remove the conditions of nondegeneracy and boundedness on the covariance of the Itô process.

Let $\mathcal{P}^{loc}_{joint} \subset \mathcal{P}_{joint}$ be the subset of probability measures under which X is Markov processes in the form of (4.9). In other words, any $\mathbb{P}' \in \mathcal{P}^{loc}_{joint}$ is characterised by $(\alpha^{\mathbb{P}'}(t, X_t), \beta^{\mathbb{P}'}(t, X_t)) := (\mathbb{E}^{\mathbb{P}}_{t,X_t} \alpha^{\mathbb{P}}_t, \mathbb{E}^{\mathbb{P}}_{t,X_t} \beta^{\mathbb{P}}_t)$ for some $\mathbb{P} \in \mathcal{P}^1$. Moreover, under $\mathbb{P}'$, X has an initial marginal δ_{X_0} and is fully calibrated to the market prices given in $\mathcal{P}_{joint}$. Applying Proposition 3.4 of [54], we have the following proposition for the joint calibration problem:

Proposition 4.3.2 (Dimension reduction)**.** *Given $\mathcal{P}_{joint}$ and $\mathcal{P}^{loc}_{joint}$, if Problem 3 is admissible, then*

$$V = \inf_{\mathbb{P} \in \mathcal{P}_{joint}} \mathbb{E}^{\mathbb{P}} \int_0^T F(\alpha^{\mathbb{P}}_t, \beta^{\mathbb{P}}_t)\, dt = \inf_{\mathbb{P} \in \mathcal{P}^{loc}_{joint}} \mathbb{E}^{\mathbb{P}} \int_0^T F(\alpha^{\mathbb{P}}(t, X_t), \beta^{\mathbb{P}}(t, X_t))\, dt.$$

4.3.2 PDE formulation

For any $\mathbb{P} \in \mathcal{P}^{loc}_{joint}$, the characteristics are function of the state variable X_t and time t. As is classical in the theory of diffusions, this allows us to leverage PDE methods to describe Problem 3 and to use conventional numerical methods to find its solutions.

Proposition 4.3.3. *If Problem 3 is admissible, then*

$$V = \inf_{\rho,\alpha,\beta} \int_0^T \int_{\mathbb{R}^2} F(\alpha(t,x), \beta(t,x))\, \rho(t,dx)\, dt, \tag{4.10}$$

among all $(\rho, \alpha, \beta) \in C([0,T], \mathcal{P}(\mathbb{R}^2)) \times L^1(d\rho_t dt, \mathbb{R}^2) \times L^1(d\rho_t dt, \mathbb{S}^2_+)$ *satisfying the following constraints in the sense of distributions:*

$$\partial_t \rho(t,x) + \nabla_x \cdot (\rho(t,x)\alpha(t,x)) - \frac{1}{2}\sum_{i,j} \partial_{ij}(\rho(t,x)\beta_{ij}(t,x)) = 0 \quad \text{in } [0,T] \times \mathbb{R}^2, \tag{4.11}$$

$$\int_{\mathbb{R}^2} G_i(x)\, \rho(\tau_i, dx) = u_i^{SPX} \quad \forall i = 1, \dots, m, \tag{4.12}$$

$$\int_{\mathbb{R}^2} J(x)\, \rho(t_0, dx) = u^{VIX,f}, \tag{4.13}$$

$$\int_{\mathbb{R}^2} (H_i \circ J)(x)\, \rho(t_0, dx) = u_i^{VIX} \quad \forall i = 1, \dots, n, \tag{4.14}$$

$$\int_{\mathbb{R}^2} \xi(x)\, \rho(T, dx) = 0, \tag{4.15}$$

and the initial condition $\rho(0, \cdot) = \delta_{X_0}$.

Proof. This proposition follows immediately from Lemma 4.3.1. The interchange of integrals in the objective is justified by Fubini's theorem. For the weak continuity of measure ρ in time we refer the reader to [82]. □

The PDE formulation can be solved by the alternating direction method of multipliers (ADMM) which was originally used in [8] to solve the classical optimal transport. This method was extended to a one-dimensional martingale optimal transport problem in [51] and to instationary mean field games with diffusion in [3]. However, for problems with diffusions, the ADMM method requires to solve a fourth-order PDE with a bi-Laplacian operator. In this chapter, we work on an alternative dual formulation derived by following the arguments in [54]. This will be presented in the next subsection.

4.3.3 Dual formulation

Although the PDE formulation is not a convex problem, it can be made convex by considering the triple of measures $(\rho, \mathcal{A}, \mathcal{B}) := (\rho, \rho\alpha, \rho\beta)$. By doing so, the objective function (4.10) is convex in $(\rho, \mathcal{A}, \mathcal{B})$. Moreover, the initial condition and the constraints (4.11) to (4.15) are linear in $(\rho, \mathcal{A}, \mathcal{B})$ and hence produce a convex feasible set. In consequence, the classical tools of convex analysis can be applied. Following Proposition 3.5 of [54], we introduce a dual formulation.

Let $\lambda^{SPX} \in \mathbb{R}^m$, $\lambda^{VIX,f} \in \mathbb{R}$, $\lambda^{VIX} \in \mathbb{R}^n$ and $\lambda^\xi \in \mathbb{R}$ be the Lagrange multipliers of constraints (4.12) to (4.15), respectively. To avoid confusion with the Dirac measure $\delta : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ used previously, we denote by $\mathcal{D} : [0,T] \to \mathbb{R} \cup \{+\infty\}$ the Dirac delta function in the sense of distributions. The dual formulation is given as follows:

Theorem 4.3.4 (Duality). *If Problem 3 is admissible, we have*

$$V = \sup_{(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^{\xi}) \in \mathbb{R}^{m+n+2}} \lambda^{SPX} \cdot u^{SPX} + \lambda^{VIX,f} u^{VIX,f} + \lambda^{VIX} \cdot u^{VIX} - \phi(0, X_0), \tag{4.16}$$

where ϕ is the viscosity solution to the HJB equation:

$$\begin{aligned}\partial_t \phi(t,x) + F^*(\nabla_x \phi(t,x), \frac{1}{2}\nabla_x^2 \phi(t,x)) &= -\sum_{i=1}^m \lambda_i^{SPX} G_i(x) \mathcal{D}(t-\tau_i) \\ - \lambda^{VIX,f} J(x) \mathcal{D}(t-t_0) - \sum_{i=1}^n \lambda_i^{VIX} (H_i \circ J)(x) \mathcal{D}(t-t_0) &- \lambda^{\xi} \xi(x) \mathcal{D}(t-T) \quad \text{in } [0,T] \times \mathbb{R}^2,\end{aligned} \tag{4.17}$$

with the terminal condition $\phi(T,\cdot) = 0$. Moreover, if Problem 3 is admissible, then the infimum in (4.10) is attained. If the supremum in (4.16) is attained by some λ^{SPX}, $\lambda^{VIX,f}$, λ^{VIX} and λ^{ξ} for which the associated solution to (4.17) is $\phi^* \in BV([0,T], C_b^2(\mathbb{R}^2))$, and if (ρ, α, β) is an optimal solution of Problem 3, then (α, β) is given by

$$(\alpha_t, \beta_t) = \nabla F^*(\nabla_x \phi^*(t,\cdot), \frac{1}{2}\nabla_x^2 \phi^*(t,\cdot)), \quad d\rho_t dt - \textit{almost everywhere}. \tag{4.18}$$

Theorem 4.3.4 is an application of the Fenchel–Rockafellar duality theorem [106, Theorem 1.9]. Due to the presence of $\mathcal{D}$ in the source terms, the viscosity solution ϕ satisfies (4.17) in the sense of distributions[3]. Moreover, ϕ has possible discontinuities at t_0, T and τ_i, $i = 1, \ldots, m$. The numerical solution to (4.17) is described in detail in Section 4.4. For the cost function F defined in (4.4), the convex conjugate F^* is given in Lemma B.1.1.

Remark 4.3.5. As mentioned in the previous work [54], the admissibility condition in Theorem 4.3.4 was imposed for fulfilling the conditions of Fenchel–Rockafellar theorem and hence simplifying the presentation and arguments. However, it is possible to remove this assumption from Proposition 4.3.3 with some modifications in the proof and still obtain the duality result in Theorem 4.3.4. Furthermore, characterising the admissibility of Problem 3 can be seen as a more elaborate analogue of Strassen's theorem for the classical optimal transport problem, which is however out of the scope of this paper.

In the dual formulation, the supremum can be solved by a standard optimisation algorithm. As pointed out in [54, Lemma 4.5], the convergence can be improved by providing the gradients of the objective.

Lemma 4.3.6. *Suppose Problem 3 is admissible and let*

$$L(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^{\xi}) := \lambda^{SPX} \cdot u^{SPX} + \lambda^{VIX,f} u^{VIX,f} + \lambda^{VIX} \cdot u^{VIX} - \phi(0, X_0).$$

Then, the gradients of the objective can be formulated as the difference between the market prices and the model prices:

$$\partial_{\lambda_i^{SPX}} L = u_i^{SPX} - \mathbb{E}^{\mathbb{P}} G_i(X_{\tau_i}), \quad i = 1, \ldots, m, \tag{4.19}$$

$$\partial_{\lambda^{VIX,f}} L = u^{VIX,f} - \mathbb{E}^{\mathbb{P}} J(X_{t_0}), \tag{4.20}$$

$$\partial_{\lambda_i^{VIX}} L = u_i^{VIX} - \mathbb{E}^{\mathbb{P}} (H_i \circ J)(X_{t_0}), \quad i = 1, \ldots, n, \tag{4.21}$$

$$\partial_{\lambda^{\xi}} L = -\mathbb{E}^{\mathbb{P}} \xi(X_T). \tag{4.22}$$

[3]For the precise definition of viscosity solutions to (4.17) and the corresponding comparison principle, we refer the reader to [54, Section 3.3].

In the optimisation process, the gradients are decreasing to zero while the solution is approaching the optimal solution, which illustrates the improving matching of model prices with the market prices. We note that the model prices, corresponding to a particular model (α, β), are obtained, via the Feynman-Kac formula, by solving linear pricing PDEs. More precisely, the model price of an instrument with payoff $\mathcal{G}$ and maturity $\mathcal{T}$ is equal to $\mathbb{E}^{\mathbb{P}}\mathcal{G}(X_{\mathcal{T}}) = \phi'(0, X_0)$, where ϕ' satisfies

$$\begin{cases} \partial_t \phi' + \alpha \cdot \nabla_x \phi' + \dfrac{1}{2}\beta : \nabla_x^2 \phi' = 0, & \text{in } [0, \mathcal{T}) \times \mathbb{R}^2, \\ \phi'(\mathcal{T}, \cdot) = \mathcal{G}. \end{cases} \tag{4.23}$$

When applying Lemma 4.3.6, we shall be using (4.23) m times for $(\mathcal{G}, \mathcal{T}) = (G_i, \tau_i)$, $i = 1, \dots, m$, once for $(\mathcal{G}, \mathcal{T}) = (J, t_0)$, n times for $(\mathcal{G}, \mathcal{T}) = (H_i \circ J, t_0)$, $i = 1, \dots, n$, and once for $(\mathcal{G}, \mathcal{T}) = (\xi, T)$. We shall simply refer to this as solving the linear pricing PDEs (4.23). Naturally, once the optimal model (α^*, β^*) is found, the above can be used not only to verify that it is indeed calibrated but also to compute other option prices under the model.

Remark 4.3.7. The most computationally expensive operation of numerically solving (4.23) is inverting a large sparse matrix. However, since the computations of all components of the gradient involve solving the same linear PDE but with different terminal conditions, the matrix inversion only need to be carried out once per time step. Alternatively, all gradients can be efficiently computed in one Monte Carlo simulation. In the numerical examples below (see Section 4.5), we choose to numerically solve (4.23) for the sake of accuracy.

4.4 Numerical methods

4.4.1 Solving the dual formulation

The numerical method proposed in [54] can be directly applied to solve the dual formulation, albeit with a number of caveats. Let us first recall the numerical method. Given an initial guess $(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^{\xi})$, we solve the HJB equation (4.17) to get $\phi(0, X_0)$ and hence to calculate the objective value. Due to the presence of the Dirac delta functions $\mathcal{D}$, ϕ might be discontinuous in time. The HJB equation can be solved in several time intervals in which, in each interval, the solution ϕ is continuous in both time and space, and the source terms with $\mathcal{D}$ can be incorporated into the terminal conditions. For example, if we consider SPX options with maturities t_0 and T, the HJB equation (4.17) can be reformulated as follows:

$$\begin{cases} \partial_t \phi + \sup_{\beta \in \mathbb{S}^2_+} \Big(-\frac{1}{2}\beta_{11}\partial_{x_1}\phi - \frac{1}{2}\beta_{11}\partial_{x_2}\phi + \frac{1}{2}\beta_{11}\partial_{x_1x_1}\phi \\ \qquad + \beta_{12}\partial_{x_1x_2}\phi + \frac{1}{2}\beta_{22}\partial_{x_2x_2}\phi - \sum_{i,j=1}^{2}(\beta_{ij} - \overline{\beta}_{ij})^2 \Big) = 0 & \text{in } [t_0, T), \\ \phi(T^-, \cdot) = \sum_{i=1}^{m} \lambda_i^{SPX} G_i \mathbb{1}(\tau_i = T) + \lambda^{\xi}\xi, \end{cases} \tag{4.24}$$

$$\begin{cases} \partial_t \phi + \sup_{\beta \in \mathbb{S}^2_+} \Big(-\frac{1}{2}\beta_{11}\partial_{x_1}\phi - \frac{1}{2}\beta_{11}\partial_{x_2}\phi + \frac{1}{2}\beta_{11}\partial_{x_1x_1}\phi \\ \qquad + \beta_{12}\partial_{x_1x_2}\phi + \frac{1}{2}\beta_{22}\partial_{x_2x_2}\phi - \sum_{i,j=1}^{2}(\beta_{ij} - \overline{\beta}_{ij})^2 \Big) = 0 & \text{in } [0, t_0), \\ \phi(t_0^-, \cdot) = \phi(t_0, \cdot) + \sum_{i=1}^{m} \lambda_i^{SPX} G_i \mathbb{1}(\tau_i = t_0) + \lambda^{VIX,f} J + \sum_{i=1}^{n} \lambda_i^{VIX}(H_i \circ J). \end{cases} \tag{4.25}$$

We then calculate the gradients of the objective by Lemma 4.3.6, in which the linear pricing PDEs (4.23) are solved by an alternating direction implicit (ADI) method (see e.g., [68]). Once we have the gradient values, we update $(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^{\xi})$ by moving them against their gradients or by supplying gradients to an optimisation algorithm. Notably, the L-BFGS algorithm [81] was employed and showed good convergence. The above steps are repeated until some optimality condition is met. When Problem 3 is not admissible, i.e., there does not exist a probability measure that calibrates the model to the given prices, we observe that the numerical solution will not converge, which is consistent with the arguments in Remark 4.3.5. The numerical method is summarised in Appendix B.2.

4.4.2 Solving HJB equations

In terms of numerical schemes for HJB equations, in their seminal work, Barles and Souganidis [6] have established a convergence that requires schemes to be monotone. Since then, a wide literature on monotone schemes has developed. For multidimensional HJB equations, it is usually difficult to construct a monotone scheme because of the cross partial derivative terms. To ensure monotonicity, the explicit wide stencil schemes were studied by Bonnans and Zidani [15] and by Debrabant and Jakobsen [34]; however, the stability of explicit schemes are restricted by some CFL condition. In [84], Ma and Forsyth proposed an implicit wide stencil finite difference scheme with a local coordinate rotation which is unconditionally stable. They also maximised the use of the fixed point stencil and the central finite difference scheme to improve the order of accuracy while preserving the monotonicity of the scheme.

In this chapter, we solve the HJB equations by a fully implicit finite difference method with central-difference schemes for approximating both first- and second-order derivatives. We discretise the time interval, and then, at each time step, we approximate β by Lemma B.1.1. Once the optimal β has been found, the fully nonlinear HJB equation reduces to a linear PDE which can be solved by the standard implicit finite difference method. When approximating β, we start with an arbitrary ϕ to approximate the derivatives of ϕ. Next we solve the linearised PDE and plug the solution back into the supremum to approximate β at the same time. The above procedure is repeated until ϕ converges, then we proceed to the next time step. This successive approximation is known as policy iteration in the literature. A good approximation to the initial ϕ is the one from the previous time step, which makes ϕ converge within a few iterations.

It is difficult to choose the boundary conditions of the HJB equations for this problem. Consider a computational domain $(x_1, x_2) \in [X^1_{min}, X^1_{max}] \times [0, X^2_{max}]$. We impose the following boundary conditions to equations (4.24) and (4.25):

$$\begin{cases} \nabla^2_x \phi(t,x) = \nabla^2_x \phi(T^-, x), & \text{for } (t,x) \in [t_0, T) \times (\{X^1_{min}, X^1_{max}\} \times [0, X^2_{max}] \cup [X^1_{min}, X^1_{max}] \times \{X^2_{max}\}) \\ \phi(t,x) = \phi(T^-, x), & \text{for } (t,x) \in [t_0, T) \times [X^1_{min}, X^1_{max}] \times \{0\} \end{cases}$$

$$\begin{cases} \nabla^2_x \phi(t,x) = \nabla^2_x \phi(t_0^-, x), & \text{for } (t,x) \in [0, t_0) \times (\{X^1_{min}, X^1_{max}\} \times [0, X^2_{max}] \cup [X^1_{min}, X^1_{max}] \times \{X^2_{max}\}) \\ \phi(t,x) = \phi(t_0^-, x), & \text{for } (t,x) \in [0, t_0) \times [X^1_{min}, X^1_{max}] \times \{0\} \end{cases}$$

In addition, we set a sufficiently large computational domain to further reduce the impact of the boundary conditions. Since the linear pricing PDEs are related to the HJB equation, we use the following boundary conditions for equations (4.23):

$$\begin{cases} \nabla^2_x \phi'(t,x) = \nabla^2_x \mathcal{G}(x), & \text{for } (t,x) \in [0, \mathcal{T}) \times (\{X^1_{min}, X^1_{max}\} \times [0, X^2_{max}] \cup [X^1_{min}, X^1_{max}] \times \{X^2_{max}\}) \\ \phi'(t,x) = \mathcal{G}(x), & \text{for } (t,x) \in [0, \mathcal{T}) \times [X^1_{min}, X^1_{max}] \times \{0\}. \end{cases}$$

As noted in [79], the standard finite difference schemes are non-monotone unless the diffusion matrix is diagonally dominated. In spite of being non-monotone in general, this scheme has the advantage of second-order accuracy for smooth solutions and ease of implementation compared to sophisticated monotone schemes. In fact, the variance of $X^2_{t,T}$ is much smaller than the variance of X^1_t, especially when t is close to T. Thus, we scale up $X^2_{t,T}$ by performing a simple change of variables: $(X^1, X^2) \mapsto (X^1, KX^2)$ with $K > 1$. In the numerical example of the next section we take $K = 40$. Although the diffusion matrix is not diagonally dominated and the scheme is still non-monotone in general, it shows good stability and convergence for this problem after the scaling.

4.4.3 Smoothing the volatility skews

It is clear from the formulation of Problem 3 that the reference $\overline{\beta}$ influences, potentially in a very significant way, the solution. This is also confirmed by our numerics, see Section 4.5.1 below. However, in practice, a good selection of the reference $\overline{\beta}$ might not be available. Assume that there exists a $\mathbb{P}_{mkt} \in \mathcal{P}^{loc}_{joint}$, characterised by $(\alpha_{mkt}, \beta_{mkt})$, which describes the real market dynamics. When $\overline{\beta}$ is far away from β_{mkt}, even though the optimised model matches all the calibrating option prices, the optimal β may still be very different from β_{mkt}. In the numerical experiment, we observed spiky volatility surfaces and hump-shaped model volatility skews. This is not surprising because the optimiser is trying to match the model prices to the calibrating option prices while keeping β close to $\overline{\beta}$.

Denote by $F^{\overline{\beta}}$ the cost function defined in (4.4) with reference $\overline{\beta}$. Let $V(\overline{\beta})$ be the optimal objective value of Problem 3 with cost function $F^{\overline{\beta}}$. If $V(\overline{\beta}) < \infty$, by Theorem 4.3.4, $V(\overline{\beta})$ is equal to the optimal objective value of the dual formulation with $(F^{\overline{\beta}})^*$ in the HJB equation (4.17). Let $R(\overline{\beta})$ be some regularisation term that measures the smoothness of $\overline{\beta}$. In order to smooth out the volatility surfaces and the model volatility skews, it is natural to consider the following problem:

$$\operatorname*{arg\,inf}_{\overline{\beta} \in L^1(d\rho_t dt, \mathbb{S}^2_+)} \quad V(\overline{\beta}) + R(\overline{\beta}). \tag{4.26}$$

While we might not actually solve this problem, it motivates our *reference measure iteration* method. We start with an initial reference $\overline{\beta}^0$ and numerically solve the dual formulation with cost function $F^{\overline{\beta}^0}$. Then an optimal $(\beta^*)^0$ is obtained as a by-product of solving (4.17). Next, we smooth $(\beta^*)^0$ by a simple moving average over (t, X^1, X^2) with bandwidths of (l_t, l_{x_1}, l_{x_2}). In the numerical examples, we set $(l_t, l_{x_1}, l_{x_2}) = (3, 5, 5)$. Next, we set the smoothed $(\beta^*)^0$ to $\overline{\beta}^1$ and solve the dual formulation with $\overline{\beta}^1$. The above steps are repeated until the model volatility skews are smooth enough.

Remark 4.4.1. When the calibrating instruments include VIX futures, the elements of $\overline{\beta}(t, x_1, x_2)$ might contain spikes around $x_2 = 0$, which might lead to numerical instability if we take a spiky $\overline{\beta}$ as the reference. In the numerical experiments below, we remove these spikes by replacing the values of $\overline{\beta}(\cdot, \cdot, x_2), x_2 < \epsilon$ with an approximation calculated by linearly extrapolating the values of $\overline{\beta}(\cdot, \cdot, x_2), x_2 \geq \epsilon$ along x_2, where ϵ is a small positive number. We find that this simple workaround effectively eliminates the numerical instability.

Let us call the optimisation of solving (4.16) as the *inner iteration* and call the optimisation of solving (4.26) as the *outer iteration*. For the outer iteration, if the optimal $\overline{\beta}$ that achieves the infimum in (4.26) is not very smooth, bandwidths (l_t, l_{x_1}, l_{x_2}) with large values might cause the optimiser to search around the optimal $\overline{\beta}$ forever. Thus, (l_t, l_{x_1}, l_{x_2}) can be intuitively interpreted as the "step size" for the outer iteration. Moreover, in practice, we can apply an early stop technique by only running for a few iterations for the inner iteration. By doing so, the optimiser is alternating between the inner iteration and the outer iteration. We include this procedure in our numerical routines presented in the next section.

4.5 Numerical experiments

4.5.1 Simulated data

In this section, we present a numerical example to demonstrate our method. We generate some calibrating options and futures prices from a Heston model with given parameters $(\kappa, \theta, \omega, \eta)$, and we call this model the *generating model*. Next, we calibrate the semimartingale X to these simulated prices by solving the dual formulation. In this case, we know that there exists such a probability measure $\mathbb{P} \in \mathcal{P}^1$ that X can be fully calibrated to the simulated prices under $\mathbb{P}$, i.e., $\mathcal{P}^{loc}_{joint} \neq \emptyset$. Recall that the interest rates and dividends are set to null. The characteristics of $\mathbb{P}$ are given by (4.7) and the calibrating options and futures prices are computed by solving the linear pricing PDEs (4.23).

Recall that Problem 3, combined with Proposition 4.3.2, looks for a Markovian diffusion model which minimises a certain distance to a *reference model* $\overline{\beta}$ subject to being calibrated. In this section we not only show that our approach is feasible but also investigate the potential influence of the choice of the reference $\overline{\beta}$. Specifically, we consider two reference models:

(a) a Heston model with a different set of parameters $(\overline{\kappa}, \overline{\theta}, \overline{\omega}, \overline{\eta})$:

$$\overline{\beta}(t, X^1_t, X^2_{t,T}) = \begin{bmatrix} \nu(t, X^2_{t,T}, \overline{\kappa}, \overline{\theta}) & \frac{1}{2}\overline{\eta}\overline{\omega}A(t, \overline{\kappa})\nu(t, X^2_{t,T}, \overline{\kappa}, \overline{\theta}) \\ \frac{1}{2}\overline{\eta}\overline{\omega}A(t, \overline{\kappa})\nu(t, X^2_{t,T}, \overline{\kappa}, \overline{\theta}) & \frac{1}{4}\overline{\omega}^2 A(t, \overline{\kappa})^2\nu(t, X^2_{t,T}, \overline{\kappa}, \overline{\theta}) \end{bmatrix}; \tag{4.27}$$

(b) a model with constant reference values:

$$\overline{\beta}(t, X^1_t, X^2_{t,T}) = \begin{bmatrix} \overline{\beta}_{11} & \overline{\beta}_{12} \\ \overline{\beta}_{12} & \overline{\beta}_{22} \end{bmatrix}. \tag{4.28}$$

Parameter	Value	Interpretation
S_0	100	SPX spot price
X_0^1	4.6052	Initial position of X^1
$X_{0,T}^2$	0.0098	Initial position of X^2
κ	0.6	Mean reversion speed of the generating model
θ	0.09	Long-term variance of the generating model
ω	0.4	Volatility-of-volatility of the generating model
η	-0.5	Correlation between SPX and variance of the generating model
$\overline{\kappa}$	0.9	Mean reversion speed of the Heston reference model
$\overline{\theta}$	0.04	Long-term variance of the Heston reference model
$\overline{\omega}$	0.6	Volatility-of-volatility of the Heston reference model
$\overline{\eta}$	-0.3	Correlation between SPX and variance of the Heston reference model
$\overline{\beta}_{11}$	0.09	Reference value of β_{11} of the constant reference model
$\overline{\beta}_{12}$	-0.01	Reference value of β_{12} of the constant reference model
$\overline{\beta}_{22}$	0.04	Reference value of β_{22} of the constant reference model

Table 4.1: Parameter values and interpretations for the simulated data example.

The optimal models (α^*, β^*) obtained using these two reference values will be referred to, respectively, as the *OT-calibrated model with a Heston reference* and the *OT-calibrated model with a constant reference*. These should not be confused with the generating (Heston) model. The idea behind the selection of candidates is to analyse the significance of $\overline{\beta}$ by comparing the results between two cases: (a) the dynamics of the reference model are close to the true dynamics, (b) the dynamics of the reference model are very different from the true dynamics. Note that in (a), if $(\overline{\kappa}, \overline{\theta}, \overline{\omega}, \overline{\eta}) = (\kappa, \theta, \omega, \eta)$, the supremum in (4.16) is achieved by a null vector $\mathbf{0} \in \mathbb{R}^{m+n+2}$ and hence $V = 0$. In this case, the OT-calibrated model quickly recovers the generating model.

Let $t_0 = 49$ days and $T = 79$ days. The calibration instruments we consider are:

1. SPX call options maturing at 44 days ($= t_0 - 5$ days) and $T = 79$ days,
2. VIX futures maturing at $t_0 = 49$ days,
3. VIX call options maturing at $t_0 = 49$ days.

Note that we also need to consider the singular contract (i.e., $\mathbb{E}^{\mathbb{P}}\xi(X_T) = 0$) to ensure that the dynamics of X are correct. All the parameter values and their interpretations are given in Table 4.1.

In this example, we consider a uniformly discretised time interval with step size $\Delta t = 0.5$ day. The numerical solutions were mainly computed on a 100×100 uniform grid points, except for that we use 100×400 (i.e., 400 grid points in X^2) grid points for the last 10 time steps for capturing the small variation of X_2 around zero when t is close to T.

Ideally, we want the calibrated model to have at most 1 basis point error in implied volatility for both SPX options and VIX options. However, in our method, we can only calibrate the model to option prices instead of implied volatility. Therefore, we scale the payoff functions and option prices by dividing them by their Black–Scholes Vegas, which roughly converts errors in option prices to errors in implied volatility. The optimisation algorithm will iterate until the maximum error between calibrating prices and model prices are below 0.0001, or until it cannot be further optimised. In addition, the volatility skews are smoothed by the reference measure iteration method introduced in Section 4.4.3.

All numerical experiments are performed in Matlab (2020a) on a standard desktop with an i7-7700K CPU (4.5 GHz) and 32GB of RAM. The example of Heston reference takes 4 hours and the example of constant reference takes 10.7 hours. The reason that the latter example takes longer to complete is that as the constant reference value is very different from the generating model, it takes more iterations to smooth the volatility surfaces and skews by using the reference measure iteration method. We must acknowledge that our method is very computationally expensive. We plan to study on reducing the computational time in future research.

The calibration results are shown in Table 4.2, and the volatility skews are given in Figure 4.1–4.2. We can see that the OT-calibrated models, both with the Heston reference and the constant reference, accurately capture the calibrating SPX options, VIX futures and VIX options prices. The errors, in implied volatility, of the SPX options are at most 1 basis point and of the VIX options are at most 10 basis points.

To verify if the model dynamics are correct, we perform a Monte Carlo simulation of X with the Euler scheme, and the results are shown in Figure 4.3–4.4. As demonstrated, $X^2_{T,T} \approx 0$ in all three models, so we consider the constraints $X^2_{T,T} = 0$ $\mathbb{P}$-a.s. are satisfied, and the model dynamics are correct.

Regarding the robustness of the method, there is no doubt that the reference value has a significant influence on the model dynamics. In Figures 4.1 and 4.2, the SPX model volatility skews show some differences between the ones with different reference values. In the intervals between any two adjacent option strikes, these difference are relative small, which is the result of the smoothing method. In the intervals that are less than the smallest strike and greater than the largest strike, these differences are relative large, because the model is penalised away from the reference values. Surprisingly, the VIX model volatility skews show only small differences. In Figures 4.3 and 4.4, we note that the dynamics of the three models are different. In fact, the OT-calibrated model with the constant reference is very different from the other two models. We further display the volatility behaviour of the three models in Appendix B.3.

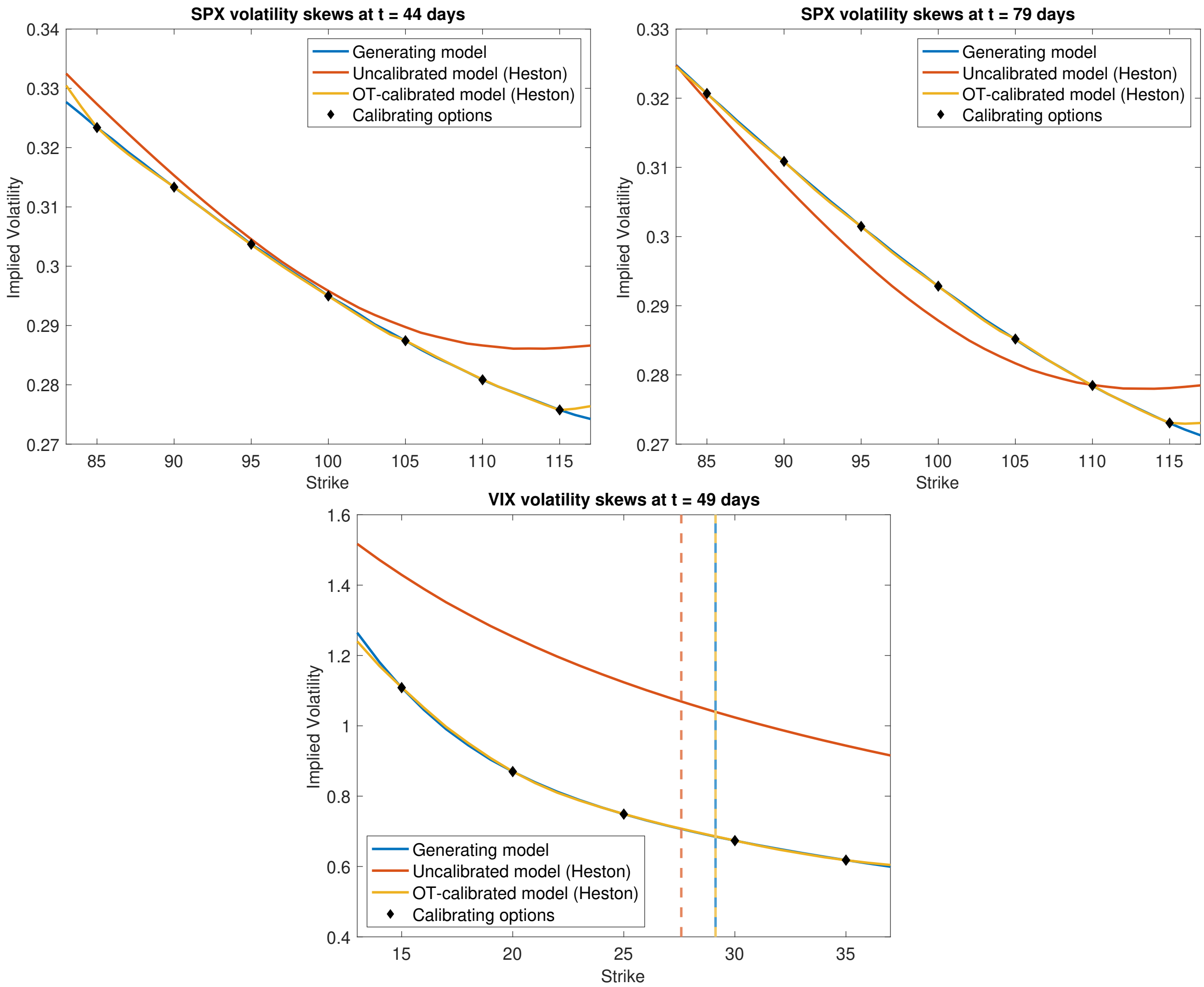

Figure 4.1: The volatility skews of SPX options at $t_0 - 5$ days $= 44$ days, SPX options at $T = 79$ days and VIX options at $t_0 = 49$ days for the simulated data example, including the implied volatility of the generating model, the uncalibrated Heston reference model and the OT-calibrated model with a Heston reference. The diamonds are the implied volatility of the calibrating options. The vertical lines are VIX futures prices.

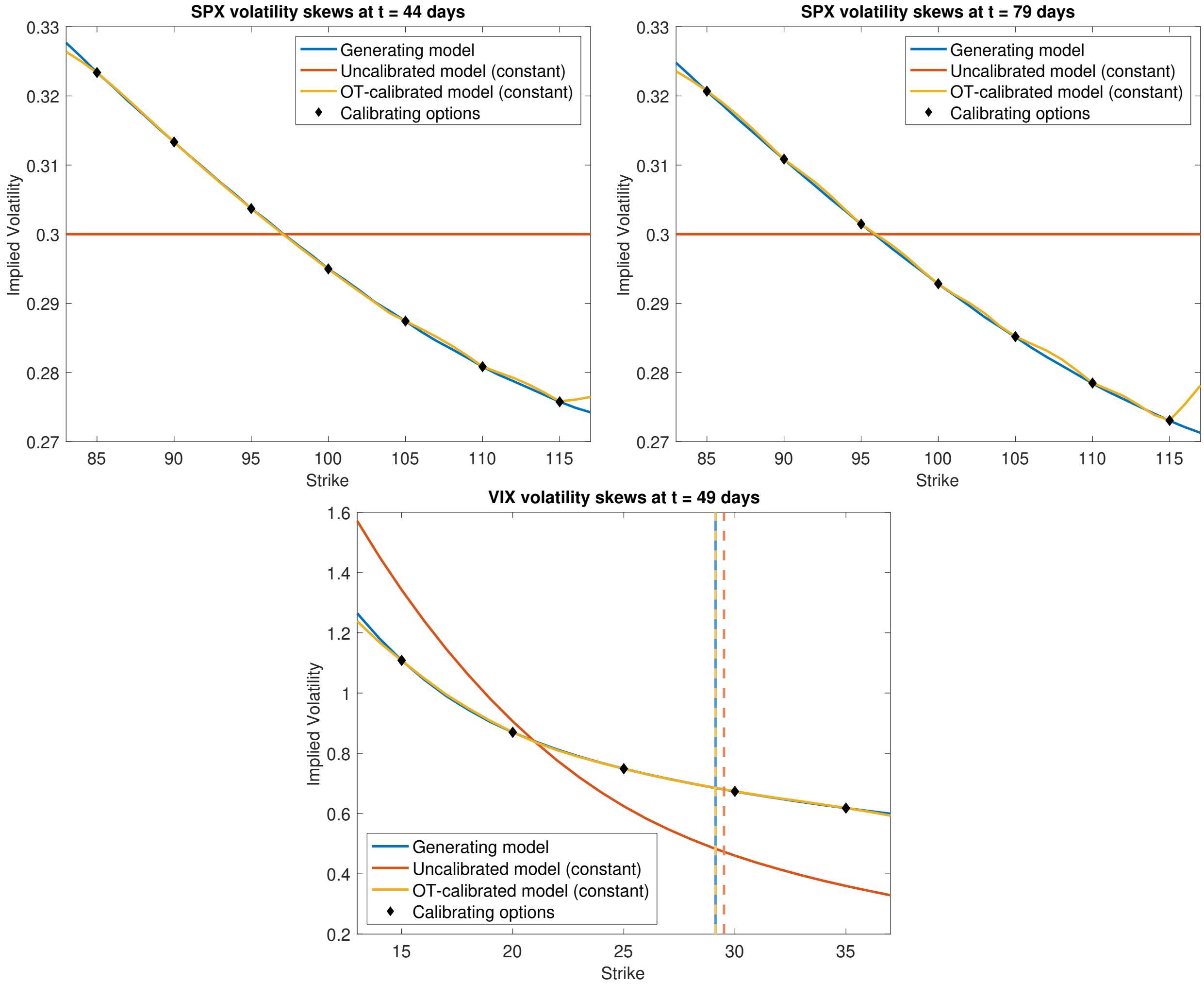

Figure 4.2: The volatility skews of SPX options at $t_0 - 5$ days $= 44$ days, SPX options at $T = 79$ days and VIX options at $t_0 = 49$ days for the simulated data example, including the implied volatility of the generating model, the uncalibrated constant reference model and the OT-calibrated model with a constant reference. The diamonds are the implied volatility of the calibrating options. The vertical lines are VIX futures prices.

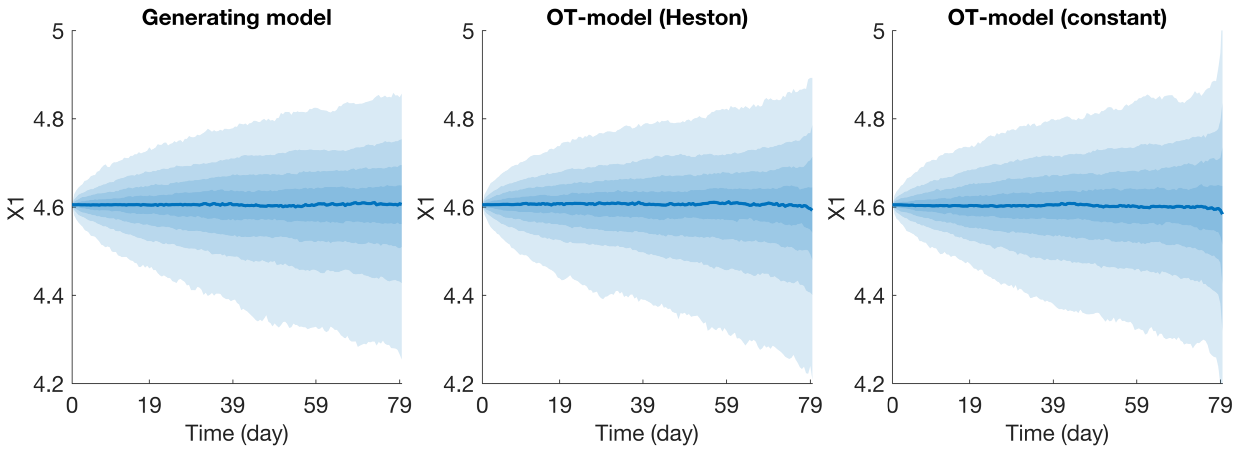

Figure 4.3: The simulations of X_t^1 for the simulated data example, including the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference.

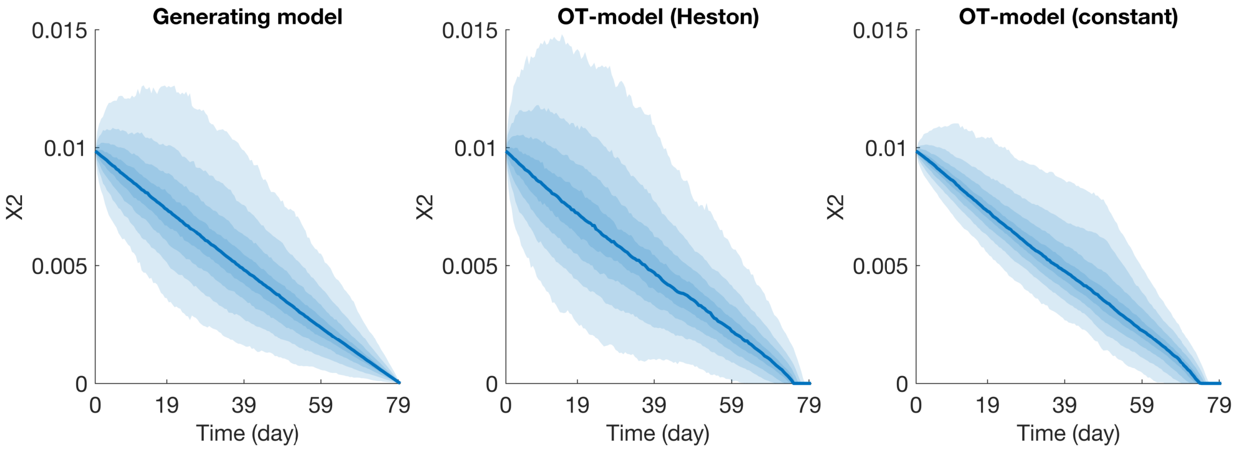

Figure 4.4: The simulations of $X^2_{t,T}$ for the simulated data example, including the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference.

			Generating model		OT-model (Heston)		OT-model (constant)	
	Maturity	Strike	Price	IV	Model price	Model IV	Model price	Model IV
SPX call options	44 days	85	15.3513	0.3234	15.3514 (0.0001)	0.3234 (0.0000)	15.3512 (-0.0001)	0.3234 (0.0000)
		90	10.9298	0.3133	10.9300 (0.0002)	0.3134 (0.0001)	10.9297 (-0.0001)	0.3133 (0.0000)
		95	7.0999	0.3037	7.0989 (-0.0010)	0.3036 (-0.0001)	7.1000 (0.0001)	0.3037 (0.0000)
		100	4.1123	0.2950	4.1121 (-0.0002)	0.2950 (0.0000)	4.1118 (-0.0005)	0.2949 (-0.0001)
		105	2.0817	0.2874	2.0819 (0.0002)	0.2875 (0.0001)	2.0818 (0.0001)	0.2874 (0.0000)
		110	0.9061	0.2808	0.9068 (0.0007)	0.2809 (0.0001)	0.9063 (0.0002)	0.2809 (0.0001)
		115	0.3392	0.2758	0.3390 (-0.0002)	0.2757 (-0.0001)	0.3395 (0.0003)	0.2758 (0.0000)
	79 days	85	15.9829	0.3207	15.9832 (0.0003)	0.3207 (0.0000)	15.9836 (0.0007)	0.3207 (0.0000)
		90	11.8931	0.3108	11.8936 (0.0005)	0.3109 (0.0001)	11.8934 (0.0003)	0.3108 (0.0000)
		95	8.3453	0.3014	8.3457 (0.0004)	0.3015 (0.0001)	8.3456 (0.0003)	0.3014 (0.0000)
		100	5.4675	0.2928	5.4680 (0.0005)	0.2928 (0.0000)	5.4678 (0.0003)	0.2928 (0.0000)
		105	3.3174	0.2851	3.3182 (0.0008)	0.2852 (0.0001)	3.3188 (0.0014)	0.2852 (0.0001)
		110	1.8524	0.2784	1.8529 (0.0005)	0.2785 (0.0001)	1.8535 (0.0011)	0.2785 (0.0001)
		115	0.9533	0.2730	0.9539 (0.0006)	0.2731 (0.0001)	0.9539 (0.0006)	0.2731 (0.0001)
VIX call options	49 days	15	14.3139	1.1086	14.3146 (0.0007)	1.1094 (0.0008)	14.3131 (-0.0008)	1.1076 (-0.0010)
		20	9.5850	0.8699	9.5856 (0.0006)	0.8702 (0.0003)	9.5854 (0.0004)	0.8701 (0.0002)
		25	5.4779	0.7489	5.4794 (0.0015)	0.7494 (0.0005)	5.4778 (-0.0001)	0.7489 (0.0000)
		30	2.5079	0.6735	2.5085 (0.0006)	0.6737 (0.0002)	2.5102 (0.0023)	0.6741 (0.0006)
		35	0.8639	0.6181	0.8632 (-0.0007)	0.6179 (-0.0002)	0.8652 (0.0013)	0.6185 (0.0004)
VIX futures	49 days		29.1285		29.1292 (0.0007)		29.1268 (-0.0017)	
Singular contract	79 days		0		5.34E-06		5.26E-08	

Table 4.2: The calibration results of the simulated data example, including prices and implied volatility (IV) of the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference. The errors are shown in the parentheses.

4.5.2 Market data

To further test the effectiveness of our method, we calibrate the model to the market data as of September 1st, 2020.

Remark 4.5.1. For simplicity, we have assumed that the interest rates and dividends are null, and the spot price is a martingale under the risk-neutral measure. However, this assumption does not apply to the market data. To overcome this issue, we let X^1 be the logarithm of the T-forward price of the SPX index instead of the spot price. Then, we are interested in T-forward measures $\mathbb{P} \in \mathcal{P}^1$ under which $\exp(X^1)$ is a martingale.

The market data consists of monthly SPX options maturing at 17 days and 45 days and monthly VIX futures and options maturing at 15 days. The model is optimised with a Heston reference (4.27) with parameters given in Table 4.3. The parameters are obtained by (roughly) calibrating a standard Heston model to the SPX option prices. It should be noted that, even with these parameters, the VIX skew generated by the Heston reference model is very unrealistic. Numerically, we have also observed that the convergence is sensitive to $\overline{\beta}$. Therefore, we apply the reference measure iteration method, developed in Section 4.4.3, to iteratively improve the reference value. The total computation time (including the reference measure iterations) is 11 hours. From a practical perspective, one way to reduce the computation time is to set the reference value to a pre-calibrated β. Nevertheless, we leave the task of finding better reference values and reducing the computation time for future research.

Parameter	X_0^1	$X_{0,T}^2$	$\overline{\kappa}$	$\overline{\theta}$	$\overline{\omega}$	$\overline{\eta}$
Value	8.17	0.0048	4.99	0.038	0.52	-0.99

Table 4.3: Parameter values for the market data example.

The OT-calibrated model volatility skews are plotted in Figure 4.5, and the simulation of X is given in Figure 4.6. From the plots, we can see that the OT-calibrated model accurately captures the market data while keeping $X_{T,T}^2 = 0$ $\mathbb{P}$-a.s. satisfied. The volatility behaviour is displayed in Appendix B.4.

Remark 4.5.2. Theoretically, the choice of $\overline{\beta}$ should affect the calibration result, but not the feasibility thereof. If there is only one model that calibrates to the constraints, e.g., when calibrating to option prices with all strikes available, the result will not depend on $\overline{\beta}$. The degree of freedom in the choice of $\overline{\beta}$ and the cost function allow us to calibrate a model even when option constraints are sparse.

Remark 4.5.3. In Figure 4.6, we observe a rapid distribution change of X^2 after the VIX options expiry. Recall that our X^2 is the forward expected quadratic variation of X^1, which is indeed the scaled variance swap. Since the market prices are from the true VIX options, this rapid distribution change could be caused by the discrepancy between the VIX value and the variance swap which does not have a listed market. This discrepancy is well known to practitioners. In our approach, the VIX is inferred from the true log-contract, coherently with the variance swap. This approximation could lead to a slight incoherence with observed market prices. We have indeed observed that the convergence of the calibration was highly sensitive to X_0^2. Note that the same approach still works if we replace X^2 by the combination of vanilla options that is used in the CBOE VIX calculation, which then allows us to potentially get better values of X_0^2 from market prices. However, we did not model X^2 that way here for the simplicity of presentation.

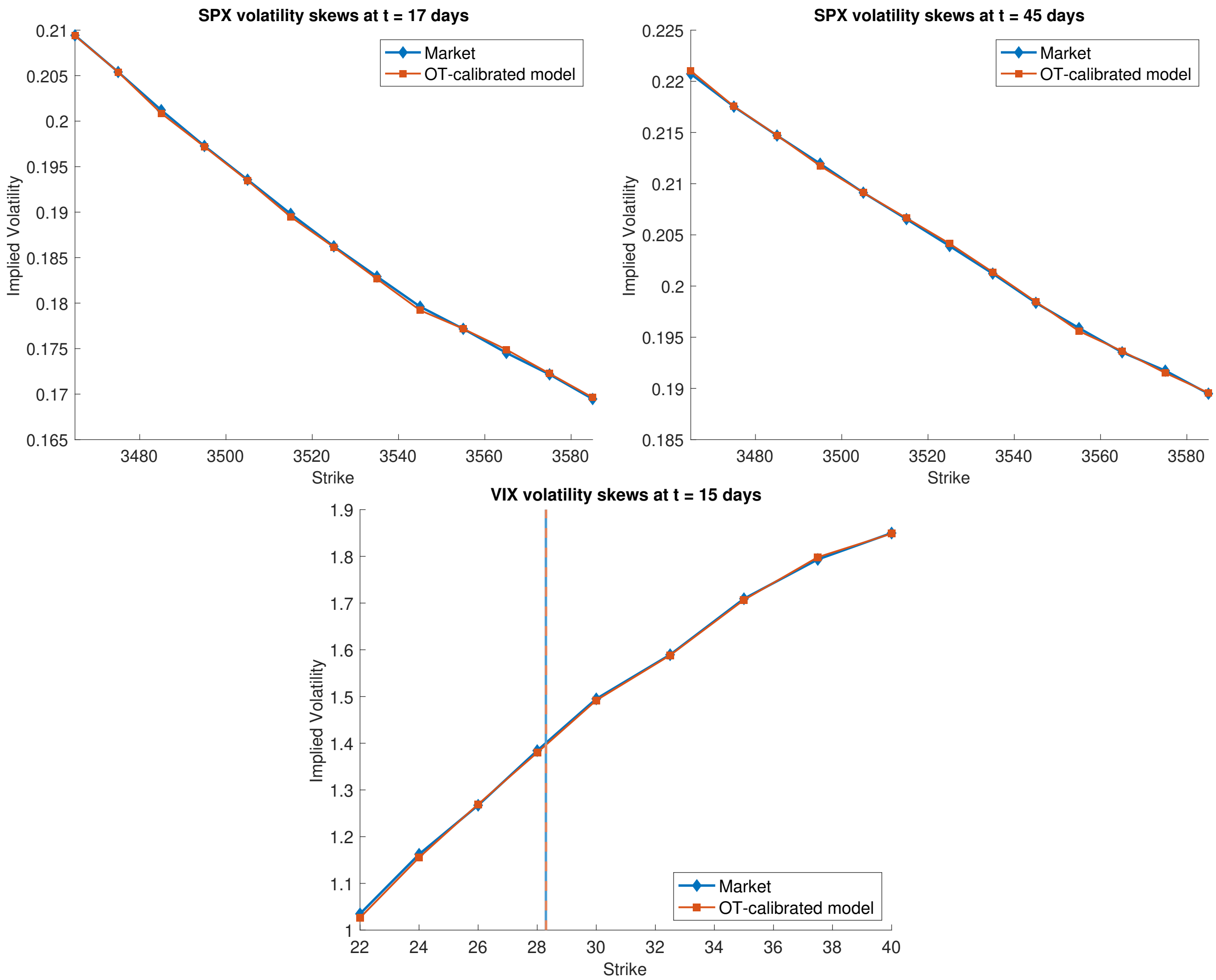

Figure 4.5: Approximated OT-calibrated model volatility skews of SPX options at $t_0 + 2$ days $= 17$ days, SPX options at $T = 45$ days and VIX options at $t_0 = 15$ days in the market data example. The vertical lines are VIX futures prices. Markers correspond to computed prices which are then interpolated with a piece-wise linear function.

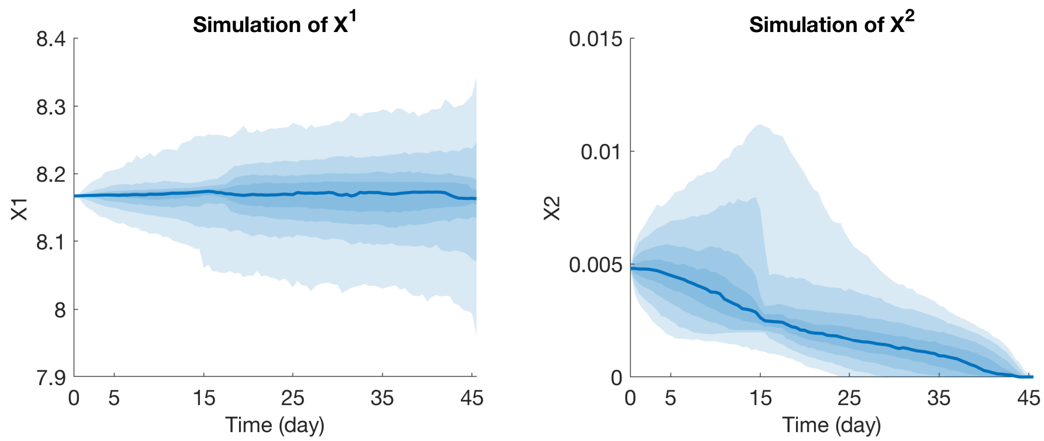

Figure 4.6: The simulations of the OT-calibrated model X in the market data example.

List of Figures

List of Tables

Bibliography

[1] F. Abergel and R. Tachet. A nonlinear partial integro-differential equation from mathematical finance. *Discrete and Continuous Dynamical Systems. Series A*, 27(3): 907–917, 2010.

[2] B. Acciaio and J. Guyon. Inversion of convex ordering: Local volatility does not maximize the price of VIX futures. *SIAM Journal on Financial Mathematics*, 11(1): SC–1, 2020.

[3] R. Andreev. Preconditioning the augmented Lagrangian method for instationary mean field games with diffusion. *SIAM Journal on Scientific Computing*, 39(6): A2763–A2783, 2017.

[4] M. Avellaneda, C. Friedman, R. Holmes, and D. Samperi. Calibrating volatility surfaces via relative-entropy minimization. *Applied Mathematical Finance*, 4(1):37–64, 1997.

[5] J. Baldeaux and A. Badran. Consistent modelling of VIX and equity derivatives using a 3/2 plus jumps model. *Applied Mathematical Finance*, 21(4):299–312, 2014.

[6] G. Barles and P. E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis*, 4(3):271–283, 1991.

[7] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.

[8] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[9] L. Bergomi. Smile dynamics. *Risk*, pages 117–123, September 2004.

[10] L. Bergomi. Smile dynamics II. *Risk*, pages 67–73, October 2005.

[11] L. Bergomi. Smile dynamics III. *Risk*, pages 90–96, October 2008.

[12] L. Bergomi. Smile dynamics IV. *Risk*, pages 94–100, December 2009.

[13] L. Bergomi. *Stochastic volatility modeling.* Chapman & Hall/CRC Financial Mathematics Series. CRC Press, Boca Raton, FL, 2016.

[14] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[15] J. F. Bonnans and H. Zidani. Consistency of generalized finite difference schemes for the stochastic HJB equation. *SIAM Journal on Numerical Analysis*, 41(3):1008–1021, 2003.

[16] B. Bouchard, G. Loeper, and Y. Zou. Almost-sure hedging with permanent price impact. *Finance and Stochastics*, 20(3):741–771, 2016.

[17] B. Bouchard, G. Loeper, and Y. Zou. Hedging of covered options with linear market impact and gamma constraint. *SIAM Journal on Control and Optimization*, 55(5): 3319–3348, 2017.

[18] B. Bouchard, G. Loeper, and Y. Zou. Hedging of covered options with linear market impact and gamma constraint. *SIAM Journal on Control and Optimization*, 55(5): 3319–3348, 2017.

[19] D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651, 1978.

[20] Y. Brenier. Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. *Communications on Pure and Applied Mathematics*, 52(4):411–452, 1999.

[21] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, and A. Sobolevskiĭ. Reconstruction of the early Universe as a convex optimization problem. *Monthly Notices of the Royal Astronomical Society*, 346(2):501–524, 2003.

[22] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, 2011.

[23] G. Brunick and S. Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *The Annals of Applied Probability*, 23(4):1584–1628, 2013.

[24] P. Carr and D. Madan. Towards a theory of volatility trading. In R. Jarrow, editor, *Volatility*, pages 417–27. Risk Publications, 1998.

[25] M. Chernov. On the role of risk premia in volatility forecasting. *Journal of Business & Economic Statistics*, 25(4):411–426, 2007.

[26] P. K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: Journal of the Econometric Society*, pages 135–155, 1973.

[27] R. Cont and T. Kokholm. A consistent pricing model for index options and volatility derivatives. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 23(2):248–274, 2013.

[28] J. C. Cox, S. A. Ross, and M. Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979.

[29] M. G. Crandall, H. Ishii, and P.-L. Lions. User's guide to viscosity solutions of second order partial differential equations. *American Mathematical Society. Bulletin. New Series*, 27(1):1–67, 1992.

[30] C. Cuchiero, W. Khosrawi, and J. Teichmann. A generative adversarial network approach to calibration of local stochastic volatility models. *Risks*, 8(4):101, 2020.

[31] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26:2292–2300, 2013.

[32] M. Davis and J. Obłój. Market completion using options. In *Advances in mathematics of finance*, volume 83 of *Banach Center Publ.*, pages 49–60. Polish Acad. Sci. Inst. Math., Warsaw, 2008.

[33] H. De March and P. Henry-Labordère. Building arbitrage-free implied volatility: Sinkhorn's algorithm and variants. *Available at SSRN 3326486*, 2019.

[34] K. Debrabant and E. R. Jakobsen. Semi-Lagrangian schemes for linear and fully non-linear diffusion equations. *Mathematics of Computation*, 82(283):1433–1462, 2013.

[35] E. Derman and I. Kani. Riding on a smile. *Risk*, 7(2):32–39, 1994.

[36] B. Dupire. Arbitrage pricing with stochastic volatility. *Preprint*, 1993.

[37] B. Dupire. Pricing with a smile. *Risk*, pages 18–20, 1994.

[38] B. Engelmann, F. Koster, and D. Oeltz. Calibration of the Heston stochastic local volatility model: A finite volume scheme. *International Journal of Financial Engineering*, page 2050048, 2021.

[39] B. Eraker, M. Johannes, and N. Polson. The impact of jumps in volatility and returns. *The Journal of Finance*, 58(3):1269–1300, 2003.

[40] A. Figalli. Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *Journal of Functional Analysis*, 254(1):109–153, 2008.

[41] M. Fortin and R. Glowinski. *Augmented Lagrangian methods*, volume 15 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1983. Applications to the numerical solution of boundary value problems, Translated from the French by B. Hunt and D. C. Spicer.

[42] J.-P. Fouque and Y. F. Saporito. Heston stochastic vol-of-vol model for joint calibration of VIX and S&P 500 options. *Quantitative Finance*, 18(6):1003–1016, 2018.

[43] A. Galichon. *Optimal transport methods in economics*. Princeton University Press, Princeton, NJ, 2016.

[44] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *The Annals of Applied Probability*, 24(1):312–336, 2014.

[45] J. Gatheral. Consistent modeling of SPX and VIX options. In *Bachelier congress*, volume 37, pages 39–51, 2008.

[46] J. Gatheral. *The volatility surface: a practitioner's guide*, volume 357. John Wiley & Sons, 2011.

[47] J. Gatheral, P. Jusselin, and M. Rosenbaum. The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. *Risk, May*, 2020.

[48] S. Goutte, A. Ismail, and H. Pham. Regime-switching stochastic volatility model: estimation and calibration to VIX options. *Applied Mathematical Finance*, 24(1): 38–75, 2017.

[49] I. Guo and G. Loeper. Pricing bounds for volatility derivatives via duality and least squares Monte Carlo. *Journal of Optimization Theory and Applications*, 179(2):598–617, 2018.

[50] I. Guo and G. Loeper. Path dependent optimal transport and model calibration on exotic derivatives. *The Annals of Applied Probability*, 31(3):1232–1263, 2021.

[51] I. Guo, G. Loeper, and S. Wang. Local volatility calibration by optimal transport. In *2017 MATRIX annals*, volume 2 of *MATRIX Book Ser.*, pages 51–64. Springer, Cham, 2019.

[52] I. Guo, G. Loeper, J. Obłój, and S. Wang. Joint modelling and calibration of SPX and VIX by optimal transport. *Available at SSRN 3568998*, 2020.

[53] I. Guo, N. Langrené, G. Loeper, and W. Ning. Deep semi-martingale optimal transport. *arXiv preprint arXiv:2103.03628*, 2021.

[54] I. Guo, G. Loeper, and S. Wang. Calibration of local-stochastic volatility models by optimal transport. *Mathematical Finance*, 32, 2021.

[55] J. Guyon. Inversion of convex ordering in the VIX market. *Quantitative Finance*, pages 1–27, 2020.

[56] J. Guyon. The joint S&P 500/VIX smile calibration puzzle solved. *Risk, April*, 2020.

[57] J. Guyon and P. Henry-Labordère. Being particular about calibration. *Risk*, 25(1): 88, 2012.

[58] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability Theory and Related Fields*, 71(4):501–516, 1986.

[59] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing smile risk. *The Best of Wilmott*, 1:249–296, 2002.

[60] J. Han, A. Jentzen, et al. Algorithms for solving high dimensional PDEs: From nonlinear monte carlo to machine learning. *arXiv preprint arXiv:2008.13333*, 2020.

[61] P. Henry-Labordère. Calibration of local stochastic volatility models to market smiles: A monte-carlo approach. *Risk*, 2009.

[62] P. Henry-Labordère. *Model-free hedging: A martingale optimal transport viewpoint.* CRC Press, 2017.

[63] P. Henry-Labordère. From (Martingale) Schrödinger bridges to a new class of Stochastic Volatility Models. *arXiv preprint arXiv:1904.04554*, 2019.

[64] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

[65] D. G. Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2 (4):329–347, 1998.

[66] M. Huesmann and D. Trevisan. A Benamou-Brenier formulation of martingale optimal transport. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 25(4A):2729–2757, 2019.

[67] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300, 1987.

[68] K. J. In 't Hout and S. Foulon. ADI finite difference schemes for option pricing in the Heston model with correlation. *International Journal of Numerical Analysis and Modeling*, 7(2):303–320, 2010.

[69] A. Jacquier, C. Martini, and A. Muguruza. On VIX futures in the rough Bergomi model. *Quantitative Finance*, 18(1):45–61, 2018.

[70] M. Jex, R. Henderson, and D. Wang. Pricing exotics under the smile. *Risk*, pages 72–75, 1999.

[71] B. Jourdain and A. Zhou. Existence of a calibrated regime switching local volatility model. *Mathematical Finance*, 30(2):501–546, 2020.

[72] L. V. Kantorovich. On the translocation of masses. *Comptes Rendus (Doklady) Academy of Science URSS*, 37:199–201, 1942.

[73] L. V. Kantorovich. On a problem of Monge (in Russian). *Uspekhi Matematicheskikh Nauk*, 3:255–226, 1948.

[74] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, second edition, 1991.

[75] T. Kokholm and M. Stisen. Joint pricing of VIX and SPX options with stochastic volatility and jump models. *The Journal of Risk Finance*, 2015.

[76] S. G. Kou. A jump-diffusion model for option pricing. *Management Science*, 48(8): 1086–1101, 2002.

[77] S. G. Kou and H. Wang. Option pricing under a double exponential jump diffusion model. *Management Science*, 50(9):1178–1192, 2004.

[78] N. V. Krylov. On the rate of convergence of finite-difference approximations for Bellman's equations with variable coefficients. *Probab. Theory Related Fields*, 117(1): 1–16, 2000.

[79] H. J. Kushner and P. Dupuis. *Numerical methods for stochastic control problems in continuous time*, volume 24 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2001.

[80] D. Lacker, M. Shkolnikov, and J. Zhang. Inverting the Markovian projection, with an application to local stochastic volatility models. *The Annals of Probability*, 48(5): 2189–2211, 2020.

[81] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(3, (Ser. B)):503–528, 1989.

[82] G. Loeper. The reconstruction problem for the Euler-Poisson system in cosmology. *Archive for Rational Mechanics and Analysis*, 179(2):153–216, 2006.

[83] G. Loeper. Option pricing with linear market impact and nonlinear Black-Scholes equations. *The Annals of Applied Probability*, 28(5):2664–2726, 2018.

[84] K. Ma and P. A. Forsyth. An unconditionally monotone numerical scheme for the two-factor uncertain volatility model. *IMA Journal of Numerical Analysis*, 37(2): 905–944, 2017.

[85] D. B. Madan, P. P. Carr, and E. C. Chang. The variance gamma process and option pricing. *Review of Finance*, 2(1):79–105, 1998.

[86] R. C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4:141–183, 1973.

[87] R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2):125–144, 1976.

[88] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.

[89] A. Neuberger. The log contract. *Journal of portfolio management*, 20:74–74, 1994.

[90] C. Pacati, G. Pompa, and R. Renò. Smiling twice: The Heston++ model. *Journal of Banking & Finance*, 96:185–206, 2018.

[91] A. Papanicolaou and R. Sircar. A regime-switching heston model for VIX and S&P 500 implied volatilities. *Quantitative Finance*, 14(10):1811–1827, 2014.

[92] G. Peyré, M. Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[93] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.

[94] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. II.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Applications.

[95] Y. Ren, D. Madan, and M. Q. Qian. Calibrating and pricing with embedded local volatility models. *Risk*, 20(9):138, 2007.

[96] R. T. Rockafellar. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[97] W. Rudin. *Real and complex analysis.* McGraw-Hill Book Co., third edition, 1987.

[98] Y. F. Saporito, X. Yang, and J. P. Zubelli. The calibration of stochastic local-volatility models: an inverse problem perspective. *Computers & Mathematics with Applications. An International Journal*, 77(12):3054–3067, 2019.

[99] W. Schoutens. *Lévy processes in finance.* Wiley, 2003.

[100] A. Sobolevskiĭ and U. Frish. Application of optimal transportation theory to the reconstruction of the early Universe. *Rossiĭskaya Akademiya Nauk. Sankt-Peterburgskoe Otdelenie. Matematicheskiĭ Institut im. V. A. Steklova. Zapiski Nauchnykh Seminarov (POMI)*, 312(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11):303–309, 317, 2004.

[101] Z. Song and D. Xiu. A tale of two option markets: State-price densities implied from S&P 500 and VIX option prices. *Unpublished working paper. Federal Reserve Board and University of Chicago*, 2012.

[102] V. Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36:423–439, 1965.

[103] X. Tan and N. Touzi. Optimal transportation under controlled stochastic dynamics. *The Annals of Probability*, 41(5):3201–3240, 2013.

[104] Y. Tian, Z. Zhu, G. Lee, F. Klebaner, and K. Hamza. Calibrating and pricing with a stochastic-local volatility model. *Journal of Derivatives*, 22(3):21, 2015.

[105] D. Trevisan. Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. *Electronic Journal of Probability*, 21:Paper No. 22, 41, 2016.

[106] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[107] C. Villani. *Optimal transport: old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

[108] M. Wyns and J. Du Toit. A finite volume–alternating direction implicit approach for the calibration of stochastic local volatility models. *International Journal of Computer Mathematics*, 94(11):2239–2267, 2017.

Appendix A

Appendix for Chapter 3

A.1 Lemma A.1.1

Lemma A.1.1. *Define* $\Phi : C_b(\Lambda, \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ *by*

$$\Phi(r, a, b) = \begin{cases} 0 & \text{if } r + F^*(a, b) \le 0, \\ +\infty & \text{otherwise.} \end{cases}$$

If we restrict the domain of its convex conjugate $\Phi^* : C_b(\Lambda, \mathcal{X})^* \to \mathbb{R} \cup \{+\infty\}$ *to* $\mathcal{M}(\Lambda, \mathcal{X})$*, then*

$$\Phi^*(\rho, \mathcal{A}, \mathcal{B}) = \begin{cases} \displaystyle\int_\Lambda F\left(\frac{d\mathcal{A}}{d\rho}, \frac{d\mathcal{B}}{d\rho}\right) d\rho & \text{if } \rho \in \mathcal{M}_+(\Lambda) \text{ and } (\mathcal{A}, \mathcal{B}) \ll \rho, \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. Let us identify the cases where $\Phi^* < +\infty$. For any $(\rho, \mathcal{A}, \mathcal{B}) \in C_b(\Lambda, \mathcal{X})^*$, using the definition of convex conjugate, we have

$$\Phi^*(\rho, \mathcal{A}, \mathcal{B}) = \sup_{(r,a,b) \in C_b(\Lambda, \mathcal{X})} \{\langle (r, a, b), (\rho, \mathcal{A}, \mathcal{B})\rangle \ ; \ r + F^*(a, b) \le 0\}.$$

If we restrict the domain of Φ^* to $\mathcal{M}(\Lambda, \mathcal{X}) \subset C_b(\Lambda, \mathcal{X})^*$, then

$$\Phi^*(\rho, \mathcal{A}, \mathcal{B}) = \sup_{(r,a,b) \in C_b(\Lambda, \mathcal{X})} \left\{ \int_\Lambda r\, d\rho + a \cdot d\mathcal{A} + b : d\mathcal{B} \ ; \ r + F^*(a, b) \le 0 \right\}.$$

To show that one can restrict to $\rho \in \mathcal{M}_+(\Lambda, \mathcal{X})$ if $\Phi^* < +\infty$, we assume that there exists a measurable set $E \subset \Lambda$ such that $\rho(E) < 0$. By the fact that C_b is dense in L^1, there exists a sequence of nonnegative functions $\zeta_n \in C_b(\Lambda)$ that converges to $\mathbb{1}_E \in L^1(d\rho_t dt)$. Let us construct a sequence $(r_n, a_n, b_n) = (-k\zeta_n, O^{d\times 1}, O^{d\times d}) \in C_b(\Lambda, \mathcal{X})$ where k is an arbitrary positive constant and $O^{m\times n}$ denotes a null matrix of size $m \times n$. It is clear that the constraint $r + F^*(a, b) \le 0$ is satisfied at $(r, a, b) = (r_n, a_n, b_n)$ as $F^*(O^{d\times 1}, O^{d\times d}) \le 0$. Then, by the dominated convergence theorem, we have

$$\begin{aligned} \Phi^*(\rho, \mathcal{A}, \mathcal{B}) &\ge \lim_{n\to+\infty} \int_\Lambda r_n\, d\rho + a_n \cdot d\mathcal{A} + b_n : d\mathcal{B} \\ &= \int_\Lambda \lim_{n\to+\infty} \left(r_n\, d\rho + a_n \cdot d\mathcal{A} + b_n : d\mathcal{B} \right) \\ &= -k \int_\Lambda \lim_{n\to+\infty} \zeta_n\, d\rho \\ &= -k\rho(E). \end{aligned}$$

If we send k to infinity, the function Φ^* becomes unbounded.

To show that it is necessary to have $(\mathcal{A},\mathcal{B}) \ll \rho$ if $\Phi^* < +\infty$, we assume that there exists a measurable set E such that $(\mathcal{A},\mathcal{B})(E) \neq 0$ but $\rho(E) = 0$. Again, by the fact that C_b is dense in L^1, there exists a sequence of functions $\zeta_n \in C_b(\Lambda)$ such that ζ_n take values between 0 and 1 and the sequence converges to $\mathbb{1}_E \in L^1(d\rho_t dt)$. Such sequence can be found by taking convolution of $\mathbb{1}_E$ with a standard regularising kernel. Let us construct a sequence $(r_n, a_n, b_n) = (-F^*(k_1 I^{d\times 1}, k_2 I^{d\times d})\zeta_n, k_1\zeta_n I^{d\times 1}, k_2\zeta_n I^{d\times d}) \in C_b(\Lambda, \mathcal{X})$ where k_1, k_2 are arbitrary constants and $I^{m\times n}$ denotes an all-ones matrix of size $m \times n$. By the convexity of F^* and the fact that $F^*(O^{d\times 1}, O^{d\times d}) \leq 0$, it is clear that the constraint $r + F^*(a,b) \leq 0$ is satisfied at $(r,a,b) = (r_n, a_n, b_n)$. Then, by the dominated convergence theorem, we have

$$
\begin{aligned}
\Phi^*(\rho,\mathcal{A},\mathcal{B}) &\geq \lim_{n\to+\infty} \int_\Lambda r_n\, d\rho + a_n \cdot d\mathcal{A} + b_n : d\mathcal{B} \\
&= \int_\Lambda \lim_{n\to+\infty} \Big(r_n\, d\rho + a_n \cdot d\mathcal{A} + b_n : d\mathcal{B} \Big) \\
&= \int_\Lambda \lim_{n\to+\infty} \Big(-F^*(k_1 I^{d\times 1}, k_2 I^{d\times d})\zeta_n\, d\rho + k_1\zeta_n I^{d\times 1} \cdot d\mathcal{A} + k_2\zeta_n I^{d\times d} : d\mathcal{B} \Big) \\
&= k_1 \sum_i (\mathcal{A}(E))_i + k_2 \sum_{i,j} (\mathcal{B}(E))_{ij}
\end{aligned}
$$

The function Φ^* goes to infinity if we send k_1, k_2 to $+\infty$ or $-\infty$, depending on the sign of $\sum_i (\mathcal{A}(E))_i$ and $\sum_{i,j} (\mathcal{B}(E))_{ij}$.

Now, since the integrand of the integral in Φ^* is linear in (r,a,b), if Φ^* is finite, the supremum must occur at the boundary. Thus, assuming that $\rho \in \mathcal{M}_+(\Lambda,\mathcal{X})$ and $(\mathcal{A},\mathcal{B}) \ll \rho$, we have

$$
\begin{aligned}
\Phi^*(\rho,\mathcal{A},\mathcal{B}) &= \sup_{r+F^*(a,b)=0} \int_\Lambda \left(r + a \cdot \frac{d\mathcal{A}}{d\rho} + b : \frac{d\mathcal{B}}{d\rho} \right) d\rho \\
&= \sup_{(a,b)} \int_\Lambda \left(a \cdot \frac{d\mathcal{A}}{d\rho} + b : \frac{d\mathcal{B}}{d\rho} - F^*(a,b) \right) d\rho \\
&\leq \int_\Lambda \sup_{(a,b)} \left(a \cdot \frac{d\mathcal{A}}{d\rho} + b : \frac{d\mathcal{B}}{d\rho} - F^*(a,b) \right) d\rho \\
&= \int_\Lambda F\left(\frac{d\mathcal{A}}{d\rho}, \frac{d\mathcal{B}}{d\rho} \right) d\rho.
\end{aligned}
$$

The last equality holds since the convex and lower semi-continuous function F coincides with its bi-conjugate F^{**} according to the Fenchel–Moreau theorem (see e.g., Brezis [22, Theorem 1.11]).

Conversely, by the density of C_b in L^1, let us choose a sequence of functions $(a_n, b_n) \in C_b(\Lambda, \mathbb{R}^d \times \mathbb{S}^d)$ converging to

$$
\nabla F\left(\frac{d\mathcal{A}}{d\rho}, \frac{d\mathcal{B}}{d\rho}\right) = \arg\sup_{(a,b)} \left(a \cdot \frac{d\mathcal{A}}{d\rho} + b : \frac{d\mathcal{B}}{d\rho} - F^*(a,b) \right) \quad \text{in } L^1(d\rho_t dt, \mathbb{R}^d \times \mathbb{S}^d).
$$

Applying the dominated convergence theorem, we have

$$\begin{aligned}
\Phi^*(\rho, \mathcal{A}, \mathcal{B}) &= \sup_{(a,b)} \int_\Lambda \left(a \cdot \frac{d\mathcal{A}}{d\rho} + b : \frac{d\mathcal{B}}{d\rho} - F^*(a,b) \right) d\rho \\
&\geq \lim_{n\to+\infty} \int_\Lambda \left(a_n \cdot \frac{d\mathcal{A}}{d\rho} + b_n : \frac{d\mathcal{B}}{d\rho} - F^*(a_n,b_n) \right) d\rho \\
&= \int_\Lambda \lim_{n\to+\infty} \left(a_n \cdot \frac{d\mathcal{A}}{d\rho} + b_n : \frac{d\mathcal{B}}{d\rho} - F^*(a_n,b_n) \right) d\rho \\
&= \int_\Lambda \sup_{(a,b)} \left(a \cdot \frac{d\mathcal{A}}{d\rho} + b : \frac{d\mathcal{B}}{d\rho} - F^*(a,b) \right) d\rho \\
&= \int_\Lambda F\left(\frac{d\mathcal{A}}{d\rho}, \frac{d\mathcal{B}}{d\rho} \right) d\rho.
\end{aligned}$$

The proof is completed. $\square$

A.2 Lemma A.2.1

In this section, we prove that the duality between spaces C_b and $\mathcal{M}$ can be extended to the non-compact space $[0,T] \times \mathbb{R}^d$ in this particular case. A similar argument for the Kantorovich duality of the classical optimal transport was made in Villani [106, Appendix 1.3].

Lemma A.2.1. *Denote by K^o the set of (r,a,b) in $C_b(\Lambda, \mathcal{X})$ that can be represented by some (ϕ, λ) in $BV([0,T], C_b^2(\mathbb{R}^d)) \times \mathbb{R}^m$ with $\phi(T,\cdot) = 0$ (see the proof of Theorem 3.5 for the definition of 'represented'). Let $\Phi^* : C_b(\Lambda, \mathcal{X})^* \to \mathbb{R} \cup \{+\infty\}$ and $\Psi^* : C_b(\Lambda, \mathcal{X})^* \to \mathbb{R} \cup \{+\infty\}$ be defined by*

$$\Phi^*(\rho, \mathcal{A}, \mathcal{B}) = \sup_{(r,a,b)\in C_b(\Lambda,\mathcal{X})} \{\langle (r,a,b), (\rho, \mathcal{A}, \mathcal{B})\rangle \ ; \ r + F^*(a,b) \leq 0\},$$

$$\Psi^*(\rho, \mathcal{A}, \mathcal{B}) = \sup_{(r,a,b)\in K^o} \left\{ \langle (r,a,b), (\rho, \mathcal{A}, \mathcal{B})\rangle - \int_{\mathbb{R}^d} \phi(0,x)\, d\mu_0 + \sum_{i=1}^m \lambda_i c_i \right\}.$$

Then,

$$\inf_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda,\mathcal{X})^*} (\Phi^* + \Psi^*)(\rho, \mathcal{A}, \mathcal{B}) = \inf_{(\rho,\mathcal{A},\mathcal{B})\in \mathcal{M}(\Lambda,\mathcal{X})} (\Phi^* + \Psi^*)(\rho, \mathcal{A}, \mathcal{B}). \tag{A.1}$$

Proof. Let $C_0(\Lambda, \mathcal{X})$ be the space of continuous functions on Λ valued in $\mathcal{X}$ that vanish at infinity. We decompose $(\rho, \mathcal{A}, \mathcal{B}) = (\tilde{\rho}, \tilde{\mathcal{A}}, \tilde{\mathcal{B}}) + (\delta\rho, \delta\mathcal{A}, \delta\mathcal{B})$ such that $(\tilde{\rho}, \tilde{\mathcal{A}}, \tilde{\mathcal{B}}) \in \mathcal{M}(\Lambda, \mathcal{X})$ and $\langle (\phi_\rho, \phi_\mathcal{A}, \phi_\mathcal{B}), (\delta\rho, \delta\mathcal{A}, \delta\mathcal{B})\rangle = 0$ for any $(\phi_\rho, \phi_\mathcal{A}, \phi_\mathcal{B}) \in C_0(\Lambda, \mathcal{X})$ (The reader can refer to Villani [106, Appendix 1.3] for the existence of such a decomposition.). Since, $\mathcal{M}(\Lambda, \mathcal{X})$ is a subset of $C_b(\Lambda, \mathcal{X})^*$, it follows that

$$\inf_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda,\mathcal{X})^*} (\Phi^* + \Psi^*)(\rho, \mathcal{A}, \mathcal{B}) \leq \inf_{(\rho,\mathcal{A},\mathcal{B})\in \mathcal{M}(\Lambda,\mathcal{X})} (\Phi^* + \Psi^*)(\rho, \mathcal{A}, \mathcal{B}).$$

Next, we show that the converse of the above inequality is also valid. If $\Phi^* \equiv +\infty$ or $\Psi^* \equiv +\infty$, then the proof is trivial. Thus, we assume that Φ^* and Ψ^* take finite values at some $(\rho, \mathcal{A}, \mathcal{B}) \in C_b(\Lambda, \mathcal{X})^*$. For Φ^*, since $C_0(\Lambda, \mathcal{X}) \subseteq C_b(\Lambda, \mathcal{X})$, we have

$$\begin{aligned}\Phi^*(\rho, \mathcal{A}, \mathcal{B}) &= \sup_{(r,a,b)\in C_b(\Lambda,\mathcal{X})} \{\langle (r,a,b), (\rho, \mathcal{A}, \mathcal{B})\rangle \; ; \; r + F^*(a,b) \leq 0\} \\ &\geq \sup_{(r,a,b)\in C_0(\Lambda,\mathcal{X})} \{\langle (r,a,b), (\rho, \mathcal{A}, \mathcal{B})\rangle \; ; \; r + F^*(a,b) \leq 0\} \\ &= \sup_{(r,a,b)\in C_0(\Lambda,\mathcal{X})} \left\{ \int_\Lambda r \, d\tilde{\rho} + a \cdot d\tilde{\mathcal{A}} + b : d\tilde{\mathcal{B}} \; ; \; r + F^*(a,b) \leq 0 \right\}.\end{aligned} \tag{A.2}$$

Let $\chi_n \in C_0(\Lambda)$ be a sequence of cutoff functions with $0 \leq \chi_n \leq 1$ on Λ and $\chi_n \to 1$ as $n \to \infty$. The existence of the sequence (χ_n) follows from the Urysohn's lemma [97, Lemma 2.12]. Let us construct a sequence $(r_n, a_n, b_n) = (-F^*(a,b)\chi_n, a\chi_n, b\chi_n) \in C_0(\Lambda, \mathcal{X})$ for some $(a,b) \in C_b(\Lambda, \mathbb{R}^d \times \mathbb{S}^d)$, then $(r_n, a_n, b_n) \to (-F^*(a,b), a, b) \in C_b(\Lambda, \mathcal{X})$ as $n \to \infty$. The finiteness of $F^*(a,b)$ is guaranteed by the coercivity of F. By the convexity of F^* and the fact that $F^*(O^{d\times 1}, O^{d\times d}) \leq 0$ where $O^{m\times n}$ denotes a null matrix of size $m \times n$, it is clear that (r_n, a_n, b_n) satisfies $r_n + F^*(a_n, b_n) \leq 0$. Since the supremum in the last line of (A.2) is taken over all $(r,a,b) \in C_0(\Lambda, \mathcal{X})$, we have

$$\begin{aligned}\Phi^*(\rho, \mathcal{A}, \mathcal{B}) &\geq \sup_{(a,b)\in C_b(\Lambda,\mathbb{R}^d\times\mathbb{S}^d)} \lim_{n\to\infty} \left\{ \int_\Lambda r_n \, d\tilde{\rho} + a_n \cdot d\tilde{\mathcal{A}} + b_n : d\tilde{\mathcal{B}} \right\} \\ &= \sup_{(a,b)\in C_b(\Lambda,\mathbb{R}^d\times\mathbb{S}^d)} \left\{ \int_\Lambda -F^*(a,b) \, d\tilde{\rho} + a \cdot d\tilde{\mathcal{A}} + b : d\tilde{\mathcal{B}} \right\} \\ &= \sup_{(r,a,b)\in C_b(\Lambda,\mathcal{X})} \left\{ \int_\Lambda r \, d\tilde{\rho} + a \cdot d\tilde{\mathcal{A}} + b : d\tilde{\mathcal{B}} \; ; \; r + F^*(a,b) \leq 0 \right\} \\ &= \Phi^*(\tilde{\rho}, \tilde{\mathcal{A}}, \tilde{\mathcal{B}}).\end{aligned}$$

The first equality above is justified by the dominated convergence theorem. The second equality above holds because if Φ^* is finite, then the supremum must occur at the boundary.

For Ψ^*, if we restrict its domain to $(\tilde{\rho}, \tilde{\mathcal{A}}, \tilde{\mathcal{B}}) \in \mathcal{M}(\Lambda, \mathcal{X})$, then $\Psi^* = 0$ if $(\tilde{\rho}, \tilde{\mathcal{A}}, \tilde{\mathcal{B}})$ satisfies (3.14) and (3.15) or $\Psi^* = +\infty$ otherwise. Recall that in K^o, $r = -\partial_t \phi - \sum_{i=1}^m \lambda_i G_i \delta_i$, $a = -\nabla_x \phi$ and $b = -\frac{1}{2}\nabla_x^2 \phi$. Whenever Ψ^* is finite, by (3.14) and (3.15), we have

$$\int_\Lambda r \, d\tilde{\rho} + a \cdot d\tilde{\mathcal{A}} + b : d\tilde{\mathcal{B}} - \int_{\mathbb{R}^d} \phi(0,x) \, d\mu_0 + \sum_{i=1}^m \lambda_i c_i = 0 \qquad \forall (r,a,b) \in K^o. \tag{A.3}$$

The equation (A.3) holds in particular for (r,a,b) in the subset $K^o \cap C_0(\Lambda, \mathcal{X})$. Also, since $K^o \cap C_0(\Lambda, \mathcal{X}) \subseteq K^o$, we have

$$\begin{aligned}\Psi^*(\rho, \mathcal{A}, \mathcal{B}) &= \sup_{(r,a,b)\in K^o} \left\{ \langle (r,a,b), (\rho, \mathcal{A}, \mathcal{B})\rangle - \int_{\mathbb{R}^d} \phi(0,x) \, d\mu_0 + \sum_{i=1}^m \lambda_i c_i \right\} \\ &\geq \sup_{(r,a,b)\in K^o\cap C_0(\Lambda,\mathcal{X})} \left\{ \langle (r,a,b), (\rho, \mathcal{A}, \mathcal{B})\rangle - \int_{\mathbb{R}^d} \phi(0,x) \, d\mu_0 + \sum_{i=1}^m \lambda_i c_i \right\} \\ &= \sup_{(r,a,b)\in K^o\cap C_0(\Lambda,\mathcal{X})} \left\{ \int_\Lambda r \, d\tilde{\rho} + a \cdot d\tilde{\mathcal{A}} + b : d\tilde{\mathcal{B}} - \int_{\mathbb{R}^d} \phi(0,x) \, d\mu_0 + \sum_{i=1}^m \lambda_i c_i \right\} \\ &= \sup_{(r,a,b)\in K^o} \left\{ \int_\Lambda r \, d\tilde{\rho} + a \cdot d\tilde{\mathcal{A}} + b : d\tilde{\mathcal{B}} - \int_{\mathbb{R}^d} \phi(0,x) \, d\mu_0 + \sum_{i=1}^m \lambda_i c_i \right\} \\ &= \Psi^*(\tilde{\rho}, \tilde{\mathcal{A}}, \tilde{\mathcal{B}}).\end{aligned}$$

Therefore,

$$\inf_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda,\mathcal{X})^*}(\Phi^*+\Psi^*)(\rho,\mathcal{A},\mathcal{B}) \geq \inf_{(\rho,\mathcal{A},\mathcal{B})\in \mathcal{M}(\Lambda,\mathcal{X})}(\Phi^*+\Psi^*)(\rho,\mathcal{A},\mathcal{B}).$$

This completes the proof. □

A.3 The LSV calibration algorithm with policy iteration

Algorithm 2: LSV calibration with policy iteration

Data: Market prices of European option
Result: A calibrated OT-LSV model that matches all market prices

Set an initial λ
do
  for $k = N_T - 1, \dots, 0$ **do**
    `/* Solving the HJB equation */`
    if t_{k+1} *is equal to the maturity of any calibrating options* **then**
      $\phi_{t_{k+1}} \leftarrow \phi_{t_{k+1}} + \sum_{i=1}^m \lambda_i G_i \mathbb{1}(t_{k+1} = \tau_i)$
    end
    `/* Policy iteration */`
    Let $\phi_{t_k}^{new} = \phi_{t_{k+1}}$
    do
      $\phi_{t_k}^{old} \leftarrow \phi_{t_k}^{new}$
      Approximate $\sigma_{t_k}^2$ by solving (3.26) with $\phi = \phi_{t_k}^{old}$
      Solve the HJB equation (3.25) by the ADI method at $t = t_k$, and set the solution to $\phi_{t_k}^{new}$
    while $\|\phi_{t_k}^{new} - \phi_{t_k}^{old}\|_2 > \epsilon_2$
    $\phi_{t_k} \leftarrow \phi_{t_k}^{new}$
  end
  `/* Calculating model prices and gradient */`
  Solve (3.28) to calculate the model prices by the ADI method
  Calculate the gradient $\nabla J(\lambda)$ by (3.27)
  Update λ by the L-BFGS algorithm
while $\|\nabla J(\lambda)\|_\infty > \epsilon_1$

A.4 FX options data

Maturity	Option type	Strike	Implied Vol	Maturity	Option type	Strike	Implied Vol
	Call	1.3006	0.0905		Call	1.4563	0.1069
	Call	1.2800	0.0898		Call	1.3627	0.1052
1m	Call	1.2578	0.0915	1Y	Call	1.2715	0.1118
	Put	1.2344	0.0966		Put	1.1701	0.1278
	Put	1.2110	0.1027		Put	1.0565	0.1491
	Call	1.3191	0.0897		Call	1.5691	0.1100
	Call	1.2901	0.0896		Call	1.4265	0.1096
2m	Call	1.2588	0.0933	2Y	Call	1.2889	0.1168
	Put	1.2243	0.1014		Put	1.1421	0.1328
	Put	1.1882	0.1109		Put	0.9863	0.1540
	Call	1.3355	0.0912		Call	1.6683	0.1109
	Call	1.2987	0.0908		Call	1.4860	0.1122
3m	Call	1.2598	0.0955	3Y	Call	1.3113	0.1200
	Put	1.2160	0.1058		Put	1.1308	0.1352
	Put	1.1684	0.1185		Put	0.9468	0.1547
	Call	1.3775	0.0960		Call	1.7507	0.1104
	Call	1.3213	0.0953		Call	1.5351	0.1127
6m	Call	1.2633	0.1013	4Y	Call	1.3306	0.1210
	Put	1.1973	0.1145		Put	1.1226	0.1365
	Put	1.1236	0.1316		Put	0.9152	0.1554
	Call	1.4068	0.1013		Call	1.8355	0.1111
	Call	1.3329	0.1005		Call	1.5835	0.1137
9m	Call	1.2583	0.1068	5Y	Call	1.3505	0.1220
	Put	1.1745	0.1215		Put	1.1180	0.1379
	Put	1.0805	0.1407		Put	0.8887	0.1571

Table A.1: The EUR/USD option data as of 23 August 2012. The spot price $S_0 = 1.257$ USD per EUR. At each maturity, the options correspond to 10-delta calls, 25-delta calls, 50-delta calls, 25-delta puts and 10-delta puts

Maturity	1m	2m	3m	6m	9m	1Y	2Y	3Y	4Y	5Y
Domestic yield	0.41	0.51	0.66	0.95	1.19	1.16	0.60	0.72	0.72	0.72
Foreign yield	0.04	0.11	0.23	0.47	1.62	0.64	0.03	0.03	0.03	0.03

Table A.2: The domestic and foreign yields (in %) as of 23 August 2012.

Appendix B

Appendix for Chapter 4

B.1 The convex conjugate F^*

Given $a \in \mathbb{R}^2, b \in \mathbb{S}^2$ and $\overline{\beta} \in \mathbb{S}^2$, define

$$\begin{aligned}
A &:= \overline{\beta}_{11} + \frac{1}{2}b_{11} - \frac{1}{4}a_1 - \frac{1}{4}a_2, \\
B &:= \overline{\beta}_{12} + \frac{1}{2}b_{12}, \\
C &:= \overline{\beta}_{22} + \frac{1}{2}b_{22}, \\
M &:= \begin{bmatrix} A & B \\ B & C \end{bmatrix}.
\end{aligned}$$

We also define

$$\begin{aligned}
x'_+ &:= \frac{A-C}{4} + \frac{A^2 - C^2}{4\sqrt{4B^2 + (A-C)^2}}, & x'_- &:= \frac{A-C}{4} - \frac{A^2 - C^2}{4\sqrt{4B^2 + (A-C)^2}}, \\
y'_+ &:= \frac{B}{2} + \frac{B(A+C)}{2\sqrt{4B^2 + (A-C)^2}}, & y'_- &:= \frac{B}{2} - \frac{B(A+C)}{2\sqrt{4B^2 + (A-C)^2}},
\end{aligned}$$

and define

$$\begin{aligned}
\lambda_+ &:= \begin{bmatrix} x'_+ + \sqrt{(x'_+)^2 + (y'_+)^2} & y'_+ \\ y'_+ & -x'_+ + \sqrt{(x'_+)^2 + (y'_+)^2} \end{bmatrix}, \\
\lambda_- &:= \begin{bmatrix} x'_- + \sqrt{(x'_-)^2 + (y'_-)^2} & y'_- \\ y'_- & -x'_- + \sqrt{(x'_-)^2 + (y'_-)^2} \end{bmatrix}.
\end{aligned}$$

Lemma B.1.1. *The convex conjugate of F is*

$$F^*(a,b) = (b_{11} - \frac{1}{2}a_1 - \frac{1}{2}a_2)\beta^*_{11} + 2b_{12}\beta^*_{12} + b_{22}\beta^*_{22} - \sum_{i,j=1}^{2}(\beta^*_{ij} - \overline{\beta}_{ij})^2,$$

where the values of β^* are determined as follows:

1. *If $M \in \mathbb{S}^2_+$, then $\beta^* = M$.*
2. *If $AC \geq B^2$ and $A + C < 0$, then β^* is the null matrix.*

3. *Otherwise,*

$$\beta^* = \underset{\beta\in\{\lambda_+,\lambda_-\}}{\arg\min}\,(\beta_{11}-A)^2+2(\beta_{12}-B)^2+(\beta_{22}-C)^2.$$

Proof. By definition, the convex conjugate of F is given by

$$\begin{aligned} F^*(a,b) &= \sup_{\beta\in\mathbb{S}^2_+}\{-\frac{1}{2}a_1\beta_{11}-\frac{1}{2}a_2\beta_{11}+b_{11}\beta_{11}+2b_{12}\beta_{12}+b_{22}\beta_{22}-\sum_{i,j=1}^{2}(\beta_{ij}-\overline{\beta}_{ij})^2\} \\ &= -\inf_{\beta\in\mathbb{S}^2_+}\{(\beta_{11}-A)^2+2(\beta_{12}-B)^2+(\beta_{22}-C)^2\}+(A^2-\overline{\beta}_{11}^2)+2(B^2-\overline{\beta}_{12}^2)+(C^2-\overline{\beta}_{22}^2). \end{aligned}$$

Finding the β that achieves the above infimum is equivalent to solving

$$(\beta_{11},\beta_{12},\beta_{22}) = \underset{(x,y,z)\in\mathbb{R}_{\geq 0}\times\mathbb{R}\times\mathbb{R}_{\geq 0}}{\arg\inf}\{(x-A)^2+2(y-B)^2+(z-C)^2 \mid xz\geq y^2\}. \tag{B.1}$$

In order to solve this problem, let us rotate the xyz-axes around y-axis clockwise through an angle of 45° into $x'y'z'$-axes, which can be described by the linear transformation:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

The inverse transformation is

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}.$$

In terms of (x',y',z'), the infimum in (B.1) can be reformulated as

$$\inf_{(x',y',z')\in W} 2(x'-\overline{x}')^2+2(y'-\overline{y}')^2+2(z'-\overline{z}')^2, \tag{B.2}$$

where $(\overline{x}',\overline{y}',\overline{z}') := (\frac{1}{2}A-\frac{1}{2}C, B, \frac{1}{2}A+\frac{1}{2}C)$, and W is a convex cone defined as

$$W = \{(x',y',z')\in\mathbb{R}^3 \mid z'\geq 0,\, x'^2+y'^2\leq z'^2\}.$$

In the $x'y'z'$-axes, the above problem can be simply described as finding the minimum Euclidean distance from the point $(\overline{x}',\overline{y}',\overline{z}')$ to W. There are three cases:

(a) If $(\overline{x}',\overline{y}',\overline{z}')\in W$, the solution is $(x',y',z')=(\overline{x}',\overline{y}',\overline{z}')$.

(b) If $\overline{x}'^2+\overline{y}'^2\leq\overline{z}'^2$, but $\overline{z}'<0$. Then the solution should be on the boundary $z'=0$, which also implies that $x'=y'=0$.

(c) Otherwise, the solution must be on the boundary of W:

$$\partial W = \{(x',y',z')\in\mathbb{R}^3 \mid z'\geq 0,\, x'^2+y'^2= z'^2\}.$$

By substituting $z'=\sqrt{x'^2+y'^2}$ into (B.2) and solving the infimum, we find two stationary points:

$$(x'_+,y'_+,z'_+) = \left(\frac{\overline{x}'}{2}+\frac{\overline{x}'\overline{z}'}{2\sqrt{\overline{x}'^2+\overline{y}'^2}}, \frac{\overline{y}'}{2}+\frac{\overline{y}'\overline{z}'}{2\sqrt{\overline{x}'^2+\overline{y}'^2}}, \sqrt{(x'_+)^2+(y'_+)^2}\right),$$

$$(x'_-,y'_-,z'_-) = \left(\frac{\overline{x}'}{2}-\frac{\overline{x}'\overline{z}'}{2\sqrt{\overline{x}'^2+\overline{y}'^2}}, \frac{\overline{y}'}{2}-\frac{\overline{y}'\overline{z}'}{2\sqrt{\overline{x}'^2+\overline{y}'^2}}, \sqrt{(x'_-)^2+(y'_-)^2}\right).$$

One of the stationary points achieves the infimum. Thus, we choose the one with the smaller objective value.

Transforming the above solutions back to the xyz-axes through the inverse transformation and replacing (x, y, z) by $(\beta_{11}, \beta_{12}, \beta_{22})$, we obtain the desired result. □

B.2 The joint calibration algorithm

Let $\pi^N := \{t_k : 0 \le k \le N\}$ be a discretisation of $[0, T]$ such that $0 = t^0 < t^1 < \ldots < t^N = T$. We assume that each of t_0 and $\tau_i, i = 1, \ldots, m$ coincides with some value in π^N. Denote by ϵ_1 the tolerance of the maximum of the gradients (4.19)–(4.22), and denote by ϵ_2 the tolerance for the policy iteration. Recall that ϵ_1 has an alternative interpretation as the tolerance of the maximum error between the calibrating prices and the model prices. In the numerical example presented in Section 4.5, $\epsilon_1 = 10^{-4}$ and $\epsilon_2 = 10^{-8}$. The numerical method described in Section 4.4 is summarised as the following algorithm.

Algorithm 3: The joint calibration algorithm

Set an initial $(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^{\xi})$
do
  `/* Solving the HJB equation */`
  for $k = N-1, \ldots, 0$ **do**
    `/* Terminal conditions */`
    if $\exists i = 1, \ldots, m,\ t_{k+1} = \tau_i$ **then**
      $\phi_{t_{k+1}} \leftarrow \phi_{t_{k+1}} + \sum_{i=1}^{m} \lambda_i^{SPX} G_i \mathbb{1}(t_{k+1} = \tau_i)$ `// SPX options`
    end
    if $t^{k+1} = t_0$ **then**
      $\phi_{t^{k+1}} \leftarrow \phi_{t^{k+1}} + \lambda^{VIX,f} J$ `// VIX futures`
      $\phi_{t^{k+1}} \leftarrow \phi_{t^{k+1}} + \sum_{i=1}^{n} \lambda_i^{VIX} (H_i \circ J)$ `// VIX options`
    end
    if $t^{k+1} = T$ **then**
      $\phi_{t^{k+1}} \leftarrow \phi_{t^{k+1}} + \lambda^{\xi} \xi$ `// Singular contract`
    end
    `/* Policy iteration */`
    $\phi_{t_k}^{new} \leftarrow \phi_{t_{k+1}}$
    do
      $\phi_{t_k}^{old} \leftarrow \phi_{t_k}^{new}$
      Approximate β^* by Lemma B.1.1 with $\phi_{t_k}^{old}$
      Solve the HJB equation (4.24) or (4.25) with β^* as a linearised PDE by the standard implicit finite difference method, and set the solution as $\phi_{t_k}^{new}$
    while $\|\phi_{t_k}^{new} - \phi_{t_k}^{old}\|_\infty > \epsilon_2$
    $\phi_{t_k} \leftarrow \phi_{t_k}^{new}$
  end
  `/* Model prices and gradients */`
  Calculate the model prices by solving equations (4.23) by the ADI method
  Calculate the gradients (4.19) to (4.22)
  Update $(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^{\xi})$ by the L-BFGS algorithm
while *The maximum of the gradients (4.19) to (4.22) is greater than* ϵ_1

B.3 The diffusion process β for the simulated data example

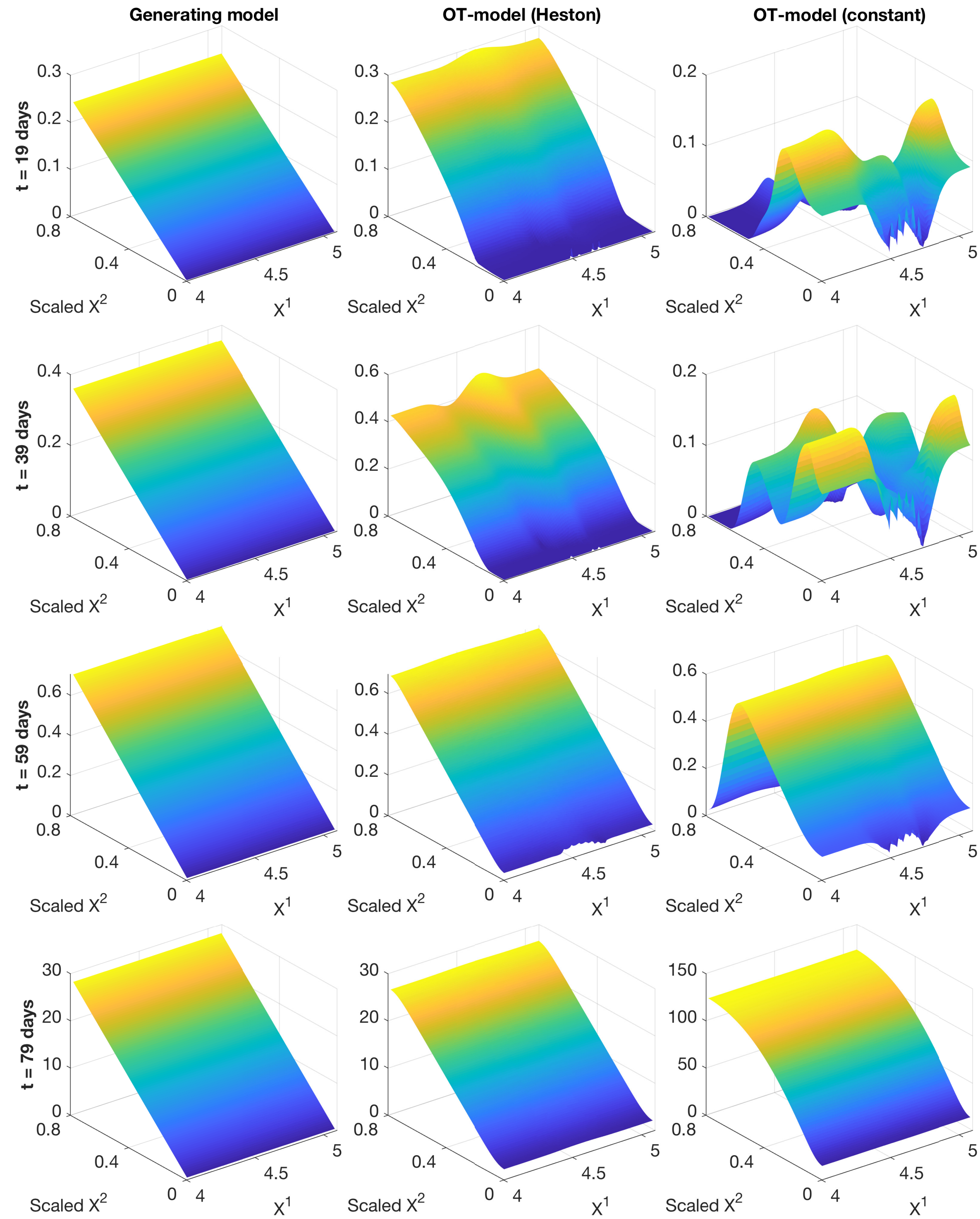

Figure B.1: The functions $\beta_{11}(t, X^1, X^2)$ of the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference for the simulated data example.

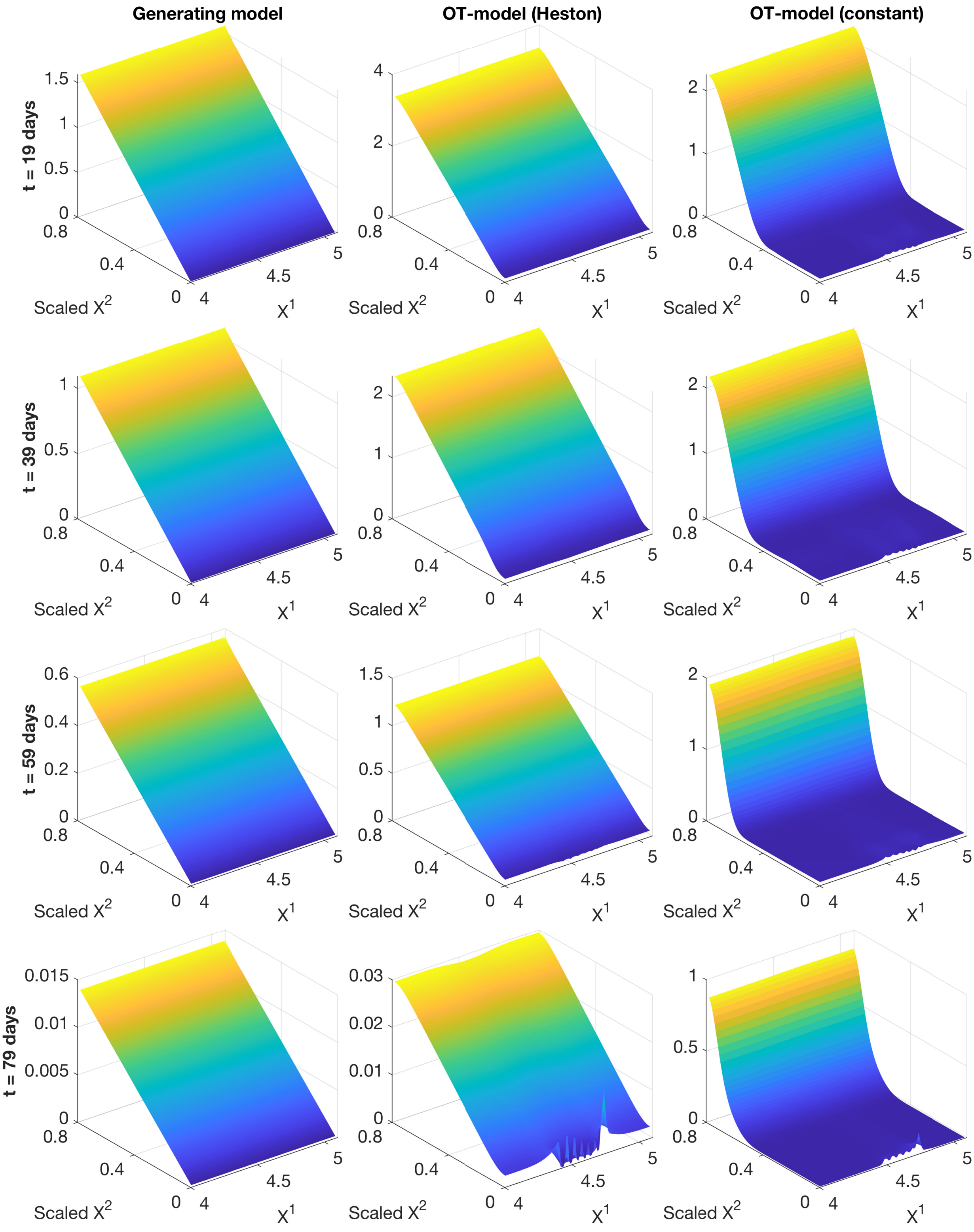

Figure B.2: The functions $\beta_{22}(t, X^1, X^2)$ of the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference for the simulated data example.

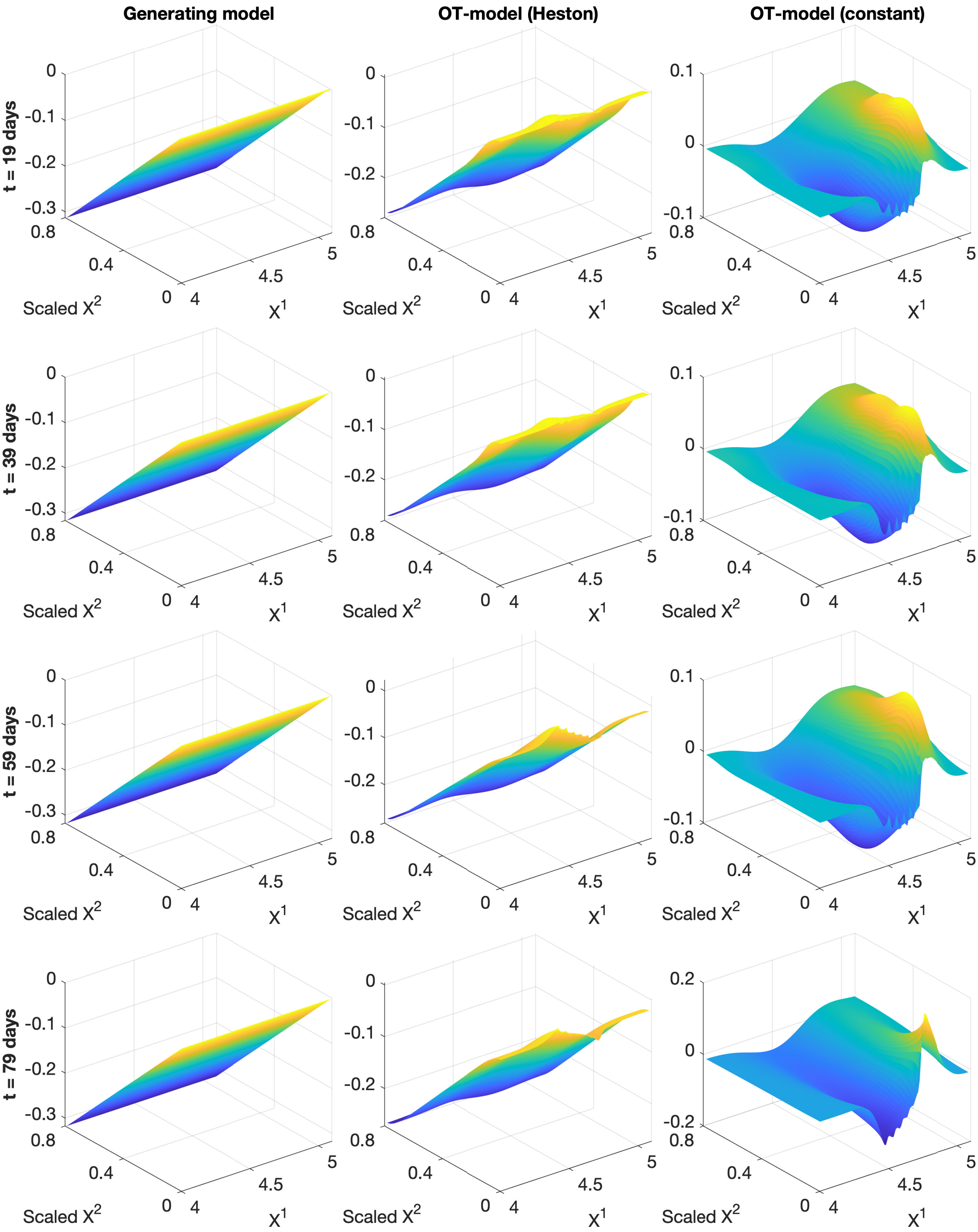

Figure B.3: The functions $\beta_{12}(t, X^1, X^2)$ of the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference for the simulated data example.

B.4 The diffusion process β for the market data example

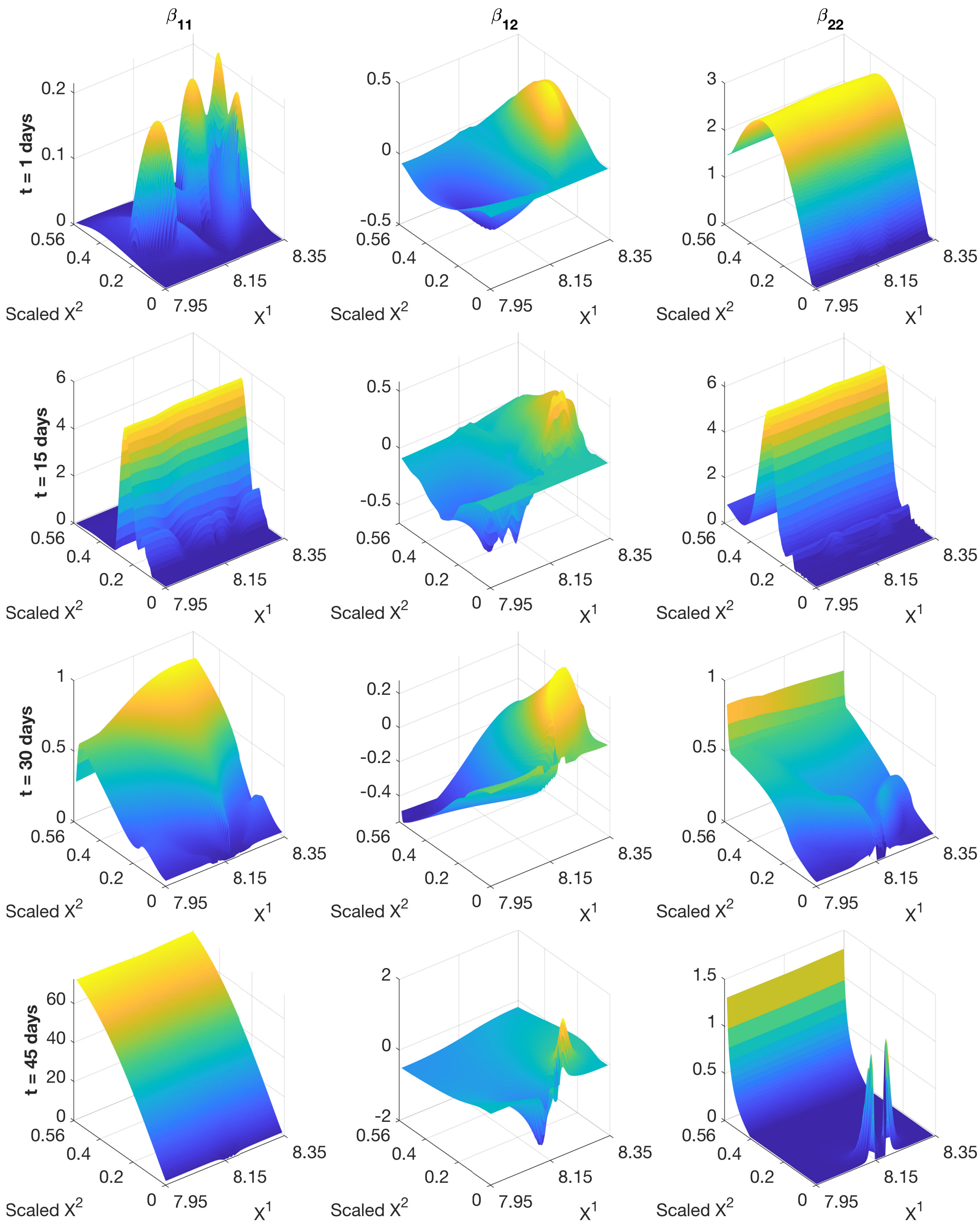

Figure B.4: The functions $\beta_{11}(t, X^1, X^2)$, $\beta_{12}(t, X^1, X^2)$ and $\beta_{22}(t, X^1, X^2)$ of the OT-calibrated model for the market data example.