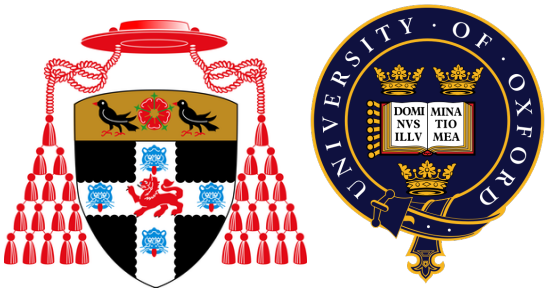
Model Calibration with Optimal Transport: Stochastic Interest Rates and Bass Interpolation



Benjamin Joseph
Christ Church
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2024

This thesis is dedicated to my mother, Francesca Joseph, for her unconditional support throughout my mathematical career.

Acknowledgements

First and foremost, I would like to thank my supervisors Grégoire Loeper and Jan Obłój for all of their support throughout the last four years. I would not have achieved anywhere near as much during my studies without your insightful guidance and rigorous sanity checks with my work. It has been a pleasure to work with you both over the last four years, and I am very grateful for the opportunity you gave me to be your student.

Thank you also to the CDT Random Systems and BNP Paribas for funding and organising this incredible opportunity to be a D.Phil student at Oxford. I have especially enjoyed the opportunities to go to conferences as part of my postgraduate studies. Thank you also to Yifan Jiang, Fang Rui Lim, and Vlad Tuchilus for all of the interesting talks during the working group.

Thank you to everyone involved in Christ Church boat club: athletes, coxes, and coaches. The community spirit fostered there has been fantastic, and it's been a uniquely wonderful experience to be a part of such a big club. While it was difficult at times to keep the balance between mathematical work, training, and committee roles in the club (particularly during Torpids or Summer VIIIs), I'm glad I took such an active part in the club. Special thanks to Harriet Anderson, Anna Betteridge, Ayman D'Souza, Kate Friesen, India Marsden, Lawrence Tatiqaj, and Jake Williams — it's been a pleasure to work alongside you all running the club. I'll genuinely miss the training, regattas, and dinners, and am very glad that I was convinced by Mike Walmsley to go to the taster session in October 2020.

Thank you to my partner, Victoria Walls, for putting up with my erratic hours while studying, and listening to far too much of me rambling on about maths and rowing. I'm still astonished you sat through many hours of me talking through various proofs or numerical bugs I was stuck on. I would not have completed my D.Phil without your emotional support and understanding throughout.

Thank you to my friends in Christ Church and the boat club: Netta Claydon, Eimear Conroy, Matthew Cooke, Sean Cooney, Delaney Dominey-Foy, Linnea Drexhage, Connor Ellis, Chris Goble, Sarah Guggemos, Jan-Hendrik Höhnk, Alex Illsley, Federico Nebuloni, Jonny Puleston, Evan Roberts, Aisling Roper, John Ryan, Tavish Traut, Marin Vukšić, Milena Vuletić, Arthur Wotton, Peter Watson, and Tom Yildirim for keeping me sane throughout my degree and making me feel welcome at Christ Church and the boat club. I'll cherish the memories we made together during my four years at Christ Church.

Thank you to my friends from LMH: Rafi Kelion, Rory Kenney-Herbert, Will Klemperer, Amy Lyons, Eve Shenkman, Sapphire Shoferpoor, Maddison Sumner, Owen Sweeney, Laura Willsmore, Joe Windo, and Seb Ziman for making me feel at home in Oxford. Thank you also to Dmitry Belyaev, Allesandra Caraceni, Paul Chleboun, Rolf Suabedissen, and Dominic Yeo for supporting my D.Phil application.

Thank you to my godparents, Jeanmarie and Jerry Hagedorn, for all of your support. I would not be able to have taken the opportunities I have done during my time at Oxford without your backing. Thank you also to Ed, Ian, Liv, and Ness Whitehouse for always being there for me.

Finally, thank you to my family: Simon Colley, Francesca and Matilda Joseph, and Brian and Stella Wray, for always having my back and supporting me throughout everything. This thesis wouldn't exist without your unconditional support and I can't put into words how grateful I am for that.

Abstract

We introduce a non-parametric, optimal transport driven method of calibrating a two dimensional log-stock and stochastic interest rate model. The method finds a fully calibrated model which is the closest (in a way that can be defined by a general cost function) to a given reference model. We first use the Markovian projection results of Gyöngy, 1986; Brunick and Shreve, 2013 to reduce to models that are Markov in the state variables and stochastic discount factor. Our main theoretical contribution is a novel conditioning argument to work with so-called "discounted densities", or sub-probability measures. We prove that the path-dependent discount factor can be eliminated without any loss of generality in the calibration problem, and thus we reduce the dimension of the problem to that of the state variables. We then derive a generic duality result to generate a numerically tractable dual problem. We first develop a "sequential" calibration approach, where the interest rate is assumed to be calibrated first and frozen, and calibrate the local volatility to European call options. We also develop a joint calibration approach where the dynamics of the interest rate are unknown, and jointly calibrate the two dimensional system to European call options and European cap options, and test on simulated and market data. We trial a different cost function and a faster approach with the joint calibration. We also develop another sequential method, called the "full sequential" calibration with relaxed assumptions on the correlation and compare all three approaches in different circumstances.

We then contribute to the recent studies of the so-called Bass martingale. Backhoff-Veraguas, Beiglböck, Huesmann, et al., 2020 showed it is the solution to the martingale Benamou-Brenier (mBB) problem, i.e., among all martingales with prescribed initial and terminal distributions it is the one closest to the Brownian motion. We link it with semi-martingale optimal transport and deduce an alternative way to derive the dual formulation recently obtained in Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023. We then consider computational methods to compute the Bass martingale. The dual formulation of the transport problem leads to an iterative scheme that mirrors to the celebrated Sinkhorn algorithm for entropic optimal transport. We call it the *measure preserving martingale Sinkhorn* (MPMS) algorithm. We prove that in any dimension, each step of the algorithm improves the value of the dual problem. Specifically, we provide two separate implementations. The first being equivalent to the fixed-point method of Conze and Henry-Labordere, 2021, studied in Acciaio, Marini, and Pammer, 2023, and performs very well on a range of examples, including real market data. The second approach uses an approximate method on one of the iterations and also rapidly solves the numerical problem in the same examples. We finally demonstrate the MPMS algorithm in the Geometric Bass setting of Backhoff-Veraguas, Loeper, and Obłój, 2024 and compare it with the standard Bass martingale in a calibration setting, along with comparing the local volatility surfaces. This is tested in a simulated Black-Scholes setting and a CEV setting on simulated data, and in a multimarginal market data setting with a large number of strikes and maturities.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A challenging problem in mathematical finance is the pricing of complex over-the-counter derivatives, such as auto-callable yield notes or other structured products. When a practitioner is given a model, this is relatively easy to do, with Monte Carlo methods often obtaining sufficiently low variance to give an accurate price. In reality, the market does not offer a universal model of a stock; therefore the problem becomes how to build a model from market prices, or how to calibrate a model to the market. One can formulate two classes of calibration problems: parametric, where a formula for the volatility (and drift) is specified in terms of unknown parameters; and non-parametric, where no assumptions are made on the functional form of the characteristics. While parametric methods have their advantages, such as interpretability of the dynamics of the system and computing Greeks for hedging purposes, the existence of an optimal model is not always guaranteed as seen in the elusive attempts at joint parametric calibration of SPX and VIX — see Guyon, 2020 for the literature review and discussion. Another issue is the core assumption that the dynamics are constrained to follow a particular model, which is impossible in general to guarantee. Our model-free, non-parametric approach avoids both of these drawbacks. We recast the calibration problem within the framework of semimartingale optimal transport (SOT), and provided that an optimal probability measure consistent with the observed data exists, our methodology will be able to find it. The output of the calibration methods developed in this thesis is a continuous time local volatility model that matches the observed market data, and as will be seen in the simulated examples also accurately interpolates the implied

volatilties between the strikes.

We now briefly summarise the content of the thesis. The rest of Chapter 1 serves as a primer on the theoretical techniques used in Chapter 2 and Chapter 4. In particular, we review martingale optimal transport (MOT) in detail, providing a background on the so-called martingale Benamou-Brenier formulation of Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023 and its connections with the Bass martingale. The duality result of Huesmann and Trevisan, 2019 is reviewed as it provides the basis of the central result duality of Chapter 2, along with well-posedness results of Trevisan, 2016. Additionally, we review the mimicking results of Gyöngy, 1986; Brunick and Shreve, 2013, which are crucial in allowing us to restrict our attention to Markov processes without loss of generality (in a sense determined in Chapter 2), in turn enabling us to take advantage of a PDE formulation of the calibration problem. We also review other techniques of model calibration, both parametric and non-parametric, to put our methodology in context.

Chapter 2 and Chapter 3 cover the non-parametric calibration of a hybrid equities-rates model to European options. Chapter 2 shows the theoretical results, and a novel treatment of the stochastic discount factor to further reduce the dimension of the problem after the mimicking results are applied. We adapt the proof of Trevisan, 2016, Theorem 2.5 to achieve this, and then prove a duality result via the standard techniques used in Huesmann and Trevisan, 2019. Chapter 3 outlines the numerical approach we use in our two dimensional calibration problem and applies it in different contexts. One being a sequential calibration, where the stochastic rate is a priori known, and another a joint calibration where the dynamics of both are unknown. We provide a way of accelerating the joint calibration and obtaining better volatility surfaces through bounds similar in spirit to those in the uncertain volatility model Avellaneda, Friedman, et al., 1997. From this faster joint calibration method, we obtain another method in-between which is sequential in nature but with relaxed assumptions on the correlation. We additionally compare the performance of all three methods.

In Chapter 4, we construct an algorithm to solve the Bass martingale problem. This allows us to express a martingale as a function of Brownian motion, where the function solves the heat equation. This in turn results in a natural interpolation of two discrete measures from which we can construct a continuous time model. We show different implementations of

the measure preserving martingale Sinkhorn (MPMS) algorithm in a variety of cases. We also demonstrate how to construct local volatility surfaces under both the Bass and Geometric Bass martingales, and verify the accuracy of the model prices and implied volatilities.

In Chapter 5 we summarise the findings of the thesis and provide avenues for further research, in particular two methods to extend the numerical results of Chapter 3 to higher dimensions, and a method to extend the MPMS algorithm to higher dimensions.

1.1 Optimal Transport

Optimal transport was first studied in Monge, 1781, famously to minimise the effort in moving piles of sand on a construction site. Reformulated in Tolstoi, 1930, the modern mathematical problem is to search for a map pushing one density onto another while minimising some cost function. We use the notation of Santambrogio, 2015, and also refer to Villani, 2003; Villani, 2009 as seminal works for the theory. Here we look at some of the key results from static optimal transport, and then turn our attention to the dynamical formulation of Benamou and Brenier, 2000. We start with the formulation of the Monge problem.

Problem 1.1.1 (Monge Problem, Monge, 1781)**.** *Let X and Y be Polish spaces, and $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ the set of probability measures on X and Y. Then, for a given cost function $c : X \times Y \to [0, \infty]$ and probability measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, the Monge Problem is given by*

$$\inf_{T} \left\{ \int_X c(x, T(x)) \, \mathrm{d}\mu \ : \ T_{\#}\mu = \nu \right\}. \tag{1.1.1}$$

If the infimum is attained for some $T^* : X \to Y$, then such a map is called an optimal transport map.

Since the constraint $T_{\#}\mu = \nu$ in (1.1.1) is not closed under weak convergence, the relaxation of Problem 1.1.1 formulated in Kantorovich, 1942 is the version widely used.

Problem 1.1.2 (Kantorovich Problem, Kantorovich, 1942)**.** *In the same setting as Prob-*

lem 1.1.1, the Kantorovich Problem is given by

$$\inf_{\pi}\left\{\int_{X\times Y} c(x,y)\,\mathrm{d}\pi \;:\; \pi\in\Pi(\mu,\nu)\right\}, \tag{1.1.2}$$

Here, $\Pi(\mu,\nu)$ is the set of probability measures on $X\times Y$ with x-marginal μ and y-marginal ν.

Kantorovich then directly proved the duality result in Kantorovich, 1942, notably prior to the development of general infinite dimensional duality results.

Problem 1.1.3 (Kantorovich Dual Problem, Kantorovich, 1942)**.** *Let $C_b(X)$ and $C_b(Y)$ be the set of continuous and bounded functions on X and Y. Then, for a given cost function $c: X\times Y\to[0,\infty)$ and probability measures $\mu\in\mathcal{P}(X)$, $\nu\in\mathcal{P}(Y)$, the Kantorovich Dual Problem is given by*

$$\max_{\varphi\in C_b(X),\psi\in C_b(Y)}\left\{\int_X \varphi\,\mathrm{d}\mu+\int_Y \psi\,\mathrm{d}\nu \;:\varphi(x)+\psi(y)\leqslant c(x,y)\right\}. \tag{1.1.3}$$

We note that under certain conditions on the cost function and spaces, we in fact have $\psi(y)=\inf_{x\in X}\{c(x,y)-\varphi(x)\}$. Additionally, with different assumptions, Problem 1.1.1 and Problem 1.1.2 can share a solution - see Gangbo and McCann, 1996; Levin, 1999; Caffarelli, 2017. Moreover, Brenier's Theorem (see Brenier, 1987; Brenier, 1991), gives equivalence between Problem 1.1.1 and Problem 1.1.2 under a quadratic cost function. As a consequence, the transportation plan is given by $\pi=(\mathrm{Id},T)_{\#}\mu$ where $T=\nabla\varphi$ for a convex function φ solving the Monge-Ampère equation

$$\det(\nabla^2\varphi)\tilde{\nu}(\nabla\varphi)=\tilde{\mu}, \tag{1.1.4}$$

where $\tilde{\mu}$ and $\tilde{\nu}$ are the densities of μ and ν respectively. We refer to Caffarelli, 1996; Gutiérrez and Brezis, 2001; X.-N. Ma, Trudinger, and X.-J. Wang, 2005; Loeper, 2009; De Philippis and Figalli, 2014 for an overview of the study of the Monge-Ampère equation. (1.1.4) is a fully nonlinear, degenerate elliptic PDE and numerically solving it presents its own challenges: see Benamou, Froese, and Oberman, 2010; Benamou, Froese, and Oberman, 2014.

A common approach is to consider the entropy regularised version of Problem 1.1.2, with Berman, 2020 providing a construction for the Schrödinger potentials that converge to the solution of (1.1.4); see also Benamou, Carlier, et al., 2015; Nutz and Wiesel, 2022; Ghosal, Nutz, and Bernton, 2022; Ghosal and Nutz, 2022 for discussion on the convergence and proporties of the Schrödinger potentials and Sinkhorn algorithm. When the distributions are discrete, this approach has applications in computer vision by efficiently computing the Earth Mover's distance of Rubner, Tomasi, and Guibas, 1998; Rubner, Tomasi, and Guibas, 2000 via the Sinkhorn-Knopp algorithm presented in Cuturi, 2013. We also point to the extensions in a multimarginal setting where Problem 1.1.2 is optimised over $\pi \in \Pi(\mu_1, \dots, \mu_n)$ which has been studied in Gangbo and Święch, 1998; Carlier, 2003 and conditions for a Monge solution to the multimarginal problem given in Kim and Pass, 2014.

We have thus far introduced the notions of discrete time optimal transport, however for applications in continuous time model calibration, we will be interested also in what happens in-between the initial and final law. We mention McCann, 1997 as the seminal work wherein the displacement interpolant $\rho_t := [(1-t)\mathrm{Id} + t\nabla\varphi]_{\#}\mu$ is studied. We consider the so-called dynamical formulation introduced in Benamou and Brenier, 2000. We take a continuous, time dependent probability measure $\rho \in C([0,1]; \mathcal{P}(\mathbb{R}^d))$ and take our marginals to be $\rho_0(\cdot) = \rho(0,\cdot)$ and $\rho_1(\cdot) = \rho(1,\cdot)$, and consider the L^2 Monge-Kantorovich problem.

Theorem 1.1.4 (Benamou and Brenier, 2000, Proposition 1.1)**.** *The square of the L^2 Kantorovich distance between two probability measures ρ_0 and ρ_1 on $\mathbb{R}^d$ is given by*

$$\inf_{T:T_{\#}\rho_0=\rho_1} \int_{\mathbb{R}^d} ||T(x)-x||^2 \rho_0(x)\,\mathrm{d}x = \inf_{\rho,v} \int_{\mathbb{R}^d}\int_0^1 \rho(t,x)|v(t,x)|^2\,\mathrm{d}x\,\mathrm{d}t, \tag{1.1.5}$$

where the infimum on the right of (1.1.5) *is taken among all ρ and v satisfying the continuity equation with initial and terminal conditions:*

$$\partial_t \rho + \nabla_x \cdot (\rho v) = 0, \quad (t,x) \in [0,1]\times\mathbb{R}^d, \tag{1.1.6}$$

$$\rho(0,\cdot) = \rho_0(\cdot), \quad \rho(1,\cdot) = \rho_1(\cdot). \tag{1.1.7}$$

Benamou and Brenier then develop a numerical scheme to solve their dual formulation

using the ADMM algorithm of Fortin and Glowinski, 1983, and this duality and numerical approach has been adapted via the inclusion of discrete constraints to local volatility model calibration in I. Guo, Loeper, and Shiyi Wang, 2019.

1.2 Martingale and Semimartingale Optimal Transport

The Benamou-Brenier theorem gives us a time-continuous formulation of optimal transport; a natural extension of this is to add the constraint that the time continuous measures must be the law of a martingale. Martingale optimal transport (MOT) was introduced in the discrete time setting in Beiglböck, Henry-Labordère, and Penkner, 2013, and in the continuous time setting in Galichon, Henry-Labordère, and Touzi, 2014; the first dual formulation of MOT in continuous time was obtained in Xiaolu Tan and Touzi, 2013 using stochastic control theory in a similar vein to Mikami and Thieullen, 2006. The later work of Huesmann and Trevisan, 2019 provides a PDE formulation and duality approch akin to Benamou and Brenier, 2000, in particular avoiding the dynamic programming principle approach of Xiaolu Tan and Touzi, 2013 or the notion of viscosity solutions to close the duality gap. Additionally, I. Guo, Loeper, and Shiyi Wang, 2022 incorporates discrete constraints into the duality result of Huesmann and Trevisan, 2019, and I. Guo and Loeper, 2021 further extends the duality theory to a path-dependent setting.

In terms of the theory discussed in the previous section, a Monge-Kantorovich duality result for MOT was studied in Beiglböck, Nutz, and Touzi, 2016, a martingale version of Brenier's theorem has been proposed in Beiglböck and Juillet, 2016; Henry-Labordère and Touzi, 2016 and a multimarginal martingale formulation has been established in Henry-Labordère, Obłój, et al., 2016; Lim, 2016; Nutz, Stebegg, and Xiaowei Tan, 2020. Numerical methods to solve the MOT problem have been studied via a discretisation of the marginals in G. Guo and Obłój, 2019, via a penalisation of the dual problem and then a neural networks approach in Eckstein and Kupper, 2021, via a time discretisation of associated entropy regularised problems in Benamou, Chazareix, and Loeper, 2024, and via direct numerical solution of the dual problem in I. Guo, Loeper, and Shiyi Wang, 2022; I. Guo, Loeper, Obłój, et al., 2022a; stability properties of MOT were a topic of study in Backhoff-Veraguas and

Pammer, 2022; Beiglböck, Jourdain, et al., 2023; Wiesel, 2023.

In addition, MOT has connections with the Skorokhod embedding problem (see Obłój, 2004), within the recent works of Beiglböck, A. Cox, and Huesmann, 2017; Källblad, Xiaolu Tan, and Touzi, 2017; Beiglböck, A. Cox, and Huesmann, 2020. MOT has a wide variety of applications in robust finance, such as the robust hedging problem introduced by Hobson, 1998, see for example Hobson and Neuberger, 2012; Dolinsky and Soner, 2014; Galichon, Henry-Labordère, and Touzi, 2014; in the context of robust pricing as well, see Campi, Laachir, and Martini, 2017; Hou and Obłój, 2018; Bartl et al., 2019; Eckstein, G. Guo, et al., 2021; Doldi and Frittelli, 2023; Engstrom, Kallblad, and Karlsson, 2024. There are, of course, a wide range of applications in model calibration with optimal transport, however we defer the discussion to the calibration review in Section 1.5.

1.2.1 Martingale Benamou-Brenier and the Bass Martingale

We now review the martingale Benamou-Brenier (mBB) formulation given in Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023 and its connections to the Bass martingale of Bass, 1983. We assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ are in convex order, i.e. that for any convex function $f : \mathbb{R}^d \to \mathbb{R}$, we have $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu \leqslant \int_{\mathbb{R}^d} f \, \mathrm{d}\nu$; note that as a consequence of Strassen, 1965, we have the existence of a martingale coupling between μ and ν. Moreover, we assume that (μ, ν) is irreducible in the sense that for any $A, B \subset \mathbb{R}^d$ such that $\mu(A) > 0, \nu(B) > 0$, there exists a martingale $(X_t)_{t\in[0,1]}$ such that $X_0 \sim \mu$, $X_1 \sim \nu$ and $\mathbb{P}[X_0 \in A, X_1 \in B] > 0$. For financial applications in one dimension, where μ and ν are supported on the positive reals, by Beiglböck and Juillet, 2016, Theorem A.4, this condition can equivalently be stated in strict separation of call options, i.e. $\mathbb{E}[(X_0 - K)^+] < \mathbb{E}[(X_1 - K)^+]$ for all in the interior of the convex hull of the support of ν. Now we can state the martingale Benamou-Brenier problem.

Problem 1.2.1 (Martingale Benamou-Brenier, Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023, Section 1.2)**.** *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be in convex order, then the Martingale*

Benamou-Brenier problem is given by

$$\mathrm{MT}(\mu,\nu) = \inf_{\substack{M_0\sim\mu, M_1\sim\nu,\\ M_t = M_0 + \int_0^t \sigma_s \,\mathrm{d}B_s}} \mathbb{E}\left[\int_0^1 |\sigma_t - \mathrm{I}_d|^2_{\mathrm{HS}}\,\mathrm{d}t\right], \tag{1.2.1}$$

where $|\cdot|_{HS}$ *is the Hilbert-Schmidt norm, given by* $|M|_{HS} = \sqrt{\mathrm{tr}(A^\intercal A)}$ *for* $A \in \mathbb{R}^{d\times d}$ *and* I_d *is the* $d \times d$ *identity matrix.*

Backhoff-Veraguas, Beiglböck, Huesmann, et al., 2020, Theorem 1.5 shows that Problem 1.2.1 has a unique in law optimiser, which is referred to as the stretched Brownian motion. This terminology arises since Backhoff-Veraguas, Beiglböck, Huesmann, et al., 2020, Proposition 6.4 established that the optimiser is in fact the closest to Brownian motion in an adapted Wasserstein sense (see Lassalle, 2018).

Definition 1.2.2 (Bass martingale, Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023, Definition 1.1)**.** *Let* $(B_t)_{t\in[0,1]}$ *be a Brownian motion on* $\mathbb{R}^d$*, with* $B_0 \sim \alpha \in \mathcal{P}_2(\mathbb{R}^d)$*, and* $v : \mathbb{R}^d \to \mathbb{R}$ *be convex with* $\nabla v(B_1)$ *square-integrable, then the following martingale is the Bass martingale:*

$$M_t \coloneqq \mathbb{E}[\nabla v(B_1)|\sigma(B_s : 0 \leqslant s \leqslant t)] = \mathbb{E}[\nabla v(B_1)|B_t]. \tag{1.2.2}$$

We say that $(M_t)_{t\in[0,1]}$ *is a Bass martingale connecting* $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ *if* $M_0 \sim \mu$ *and* $M_1 \sim \nu$*.*

Remark 1.2.3. *If* (μ,ν) *are irreducible then Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023, Theorem 1.3 shows that the stretched Brownian motion is the Bass martingale. As the cost function in Problem 1.2.1 is quadratic, and the law of* B_1 *is given by* $\alpha * \mathbf{R}_1^d$ *where* $\mathbf{R}_t^d = (2\pi t)^{-d/2}\exp(-|x|^2/2t)$ *is the d-dimensional heat kernel, then in view of Brenier's Theorem (Brenier, 1987; Brenier, 1991), we have that* v *is given by the solution of* (1.1.4) *with marginals* $(\nu, \alpha * \mathbf{R}_1^d)$*, and moreover in the one dimensional case that* $\partial v = G_\nu^{-1} \circ G_{\alpha * \mathbf{R}_1^1}$*.*

The connection of the Bass martingale to martingale optimal transport has generated a large amount of recent interest. Backhoff-Veraguas, Schachermayer, and Tschiderer, 2023 characterised Problem 1.2.1 in terms of the initial condition of the Bass martingale, and

Backhoff-Veraguas, Pammer, and Schachermayer, 2024 studied Problem 1.2.1 from a gradient flow perspective. Conze and Henry-Labordere, 2021 developed a fast algorithm to construct the Bass martingale using a fixed point iteration method, with the convergence being analysed in Acciaio, Marini, and Pammer, 2023. We mention a different approach: the so-called Geometric Bass Martingale introduced in Backhoff-Veraguas, Loeper, and Obłój, 2024 with a PDE and probabilistic construction, and Beiglböck, Pammer, and Riess, 2024 in a weak optimal transport setting. Other generalisations include Tschiderer, 2024 where the driving random variable is no longer assumed to be Gaussian, but follow any $q \in \mathcal{P}_2(\mathbb{R}^d)$, and Schachermayer and Tschiderer, 2024 where (μ, ν) is no longer assumed to be irreducible.

1.2.2 Fokker-Planck Equation and SOT Duality

We review the dualitly result of Huesmann and Trevisan, 2019 in more detail, along with the well-posedness results and superposition principle of Trevisan, 2016 (itself using the techniques of Ambrosio and Trevisan, 2014), since the theoretical results of Chapter 2 are adaptations of these works. We first define the notion of weak solution to the Fokker-Planck equation that will be implicitly used throughout Chapter 2.

Definition 1.2.4 (Diffusion Operator, Trevisan, 2016, Definition 2.1)**.** *Define $\mathcal{L} = \mathcal{L}(a, b)$ with $\mathcal{L} : C_b^{1,2}((0,T) \times \mathbb{R}^d; \mathbb{R}) \to \mathcal{B}((0,T) \times \mathbb{R}^d; \mathbb{R})$ to be the linear differential operator defined by*

$$f \mapsto \mathcal{L}f := a \cdot \nabla_x f + \frac{1}{2} b : \nabla_x^2 f. \tag{1.2.3}$$

Where $a : (0,T) \times \mathbb{R}^d \to \mathbb{R}^d$, $b : (0,T) \times \mathbb{R}^d \to \mathbb{S}_+^d$ are Borel measurable. Further, write $\mathcal{L}_t f = (\mathcal{L}f)_t$ for $t \in (0,T)$.

Definition 1.2.5 (Narrow continuity, Trevisan, 2016, Section 2.1)**.** *Let $\nu = (\nu_t)_{t \in [0,T]} \subset \mathcal{M}(\mathbb{R}^d)$, then ν is a Borel curve if for every Borel $A \subset \mathbb{R}^d$, the curve $t \mapsto \nu_t(A)$ is Borel. Moreover, ν is narrowly continuous if for every $f \in C_b(\mathbb{R}^d)$, $t \mapsto \int_{\mathbb{R}^d} f \, \mathrm{d}\nu_t$ is continuous.*

Note that if $\nu \in \mathcal{P}(C([0,T]; \mathbb{R}^d))$ and $\nu_t = (e_t)_{\#}\nu$, then the family $\nu = (\nu_t)_{t \in [0,T]} \subset \mathcal{P}(\mathbb{R}^d)$ is narrowly continuous, where e_t is the evaluation map at time $t \in [0,T]$.

Definition 1.2.6 (Weak solution of FPE, Trevisan, 2016, Definition 2.2)**.** *A Borel curve taking values in the set of measures on $\mathbb{R}^d$, $\nu = (\nu_t)_{t\in(0,T)} \subset \mathcal{M}(\mathbb{R}^d)$ is a weak solution of the Fokker-Planck equation*

$$\partial_t \nu_t = \mathcal{L}_t^* \nu_t, \qquad \text{on } (0,T)\times\mathbb{R}^d, \tag{1.2.4}$$

if the following integrability condition holds

$$\int_0^T \int_{\mathbb{R}^d} (|a_t| + |b_t|)\, \mathrm{d}\nu_t\, \mathrm{d}t < \infty, \tag{1.2.5}$$

and for all $f \in C_c^{1,2}((0,T)\times\mathbb{R}^d)$,

$$\int_0^T \int_{\mathbb{R}^d} [\partial_t f(t,x) + \mathcal{L} f(t,x)]\, \mathrm{d}\nu_t(x)\, \mathrm{d}t = 0. \tag{1.2.6}$$

Trevisan, 2016, Remark 2.3 (also Ambrosio, Gigli, and Savaré, 2005, Lemma 8.1.2) shows that any solution $(\nu_t)_{t\in(0,T)}$ of (1.2.4) has a narrowly continuous representative $(\tilde{\nu}_t)_{t\in[0,T]}$ such that $\nu_t = \tilde{\nu}_t$ for almost all $t \in (0,T)$. Moreover,

$$\int f_{t_2}\, \mathrm{d}\tilde{\nu}_{t_2} - \int f_{t_1}\, \mathrm{d}\tilde{\nu}_{t_1} = \int_{t_1}^{t_2} \int \partial_t f + \mathcal{L}_t f\, \mathrm{d}\nu_t\, \mathrm{d}t, \quad \text{for } 0 \leqslant t_1 < t_2 \leqslant T.$$

Definition 1.2.7 (Solution of the Martingale Problem, Trevisan, 2016, Definition 2.4)**.** *A probability measure $\boldsymbol{\eta} \in \mathcal{P}(C([0,T];\mathbb{R}^d))$ is a solution of the martingale problem associated to $\mathcal{L}$ if*

$$\int \int_0^T (|a_t| \circ e_t + |b_t| \circ e_t)\, \mathrm{d}t\mathrm{d}\boldsymbol{\eta} < \infty, \tag{1.2.7}$$

and further for all $f \in C_b^{1,2}([0,T]\times\mathbb{R}^d)$ the process:

$$t \mapsto f_t \circ e_t - \int_0^t [\partial_t f_s + \mathcal{L}_s f] \circ e_s\, \mathrm{d}s, \tag{1.2.8}$$

is a martingale with respect to the natural filtration on $C([0,T];\mathbb{R}^d)$.

We remark that the link between the martingale problem and Fokker-Planck equation has been well studied, with criteria for well-posedness and uniqueness of the stochastic

process corresponding to the Fokker-Planck equation given in Stroock and Varadhan, 1979; Ethier and Kurtz, 1986; Karatzas and Shreve, 2014. These definitions lead us to the central well-posedness result of Trevisan, 2016.

Theorem 1.2.8 (Superposition Principle, Trevisan, 2016, Theorem 2.5)**.** *Let $\nu = (\nu_t)_{t\in[0,T]} \subset C([0,T];\mathcal{P}(\mathbb{R}^d))$ be a narrowly continuous solution of* (1.2.4)*. Then there exists $\boldsymbol{\eta}$ which is a solution to the martingale problem associated with the diffusion operator $\mathcal{L}$ such that for all $t \in [0,T]$, we have $\eta_t = \nu_t$.*

Equipped with the above terminology, we can now consider the MOT formulation corresponding to the Fokker-Planck equation studied in Huesmann and Trevisan, 2019.

Problem 1.2.9 (Huesmann and Trevisan, 2019, Equation 3.4)**.** *The formulation of optimal transport among all curves with the Fokker-Planck continuity equation is given by*

$$c_{\mathrm{FPE}}(\mu,\nu) \coloneqq \inf\left\{\int_0^1 \int_{\mathbb{R}^d} c(t,x,a_t(x))\,\mathrm{d}\rho_t(x)\,\mathrm{d}t \;:\; \partial_t\rho = \mathrm{Tr}\left(\frac{1}{2}\nabla_x^2 a\rho_t\right),\ \rho_0=\mu,\ \rho_1=\nu\right\}, \tag{1.2.9}$$

where the infimum is taken across $(\rho_t, a) \in C([0,1];\mathcal{P}_2(\mathbb{R}^d)) \times L^1(\mathrm{d}\rho_t\,\mathrm{d}t;\mathbb{S}_+^d)$. Note that the PDE constraint in the minimisation problem above holds in the sense of distributions.

Now, we state a formulation of the dual problem of Problem 1.2.9, which is crucial in the duality arguments of I. Guo, Loeper, and Shiyi Wang, 2022 and I. Guo, Loeper, Obłój, et al., 2022a. In Chapter 2, this is a key part of transforming the optimal transportation formulation of the calibration into a numerically solvable problem.

Theorem 1.2.10 (Dual of c_{FPE}, Huesmann and Trevisan, 2019, Theorem 4.3)**.** *Let $p \in (1,\infty)$ and $c : (0,1)\times\mathbb{R}^d\times\mathbb{S}_+^d \to \mathbb{R}\cup\{+\infty\}$ be p-admissible. For all $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$, if $c_{\mathrm{FPE}} < \infty$, then*

$$c_{\mathrm{FPE}} = \sup\left\{\int_{\mathbb{R}^d}\varphi(1,x)\,\mathrm{d}\nu(x) - \int_{\mathbb{R}^d}\varphi(0,x)\,\mathrm{d}\mu(x)\right\}. \tag{1.2.10}$$

Where the supremum runs over all super-solutions $\varphi \in C_b^{1,2}([0,1]\times\mathbb{R}^d)$ to

$$\partial_t\varphi(t,x) + c^*\left(t,x,\frac{1}{2}\nabla^2\varphi(t,x)\right) = 0. \tag{1.2.11}$$

Moreover the infimum in (1.2.9) *is attained and is therefore a minimum.*

We can also state an explicit form of $a_t(x)$ such that the corresponding Borel curve $(\rho_t)_{t\in[0,1]}$ is a minimiser of Problem 1.2.9.

Corollary 1.2.11 (Form of Optimiser, Huesmann and Trevisan, 2019, Corollary 4.6)**.** *Let c be p-admissible and $\varphi \in C_b^{1,2}([0,1]\times\mathbb{R}^d)$ solve* (1.2.11)*, and define*

$$a_t(x) \coloneqq \nabla_u c^*\left(t, x, \frac{1}{2}\nabla_x^2\varphi(t,x)\right), \quad \text{for } (t,x)\in[0,1]\times\mathbb{R}^d. \tag{1.2.12}$$

Let $(\rho_t)_{t\in[0,1]} \subset C([0,1];\mathcal{P}(\mathbb{R}^d))$ be a solution to (1.2.4) *with a_t defined by* (1.2.12)*. Then $(\rho_t)_{t\in[0,1]}$ is a minimiser in Problem 1.2.9.*

1.3 Markovian Projection

Here we outline the key results surrounding the idea of "Markovian Projection", which will play a core role in formulating the calibration problem in Chapter 2. Suppose we have an Itô process ξ defined on a finite time interval $[0,T]$, $T>0$. Our aim is to find a Markov process, X, such that $\xi_t = X_t$ in law for all $t\in[0,T]$. Such an X satisfying this form of equality in law mimicks the process ξ and is referred to as the Markovian projection of ξ. The idea was first formulated in Krylov, 1984, where he considered the Green λ-measure of an Itô process X (see Øksendal, 2003, Chapter 9 for a definition). Given X, which may have drift and volatility defined by random functions, Krylov then was able to construct a process with the same Green λ-measure satisfying a time-independent diffusion equation, which in particular has deterministic drift and volatility. He further postulated that one could construct a process satisfying a time-dependent diffusion equation that matches the one-dimensional marginal of an Itô process. Gyöngy, 1986 fixed a gap in Krylov, 1984, which made the stronger assumption that the diffusion had bounded second derivatives, by employing approximation and limit theorem techniques from Gyöngy and Krylov, 1980 to overcome this issue. The second result in Gyöngy, 1986 concerns the extension postulated in Krylov, 1984, which is that the one dimensional distributions of the Itô process and the Markov process coincide.

This result was recovered via a financial argument in Dupire, 1996, which is referred to in Brunick and Shreve, 2013 as the Gyöngy-Dupire formula. Gyöngy, 1986, Theorem 4.6 was extended to the case of right continuous semimartingales with jump discontinuities in Bentata and Cont, 2009. There, they constructed a probability measure on the Skorokhod space of right-continuous functions with left limits, $D([0,T];\mathbb{R}^d)$, such that the canonical process $X_t(\omega) = \omega(t)$ on $D([0,T];\mathbb{R}^d)$ is a Markov process mimicking the Itô process ξ. Their formulation of the Markovian projection in this setting uses the same conditional expectation formulation of the coefficients in the SDE of the mimicking process X, including in the jump compensator, but at ξ_{t-} instead.

Brunick and Shreve, 2013 extended Gyöngy, 1986, Theorem 4.6 to cover far weaker assumptions on the characteristics of the Itô process, in particular removing the boundedness condition and non-degeneracy of the covariance. Moreover, they showed also that functionals of the process ξ are preserved by the mimicking process X. Let Ω be a Polish space, and $\mathcal{E} \subset \Omega$ be closed, denote by $C([0,\infty);\mathcal{E})$ the space of continuous functions from $[0,\infty)$ to $\mathcal{E}$ and $C_0([0,\infty);\mathcal{E}) \subset C([0,\infty);\mathcal{E})$ the subset of continuous functions with initial value 0.

Definition 1.3.1 (Brunick and Shreve, 2013, Section 3)**.** *Define the following operators by:*

- *The shift operator* $\Theta : C([0,\infty);\mathcal{E}) \times \mathbb{R} \to C([0,\infty);\mathcal{E})$ *by:* $\Theta(x,t) \coloneqq x\left((t+\cdot)^+\right)$,
- *The stopping operator* $\mathcal{S} : C([0,\infty);\mathcal{E}) \times [0,\infty) \to C([0,\infty);\mathcal{E})$ *by:* $\mathcal{S}(x,t) \coloneqq x(\cdot \wedge t)$,
- *If* $\mathcal{E}$ *is a vector space, the difference operator* $\Delta : C([0,\infty);\mathcal{E}) \times [0,\infty) \to C([0,\infty);\mathcal{E})$ *by:* $\Delta(x,t) \coloneqq x(t+\cdot) - x(t)$.

For $\omega \in \mathcal{E} \times C_0([0,\infty);\mathbb{R}^d)$ and a random time T, use the shorthand for the stopped process as $X_t^T(\omega) = \mathcal{S}_t(X(\omega), T(\omega))$.

Definition 1.3.2 (Updating function, Brunick and Shreve, 2013, Definition 3.1)**.** $\Phi : \mathcal{E} \times C_0([0,\infty);\mathbb{R}^d) \to C([0,\infty);\mathcal{E})$ *is called an updating function if:*

$$\Phi_0(e,x) = e, \quad \textit{for } e \in \mathcal{E},\ x \in C_0([0,\infty);\mathbb{R}^d), \tag{1.3.1}$$

$$\Phi^t(e,x) = \Phi^t(e,\mathcal{S}(x,t)), \quad \textit{for } t \geqslant 0,\ e \in \mathcal{E},\ x \in C_0([0,\infty);\mathbb{R}^d), \tag{1.3.2}$$

$$\Theta(\Phi(e,x),t) = \Phi(\Phi_t(e,x),\Delta(x,t)), \quad \textit{for } t \geqslant 0,\ e \in \mathcal{E},\ x \in C_0([0,\infty);\mathbb{R}^d). \tag{1.3.3}$$

This updating function takes an initial condition $e \in \mathcal{E}$ and a path $x \in C_0([0,\infty);\mathbb{R}^d)$ and generates a path in $C([0,\infty);\mathcal{E})$ in condition (1.3.1). Condition (1.3.2) gives us that the path $\Phi(e,x) \in \mathcal{E}$ stopped at $t > 0$ only depends on the initial condition $e \in \mathcal{E}$ and the path $x \in C_0([0,\infty);\mathbb{R}^d)$ stopped at t. Condition (1.3.3) gives us some form of a Markov property such that the path of $\Phi(e,x)$ from time t onwards only depends on the value of the path at t and the increments of x from t onwards. Many examples of such updating functions can be found in Brunick and Shreve, 2013. Equipped with the above, we may now quote Brunick and Shreeve's more general formulation of Markovian projection.

Theorem 1.3.3 (Brunick and Shreve, 2013, Theorem 7.1)**.** *Let $\mathcal{E}$ be a Polish space and $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ a filtered probability space supporting an $\mathcal{E}$-valued random variable Z_0 and an adapted continuous $\mathbb{R}^d$-valued semimartingale Y with $Y_0 = 0$ and characteristic pair (A,B), such that*

$$A_t = \int_0^t \alpha_s \, \mathrm{d}s, \qquad B_t = \int_0^t \beta_s \, \mathrm{d}s, \tag{1.3.4}$$

and the adapted $\mathbb{R}^d$-valued process α and the adapted $\mathbb{R}^d \times \mathbb{R}^d$-valued positive semidefinite process β satisfy

$$\mathbb{E}\left[\int_0^t (||\alpha_s|| + ||\beta_s||) \, \mathrm{d}s\right] < \infty, \qquad t \geqslant 0. \tag{1.3.5}$$

Let $\hat{\alpha}$ and $\hat{\beta}$ be measurable functions defined on $[0,\infty) \times \mathcal{E}$ with $\hat{\alpha}$ taking values in $\mathbb{R}^d$ and $\hat{\beta}$ taking values in the space of $d \times d$ positive semidefinite matrices, and let $N \subset [0,\infty)$ be a Lebesgue-null set such that

$$\hat{\alpha}(t, Z_t) = \mathbb{E}[\alpha_t | Z_t], \qquad \hat{\beta}(t, Z_t) = \mathbb{E}[\beta_t | Z_t], \qquad \mathbb{P} - \text{a.s.}, t \in N^c. \tag{1.3.6}$$

Let $\Phi : \mathcal{E} \times C_0([0,\infty);\mathbb{R}^d) \to C([0,\infty);\mathcal{E})$ be a continuous updating function and let Z be the continuous, $\mathcal{E}$-valued process given by $Z = \Phi(Z_0, Y)$. Let $\hat{Y} : \mathcal{E} \times C_0([0,\infty);\mathbb{R}^d) \to C_0([0,\infty);\mathbb{R}^d)$ be given by $\hat{Y}(e,x) = x$ and $\hat{Z} : \mathcal{E} \times C_0([0,\infty);\mathbb{R}^d) \to C([0,\infty);\mathcal{E})$ be given by $\hat{Z} = \Phi(e,x)$. Then there exists a measure $\hat{\mathbb{P}}$ on $\Omega^{\mathcal{E},d}$ such that:

- *$\hat{Y}$ is a semimartingale with characteristic pair $(\hat{A}, \hat{B})$ under $\hat{\mathbb{P}}$, where $\hat{A}_t \coloneqq \int_0^t \hat{\alpha}(s, \hat{Z}_s) \, \mathrm{d}s$ and $\hat{B}_t \coloneqq \int_0^t \hat{\beta}(s, \hat{Z}_s) \, \mathrm{d}s$,*

- *For each $t \geqslant 0$, the distribution of $\hat{Z}_t$ under $\hat{\mathbb{P}}$ agrees with the distribution of Z_t under $\mathbb{P}$.*

By setting the updating function of Definition 1.3.2 to $\Phi_t(e; x) = e + x(t)$, we can then recover Gyöngy, 1986, Theorem 4.6 result as a corollary.

1.4 Entropic Optimal Transport and Sinkhorn

To contextualise the algorithm developed in Chapter 4, we recall Sinkhorn in the setting of a general cost function. For this section we follow Léonard, 2014; De March, 2018; Nutz, 2022 to give a brief summary of the results required to obtain the Schrödinger system. For simplicity, we take $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R})$, a convex, lower semicontinuous cost function $c : \mathbb{R} \times \mathbb{R} \to (-\infty, \infty]$ that is bounded below, and a regularisation parameter $\varepsilon > 0$. If $\Pi(\mu_0, \mu_1)$ is the set of couplings on $\mathbb{R} \times \mathbb{R}$ with marginals μ_0, μ_1, then the EOT problem can be formulated as

$$V_\varepsilon(\mu_0, \mu_1, c) \coloneqq \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}\times\mathbb{R}} c(x, y)\pi(\mathrm{d}x, \mathrm{d}y) + \varepsilon \mathbf{KL}(\pi, \mathbf{P}_{\text{ref}}), \tag{1.4.1}$$

where $\mathbf{KL}$ is the Kullback-Leibler divergence (see Kullback and Leibler, 1951) and $\mathbf{P}_{\text{ref}}$ is a reference measure. The canonical example used in EOT is the quadratic cost with a product measure as reference, i.e. $c(x, y) = ||x - y||^2$ and $\mathbf{P}_{\text{ref}} = \mu_0 \otimes \mu_1$. Under suitable conditions, it can be shown also that (1.4.1) converges to the classical optimal transport problem (see (1.1.2)) as $\varepsilon \to 0$. This motivates the Entropic penalisation as a method of approximating the solution to classical optimal transport due to the fast Sinkhorn algorithm available in the Entropic setting. Since $V_\varepsilon(\mu_0, \mu_1, c) = \varepsilon V_1\left(\mu_0, \mu_1, \frac{c}{\varepsilon}\right)$, we could take $\varepsilon = 1$ without loss of generality in all of the subsequent calculations (see Nutz, 2022, Remark 4.12). Indeed, the similarities between (1.4.1) and the Schrödinger problem (see (4.2.1)) are not just aesthetic: it can be shown that both problems have the same minimisers when $\varepsilon = 1$ and further if $\int e^{-c}\, \mathrm{d}\mathbf{P}_{\text{ref}} = 1$, then the problems are identical. From now, we will take $\mathbf{P}_{\text{ref}} \coloneqq \mu_0 \otimes \mu_1$.

Theorem 1.4.1 (Optimisers for (1.4.1), Nutz, 2022, Theorem 4.2)**.** *Assume $V_\varepsilon(\mu_0, \mu_1, c) < \infty$, then there is a unique minimiser π^*_ε of* (1.4.1)*.*

- *Let $c < \infty$ $\mu_0 \otimes \mu_1$-a.s. Then $\pi^*_\varepsilon \sim \mu_0 \otimes \mu_1$ and there are measurable $\varphi_\varepsilon : \mathbb{R} \to \mathbb{R}$,*

$\psi_\varepsilon : \mathbb{R} \to \mathbb{R}$ *called EOT potentials unique up to an additive constant such that*

$$\frac{\mathrm{d}\pi_\varepsilon^*}{\mathrm{d}\mu_0 \otimes \mu_1} = e^{\frac{\varphi_\varepsilon \oplus \psi_\varepsilon - c}{\varepsilon}}, \qquad \mu_0 \otimes \mu_1\text{−a.s.} \tag{1.4.2}$$

Further, if $c \in L^1(\mu_0 \otimes \mu_1)$ *then* $\varphi_\varepsilon \in L^1(\mu_0)$ *and* $\psi_\varepsilon \in L^1(\mu_1)$.

- *Conversely, let* $\tilde{\pi}_\varepsilon \in \Pi(\mu_0, \mu_1)$ *admit a density of the form*

$$\frac{\mathrm{d}\tilde{\pi}_\varepsilon}{\mathrm{d}\mu_0 \otimes \mu_1} = e^{\frac{\varphi_\varepsilon \oplus \psi_\varepsilon - c}{\varepsilon}}, \qquad \mu_0 \otimes \mu_1\text{−a.s.} \tag{1.4.3}$$

for $\varphi_\varepsilon : \mathbb{R} \to [-\infty, \infty)$ *and* $\psi_\varepsilon : \mathbb{R} \to [-\infty, \infty)$. *Then* $\tilde{\pi}_\varepsilon = \pi_\varepsilon^*$, *and* φ_ε *and* ψ_ε *are the EOT potentials.*

Now we are posed to state the dual formulation of EOT.

Theorem 1.4.2 (EOT Dual, Nutz, 2022, Theorem 4.7)**.** *Let* $c \in L^1(\mu_0 \otimes \mu_1)$. *Then*

$$V_\varepsilon(\mu_0, \mu_1, c) = \sup_{\varphi_\varepsilon \in L^1(\mu_0), \psi_\varepsilon \in L^1(\mu_1)} \mu_0(\varphi_\varepsilon) + \mu_1(\psi_\varepsilon) - \varepsilon \int e^{\frac{\varphi_\varepsilon \oplus \psi_\varepsilon - c}{\varepsilon}} \,\mathrm{d}(\mu_0 \otimes \mu_1) + 1, \tag{1.4.4}$$

the supremum in (1.4.4) *is attained by the EOT potentials* $\varphi_\varepsilon^* \in L^1(\mu_0)$ *and* $\psi_\varepsilon^* \in L^1(\mu_1)$ *a.s. unique up to an additive constant and*

$$V_\varepsilon(\mu_0, \mu_1, c) = \mu_0(\varphi_\varepsilon^*) + \mu_1(\psi_\varepsilon^*). \tag{1.4.5}$$

Due to the form of the density of π_ε^* in (1.4.2), it can therefore be shown that the EOT potentials φ_ε and ψ_ε solve the Schrödinger equations:

$$\varphi_\varepsilon(x) = -\varepsilon \log \int_{\mathbb{R}} e^{\frac{\psi_\varepsilon(y) - c(x,y)}{\varepsilon}} \mu_1(\mathrm{d}y), \qquad \mu_0\text{−a.s.}, \tag{1.4.6}$$

$$\psi_\varepsilon(y) = -\varepsilon \log \int_{\mathbb{R}} e^{\frac{\varphi(x) - c(x,y)}{\varepsilon}} \mu_0(\mathrm{d}x), \qquad \mu_1\text{−a.s.} \tag{1.4.7}$$

This leads us to the Sinkhorn algorithm for optimal transport. The iterations in Algorithm 1, and form Bregman projections computing (see Bregman, 1967) the dual optimiser. The algorithm has linear convergence in the discrete case (see Franklin and Lorenz, 1989; Knight,

Algorithm 1: Sinkhorn algorithm for optimal transport

Input: $\mu_0, \mu_1, \varepsilon$

Output: φ^*, ψ^*

Set $\varphi_0 = 0$ **for** $n = 0$ *to* `max_iter` **do**

  $\psi_n(y) \leftarrow -\varepsilon \log \int_{\mathbb{R}} e^{\frac{\psi_n(x) - c(x,y)}{\varepsilon}} \mu_0(\mathrm{d}x)$

  $\varphi_{n+1}(y) \leftarrow -\varepsilon \log \int_{\mathbb{R}} e^{\frac{\psi_n(y) - c(x,y)}{\varepsilon}} \mu_1(\mathrm{d}x)$

end

return *EOT potentials* (φ^*, ψ^*)

2008) which motivated the seminal work of Cuturi, 2013.

1.4.1 Martingale Sinkhorn

We next review the extension of Algorithm 1 to the setting of MOT. It was first introduced in De March, 2018, and later saw an application in fast SPX/VIX calibration in Guyon and Bourgey, 2022. Let $\Pi_{\text{mart}}(\mu_0, \mu_1)$ denote the set of martingale couplings between $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R})$. Assume also that μ_0, μ_1 are in convex order so that by Strassen, 1965 we always have $\Pi_{\text{mart}}(\mu_0, \mu_1) \neq \emptyset$. The martingale EOT problem then becomes

$$V_\varepsilon(\mu_0, \mu_1) \coloneqq \inf_{\pi \in \Pi_{\text{mart}}(\mu_0, \mu_1)} \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\pi(\mathrm{d}x, \mathrm{d}y) + \varepsilon \mathbf{KL}(\pi, \mathbf{P}_{\text{ref}}). \tag{1.4.8}$$

For ease, we again take $\mathbf{P}_{\text{ref}} = \mu_0 \otimes \mu_1$. Then, in the same vein as before, and as in De March, 2018, Section 3.3.3, the dual of (1.4.8) is given by

$$V_\varepsilon(\mu_0, \mu_1, c) = \sup \mu_0(\varphi_\varepsilon) + \mu_1(\psi_\varepsilon) - \varepsilon \int e^{\frac{\varphi_\varepsilon \oplus \psi_\varepsilon + h_\varepsilon^{\otimes} - c}{\varepsilon}} \, \mathrm{d}(\mu_0 \otimes \mu_1), \tag{1.4.9}$$

where the supremum in (1.4.9) is taken over $(\varphi_\varepsilon, \psi_\varepsilon, h_\varepsilon) \in L^1(\mu_0) \times L^1(\mu_1) \times L^0(\mathbb{R})$, and $h_\varepsilon^{\otimes} = h_\varepsilon(x) \cdot (y - x)$. We therefore obtain the following Schrödinger system for the EOT potentials $(\varphi_\varepsilon, \psi_\varepsilon, h_\varepsilon)$ attaining the supremum in (1.4.9):

$$\varphi_\varepsilon(x) = -\varepsilon \log \int_{\mathbb{R}} e^{\frac{\psi_\varepsilon(y) + h_\varepsilon(x)\cdot(y-x) - c(x,y)}{\varepsilon}} \mu_1(\mathrm{d}y), \qquad \mu_0\text{−a.s.}, \tag{1.4.10}$$

$$\psi_\varepsilon(y) = -\varepsilon \log \int_{\mathbb{R}} e^{\frac{\varphi_\varepsilon(y) + h_\varepsilon(x)\cdot(y-x) - c(x,y)}{\varepsilon}} \mu_0(\mathrm{d}y), \qquad \mu_1\text{−a.s.}, \tag{1.4.11}$$

$$0 = \int_{\mathbb{R}} e^{\frac{\varphi_\varepsilon(x) + \psi_\varepsilon(y) + h_\varepsilon(x)\cdot(y-x) - c(x,y)}{\varepsilon}} (y - x) \mu_1(\mathrm{d}y), \qquad \text{for all } x \in \mathbb{R}. \tag{1.4.12}$$

In analogy with (1.4.6)-(1.4.7), the iterates (1.4.10)-(1.4.11) are explicit. On the other hand, the iterate (1.4.12) is implicit so must be numerically solved. See De March, 2018; Guyon and Bourgey, 2022 for implementations of Newton methods solving (1.4.12). We summarise this in an algorithm below:

Algorithm 2: Sinkhorn algorithm for MOT

Input: $\mu_0, \mu_1, \varepsilon$

Output: φ^*, ψ^*, h^*

Set $\varphi_0 = 0$, $h_0 = \mathrm{Id}$ **for** $n = 0$ *to* `max_iter` **do**

  $\psi_n(y) \leftarrow -\varepsilon \log \int_{\mathbb{R}} e^{\frac{\psi_n(x)+h_n(x)\cdot(y-x)-c(x,y)}{\varepsilon}} \mu_0(\mathrm{d}x)$

  $\varphi_{n+1}(y) \leftarrow -\varepsilon \log \int_{\mathbb{R}} e^{\frac{\psi_n(y)+h_n(x)\cdot(y-x)-c(x,y)}{\varepsilon}} \mu_1(\mathrm{d}x)$

  Solve implicitly for $h(x)$: $0 = \int_{\mathbb{R}} e^{\frac{\varphi_{n+1}(x)+\psi_n(y)+h(x)\cdot(y-x)-c(x,y)}{\varepsilon}} (y-x)\mu_1(\mathrm{d}y)$

  $h_{n+1}(x) \leftarrow h(x)$

end

return *EOT potentials* (φ^*, ψ^*, h^*)

Note that $\mathbf{P}_{\text{ref}} = \mu_0 \otimes \mu_1$ is only chosen for the convenience of presentation. The Schrödinger systems can clearly be equivalently formulated for different reference measures, and as remarked in Guyon and Bourgey, 2022, Remark 2, for financial applications it would be better motivated to use a lognormal prior as the reference.

1.5 Model Calibration

The world of option pricing saw a paradigm shift with the derivation of the Black-Scholes formula in Black and Scholes, 1973 for options, the Black formula for futures (in particular for interest rate derivatives) in Black, 1976, and Merton's work on optimal investment, see Merton, 1969; Merton, 1975. The main contribution here is the idea that one can construct a risk-free portfolio by dynamically hedging an option against its underlying. In addition, they use the idea that the markets fundamentally possessed some randomness, an idea originally explored in Bachelier, 1900; Samuelson, 1965; Samuelson, 1973. Other seminal works include the theory of no arbitrage, see Harrison and Kreps, 1979, and the first fundamental theorem of asset pricing, see Delbaen and Schachermayer, 1994, which in particular shows that an underlying asset has to be a semimartingale via the efficient market hypothesis.

The Black-Scholes formula was not without its own drawbacks, in particular the volatility

being constant is not a realistic assumption which can immediately be seen by computing the implied volatilities of options, and having the volatility only depend on time does not remedy this issue. An immediate extension is the class of parametric local volatility models where the volatility depends on time and the underlying, such as the displaced diffusion model Rubinstein, 1983, the CEV model J. Cox and Ross, 1976; J. Cox, 1996, or the lognormal mixture model Brigo and Mercurio, 2002. Another class of models are stochastic volatility models where the volatility depends on an extra state variable given by a stochastic process, with examples being Hull and A. White, 1987, Heston, 1993, and Hagan et al., 2002. A combination of both of these approaches was later developed in Jex, Henderson, and D. Wang, 1999, known as local-stochastic volatility. We point to Bakshi, Cao, and Chen, 1997 for an empirical comparison of all of these models.

In addition, since any financial model involves future payments, the effect of a stochastic interest rate cannot be ignored. The subject of interest rate models is vast and for brevity we refer to Filipovic, 2009; Cairns, 2018; Rebonato, McKay, and R. White, 2009; Brigo and Mercurio, 2007 for more comprehensive coverage of the properties and parametric calibration of specific classes of models. The simplest interest rate model is a short rate model, which directly models the annualised and continuously compounded rate at which one can borrow for an instantaneous time to maturity. This is often modelled as a Markov process, with one of the simplest models being the Vasicek model (see Vasicek, 1977) which models the short rate as an Ornstein-Uhlenbeck process. A relatively simple extension of this is the Ho-Lee model (see Ho and Lee, 1986, and Kalotay, Williams, and Fabozzi, 1993 for its lognormal version); the Hull-White model (see Hull and A. White, 1990 and Black and Karasinski, 1991 for its lognormal extension), which adds time-dependent drift to fit the term structure; or the CIR model (see J. Cox, Ingersoll Jr, and Ross, 1985), which instead has a local-volatility. These are all examples of one-factor short rate models, i.e. they are governed by one state variable. Multi-factor models, such as Longstaff and Schwartz, 1992 or the two-factor Hull-White model (see Hull, 1994) can also account for stochastic mean and volatility at the expense of more state variables. Another perspective is to instead consider modelling the whole forward curve via the well known Heath-Jarrow-Morton framework introduced in Heath, Jarrow, and Morton, 1992. When combined with Ritchken and Sankarasubramanian,

1995, this gives necessary and sufficient conditions on the volatility function of each forward rate for the induced short rate process to be Markov.

A different class of models are market models, which model the inter-bank offered rates (e.g. LIBOR or EURIBOR) directly. The lognormal forward model was separately introduced in both Brace, Gątarek, and Musiela, 1997 and Miltersen, Sandmann, and Sondermann, 1997, and models the LIBOR rates for fixed maturity times under different forward measures, which are calibrated to caps and floors. Jamshidian, 1997 also constructed an analogous market model that is calculated using swaption pricing; however, this is incompatible with the previous one in the sense that lognormal dynamics cannot be attained using different models for the instruments. However, as noted in Brace, Dun, and Barton, 2001, this is more of a theoretical incompatibility since the distributions are "close" to lognormal. We also mention the multi-curve framework of Henrard, 2007; Henrard, 2010, where the post-crisis effects of the LIBOR-OIS spreads are taken into account with multiple discounting curves. Kenyon, 2010 constructed a short rate model in the multi-curve framework with a focus on constructing prices for Swaptions, with a single discounting curve and multiple fixing curves for different EURIBOR tenors – the discount curve here is modelled as a short rate. Kijima, Tanaka, and Wong, 2009 also constructed a short rate model in the multi-curve framework through three different curves modelling a discount curve under a no-arbitrage setup, the curve for a floating LIBOR rate and another curve associated with government bonds, based on Boenkost and Schmidt, 2005. Here, three different short rates are considered in the natural way and their corresponding bonds are set up analogously to the one curve case. The HJM model was investigated within the multi-curve framework in Pallavicini and Tarenghi, 2010, and provided a method of calibrating this HJM construction to interest rate derivatives. It has also been explored in Crépey, Grbac, and Nguyen, 2012, and extended in Crépey, Grbac, Ngor, et al., 2015 to a Lévy processes setting. Ametrano and Bianchetti, 2013 develop a bootstrapping technique for the multicurve LIBOR market model, that is bootstrapping the discounting curve and yield curves consistently with market instruments. Bianchetti and Carlicchi, 2011 also studied the extension of the SABR model of Hagan et al., 2002 to the multi-curve framework, with it being demonstrated to fit the volatility smile well.

In practice, models for a key underlying, such as the S&P500 index, will need to be calibrated to a large number of options with different maturities and strikes. This may be nigh impossible for the parametric models described earlier. Under the assumption of a continuum of strikes, one approach is the Breeden-Litzenberger formula (see Breeden and Litzenberger, 1978) where the density of the underlying is expressed in terms of the second derivative of a call option with respect to its strikes. Other developments are the seminal Dupire formula, see Dupire, 1994; Dupire, 1997, and Atlan, 2006 for its stochastic rate extension, or Derman and Kani, 1994 for an independent derivation, and Derman and Kani, 1998 for an HJM-like construction of stochastic volatility. Monte Carlo methods to calibrate with Dupire's formula have been used in Deelstra and Rayée, 2013; Hok and S.-H. Tan, 2019; Ögetbil, Ganesan, and Hientzsch, 2022. Dupire, 1996 derives a non-linear McKean SDE for the local volatility, which in fact recovered the mimicking result of Gyöngy, 1986 using a financial argument. This has been solved via Monte Carlo methods developed in Henry-Labordère, 2009 for an approximate calibration, and an exact calibration by the particle method of Guyon and Henry-Labordere, 2011; Guyon and Henry-Labordère, 2012. Cozma, Mariapragassam, and Reisinger, 2019 used variance reduction techniques to implement the particle method in a four-factor model also. Another approach is to discretise the non-linear Fokker-Planck equation and directly solve it by finite differences as in Ren, Madan, and Qian, 2007, or finite volume methods as in Wyns and Du Toit, 2017; Engelmann, Koster, and Oeltz, 2021. All of these methods require a continuum of strikes and maturities to construct a surface of implied volatilities, which in practice results in interpolation and extrapolation of market data to build these surfaces.

In Chapter 2, using the SOT framework described in Section 1.2, we propose a different approach without this requirement by formulating the exact calibration of European options as discrete constraints within a convex optimisation problem. The SOT framework has already been used in several contexts: for local volatility calibration in I. Guo, Loeper, and Shiyi Wang, 2019, for local-stochastic volatility calibration in I. Guo, Loeper, and Shiyi Wang, 2022, for joint SPX/VIX calibration in I. Guo, Loeper, Obłój, et al., 2022a, and for optimal investment in I. Guo, Langrené, et al., 2022. The most general description so far of the method and the breadth of its applications is I. Guo and Loeper, 2021. An overview of

these results is given in I. Guo, Loeper, Obłój, et al., 2022b. It is worth noting that whilst the link with optimal transport was not recognized, a variational approach to calibrating local volatility models was first constructed in the uncertain volatility model of Avellaneda, Friedman, et al., 1997 (see also Lyons, 1995; Avellaneda, Levy, and Parás, 1995; Karoui, Jeanblanc-Picquè, and Shreve, 1998).

Another OT framework, namely entropic optimal transport, has recently been used to solve calibration problems. Guyon, 2020; Guyon, 2024 jointly calibrated SPX/VIX by using Sinkhorn iterations to construct a discrete time model. Another approach is that of Benamou, Chazareix, and Loeper, 2024, where the continuous time model was discretised, then solved via entropy regularisation to produce a calibrated local volatility model; and De March and Henry-Labordere, 2019 where the implied volatility surface is directly built from bid-ask data using Sinkhorn iterations. An accelerated Newton-Sinkhorn algorithm was developed in Guyon and Bourgey, 2022, with convergence of the arbitrage free discrete time model taking minutes. From the discrete time model, a continuous time model was then constructed using a simulation method. Another method of contructing a continuous time model from a discrete time one is using the fixed point iteration scheme of Conze and Henry-Labordere, 2021 to construct the Bass martingale between two discrete measures in one dimension; the convergence of this a topic of study in Acciaio, Marini, and Pammer, 2023.

Chapter 2

SOT Calibration with Stochastic Interest Rates: Duality

Modelling involves inevitable trade-offs: "All models are wrong, some models are useful" as Box and Draper, 1987 put it. Models need to capture the important aspects of the system they represent but they also need to be tractable, and analytically and/or numerically solvable. In particular, calibration – picking model parameters which recover known outputs – is an essential part of any modelling process. It is a key challenge faced by financial industry practitioners on a daily basis, as their pricing models need to match market prices of liquid instruments before they can be used to price any bespoke or illiquid products. The need for a calibrated interest rates model is common to all financial products involving future payments. While in a very low interest rate environment, which roughly held in the financial markets between 2009 and 2021, one may be tempted to ignore this need for short dated products, it is no longer feasible for the current market conditions. It is therefore natural to extend the semi-martingale optimal transport approach to a setup that includes stochastic rates. The core contribution of the method is to build a fully-calibrated model while trying to preserve desirable features of a given model. In effect, we project a given reference model onto the set of calibrated models. As the calibration constraints depend on one-dimensional marginals, classical mimicking results, see Gyöngy, 1986 and Brunick and Shreve, 2013, allow us to restrict to Markovian models. This in turn allows us to use PDE

methods to solve the dual problem.

Our contribution here is to fill this gap, and understand how to develop and calibrate joint models for rates and equities. This extension is not trivial, since the computation of the discount factor involves the whole path of the short rate, which renders the problem path dependent. We will show how to overcome this difficulty. We provide a novel conditioning argument to reduce the dimension of the problem, by switching to subprobability measures solving a Fokker-Planck equation with a source term. We prove that in using this discounting transformation, we lose no generality in the class of models we calibrate over by proving a superposition principle result in the spirit of Trevisan, 2016, Theorem 2.5. We then prove a generic duality result to transform the calibration problem into a numerically solvable problem, with the numerical results studied in Chapter 3. This chapter is formed of the theoretical results of Joseph, Loeper, and Obłój, 2023a.

2.1 Preliminaries and Notation

We adopt the setup of I. Guo, Loeper, and Shiyi Wang, 2022 and I. Guo, Loeper, Obłój, et al., 2022a, who in turn used the formulation of Xiaolu Tan and Touzi, 2013. Let E be a Polish space equipped with the Borel σ-algebra, let $C(E)$ be the space of continuous functions on E and $C_b(E)$ be the space of continuous bounded functions on E. Let $\mathcal{M}(E)$ be the space of finite signed Borel measures endowed with the weak-$*$ topology, let $\mathcal{M}_+(E) \subset \mathcal{M}(E)$ be the subset of non-negative finite Borel measures, and $\mathcal{P}(E)$ be the set of Borel probability measures also under the weak-$*$ topology. Note that if E is compact, then the topological dual of $C_b(E)$ is given by $C_b(E)^* = \mathcal{M}(E)$, but if E is non-compact, then $C_b(E)^*$ is larger than $\mathcal{M}(E)$. Let $\mathrm{BV}(E)$ be the set of bounded variation functions on E and $L^1(\mathrm{d}\mu)$ be the space of μ-integrable functions. For unambiguity we write $C_b(E;\mathbb{R}^d)$, $\mathcal{M}(E;\mathbb{R}^d)$, $\mathrm{BV}(E;\mathbb{R}^d)$, and $L^1(\mathrm{d}\mu;\mathbb{R}^d)$ for the vector valued versions of those spaces (with an analogous definition for the matrix valued versions).

Write $\mathbb{S}^d$ for the set of $d \times d$ symmetric matrices and $\mathbb{S}^d_+ \subset \mathbb{S}^d$ as the subset of positive semidefinite symmetric matrices. For $a, b \in \mathbb{R}^d$ write $a \cdot b$ for the inner product $a^\intercal b$ and for $A, B \in \mathbb{S}^d$ write $A : B$ for their inner product $\mathrm{Tr}(A^\intercal B)$. As a shorthand, we define

$\Lambda \coloneqq [0,T] \times \mathbb{R}^d$ and $\mathcal{X} \coloneqq \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d$, which will be used for the domain and range of the triple representing the law of the semimartingale, the drift and the volatility. Finally, denote the duality bracket between $C_b(E)$ and $C_b(E)^*$ by $\langle \cdot, \cdot \rangle$.

We fix a time horizon $T > 0$ and dimension $d > 1$. We mainly work on the canonical space $\Omega \coloneqq C([0,T], \mathbb{R}^d)$ of continuous $\mathbb{R}^d$-valued paths on $[0,T]$ with its canonical process and canonical filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \leqslant t \leqslant T}$. However, when needed we work on a larger probability space and highlight it accordingly. For continuous paths, e_t is the evaluation operator, $e_t(\omega) = \omega_t$, for $\omega \in C([0,T], \mathbb{R}^d)$. We consider all probability measures $\mathbb{P}$ on $(\Omega, \mathcal{F}_T)$ such that X is an $(\mathbb{F}, \mathbb{P})$-semimartingale with decomposition

$$X_t = X_0 + A_t^{\mathbb{P}} + M_t^{\mathbb{P}}, \quad t \in [0,T], \quad \mathbb{P} - \text{a.s.}, \tag{2.1.1}$$

where $(M_t^{\mathbb{P}})_{t \geqslant 0}$ is an $(\mathbb{F}, \mathbb{P})$-martingale and $(A_t^{\mathbb{P}})_{t \in [0,T]}$ is a finite variation process, both are absolutely continuous relative to the Lebesgue measure and can be characterised in the following sense.

Definition 2.1.1. *We say that $\mathbb{P}$ is characterised by $(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})$ if*

$$\alpha_t^{\mathbb{P}} = \frac{\mathrm{d}A_t^{\mathbb{P}}}{\mathrm{d}t}, \qquad \beta_t^{\mathbb{P}} = \frac{\mathrm{d}\langle M^{\mathbb{P}} \rangle_t}{\mathrm{d}t} \quad dt \times \mathbb{P}(\mathrm{d}\omega)\text{-}a.e., \tag{2.1.2}$$

where $(\alpha_t, \beta_t)_{t \in [0,T]}$ is an $\mathbb{R}^d \times \mathbb{S}^d$-valued, progressively measurable process.

The existence of progressively measurable (α_t, β_t) is guaranteed since $A_t^{\mathbb{P}}$ and $M_t^{\mathbb{P}}$ are assumed to be Lebesgue absolutely continuous in t, $\mathbb{P}$-a.s., see Jacod and Shiryaev, 2013, Proposition I.3.13, Proposition II.2.9. Note that $(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})$ is only determined up to $\mathrm{d}\mathbb{P} \times \mathrm{d}t$-almost everywhere. The set of probability measures $\mathbb{P}$ satisfying the conditions above is denoted $\mathcal{P}$. We note that regular conditional probabilities exist on Ω and we will use these implicitly, e.g., $\mathbb{E}^{\mathbb{P}}_{t,x}[\alpha_t^{\mathbb{P}}]$ or $\mathbb{E}^{\mathbb{P}}[\alpha_t^{\mathbb{P}} | X_t = x]$ will denote the conditional expectation $\mathbb{E}^{\mathbb{P}}[\alpha_t^{\mathbb{P}} \mid \mathcal{F}_t]$ seen as a measurable function of (t, X_t) and evaluated at $X_t = x$. Finally, we note that, by Doob's martingale representation Theorem (see Karatzas and Shreve, 2014, Theorem 3.4.2),

possibly on an enlarged probability space, there exists a Brownian motion $W^{\mathbb{P}}$ such that

$$X_t = X_0 + \int_0^t \alpha_s^{\mathbb{P}} \, \mathrm{d}s + \int_0^t (\beta_s^{\mathbb{P}})^{1/2} \, \mathrm{d}W_s^{\mathbb{P}}, \quad t \leqslant T. \tag{2.1.3}$$

The process $X_t := (\mathbf{S}_t, X_t^r)$, is composed of the short rate $(X_t^r)_{t\in[0,T]}$ and a $(d-1)$-dimensional process $(\mathbf{S}_t)_{t\in[0,T]}$ corresponding to the underlying asset, such as the S&P 500, and extra state variables, e.g., extra assets, stochastic factors in the volatility functions, or multi-factors in the short rate. In specific examples, e.g. in Chapter 3, the short rate is also denoted r_t. The state variable's x coordinate corresponding to X^r is denoted x_r to stress that it represents the short rate. The stochastic discount factor is derived from the short rate, $Y_t := \exp\left(-\int_0^t X_s^r \, \mathrm{d}s\right)$, and we refer to (X_t, Y_t) as the augmented process. In Proposition 2.1.2, we verify that the augmented process can be expressed as an updating function in the sense of Definition 1.3.2. We further consider the subset $\mathcal{P}^1 \subset \mathcal{P}$ of measures $\mathbb{P} \in \mathcal{P}$ which satisfy

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^T |\alpha_t^{\mathbb{P}}| + |\beta_t^{\mathbb{P}}| \, \mathrm{d}t\right] < +\infty \quad \text{and} \quad X_t^r > -1 \, \mathrm{d}t \otimes \mathrm{d}\mathbb{P}\text{-a.e.}, \tag{2.1.4}$$

where $|\cdot|$ is the L^1 norm. We are not aware of any relevant model which would require the short rate to be unbounded from below so this part of the assumption pauses no problem. It implies in particular that $0 < Y_t < \mathrm{e}^T$, $t \in [0,T]$, under all measures considered in this Chapter. The integrability assumption would equally be satisfied under any reasonable model and it is imposed so that we can apply Brunick and Shreve, 2013, Theorem 3.6. These mimicking results, extending earlier works of Krylov, 1984 and Gyöngy, 1986, allow us to construct a Markov process with the same one-dimensional marginals as a given semi-martingale. As our constraints and cost will only depend on these marginals, it will allow us restrict our attention to Markov processes. Markovian projection relies on the classical result that the marginal law of a diffusion process is a distributional solution to the corresponding Fokker-Planck equation, with the converse result given in Figalli, 2008 where existence and uniqueness results are constructed for the corresponding SDE satisfied by a process with marginal law that is a weak solution to a Fokker-Planck equation. In addition,

since we are considering a calibration problem with a stochastic interest rate, to avoid path dependent terms we augment the process to include stochastic discount factor terms. For simplicity, our argument will consider a one-factor short rate, but it is easy to adapt to a multi-factor or multi-curve setting.

Proposition 2.1.2. *The augmented process (X, Y) can be expressed as an updating function of X.*

Proof. We use the shift, stopping, and difference operators of Definition 1.3.1 throughout. For $e_1 \in \mathbb{R}^d$, $e_2 \in \mathbb{R}$, $x \in C_0([0,T];\mathbb{R}^d)$, define $\Phi : \mathbb{R}^{d+1} \times C_0([0,T];\mathbb{R}^d) \to C([0,T];\mathbb{R}^{d+1})$ by

$$\Phi_t(e_1, e_2; x) = \left[e_1 + x(t), e_2 e^{-\int_0^t e_1 + x(s)\,\mathrm{d}s}\right]. \tag{2.1.5}$$

The first property is clearly satisfied since $\Phi_0(e_1, e_2; x) = (e_1, e_2)$ for all $e_1 \in \mathbb{R}^d$, $e_2 \in \mathbb{R}$, $x \in C_0([0,T];\mathbb{R}^d)$. For the second property, we have for any $t, s \geqslant 0$

$$\Phi_s^t(e_1, e_2; x) = \left[e_1 + \mathcal{S}_s(x,t), e_2 e^{-\int_0^s e_1 + \mathcal{S}_u(x,t)\,\mathrm{d}u}\right] = \Phi^t(e_1, e_2; x). \tag{2.1.6}$$

Finally, for the third property we have:

$$\begin{aligned}
\Theta\left(\Phi(e_1, e_2; x), t\right) &= \left[e_1 + x(t+s), e_2 e^{-\int_0^{t+s} e_1 + x(u)\,\mathrm{d}u}\right], \\
&= \left[e_1 + x(t) + \Delta_s(x,t), e_2 e^{\int_0^t e_1 + x(u)\,\mathrm{d}u - \int_0^s e_1 + \Delta_u(x,t)\,\mathrm{d}u}\right], \\
&= \Phi\left(\Phi_t(e, x), \Delta(x,t)\right).
\end{aligned} \tag{2.1.7}$$

□

Theorem 2.1.3 (Markovian Projection)**.** *Let $\mathbb{P} \in \mathcal{P}^1$ be a candidate model. There exist jointly measurable versions of the conditional expectations, $\alpha_t(x,y) = \mathbb{E}^{\mathbb{P}}_{t,x,y}[\alpha_t^{\mathbb{P}}]$, $\beta_t(x,y) = \mathbb{E}^{\mathbb{P}}_{t,x,y}[\beta_t^{\mathbb{P}}]$, $\mathrm{d}t \times \mathrm{d}\mathbb{P}$-a.e., and $\tilde{\mathbb{P}} \in \mathcal{P}^1$ such that $\alpha_t^{\tilde{\mathbb{P}}} = \alpha_t(X_t, Y_t)$ and $\beta_t^{\tilde{\mathbb{P}}} = \beta_t(X_t, Y_t)$, $\mathrm{d}t \times \mathrm{d}\tilde{\mathbb{P}}$ - a.e., and $\tilde{\mathbb{P}} \circ (X_t, Y_t)^{-1} = \mathbb{P} \circ (X_t, Y_t)^{-1}$. Moreover, possibly on some other probability space,*

there exists a Markov process (X, Y) *satisfying:*

$$\begin{cases} \mathrm{d}X_t = \alpha_t(X_t, Y_t)\,\mathrm{d}t + (\beta_t(X_t, Y_t))^{\frac{1}{2}}\,\mathrm{d}W_t^{\mathbb{P}'}, & 0 \leqslant t \leqslant T, \\ \mathrm{d}Y_t = -Y_t X_t^r\,\mathrm{d}t, & 0 \leqslant t \leqslant T, \\ X_0 = x_0, \\ Y_0 = 1. \end{cases} \tag{2.1.8}$$

such that $(X_t, Y_t) \sim \mathbb{P} \circ (X_t, Y_t)^{-1}$ *for all* $t \in [0, T]$*, and where* $W^{\mathbb{P}'}$ *is a Brownian motion. Finally, the marginal distributions,* $\rho_t = \rho(t, \cdot) = \mathbb{P} \circ (X_t, Y_t)^{-1}$*, are a weak solution to the Fokker-Planck equation*

$$\begin{cases} \partial_t \rho_t + \nabla_x \cdot (\rho_t \alpha_t) - \frac{1}{2}\nabla_x^2 : (\rho_t \beta_t) - \partial_y(x_r y \rho_t) = 0, & \text{on } [0, T] \times \mathbb{R}^{d+1}, \\ \rho_0^{\mathbb{P}} = \delta_{X_0, Y_0}, & (x, y) \in \mathbb{R}^{d+1}. \end{cases} \tag{2.1.9}$$

Definition 2.1.4. *We let* $\mathcal{P}^1_{\mathrm{loc}}$ *denote the subset of measures in* $\mathcal{P}^1$ *under which the semimartingale characteristics of* X *are given as measurable functions in* (t, X_t, Y_t)*.*

Remark 2.1.5. *The first assertion in Theorem 2.1.3 follows from Brunick and Shreve, 2013, Theorem 7.1. The second one then follows via martingale representation theorem, and is Theorem 3.6 therein. The last assertion is obtained using Itô's formula. Note that* $\tilde{\mathbb{P}} \in \mathcal{P}^1_{\mathrm{loc}}$ *and is simply the distribution of* X *solving* (2.1.8)*,* $\tilde{\mathbb{P}} = \mathbb{P}' \circ X^{-1}$*. With a slight abuse of language, we will refer to* $\mathcal{P}^1_{\mathrm{loc}}$ *as measures under which* (X, Y) *is a Markov process solving* (2.1.8)*.*

Our calibration problem will be written as a minimization of a cost functional which represents a "distance" to a given favourite reference model and also ensures perfect calibration. We first define the function $\hat{F} : \Lambda \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d \to [0, +\infty]$ as $\hat{F}(t, x, y, \alpha, \beta) = f(y)F(t, x, \alpha, \beta)$, where the function f is specified by the discounting term in the calibrating options. We consider a strongly convex cost function $F : \Lambda \times \mathbb{R}^d \times \mathbb{S}^d \to [0, +\infty]$, proper and lower semicontinuous in (α, β). It will be set to take value $+\infty$ outside of a set Γ, see for example (3.2.3). In particular we will always set $F = +\infty$ if $\beta \notin \mathbb{S}^d_+$ to ensure β is a

legitimate covariance matrix. We use these properties implicitly, e.g., whenever we restrict to $\beta \in \mathbb{S}_+^d$. The strong convexity assumption of F (see Nesterov, 2018, Definition 2.1.3) means that for any subderivative ∇ performed over $(\alpha, \beta) \in \mathbb{R}^d \times \mathbb{S}_+^d$ there exists $C > 0$ such that for all $(t, x, \alpha, \beta, \alpha', \beta') \in \Lambda \times \mathbb{R}^d \times \mathbb{S}_+^d \times \mathbb{R}^d \times \mathbb{S}_+^d$, when $F(t, x, \alpha, \beta) < \infty$ we have

$$F(t, x, \alpha', \beta') \geqslant F(t, x, \alpha, \beta) + \langle \nabla F(t, x, \alpha, \beta), (\alpha' - \alpha, \beta' - \beta) \rangle + C(||\alpha - \alpha'||_2^2 + ||\beta - \beta'||_{\text{Fro}}^2). \tag{2.1.10}$$

Here $|| \cdot ||_{\text{Fro}}$ denotes the Frobenius norm, which for a matrix M is given by $||M||_{\text{Fro}} = \sqrt{\sum_{i,j} |m_{i,j}|^2}$. We additionally assume that F is p-coercive, that is there exists $p > 1$ and $C > 0$ such that for all $(t, x, \alpha, \beta) \in \Lambda \times \mathbb{R}^d \times \mathbb{S}_+^d$ we have

$$||\alpha||^p + ||\beta||^p \leqslant C(1 + F(t, x, \alpha, \beta)). \tag{2.1.11}$$

The Legendre-Fenchel transform of F (see for example Rockafellar, 1970, §12) is given by

$$F^*(t, x, a, b) \coloneqq \sup_{\alpha \in \mathbb{R}^d, \beta \in \mathbb{S}_+^d} \{\alpha \cdot a + \beta : b - F(t, x, \alpha, \beta)\}, \tag{2.1.12}$$

where the supremum is a priori over $(\alpha, \beta) \in \mathbb{R}^d \times \mathbb{S}^d$ but can be restricted $\beta \in \mathbb{S}_+^d$ by the comment above, it indeed can be later restricted to the set Γ as $F = +\infty$ elsewhere. While F itself may not be differentiable with respect to (α, β), since F is strictly convex in (α, β) we therefore have that $F^*(t, x, a, b)$ is differentiable in (a, b) (see Rockafellar, 1970, Theorem 26.3). For convenience, we will denote $F(\alpha, \beta) \coloneqq F(t, x, \alpha, \beta)$ and $F^*(a, b) \coloneqq F^*(t, x, a, b)$.

2.2 Problem formulation

We want to calibrate our model to n market prices of options, the i^{th} option has maturity $\tau_i \in (0, T)$, payoff $G_i \in C_b(\mathbb{R}^d; \mathbb{R})$ and price u_i. While European call option payoffs do not satisfy boundedness, as calibrating instruments they are equivalent to put options, via the usual call-put parity, which have bounded and continuous payoffs. We let $\tau = (\tau_1, \ldots, \tau_n)$, $G(x) = (G_1(x), \ldots, G_n(x))$ and $u = (u_1, \ldots, u_n)$. We are only interested in modelling on

the horizon covered by the market instruments, so we assume that $T = \max_{i \leqslant n} \tau_i$.

Definition 2.2.1. *Given an initial distribution μ_0, expiry times τ, market prices u corresponding to payoffs G, we introduce the set of calibrated measures*

$$\mathcal{P}(\mu_0, \tau, u) = \left\{ \mathbb{P} \in \mathcal{P}^1 : \mathbb{P} \circ X_0^{-1} = \mu_0, \, \mathbb{E}^{\mathbb{P}} \left[Y_{\tau_i} G_i(X_{\tau_i}) \right] = u_i \text{ for } i = 1, \dots, n \right\}. \tag{2.2.1}$$

We denote $\mathcal{P}_{\mathrm{loc}}(\mu_0, \tau, u) = \mathcal{P}(\mu_0, \tau, u) \cap \mathcal{P}^1_{\mathrm{loc}}$ the calibrated models under which the semimartingale characteristics of X are given as measurable functions of (t, X_t, Y_t), see Remark 2.1.5.
We denote $\widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0, \tau, u) \subseteq \mathcal{P}_{\mathrm{loc}}(\mu_0, \tau, u)$ the subset of measures under which the semimartingale characteristics of X are given as measurable functions of (t, X_t).

We remark that we will usually take $\mu_0 = \delta_{(X_0, 1)} \in \mathcal{P}(\mathbb{R}^{d+1})$ where $X_0 \in \mathbb{R}^d$ are the observed initial values of our state variables.

Our primal problem consists of selecting one of the possible calibrated models in $\mathcal{P}(\mu_0, \tau, u)$ by minimising a cost functional. The following key results asserts that when doing so, we can in fact restrict to Markov processes.

Proposition 2.2.2. *Given an initial distribution μ_0, expiration times τ, and market prices u, we have*

$$V \coloneqq \inf_{\mathbb{P} \in \mathcal{P}(\mu_0, \tau, u)} \mathbb{E}^{\mathbb{P}} \left[\int_0^T Y_t F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}}) \, \mathrm{d}t \right] = \inf_{\mathbb{P} \in \widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0, \tau, u)} \mathbb{E}^{\mathbb{P}} \left[\int_0^T Y_t F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}}) \, \mathrm{d}t \right]. \tag{2.2.2}$$

While the appearance of Y in the constraints in (2.2.1) is clearly motivated by the fact that the payoff must be discounted to give the monetary value of an option today, this is not the case in (2.2.2). Ultimately this is present as an analytical necessity to apply our conditioning arguments later on to eliminate the path dependent discount factor. If the constraints in (2.2.1) were instead given by $\mathbb{E}^{\mathbb{P}}[g(Y_{\tau_i}) G(X_{\tau_i})] = u_i$, then we would replace Y_t with $g(Y_t)$ in (2.2.2).

2.2.1 First Markovian reduction

Similarly to I. Guo, Loeper, and Shiyi Wang, 2022, Proposition 3.4, since the market constraints only depend on the marginal distributions and the cost functional is convex, a com-

bination of Markovian projection in Theorem 2.1.3 and Jensen's inequality readily shows that we can restrict our attention to Markov process in (t, X_t, Y_t).

Proof of Proposition 2.2.2 (Part I). Note that $\widetilde{\mathcal{P}}_{\text{loc}}(\mu_0, \tau, u) \subseteq \mathcal{P}(\mu_0, \tau, u)$ so the "$\leqslant$" is trivial and also if $\mathcal{P}(\mu_0, \tau, u)$ is empty then both sides are equal $+\infty$. We show now the reverse inequality but with $\mathcal{P}_{\text{loc}}(\mu_0, \tau, u)$ replacing $\widetilde{\mathcal{P}}_{\text{loc}}(\mu_0, \tau, u)$. Take $\mathbb{P} \in \mathcal{P}(\mu_0, \tau, u)$ and use Theorem 2.1.3, and Remark 2.1.5, to find the corresponding $\tilde{\mathbb{P}} \in \mathcal{P}_{\text{loc}}(\mu_0, \tau, u)$ such that (X, Y) has the same marginals under $\mathbb{P}$ and $\tilde{\mathbb{P}}$. Using the tower property and Jensen's inequality we have:

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\int_0^T Y_t F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})\ \mathrm{d}t\right] &= \mathbb{E}^{\mathbb{P}}\left[\int_0^T Y_t \mathbb{E}^{\mathbb{P}}_{t,X_t,Y_t}\left[F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})\right]\ \mathrm{d}t\right] \\
&\geqslant \mathbb{E}^{\mathbb{P}}\left[\int_0^T Y_t F(\mathbb{E}^{\mathbb{P}}_{t,X_t,Y_t}[\alpha_t^{\mathbb{P}}], \mathbb{E}^{\mathbb{P}}_{t,X_t,Y_t}[\beta_t^{\mathbb{P}}])\ \mathrm{d}t\right] \\
&= \mathbb{E}^{\tilde{\mathbb{P}}}\left[\int_0^T Y_t F(\alpha_t^{\tilde{\mathbb{P}}}, \beta_t^{\tilde{\mathbb{P}}})\ \mathrm{d}t\right]. \qquad (2.2.3)
\end{aligned}$$

This gives the desired inequality. □

Remark 2.2.3 (A caveat on generality)**.** *In the sense of solving the calibration problem — i.e. finding a model $\mathbb{P}$ that minimises a cost function subject to $\mathbb{P} \in \mathcal{P}(\mu_0, \tau, u)$, we can consider the restriction to Markov models to be without loss of generality. It is clear that the mimicking of the marginals of the Itô process would not be sufficient for the calibration of path-dependent derivatives, as they would additionally require their own state variable if expressible as an updating function. Indeed, this projection argument would certainly not suffice for derivatives with early exercise features. In addition, the Greeks may not be preserved in the Markovian projection, rendering the model unsuitable for hedging. Clearly, the model will need recalibration as up-to-date market data becomes available, however prior calibrated models can be used as a starting point, or reference, to mitigate this drawback.*

2.2.2 A 'discounted density' transformation and superposition principle

While we could easily reduce to Markovian models in the state process (X_t, Y_t), this is not satisfactory. In particular, in Chapter 3 where we consider a local volatility model with

a short rate, we would obtain a three dimensional fully nonlinear PDE instead of a two dimensional one, which involves a substantially increased computational effort. We thus want to further reduce to Markov process X, i.e., to obtain (2.2.2). This requires novel arguments to deal with the stochastic discount term. Instead of working with probability measures, we will work with discounted densities, i.e., sub-probability measures. When a $\mathbb{P} \in \mathcal{P}^1_{\text{loc}}$ is fixed, we write ρ, α, β, for $\rho^{\mathbb{P}}, \alpha^{\mathbb{P}}, \beta^{\mathbb{P}}$.

Definition 2.2.4. *Let $\mathbb{P} \in \mathcal{P}^1_{\text{loc}}$ with the semimartingale characteristics of X given by $\alpha^{\mathbb{P}}_t = \alpha_t(X_t, Y_t)$, $\beta^{\mathbb{P}}_t = \beta_t(X_t, Y_t)$. Let $\rho_t(\mathrm{d}x, \mathrm{d}y)$ and $\rho^X(\mathrm{d}x)$ be, respectively, the marginal distribution of (X_t, Y_t) and of X_t , $t \in [0, T]$. Let $\rho_t(\mathrm{d}x, \mathrm{d}y) = \zeta_{t,x}(\mathrm{d}y)\rho^X(\mathrm{d}x)$, i.e., $\zeta_{t,x}$ is the law of Y_t conditional on $\{X_t = x\}$. Define the 'discounted' density, drift and volatility for $(t, x) \in [0, T] \times \mathbb{R}^d$:*

$$\rho^D_t(\mathrm{d}x) = D_t(x)\rho^X(\mathrm{d}x), \qquad \text{where} \quad D_t(x) = \int_{\mathbb{R}} y\zeta_{t,x}(\mathrm{d}y), \tag{2.2.4}$$

$$\alpha^D_t(x) = \frac{\mathbb{E}[Y_t\alpha_t(X_t, Y_t)|X_t = x]}{\mathbb{E}[Y_t|X_t = x]} = \int_{\mathbb{R}} y\alpha_t(x, y)\frac{\zeta_{t,x}(\mathrm{d}y)}{D_t(x)}, \tag{2.2.5}$$

$$\beta^D_t(x) = \frac{\mathbb{E}[Y_t\beta_t(X_t, Y_t)|X_t = x]}{\mathbb{E}[Y_t|X_t = x]} = \int_{\mathbb{R}} y\beta_t(x, y)\frac{\zeta_{t,x}(\mathrm{d}y)}{D_t(x)}. \tag{2.2.6}$$

By construction, ρ^D_t inherits narrow continuity from ρ_t. In addition, using (2.1.4) and (2.2.4)–(2.2.6), we have

$$\begin{aligned}\int_0^T \int |\alpha^D(x)| + |\beta^D(x)|\rho^D_t(\mathrm{d}x)\,\mathrm{d}t &= \int_0^T \int |y\alpha(x, y)| + |y\beta(x, y)|\rho_t(\mathrm{d}x, \mathrm{d}y)\,\mathrm{d}t \\ &\leqslant e^T\mathbb{E}^{\mathbb{P}}\left[\int_0^T |\alpha^{\mathbb{P}}_t| + |\beta^{\mathbb{P}}_t|\,\mathrm{d}t\right] < +\infty, \end{aligned} \tag{2.2.7}$$

so that α^D and β^D are $\mathrm{d}t \times \rho^D$-integrable.

Remark 2.2.5. *Here, the assumption on the interest rate being bounded below is necessary since Y must be bounded as α^D and β^D must be integrable under any discounted density corresponding to $\mathbb{P} \in \mathcal{P}_{\text{loc}}(\mu_0, \tau, u)$. The constant chosen for the interest rate was -1, corresponding to $Y_t < e^T$, however this can of course be relaxed so long as $Y_t \in L^\infty$ for all $t \in [0, T]$ under all models $\mathbb{P} \in \mathcal{P}_{\text{loc}}(\mu_0, \tau, u)$.*

A key tool in the dimension reduction is a "discounted" version of the superposition

principle of Trevisan, 2016, Theorem 2.5. We first define the augmented martingale problem.

Definition 2.2.6 (Augmented Martingale Problem)**.** *Let $f \in C_b^{1,2}([0,T] \times \mathbb{R}^{d+1}; \mathbb{R})$, and given measurable functions $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $b : [0,T] \times \mathbb{R}^d \to \mathbb{S}_+^d$, and $c : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, define the operator*

$$\mathcal{L}f = a \cdot \nabla_x f + \frac{1}{2} b : \nabla_x^2 f - yc\partial_y f.$$

Then $\boldsymbol{\eta} \in \mathcal{P}(C([0,T]; \mathbb{R}^{d+1}))$ is a solution to the augmented martingale problem, $\mathrm{aMP}(a,b,c)$ *if*

$$\int \int_0^T (|a_t| \circ e_t + |b_t| \circ e_t + |c_t| \circ e_t) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{\eta} < \infty \tag{2.2.8}$$

and for all $f \in C_b^{1,2}([0,T] \times \mathbb{R}^{d+1}; \mathbb{R})$, the process

$$[0,T] \ni t \mapsto f(t,\cdot) \circ e_t - \int_0^t (\partial_t f(s,\cdot) + (\mathcal{L}f)(s,\cdot)) \circ e_s \, \mathrm{d}s \tag{2.2.9}$$

is a martingale with respect to the natural filtration on $C([0,T]; \mathbb{R}^{d+1})$.

Theorem 2.2.7 (Discounted Superposition Principle)**.** *Let $\boldsymbol{\nu} = (\nu_t)_{t \in [0,T]} \in C([0,T]; \mathcal{M}(\mathbb{R}^d))$ be a narrowly continuous solution of the general discounted Fokker-Planck equation on $[0,T] \times \mathbb{R}^d$:*

$$\partial_t \nu_t + \nabla_x \cdot (a_t(x)\nu_t(x)) - \frac{1}{2} \nabla_x^2 : (b_t(x)\nu_t(x)) + c_t(x)\nu_t(x) = 0, \tag{2.2.10}$$

where $a \in L^1(\mathrm{d}|\nu_t| \, \mathrm{d}t; \mathbb{R}^d)$, $b \in L^1(\mathrm{d}|\nu_t| \, \mathrm{d}t; \mathbb{S}_+^d)$, and $c \in L^1(\mathrm{d}|\nu_t| \, \mathrm{d}t; \mathbb{R})$. Then there exists $\boldsymbol{\eta} \in \mathcal{P}(C([0,T]; \mathbb{R}^{d+1}))$ a solution to the $\mathrm{aMP}(a,b,c)$*, such that the discounted version of its narrowly continuous curve of marginals $\eta_t = \boldsymbol{\eta} \circ e_t^{-1}$ coincides with ν_t, that is $\int_{\mathbb{R}} y\eta_t(\cdot, \mathrm{d}y) = \nu_t$ for all $t \in [0,T]$.*

The proof of Theorem 2.2.7 follows closely the steps in the proof of Trevisan, 2016, Theorem 2.5, which itself has a similar structure to Ambrosio, Gigli, and Savaré, 2005, Theorem 8.2.1, Ambrosio and Figalli, 2009, Theorem 4.5, Figalli, 2008, Theorem 2.6, Ambrosio and Trevisan, 2014, Theorem 7.1. The idea of the proof is to demonstrate the result holds on smooth coefficients a, b, c, and to then provide suitable smooth sequences on which we can show compactness and convergence (in a suitable sense). Using this machinery,

we can generalise the regularity of the coefficients successively up to the case of integrable coefficients.

We defer the proof to Section 2.3, but summarise the main steps here. Section 2.3.1 shows the result in the smooth case, and constructs smooth approximating sequences in an analogous way to Trevisan, 2016. Section 2.3.2 asserts the compactness of solutions for the augmented martingale problem as a direct consequence of Trevisan, 2016, Section A.2, with the proof of convergence an easy modification of Trevisan, 2016, Section A.3. For the generalisation to smooth and bounded coefficients, the adaptation of the proof is easy with the existence of a solution to aMP(a, b, c) with a given initial law following directly from the Trevisan, 2016, Theorem A.6, and uniqueness following from a simple adaptation of Trevisan, 2016, Theorem A.7, with the approach in obtaining estimates an easy modification. The generalisation to bounded coefficients relies on the de la Vallée Poussin criterion and Trevisan, 2016, Corollary A.5, with our modification again being the handling of the c term — here we rely on the already established tightness and superposition results for the augmented martingale problem (as it is just the usual martingale problem for (X, Y)), and the fact that Y is strictly positive. The estimates for the x-coordinates are obtained in the same way as Trevisan, 2016, and the estimates for the y-coordinate use the function $\log(y\chi_R)$ where χ_R is some cutoff function in order to eliminate the y term in the augmented diffusion operator. The locally bounded coefficients case then follows a similar modified approach to the bounded coefficients case, as in Trevisan, 2016.

2.2.3 Second Markovian reduction

We use now the tools introduced above to finish the proof of Proposition 2.2.2.

Proposition 2.2.8. *Let $\mathbb{P} \in \mathcal{P}^1_{loc}$ and $\rho^D, \alpha^D, \beta^D$ be given as in Definition 2.2.4. Then ρ^D is a weak solution to the 'discounted' version of the Fokker-Planck equation for $(t, x) \in [0, T] \times \mathbb{R}^d$:*

$$\partial_t \rho^D_t(x) + \nabla_x \cdot \left(\alpha^D_t(x)\rho^D_t(x)\right) - \frac{1}{2}\nabla^2_x : \left(\beta^D_t(x)\rho^D_t(x)\right) + x_r \rho^D_t(x) = 0. \tag{2.2.11}$$

Moreover, there exists a local volatility model $\tilde{\mathbb{P}} \in \mathcal{P}^1$ under which the semimartingale char-

acteristics of X are given by $\alpha_t^{\tilde{\mathbb{P}}} = \alpha_t^D(X_t)$ and $\beta_t^{\tilde{\mathbb{P}}} = \beta_t^D(X_t)$ $\mathrm{d}t \times \mathrm{d}\tilde{\mathbb{P}}$- a.e. and such that for any payoff function $G \in C_b(\mathbb{R}^d;\mathbb{R})$ and any $t \in [0,T]$:

$$\mathbb{E}^{\mathbb{P}}[Y_t G(X_t)] = \mathbb{E}^{\tilde{\mathbb{P}}}[Y_t G(X_t)]. \tag{2.2.12}$$

Furthermore, the value of the objective function decreases:

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^T Y_t F(\alpha_t(X_t,Y_t),\beta_t(X_t,Y_t))\,\mathrm{d}t\right] \geqslant \mathbb{E}^{\tilde{\mathbb{P}}}\left[\int_0^T Y_t F(\alpha_t^D(X_t)\beta_t^D(X_t))\,\mathrm{d}t\right]. \tag{2.2.13}$$

The notion of solution to (2.2.11) is in the spirit of Definition 1.2.6 but with an extra $-x_r f$ term in the diffusion operator of Definition 1.2.4.

Proof of Proposition 2.2.2 (Part II). Proposition 2.2.8 instantly gives us that

$$\inf_{\mathbb{P}\in\mathcal{P}_{\mathrm{loc}}(\mu_0,\tau,u)} \mathbb{E}^{\mathbb{P}}\left[\int_0^T Y_t F(\alpha_t^{\mathbb{P}},\beta_t^{\mathbb{P}})\,\mathrm{d}t\right] \geqslant \inf_{\mathbb{P}\in\widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0,\tau,u)} \mathbb{E}^{\mathbb{P}}\left[\int_0^T Y_t F(\alpha_t^{\mathbb{P}},\beta_t^{\mathbb{P}})\,\mathrm{d}t\right], \tag{2.2.14}$$

and hence we have equality since $\widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0,\tau,u) \subseteq \mathcal{P}_{\mathrm{loc}}(\mu_0,\tau,u)$, which completes the proof of Proposition 2.2.2. □

Proof of Proposition 2.2.8. Recall the notation of Definition 2.2.4. Since ρ_t solves (2.1.9), we have for $(t,x) \in (0,T)\times\mathbb{R}^d$:

$$\int_{\mathbb{R}} y\left(\partial_t\rho_t(x,y) + \nabla_x\cdot(\rho_t(x,y)\alpha_t(x,y)) - \frac{1}{2}\nabla_x^2 : (\rho_t(x,y)\beta_t(x,y)) - \partial_y(x_r y\rho_t(x,y))\right)\mathrm{d}y = 0. \tag{2.2.15}$$

By construction, the first term is given by $\int_{\mathbb{R}} y\partial_t\rho_t(x,y)\,\mathrm{d}y = \partial_t\rho_t^D(x)$. Then, for the drift we have

$$\begin{aligned}
\int_{\mathbb{R}} y\nabla_x\cdot(\rho_t(x,y)\alpha_t(x,y))\,\mathrm{d}y &= \nabla_x\cdot\left(\int_{\mathbb{R}}\alpha_t(x,y)\frac{y}{D_t(x)}\zeta_{t,x}(y)\rho_t^X(x)D_t(x)\,\mathrm{d}y\right),\\
&= \nabla_x\cdot\left(\frac{\int_{\mathbb{R}}\alpha_t(x,y)y\zeta_{t,x}(y)\,\mathrm{d}y}{D_t(x)}\rho_t^X(x)D_t(x)\right),\\
&= \nabla_x\cdot(\rho_t^D(x)\alpha_t^D(x)).
\end{aligned} \tag{2.2.16}$$

The same calculation gives us

$$\int_{\mathbb{R}} y\nabla_x^2 : (\rho_t(x,y)\beta_t(x,y))\,\mathrm{d}y = \nabla_x^2 : (\rho_t^D(x)\beta_t^D(x)). \tag{2.2.17}$$

Finally, integration by parts gives $\int_{\mathbb{R}} y\partial_y(x_r y\rho_t(x,y))\,\mathrm{d}y = -x_r\rho_t^D(x)$, so we have that ρ_t^D solves (2.2.11). Recall also the integrability condition (2.2.7) holds. We can thus apply Theorem 2.2.7 with $\nu = \rho^D$ to obtain another probability measure, $\tilde{\boldsymbol{\rho}} \in \mathcal{P}(C([0,T];\mathbb{R}^{d+1}))$ solving aMP(α^D, β^D, x_r) such that $\tilde{\rho}_t^D(\cdot) = \int_{\mathbb{R}} y\tilde{\rho}_t(\cdot,\mathrm{d}y) = \rho^D(\cdot)$. By I. Guo and Loeper, 2021, Lemma 3.4, for the canonical process $(\tilde{X},\tilde{Y})$ under $\tilde{\boldsymbol{\rho}}$, $\tilde{X}$ is a local volatility process with the semimartingale characteristics $\alpha_t^D(\tilde{X}_t)$ and $\beta_t^{\tilde{\mathbb{P}}} = \beta_t^D(\tilde{X}_t)$, and further $\tilde{Y}_t = \tilde{Y}_0 \exp\left(-\int_0^t \tilde{X}_s^r\,\mathrm{d}s\right)$, $\mathrm{d}t \times \mathrm{d}\tilde{\boldsymbol{\rho}}$-a.e. We let $\tilde{\mathbb{P}}$ be the projection of $\tilde{\boldsymbol{\rho}}$ on the first d coordinates. Note that $\tilde{\mathbb{P}} \in \mathcal{P}^1$ by (2.2.8), X has the desired semimartingale characteristics under $\tilde{\mathbb{P}}$ and also $\tilde{\rho}_t$ is the distribution of (X_t, Y_t) under $\tilde{\mathbb{P}}$. Then, given a payoff function $G \in C_b(\mathbb{R}^d;\mathbb{R})$, we have for any fixed $t \in [0,T]$:

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}[Y_t G(X_t)] = \int_{\mathbb{R}^{d+1}} yG(x)\rho_t(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \int_{\mathbb{R}^d} G(x)\rho_t^D(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}^d} G(x)\tilde{\rho}_t^D(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}^{d+1}} yG(x)\tilde{\rho}_t(x,y)\,\mathrm{d}x\,\mathrm{d}y = \mathbb{E}^{\tilde{\mathbb{P}}}[Y_t G(X_t)],
\end{aligned} \tag{2.2.18}$$

so (2.2.12) holds. Finally, we also have

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\left[\int_0^T Y_t F(\alpha_t(X_t,Y_t),\beta_t(X_t,Y_t))\,\mathrm{d}t\right] \\
&\qquad = \int_0^T \int_{\mathbb{R}^{d+1}} yF(\alpha_t(x,y),\beta_t(x,y))\zeta_{t,x}(\mathrm{d}y)\rho_t^X(\mathrm{d}x)\,\mathrm{d}t, \\
&\qquad = \int_0^T \int_{\mathbb{R}^{d+1}} F(\alpha_t(x,y),\beta_t(x,y))\frac{y}{D_t(x)}\zeta_{t,x}(\mathrm{d}y)\,D_t(x)\rho_t^X(\mathrm{d}x)\,\mathrm{d}t, \\
&\qquad \geqslant \int_0^T \int_{\mathbb{R}^d} F\left(\frac{\int_{\mathbb{R}} y\alpha_t(x,y)\zeta_{t,x}(\mathrm{d}y)}{D_t(x)}, \frac{\int_{\mathbb{R}} y\beta_t(x,y)\zeta_{t,x}(\mathrm{d}y)}{D_t(x)}\right) D_t(x)\rho_t^X(\mathrm{d}x)\,\mathrm{d}t, \\
&\qquad = \int_0^T \int_{\mathbb{R}^d} F(\alpha_t^D(x),\beta_t^D(x))\rho_t^D(\mathrm{d}x)\,\mathrm{d}t, \\
&\qquad = \int_0^T \int_{\mathbb{R}^d} F(\alpha_t^D(x),\beta_t^D(x))\tilde{\rho}_t^D(\mathrm{d}x)\,\mathrm{d}t,
\end{aligned} \tag{2.2.19}$$

$$= \mathbb{E}^{\tilde{\mathbb{P}}}\left[\int_0^T Y_t F(\alpha_t^D(X_t), \beta_t^D(X_t))\,\mathrm{d}t\right]. \tag{2.2.20}$$

□

2.3 Proof of the Discounted Superposition Principle

First, we define two operators associated with the discounted Fokker-Planck equation (2.2.11) and the augmented Fokker-Planck equation (2.1.9). Given $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $b : [0,T] \times \mathbb{R}^d \to \mathbb{S}^d_+$, $c : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, define:

$$C([0,T];\mathbb{R}^d) \ni f \mapsto \mathcal{L}^D f = a \cdot \nabla_x f + \frac{1}{2} b : \nabla_x^2 f - cf, \tag{2.3.1}$$

$$C([0,T];\mathbb{R}^{d+1}) \ni f \mapsto \mathcal{L}^A f = a \cdot \nabla_x f + \frac{1}{2} b : \nabla_x^2 f + yc\partial_y f. \tag{2.3.2}$$

Note that if $(\nu_t) \subset \mathcal{P}(\mathbb{R}^{d+1})$ solves the augmented Fokker-Planck equation associated with $\mathcal{L}^A$, denoted aFP(a,b,c), then $(\nu_t^D) \subset \mathcal{M}(\mathbb{R}^d)$ with $\nu_t^D = \int_{\mathbb{R}} y\nu_t(\cdot, \mathrm{d}y)$ solves the discounted Fokker-Planck equation associated with $\mathcal{L}^D$, denoted dFP(a,b,c). This is easily seen since the coefficients a, b, c have no y-dependence.

We construct a sequence of smooth solutions to (2.2.11), $(\nu_t^n)_{t\in[0,T]} \subset \mathcal{M}(\mathbb{R}^d)$, with coefficients $a^n \in C^\infty([0,T]\times\mathbb{R}^d;\mathbb{R}^d)$, $b^n \in C^\infty([0,T]\times\mathbb{R}^d;\mathbb{S}^d_+)$, and $c^n \in C^\infty([0,T]\times\mathbb{R}^d;\mathbb{R})$ for which we can easily show Theorem 2.2.7. We therefore can construct a sequence of diffusion operators $\mathcal{L}^{A,n}$ corresponding to aMP(a^n, b^n, c^n) which has solution $\boldsymbol{\eta}^n$ such that $\int_{\mathbb{R}} y\eta_t^n(\cdot, \mathrm{d}y) = \nu_t^n$. The tightness and convergence of $\boldsymbol{\eta}^n$ is then shown, and then we use the arguments to generalise the coefficients a, b, c.

2.3.1 Smooth Approximation of the Augmented Martingale Problem

Lemma 2.3.1. *Let ν_t be a smooth solution of* dFP(a,b,c) *with coefficients $a \in C^\infty([0,T]\times\mathbb{R}^d;\mathbb{R}^d)$, $b \in C^\infty([0,T]\times\mathbb{R}^d;\mathbb{S}^d_+)$, $c \in C^\infty([0,T]\times\mathbb{R}^d;\mathbb{R})$. Then there exists a solution $\boldsymbol{\eta}$ to* aMP(a,b,c) *such that $\int_{\mathbb{R}} y\eta_t(\cdot, \mathrm{d}y) = \nu(\cdot)$.*

Proof. Let $\mathbb{P} \in \mathcal{P}_1$ and define the process:

$$\begin{cases} \mathrm{d}\tilde{X}_t = a_t(\tilde{X}_t)\,\mathrm{d}t + (b_t(\tilde{X}_t))^{\frac{1}{2}}\,\mathrm{d}W_t^{\tilde{\mathbb{P}}}, \quad 0 \leqslant t \leqslant T, \\ \tilde{X}_0 = x_0. \end{cases} \tag{2.3.3}$$

Since a and b are taken to be smooth, we automatically have existence and uniqueness of (2.3.3) (see Stroock and Varadhan, 1979, Corollary 6.3.3). Define $\tilde{Y}_t = \exp\left(-\int_0^t c_s(\tilde{X}_s)\,\mathrm{d}s\right)$. Then, by the same reasoning as Theorem 2.1.3, we have that the law of the augmented process $(\tilde{X}, \tilde{Y})$ given by η solves the Fokker-Planck equation for $(t, x, y) \in [0, T] \times \mathbb{R}^{d+1}$.

$$\partial_t \eta_t(x, y) + \nabla_x \cdot (\eta_t(x, y) a_t(x)) - \frac{1}{2}\nabla_x^2 : (\eta_t(x, y) b_t(x)) - \partial_y(y c_t(x) \eta_t(x, y)) = 0. \tag{2.3.4}$$

We now define the discounted version of η by $\eta^D(t, x) \coloneqq \int_{\mathbb{R}} y\eta(t, x, y)\,\mathrm{d}y$, then by a similar calculation to that in the proof of Lemma 2.2.8, we have that η^D solves the discounted Fokker-Planck equation (2.2.11). By a similar argument to Figalli, 2008, Proposition 4.1, we have uniqueness of (2.2.11), so $\eta_t^D = \nu_t$. By Itô's formula, it is immediate that $\boldsymbol{\eta} \in \mathcal{P}(C([0, T]; \mathbb{R}^{d+1}))$ solves aMP(a, b, c), so the result follows. □

We let $\nu = (\nu_t)_{t\in[0,T]} \subset \mathcal{M}(\mathbb{R}^d)$ be a narrowly continuous solution of dFP(a, b, c) for $a \in L^1(\mathrm{d}\nu_t\,\mathrm{d}t; \mathbb{R}^d)$, $b \in L^1(\mathrm{d}\nu_t\,\mathrm{d}t; \mathbb{S}^d)$, and $c \in L^1(\mathrm{d}\nu_t\,\mathrm{d}t; \mathbb{R})$. We first build a smooth sequence ν^n approximating ν, associated with diffusion operators $\mathcal{L}^{D,n}$. Then by Lemma 2.3.1, there exists a sequence of solutions to the augmented martingale problem, $\boldsymbol{\eta}^n$, associated with $\mathcal{L}^{A,n}$, such that $\int_{\mathbb{R}} y\eta_t^n(\cdot, \mathrm{d}y) = \nu_t^n$. As in Trevisan, 2016, Section A.1, we construct the approximations in two different ways for the proof of the general case.

Pushforward via Smooth Maps:

Let $\pi = (\pi^1, \ldots, \pi^d) \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ with uniformly bounded first and second derivatives. Define the standard diffusion operator, $\mathcal{L}^S$ by $C([0, T]; \mathbb{R}^d) \ni f \mapsto \mathcal{L}^S = a \cdot \nabla_x f + \frac{1}{2} b : \nabla_x^2 f$, then we can define the approximations of (a, b, c) by:

$$\pi(a)_t^i = \frac{\mathrm{d}\pi_\#[\mathcal{L}^S(\pi^i)\nu_t]}{\mathrm{d}\pi_\#\nu_t}, \quad \pi(b)_t^{i,j} \coloneqq \frac{\mathrm{d}\pi_\#[b_t : (\nabla\pi^i \otimes \nabla\pi^j)\nu_t]}{\mathrm{d}\pi_\#\nu_t}, \quad \pi(c)_t \coloneqq \frac{\mathrm{d}\pi_\#[c_t\nu_t]}{\mathrm{d}\pi_\#\nu_t}.$$

Note that π preserves uniform bounds on a, b, c. Additionally, by an application of the chain rule to compute $\mathcal{L}^D(f \circ \pi)$, it is easy to see that $\pi_{\#}\nu$ solves dFP$(\pi(a), \pi(b), \pi(c))$, and therefore by Lemma 2.3.1 there exists $\boldsymbol{\eta}$ a solution to aMP$(\pi(a), \pi(b), \pi(c))$ such that $\pi_{\#}\nu_t = \int_{\mathbb{R}} y\eta_t(\cdot, \mathrm{d}y)$.

Mollification by Convolutions:

Let $\kappa \geqslant 0$ be a smooth probability density on $\mathbb{R}^d$. Define

$$a_t^\kappa := \frac{\mathrm{d}((a_t\nu_t) * \kappa)}{\mathrm{d}(\nu_t * \kappa)}, \qquad b_t^\kappa := \frac{\mathrm{d}((b_t\nu_t) * \kappa)}{\mathrm{d}(\nu_t * \kappa)}, \qquad c_t^\kappa := \frac{\mathrm{d}((c_t\nu_t) * \kappa)}{\mathrm{d}(\nu_t * \kappa)} \tag{2.3.5}$$

By Trevisan, 2016, Lemma A.1 we have that a^κ, b^κ, and c^κ preserve integrability, and have uniformly bounded first and second spatial derivatives on compact sets. Therefore, $\nu_t * \kappa$ solves dFP$(a^\kappa, b^\kappa, c^\kappa)$. Additionally, by Lemma 2.3.1 there exists $\boldsymbol{\eta}$ a solution to aMP$(a^\kappa, b^\kappa, c^\kappa)$ such that $\nu_t * \kappa = \int_{\mathbb{R}} y\eta_t(\cdot, \mathrm{d}y)$.

2.3.2 Compactness and Convergence of the Augmented Martingale Problem

We now assume that $\boldsymbol{\eta}^n$ is a sequence of superposition solutions corresponding to an approximation $\nu^n = \pi^n_{\#}\nu$ in the pushforward case or $\nu * \kappa^n$ in the mollification case. In the mollification case, we assume that $\kappa^n \to \delta_0$ narrowly as $n \to \infty$. In the pushforward case we assume that $\pi_n \to \mathrm{Id}$ locally uniformly, and $\nabla\pi^n \to \mathrm{Id}$, and $\nabla^2\pi^n \to 0$ pointwise. Assume also that the sequences of derivatives are uniformly bounded so there exists $C \geqslant 0$ such that $|\nabla\pi^n|, |\nabla^2\pi^n| \leqslant C$. Compactness for solutions of the augmented martingale problem then follows directly from Trevisan, 2016, Section A.2 since we could recast the superposition principle in the setting of Trevisan, 2016, Theorem 2.5 by considering superposition solutions to aFP(a, b, c) and then the tightness arguments are directly applicable.

We are looking for solutions of aMP(a, b, c) with a discounted density coinciding with a given solution of dFP(a, b, c). We therefore consider the class of test functions $C_b^{1,2}([0, T] \times \mathbb{R}^{d+1}; \mathbb{R}) \ni \tilde{f}(x, y) = yf(x)$ where $f \in C_b^{1,2}([0, T] \times \mathbb{R}^d; \mathbb{R})$. We therefore note that we always

have

$$\int \int_s^t (\mathcal{L}_r^A \tilde{f}) \circ e_r \, \mathrm{d}r \, \mathrm{d}\boldsymbol{\eta} = \int_s^t \int_{\mathbb{R}^d} (\mathcal{L}_r^D f_r) \int_{\mathbb{R}} y \eta_r(\mathrm{d}x, \mathrm{d}y) \, \mathrm{d}r. \tag{2.3.6}$$

Now, let $\overline{\mathcal{L}^D}$ be a discounted diffusion operator with coeffifients $\overline{a}, \overline{b}, \overline{c}$ continuous and compactly supported. In view of the arguments of Trevisan, 2016, Section A.3, due to narrow convergence of $\boldsymbol{\eta}^n$, we only need to provide bounds on

$$\limsup_{n\to\infty} \int |\mathcal{L}^{D,n} f - \overline{\mathcal{L}^D} f| \, \mathrm{d}\nu^n + \int |\mathcal{L}^D f - \overline{\mathcal{L}^D} f| \, \mathrm{d}\nu. \tag{2.3.7}$$

Pushforward via Smooth Maps:

Write $\pi(\mathcal{L}^D)$ for the discounted diffusion operator associated with coefficients $(\pi(a), \pi(b), \pi(c))$. Then for a fixed n, we have

$$\begin{aligned}
\int |\pi^n(\mathcal{L}^{D,n}) f - \overline{\mathcal{L}^D} f| \, \mathrm{d}\pi^n_{\#}\nu &\leqslant \int |\mathcal{L}^{D,n}(f \circ \pi^n) - (\overline{\mathcal{L}^D} f) \circ \pi^n| \, \mathrm{d}\nu, \\
&\leqslant \int \frac{1}{2} \sum_{i,j=1}^d |a : (\nabla \pi_i \otimes \nabla \pi_j) - \overline{a}^{i,j} \circ \pi| \, \mathrm{d}\nu \\
&\quad + \int \sum_{i=1}^d |\mathcal{L}^S(\pi_i) - \overline{a}^i \circ \pi| \, \mathrm{d}\nu + \int |c - \overline{c} \circ \pi| \, \mathrm{d}\nu.
\end{aligned} \tag{2.3.8}$$

Since π^n converges to the identity, we have that (2.3.7) is bounded above by

$$\int \sum_{i,j=1}^d |a : (\nabla \pi_i \otimes \nabla \pi_j) - \overline{a}^{i,j} \circ \pi| \, \mathrm{d}\nu + 2 \int \sum_{i=1}^d |\mathcal{L}^S(\pi_i) - \overline{a}^i \circ \pi| \, \mathrm{d}\nu + 2 \int |c - \overline{c} \circ \pi| \, \mathrm{d}\nu. \tag{2.3.9}$$

Since continuous and compactly supported functions are dense in $L^1(\nu)$, the above can be made arbitrarily small by optimising over $\overline{a}, \overline{b}, \overline{c}$.

Mollification by Convolution:

Let $\overline{\omega}$ be a common bounded and continuous modulus of continuity for $\overline{a}, \overline{b}, \overline{c}$. Let $\overline{\mathcal{L}^{D,n}}$ be the discounted diffusion operator associated with coefficients:

$$\overline{a}^n := \frac{\mathrm{d}(\overline{a}\nu * \kappa^n)}{\mathrm{d}(\nu * \kappa^n)}, \qquad \overline{b}^n := \frac{\mathrm{d}(\overline{b}\nu * \kappa^n)}{\mathrm{d}(\nu * \kappa^n)}, \qquad \overline{c}^n := \frac{\mathrm{d}(\overline{c}\nu * \kappa^n)}{\mathrm{d}(\nu * \kappa^n)}. \tag{2.3.10}$$

We first show $\lim_{n\to\infty}\int|\overline{\mathcal{L}^{D,n}}f-\overline{\mathcal{L}^{D}}f|\,\mathrm{d}\nu^n=0$. Since $||f||_{C^{1,2}}\leqslant 1$, we have

$$\begin{aligned}\int|\overline{\mathcal{L}^{D,n}}f-\overline{\mathcal{L}^{D}}f|\,\mathrm{d}\nu^n&\leqslant\int|\overline{a}^n-\overline{a}|\,\mathrm{d}\nu^n+\int|\overline{b}^n-\overline{b}|\,\mathrm{d}\nu^n+\int|\overline{c}^n-\overline{c}|\,\mathrm{d}\nu^n,\\&\leqslant 3\int\overline{\omega}(z)\kappa^n(z)\,\mathrm{d}z\to 0.\end{aligned}\tag{2.3.11}$$

Therefore, we have

$$\limsup_{n\to\infty}\int|\mathcal{L}^{D,n}f-\overline{\mathcal{L}^{D}}f|\,\mathrm{d}\nu^n=\limsup_{n\to\infty}\int|\mathcal{L}^{D,n}f-\overline{\mathcal{L}^{D,n}}f|\,\mathrm{d}\nu^n\leqslant\int|a-\overline{a}|+|b-\overline{b}|+|c-\overline{c}|\,\mathrm{d}\nu.\tag{2.3.12}$$

By the same argument as before, we can make the right hand side arbitrarily small which concludes the proof.

2.3.3 Generalisation of Drift and Diffusion Coefficients

We now use the approximation and convergence described in Section 2.3.1 and Section 2.3.2 to generalise the coefficients a, b and prove Theorem 2.2.7.

2.3.3.1 Smooth and Bounded Coefficients:

Let $a:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$, $b:[0,T]\times\mathbb{R}^d\to\mathbb{S}^d_+$, and $c:[0,T]\times\mathbb{R}^d\to\mathbb{R}$ be Borel maps such that

$$\int_0^T||a_t||_{C_b^2(\mathbb{R}^d)}+||b_t||_{C_b^2(\mathbb{R}^d)}+||c_t||_{C_b^2(\mathbb{R}^d)}\,\mathrm{d}t<\infty.\tag{2.3.13}$$

Lemma 2.3.2. *Let a, b, c be Borel maps satisfying* (2.3.13)*. Then for every $\overline{\nu}\in\mathcal{P}(\mathbb{R}^d)$ there exists $\boldsymbol{\eta}\in\mathcal{P}(C([0,T];\mathbb{R}^{d+1}))$ solving* aMP(a,b,c) *such that $\int_{\mathbb{R}}y\eta_0(\cdot,\mathrm{d}y)=\overline{\nu}(\cdot)$.*

The proof is the same as Trevisan, 2016, Theorem A.6, but by considering the augmented process (X, Y) where the regular martingale problem (see Trevisan, 2016, Definition 2.4) is given by aMP(a, b, c).

Lemma 2.3.3. *Let $a:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$, $b:[0,T]\times\mathbb{R}^d\to\mathbb{S}^d_+$, and $c:[0,T]\times\mathbb{R}^d\to\mathbb{R}$ be Borel maps such that*

$$\int_0^T||a_t||_{C^2(B)}+||b_t||_{C^2(B)}+||c_t||_{C^2(B)}\,\mathrm{d}t<\infty,\quad\text{for every bounded open } B\subset\mathbb{R}^d\tag{2.3.14}$$

Let $\nu = (\nu_t)_{t\in[0,T]} \subset \mathcal{M}(\mathbb{R}^d)$ be a narrowly continuous solution of $\mathrm{dFP}(a,b,c)$. *If $\nu_0 \leqslant 0$, then $\nu_t \leqslant 0$ for all $t \in [0,T]$. Therefore for $\overline{\nu} \in \mathcal{M}(\mathbb{R}^d)$ there exists a unique narrowly continuous solution ν such that $\nu_0 = \overline{\nu}$.*

The proof is similar to that of Trevisan, 2016, Theorem A.7, so we provide a sketch.

Proof. Let $g \in C_c^\infty([0,T]\times\mathbb{R}^d)$ be non-negative. Let $\chi_R : \mathbb{R}^d \to [0,1]$ be a cutoff function and let a_R^ε, b_R^ε, c_R^ε be a mollification with respect to time and space of $a\chi_R$, $b\chi_R$, and $c\chi_R$, so that a_R^ε, b_R^ε, c_R^ε satisfy (2.3.13). Let $\mathcal{L}_R^{D,\varepsilon}$ be the diffusion operator associated with $\mathrm{dFP}(a_R^\varepsilon, b_R^\varepsilon, c_R^\varepsilon)$. Let f^ε solve

$$\partial_t f^\varepsilon = -\mathcal{L}_R^{D,\varepsilon} f^\varepsilon + g, \qquad f_T^\varepsilon = 0. \tag{2.3.15}$$

Recall the definition of $\mathcal{L}^S$ in Section 2.3.1 in the construction of the pushforward via smooth maps approximation. Since ν solves $\mathrm{dFP}(a,b,c)$, $f^\varepsilon \in C_b^{1,2}(\mathbb{R}^d)$ and $f^\varepsilon, \nu_0 \leqslant 0$, we have

$$\begin{aligned}
0 &\geqslant -\int f^\varepsilon \chi_R \,\mathrm{d}\nu_0, \\
&= \int \chi_R \partial_t f^\varepsilon + \mathcal{L}^D(f^\varepsilon\chi_R)\,\mathrm{d}\nu, \\
&= \int \chi_R(g - \mathcal{L}_R^{D,\varepsilon} f^\varepsilon) + \mathcal{L}^S f^\varepsilon) + f^\varepsilon \mathcal{L}^S\chi_R + c f^\varepsilon \chi_R + b : \nabla f^\varepsilon \otimes \nabla\chi_R \,\mathrm{d}\nu, \\
&\geqslant \int g\,\mathrm{d}\nu - \sup_{t\in[0,T]} ||f_t^\varepsilon||_{C_b^2(\mathbb{R}^d)} \int \chi_R(|a_R^\varepsilon - a| + |b_R^\varepsilon - b| + |c_R^\varepsilon - c|) + |\mathcal{L}^S\chi_R| + |b||\nabla\chi_R|)\,\mathrm{d}\nu.
\end{aligned} \tag{2.3.16}$$

The sup is uniformly bounded in $\varepsilon > 0$, and as $\varepsilon \to 0$ $(a_R^\varepsilon, b_R^\varepsilon, c_R^\varepsilon) \to (a,b,c)$ for $|x| \leqslant R$. Then letting $R \to \infty$, since $\nabla\chi_R \to 0$ and $\mathcal{L}^S\chi_R \to 0$ the right hand side converges to $\int g\,\mathrm{d}\nu$. □

2.3.3.2 Bounded Coefficients:

This case follows in the same manner as the bounded case in Trevisan, 2016. We now assume that a, b, c satisfy

$$\int_0^T \sup_{x\in\mathbb{R}^d} |a_t(x)| + \sup_{x\in\mathbb{R}^d} |b_t(x)| + \sup_{x\in\mathbb{R}^d} |c_t(x)|\,\mathrm{d}t < \infty. \tag{2.3.17}$$

Using the regularisation kernel $\kappa = \exp(-\sqrt{1+|x|^2})$ and $\kappa^\varepsilon = \varepsilon^{-n}\kappa(x/\varepsilon)$, we have that $\nu^\varepsilon = \nu * \kappa^\varepsilon$ solves dFP$(a^\varepsilon, b^\varepsilon, c^\varepsilon)$ with coefficients satisfying (2.3.13). Thus, existence of $\boldsymbol{\eta}^\varepsilon \in \mathcal{P}(C([0,T];\mathbb{R}^{d+1}))$ solving aMP$(a^\varepsilon, b^\varepsilon, c^\varepsilon)$ as a discounted superposition solution to ν^ε follows. Since ν_0^ε is a narrowly convergent sequence, there exists $\theta : \mathbb{R} \to \mathbb{R}$ increasing with $\lim_{x\to\infty}\theta(x) = \infty$ such that $\sup_{\varepsilon>0}\int \theta(|x|)\,\mathrm{d}\nu_0^\varepsilon \leqslant 1$. The proof of tightness of $\boldsymbol{\eta}^\varepsilon$ is then an easy modification of that in Trevisan, 2016, where the de la Vallée Poussin criterion modifies (2.3.17) to

$$\int_0^T \Theta\left(\sup_{x\in\mathbb{R}^d}|a_t(x)|\right) + \Theta\left(\sup_{x\in\mathbb{R}^d}|b_t(x)|\right) + \Theta\left(\sup_{x\in\mathbb{R}^d}|c_t(x)|\right)\mathrm{d}t < \infty, \tag{2.3.18}$$

for $\Theta : [0,\infty) \to [0,\infty)$ convex non-decreasing with $\lim_{x\to\infty} x^{-1}\Theta(x) = \infty$. Thus, in analogy to Trevisan, 2016, Corollary 2.5, with $f^i = (x_i\chi_R)\circ e_t$ for $i = 1,\dots,d$ and $f^{d+1} = (\log(y\chi_R))\circ e_t$. Note that the function f^{d+1} is always well-defined and real valued since Y is assumed to be supported on $(0, e^T)$. Then, for the functional $\Psi : C([0,T];\mathbb{R}) \to [0,\infty]$, we have for $i = 1,\dots,d$, with $x_R^i = x_i\chi_R$:

$$\int \Psi(x_R^i\circ\gamma)\,\mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) \leqslant \int \theta(|x_R^i|)\,\mathrm{d}\eta_0^\varepsilon + \int_0^T\int \Theta(|\mathcal{L}_t^{S,\varepsilon}x_R^i|) + \Theta(b_t^\varepsilon : \nabla x_R^i\otimes\nabla x_R^i)\,\mathrm{d}\eta_t^\varepsilon\,\mathrm{d}t, \tag{2.3.19}$$

$$\int \Psi(f^{d+1}\circ\gamma)\,\mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) \leqslant \int \theta(|f^{d+1}|)\,\mathrm{d}\eta_0^\varepsilon + \int_0^T\int \Theta\left(|c_t^\varepsilon|\right)\mathrm{d}\eta_t^\varepsilon\,\mathrm{d}t \tag{2.3.20}$$

We now note that $\eta_t^\varepsilon = \nu_t^{A,\varepsilon}$ where $\nu_t^{A,\varepsilon}$ solves aFP$(a^\varepsilon, b^\varepsilon, c^\varepsilon)$, and further that by construction $\nu_0^{A,\varepsilon} = \nu_0^\varepsilon$. Thus, letting $R\to\infty$, noting that the y-marginal of η_0^ε is δ_1, by the regular superposition principle (Trevisan, 2016, Theorem 2.5), we have the uniform bounds:

$$\int \Psi(x_R^i\circ\gamma)\,\mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) \leqslant \int \theta(|x_R^i|)\,\mathrm{d}\nu_0^\varepsilon + \int_0^T\int \Theta(|\mathcal{L}_t^{S,\varepsilon}x_R^i|) + \Theta(b_t^\varepsilon : \nabla x_R^i\otimes\nabla x_R^i)\,\mathrm{d}\nu_t^{A,\varepsilon}\,\mathrm{d}t, \tag{2.3.21}$$

$$\int \Psi(f^{d+1}\circ\gamma)\,\mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) \leqslant \theta(0) + \int_0^T\int \Theta\left(|c_t^\varepsilon|\right)\mathrm{d}\nu_t^{A,\varepsilon}\,\mathrm{d}t \tag{2.3.22}$$

Thus, by Trevisan, 2016, Lemma A.1, we have the uniform bounds

$$\int \Psi(\gamma^i)\,\mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) \leqslant 1 + \int_0^T \Theta\left(\sup_{x\in\mathbb{R}^d}|a_t(x)|\right) + \Theta\left(\sup_{x\in\mathbb{R}^d}|b_t(x)|\right)\mathrm{d}t, \tag{2.3.23}$$

$$\int \Psi(\log(\gamma^{d+1}))\,\mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) \leqslant \theta(0) + \int_0^T \Theta\left(\sup_{x\in\mathbb{R}^d}|c_t(x)|\right)\mathrm{d}t. \tag{2.3.24}$$

Since $\gamma \mapsto \Psi(\log(\gamma^{d+1})) + \sum_{i=1}^d \Psi(\gamma^i)$ is coercive in $C([0,T];\mathbb{R}^{d+1})$, we have tightness of $\boldsymbol{\eta}^\varepsilon$. The limit of the approximations obtaining a discounted superposition solution follows directly by the results of Section 2.3.2.

2.3.3.3 Locally Bounded Coefficients:

We now assume that a, b, c satisfy

$$\int_0^T \sup_{x\in B}\{|a_t(x)| + |b_t(x)| + |c_t(x)|\}\,\mathrm{d}t < \infty, \qquad \text{for every bounded borel } B \subset \mathbb{R}^d. \tag{2.3.25}$$

Let $M \geqslant 1$ and χ_M be a cutoff function. Define $\pi_M : \mathbb{R}^d \to \mathbb{R}^d$ by $\pi_M(x) = x\chi_M(x)$. Thus, by (2.3.25), we have $|\mathcal{L}^D(\pi_M^i)| \leqslant ||\pi_M^i||_{C^2}\sup_{|x|\leqslant 2M}\{|a(x)| + |b(x)| + |x_r c(x)|\}$ and $|b : \nabla\pi_M^i \otimes \nabla\pi_M^j| \leqslant ||\pi_M^i||_{C^1}\sup_{|x|\leqslant MR}|b(x)|$. Thus $\nu^M := \pi_\#^M\nu$ solves $\mathrm{dFP}(\pi^M(a), \pi^M(b), \pi^M(c))$ with coefficients satisfying (2.3.17) so we have $\boldsymbol{\eta}^M \in \mathcal{P}(C([0,T];\mathbb{R}^{d+1}))$ solving $\mathrm{aMP}(\pi^M(a), \pi^M(b), \pi^M(c))$ as a discounted superposition solution to ν^M. The argument of tightness is similar to that of the previous section with the modified bound to be

$$\int_0^T \int \Theta_1(|a_t|) + \Theta_2(|b_t|) + \Theta_3(|c_t|)\,\mathrm{d}\nu_t\,\mathrm{d}t < \infty \tag{2.3.26}$$

As before, we obtain the following inequalities:

$$\int \Psi(x_R^i \circ \gamma)\,\mathrm{d}\boldsymbol{\eta}^M(\gamma) \leqslant \int \theta(|x_R^i|)\,\mathrm{d}\eta_0^M + \int_0^T \int \Theta(|\mathcal{L}_t^{S,M} x_R^i|) + \Theta(b_t^M : \nabla x_R^i \otimes \nabla x_R^i)\,\mathrm{d}\eta_t^M\,\mathrm{d}t, \tag{2.3.27}$$

$$\int \Psi(f^{d+1} \circ \gamma)\,\mathrm{d}\boldsymbol{\eta}^M(\gamma) \leqslant \int \theta(|f^{d+1}|)\,\mathrm{d}\eta_0^M + \int_0^T \int \Theta\left(|c_t^M|\right)\mathrm{d}\eta_t^M\,\mathrm{d}t. \tag{2.3.28}$$

By a similar reasoning to the smooth and bounded coefficients case, using Jensen's inequality, we obtain the following uniform bounds:

$$\int \Psi(\gamma^i)\,\mathrm{d}\boldsymbol{\eta}^M(\gamma) \leqslant 1 + \int_0^T \int \Theta_1(|a_t^i|) + \Theta_2(|b_t^{i,i}|)\,\mathrm{d}\nu_t^A\,\mathrm{d}t \tag{2.3.29}$$

$$\int \Psi(\log(\gamma^{d+1}))\,\mathrm{d}\boldsymbol{\eta}^M(\gamma) \leqslant \theta(0) + \int_0^T \int \Theta_3(|c_t|)\,\mathrm{d}\nu_t^A\,\mathrm{d}t. \tag{2.3.30}$$

The limit of approximations obtaining a discounted superposition solution follows directly by Section 2.3.2.

2.3.3.4 General Case:

The approximation is achieved via convolution, with the approximating ν^ε having superposition solutions since the approximating coefficients satisfy (2.3.25). The tightness follows an identical argument, with similar bounds to the bounded coefficient case obtained. The limit of the approximations obtaining a discounted superposition solution follows directly by Section 2.3.2. This concludes the proof of Theorem 2.2.7.

2.4 The primal problem

Proposition 2.2.2 asserted that without any loss of generality, we can restrict our attention to $\mathbb{P} \in \widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0, \tau, u)$. By conditioning on X_t, we can express the value function as an integral against the discounted density, as shown in (2.2.19)-(2.2.20). Accordingly, from now on, we always work with the discounted versions of the density, drift, and volatility. We drop the superscript D for notational ease. The problem we need to solve in the RHS of (2.2.2) is thus given equivalently as follows:

Problem 2.4.1 (Primal Problem)**.** *The value function for the Primal Problem is given by*

$$V := \inf_{\rho,\alpha,\beta} \int_0^T \int_{\mathbb{R}^d} F(\alpha_t(x), \beta_t(x))\rho_t(\mathrm{d}x)\,\mathrm{d}t, \tag{2.4.1}$$

where the infimum is taken over $(\rho, \alpha, \beta) \in C([0,T]; \mathcal{M}(\mathbb{R}^d)) \times L^1(\mathrm{d}\rho_t\,\mathrm{d}t; \mathbb{R}^d) \times L^1(\mathrm{d}\rho_t\,\mathrm{d}t; \mathbb{S}^d)$,

subject to the following constraints in the sense of distributions for $(t,x) \in [0,T] \times \mathbb{R}^d$*:*

$$\partial_t \rho_t(x) + \nabla_x(\rho_t(x)\alpha_t(x)) - \frac{1}{2}\nabla_x^2 : (\rho_t(x)\beta_t(x)) + x_r \rho_t(x) = 0, \tag{2.4.2}$$

$$\text{for } i = 1, \dots, n, \quad \int_{\mathbb{R}^d} G_i(x)\rho_{\tau_i}(\mathrm{d}x) = u_i, \tag{2.4.3}$$

$$\rho_0(\cdot) = \mu_0. \tag{2.4.4}$$

We note that Problem 2.4.1 is in fact an example of unbalanced semimartingale optimal transport, with the source term in the continuity equation being $-x_r\rho$, see Chizat et al., 2018; Séjourné, Peyré, and Vialard, 2023.

Choosing the cost function F in (2.4.1) will determine the features of the calibration. Namely, it could be desirable to penalise deviations from a pre-calibrated reference model and use the convexity of Problem 2.4.1 to further improve the calibration, or add more constraints to an already calibrated SOT model. A previously calibrated SOT model could also be taken as a reference as a starting point to update the model periodically to match the changing options values in the market. The function F can also be used to choose which characteristics to calibrated — i.e. if we had a good stochastic rate model and wished to calibrate a local volatility while taking into account the stochastic interest rate. This can be achieved by freezing a subset of the characteristics $\{\alpha_i, \beta_{i,j}\}$. On the one hand, the conditions on F are kept as general as possible to permit a wide range of calibration techniques in this framework, however as will be seen in Lemma 2.5.8, the optimisers (α^*, β^*) are given as a function of ∇F^*, thus the smoothness of F^* will determine the smoothness of (α^*, β^*).

We recall from the Markovian projection Theorem 2.1.3 that the one dimensional marginals are preserved in reducing from an Itô process to a Markov process, and the first Markovian reduction of Section 2.2.1 shows that the value V is preserved when restricting to Markovian models. The discounted superposition principle of Theorem 2.2.7 shows that the second Markovian reduction of Section 2.2.3 applied with the discounted density transformation of Definition 2.2.4 preserves the same properties highlighted here as the first Markovian reduction. Therefore this means that the constrained SOT problem (Problem 2.4.1) attains the same value as the original general optimisation over $\mathbb{P} \in \mathcal{P}(\mu_0, \tau, u)$ in Proposition 2.2.2,

with the constraints (2.4.3)-(2.4.4) since the European option (and known initial condition) constraints only require matching of the one dimensional marginals.

All of this together means that the SOT method for model calibration gives us a framework that is flexible in a variety of use cases, and obtains the same cost when considering a minimisation over a larger set of models, with the mimicking result giving us consistent European prices. Convexity in SOT ensures that if a solution to Problem 2.4.1 exists, then it is attained and is unique. Finally, the SOT method outputs a continuous time model, which can be used to price exotic options jointly dependent on the state variables.

2.5 The dual problem

To solve the constrained optimisation in Problem 2.4.1, we will use a duality method inspired by Huesmann and Trevisan, 2019; I. Guo, Loeper, and Shiyi Wang, 2022; I. Guo, Loeper, Obłój, et al., 2022a. This proof is similar in spirit to the duality argument in that proves Brenier, 1999, Proposition 2.7, which considers dual of a relaxation of the minimal geodisic problem as an alternative approach to solving the Euler equations of an incompressible fluid. This technique had previously been applied in the proof of in Brenier, 1997, Theorem 3.2 when seeking for the existence variational solutions of "homogenised vortex sheet equations". The technique is also used to formulate the dual of a variational problem involving the Euler-Poisson system of equations in Loeper, 2006. The proof relies mainly on the Fenchel-Rockafellar duality theorem (see Rockafellar, 1970, Chapter 31), and an adjustment to make the problem convex. As our primal problem is quite similar to I. Guo, Loeper, and Shiyi Wang, 2022, the approach used there can be adapted to our setting. The following result uses the notion of viscosity solution from Definition 2.5.7 below.

Theorem 2.5.1 (Dual Problem)**.** *The dual expression for the value function V is*

$$V = \sup_{\lambda} \left(\lambda \cdot u - \int_{\mathbb{R}^d} \varphi^{\lambda}(0,x) \, \mathrm{d}\mu_0 \right), \tag{2.5.1}$$

where $\lambda \in \mathbb{R}^n$ *and* $\varphi^\lambda = \varphi$ *is the viscosity solution to the HJB equation:*

$$\partial_t \varphi - x_r \varphi + \sum_{i=1}^{n} \lambda_i G_i(x) \delta_{\tau_i} + F^* \left(\nabla_x \varphi, \frac{1}{2} \nabla_x^2 \varphi \right) = 0, \quad (t,x) \in [0,T] \times \mathbb{R}^d, \tag{2.5.2}$$

with terminal condition $\varphi(T, \cdot) = 0$*. If* V *is finite, then the infimum in Problem 2.4.1 is attained. If the supremum is attained for some* $\lambda^* \in \mathbb{R}^n$*, with* $\varphi^* \in \mathrm{BV}([0,T]; C_b^2(\mathbb{R}^d))$ *such that* $\varphi^*(T, \cdot) = 0$ *solving the corresponding HJB equation, and* $(\rho^*, \alpha^*, \beta^*)$ *being the optimal solution of Problem 2.4.1, then* (α^*, β^*) *is given by:*

$$(\alpha_t^*, \beta_t^*) = \nabla F^* \left(\nabla_x \varphi^*(t, \cdot), \frac{1}{2} \nabla_x^2 \varphi^*(t, \cdot) \right), \qquad \mathrm{d}\rho_t^* \, \mathrm{d}t \text{ - almost everywhere.} \tag{2.5.3}$$

Remark 2.5.2. *Solving the dual problem in Theorem 2.5.1 yields the primal optimisers* (α^*, β^*)*. These characterise the distribution of the Markov process* (X, Y) *solving* (2.1.8) *with these coefficients , i.e.,* ρ^* *solves* (2.1.9)*, and the associated measure* $\mathbb{P}^*$ *on* Ω*,* $\rho_t^* = \mathbb{P}^* \circ (X_t, Y_t)^{-1}$*. We can then compute the associated discount factor* $D_t^*(x)$ *and apply the transformation* (2.2.4) *to obtain that* $D_t^*(x)\rho_t^*(x)$ *is a solution of* (2.2.11)*.*

Remark 2.5.3 (Attainability of the Dual)**.** *It is not presently known if the supremum in* (2.5.1) *is attained for a generic cost function* F *— in general, this is an open problem in (semi)martingale optimal transport. Beiglböck, Lim, and Obłój, 2019 prove results in one dimension in the case of compactly supported marginals, and a twice continuously differentiable or Lipschitz cost function. In the Bass martingale setting of Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023, Section 7, that is under the assumption of irreducibility of marginals and with a particular cost function, the attainability of the dual in the static optimal transport problem is proven. We note Chapter 4 establishes a link in the duality of the static problem of Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023 and the dynamical formulation of Huesmann and Trevisan, 2019.*

We now provide an interpretation of the dual formulation in the context of MOT robust pricing and hedging results (e.g. Soner, Touzi, and Zhang, 2013; Dolinsky and Soner, 2014; Galichon, Henry-Labordère, and Touzi, 2014 for a duality proof in the case of full prescribed marginals). If we take V in (2.4.1) to be a function of the market prices u, then since $V(u)$

is convex and lower semicontinuous, by the Fenchel-Moreau theorem (see e.g. Brézis, 2011, Theorem 1.11) we have $V(u) = V^{**}(u)$, and thus by taking the Legendre transform of $V^*(\lambda)$, we have

$$V(u) = \sup_{\lambda \in \mathbb{R}^n} \left(\lambda \cdot u - V^*(\lambda) \right). \tag{2.5.4}$$

In view of (2.5.1), we have for φ solving (2.5.2) that

$$\begin{aligned}
\int_{\mathbb{R}^d} \varphi(0,x)\,\mathrm{d}\mu_0 &= V^*(\lambda), \\
&= \sup_{\rho} \int_\Lambda \sum_{i=1}^n \lambda_i G_i(x)\delta_{\tau_i} - F(\alpha(t,x),\beta(t,x))\,\mathrm{d}\rho, \qquad (2.5.5) \\
&= \sup_{\mathbb{P}\in\tilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0,\tau,u)} \mathbb{E}^{\mathbb{P}} \left[\sum_{i=1}^n \lambda_i Y_{\tau_i} G_i(X_{\tau_i}) - \int_0^T Y_t F(\alpha(t,X_t),\beta(t,X_t))\,\mathrm{d}t \right], \qquad (2.5.6)
\end{aligned}$$

where the supremum in (2.5.5) is taken over $\rho \in C([0,T];\mathcal{M}(\mathbb{R}^d))$ subject to the constraints (2.4.2)-(2.4.4). The right hand side of (2.5.6) can be interpreted as the robust price of hedging an exotic that continuously pays $F(\alpha(t,X_t),\beta(t,X_t))\delta t$ over each infinitesimal time interval $[t,t+\delta t)$ for $t \in [0,T]$ with a static position in the calibration options weighted by λ. To see this, we adapt the calculation of Loeper, 2018, Section 4. For simplicity of presentation, take $\mu_0 = \delta_{X_0}$ and assume we have one option with payoff $G(x)$ and maturity at T, then given characteristics (α,β), we have:

$$\varphi(0,X_0) = \mathbb{E}^{\mathbb{P}} \left[Y_T G(X_T) - \int_0^T Y_t F(\alpha(t,X_t),\beta(t,X_t))\,\mathrm{d}t \right] \tag{2.5.7}$$

Since $\varphi(T,X_T) = G(X_T)$, integration by parts of $Y_t\varphi(t,X_t)$ then yields

$$\begin{aligned}
Y_T\varphi(T,X_T) = \varphi(0,X_0) &+ \int_0^T Y_t \left(\partial_t\varphi(t,X_t) + \alpha(t,X_t)\cdot\nabla_x\varphi(t,X_t) + \frac{1}{2}\beta(t,X_t) : \nabla_x^2\varphi(t,X_t) \right)\mathrm{d}t \\
&- \int_0^T \varphi(t,X_t)X_t^r Y_t\,\mathrm{d}t + \int_0^T Y_t\nabla_x\varphi(t,X_t)\beta^{\frac{1}{2}}(t,X_t)\,\mathrm{d}W_t^{\mathbb{P}}. \qquad (2.5.8)
\end{aligned}$$

The use of integration by parts here is justified by the notion of solution of $\varphi \in$

$\mathrm{BV}([0,T];C_b^2(\mathbb{R}^d))$ given in Definition 2.5.7. Since φ solves (2.5.2) and

$$\alpha(t,x)\cdot\nabla_x\varphi(t,x)+\frac{1}{2}\beta(t,x):\nabla_x^2\varphi(t,x)-F^*\left(\nabla_x(t,x),\frac{1}{2}\nabla^2\varphi(t,x)\right)\leqslant F(\alpha(t,x),\beta(t,x)) \tag{2.5.9}$$

holds everywhere, with equality if and only if $(\alpha,\beta)=(\alpha^*,\beta^*)$, by rearranging (2.5.8) and applying (2.5.9), we obtain the following inequality pathwise

$$Y_TG(X_T)\leqslant\varphi(0,X_0)+\int_0^T Y_tF(\alpha(t,X_t),\beta(t,X_t))\,\mathrm{d}t+\int_0^T Y_t\nabla_x\varphi(t,X_t)\beta^{\frac{1}{2}}(t,X_t)\,\mathrm{d}W_t^{\mathbb{P}}. \tag{2.5.10}$$

Since $\beta^{\frac{1}{2}}(t,X_t)\,\mathrm{d}W_t^{\mathbb{P}}=\mathrm{d}X_t-\alpha(t,X_t)\,\mathrm{d}t$, the stochastic integral in (2.5.10) is the profit and loss of (jointly in the state variables) Δ-hedging after taking into account the cost of funding the Δ-hedge included in α. We therefore obtain in (2.5.10) a discounted superhedging portfolio for the exotic with a static position in the European option. Taking expectations, we obtain:

$$\mathbb{E}^{\mathbb{P}}\left[Y_TG(X_T)-\int_0^T Y_tF(\alpha(t,X_t),\beta(t,X_t))\,\mathrm{d}t\right]\leqslant\varphi(0,X_0), \tag{2.5.11}$$

with equality if and only if $(\alpha,\beta)=(\alpha^*,\beta^*)$. This means that the optimiser of the dual in (2.5.1) can be viewed as a discounted robust hedging portfolio for the exotic corresponding to the cost function with static positions in the calibrating constraints weighted by the optimal λ^*.

2.5.1 Proof of the duality theorem

We prove the duality result above in the remainder of this section through a series of lemmas. Our first observation is that the objective function (2.4.1) is not jointly convex in (ρ,α,β). We define the measures $\mathcal{A}\coloneqq\rho\alpha$, $\mathcal{B}\coloneqq\rho\beta$, so $\mathcal{A}$ and $\mathcal{B}$ are absolutely continuous with respect to ρ. Then, the objective function is convex in $(\rho,\mathcal{A},\mathcal{B})$ with constraints that are affine in $(\rho,\mathcal{A},\mathcal{B})$. This arises from the classical notion that the function $\bar{f}(z_1,z_2,z_3)\coloneqq z_3f\left(\frac{z_1}{z_3},\frac{z_2}{z_3}\right)$ is convex in (z_1,z_2,z_3) whenever f is convex in (z_1,z_2) on the set $\{z_3>0\}$. Note also that we write $\mathrm{d}\mathcal{A}$ for $\alpha_t(x)\rho_t(\mathrm{d}x)\,\mathrm{d}t$ and $\mathrm{d}\mathcal{B}$ for $\beta_t(x)\rho_t(\mathrm{d}x)\,\mathrm{d}t$. Moreover, our constraints in

Problem 2.4.1 can be formulated in the weak sense as:

$$\int_\Lambda \partial_t \varphi \mathrm{d}\rho + \nabla\varphi \cdot \mathrm{d}\mathcal{A} + \frac{1}{2}\nabla^2\varphi : \mathrm{d}\mathcal{B} - x_r\varphi \mathrm{d}\rho + \int_{\mathbb{R}^d} \varphi \mathrm{d}\mu_0 = 0, \tag{2.5.12}$$

$$\int_\Lambda \sum_{i=1}^n \lambda_i G_i(x)\delta_{\tau_i}\mathrm{d}\rho - \sum_{i=1}^n \lambda_i u_i = 0. \tag{2.5.13}$$

for any smooth compactly supported test function $\varphi \in C_c^\infty(\Lambda)$ with $\varphi(T,\cdot) = 0$ and $\lambda \in \mathbb{R}^n$. The terminal condition on φ arises when performing an integration by parts to derive (2.5.12), since we need the $\rho_T(\cdot)$ boundary term to vanish as we do not have a priori knowledge of $\rho_T(\cdot)$. Therefore we can write Problem 2.4.1 as the following saddle point problem:

Problem 2.5.4 (Saddle Point Problem)**.**

$$\begin{aligned} V = \inf_{\rho,\mathcal{A},\mathcal{B}} \sup_{\varphi,\lambda} \Bigg\{ & \int_\Lambda F\left(\frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\rho}, \frac{\mathrm{d}\mathcal{B}}{\mathrm{d}\rho}\right)\mathrm{d}\rho - \partial_t\varphi\mathrm{d}\rho - \nabla\varphi\cdot\mathrm{d}\mathcal{A} - \frac{1}{2}\nabla^2\varphi : \mathrm{d}\mathcal{B} \\ & + x_r\varphi\mathrm{d}\rho - \int_{\mathbb{R}^d}\varphi\mathrm{d}\mu_0 - \int_\Lambda \sum_{i=1}^n \lambda_i G_i(x)\delta_{\tau_i}\mathrm{d}\rho + \sum_{i=1}^n \lambda_i u_i \Bigg\}, \end{aligned} \tag{2.5.14}$$

where the infimum is taken across $(\rho,\alpha,\beta) \in C([0,T];\mathcal{M}(\mathbb{R}^d)) \times L^1(\mathrm{d}\rho_t\,\mathrm{d}t;\mathbb{R}^d) \times L^1(\mathrm{d}\rho_t\,\mathrm{d}t;\mathbb{S}^d)$ *and the supremum is taken across* $(\varphi,\lambda) \in C_c^\infty(\Lambda;\mathbb{R}) \times \mathbb{R}^n$.

We now want to find a functional with convex conjugate equal to (2.4.1), and another that is the remainder of the infimum in Problem 2.5.4. To do this, we use the following terminology from Huesmann and Trevisan, 2019 in the proof of the duality theorem. Denote $\mathrm{BV}_T([0,T];C_b^2(\mathbb{R}^d))$ as the set of $\varphi \in \mathrm{BV}([0,T];C_b^2(\mathbb{R}^d))$ such that $\varphi(T,\cdot) = 0$.

Definition 2.5.5. *We say that the triple* $(\gamma,a,b) \in C_b(\Lambda;\mathcal{X})$ *is represented by* $(\varphi,\lambda) \in \mathrm{BV}_T([0,T];C_b^2(\mathbb{R}^d)) \times \mathbb{R}^n$ *if*

$$\gamma + \partial_t\varphi - x_r\varphi + \sum_{i=1}^n \lambda_i G_i(x)\delta_{\tau_i} = 0, \tag{2.5.15}$$

$$a + \nabla\varphi = 0, \tag{2.5.16}$$

$$b + \frac{1}{2}\nabla^2\varphi = 0. \tag{2.5.17}$$

Since $(\gamma, a, b) \in C_b(\Lambda; \mathcal{X})$, the presence of the dirac delta functions give that $t \mapsto \varphi(t, \cdot)$ is of bounded variation on $[0, T]$ with jump discontinuities at $t = \tau_i$, which we denote as $\varphi \in \mathrm{BV}_T([0, T]; C_b^2(\mathbb{R}^d))$, since we require φ to be at least C^2 in space. Proceding in an analogous way to I. Guo, Loeper, and Shiyi Wang, 2022 to obtain the dual problem, first define functionals $\Phi : C_b(\Lambda; \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ and $\Psi : C_b(\Lambda; \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ by:

$$\Phi(\gamma, a, b) = \begin{cases} 0, & \text{if } \gamma + F^*(a, b) \leqslant 0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.5.18}$$

$$\Psi(\gamma, a, b) = \begin{cases} \int_{\mathbb{R}^d} \varphi(0, x)\, \mathrm{d}\mu_0 - \sum_{i=1}^n \lambda_i u_i, & \text{if } (\gamma, a, b) \text{ is represented by} (\varphi, \lambda) \in \\ & \mathrm{BV}_T([0, T]; C_b^2(\mathbb{R}^d)) \times \mathbb{R}^n, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.5.19}$$

Lemma 2.5.6. *The objective function V can be expressed in terms of Φ and Ψ as*

$$V = \inf_{\rho, \mathcal{A}, \mathcal{B}} \left\{ \Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B}) \right\}, \tag{2.5.20}$$

where the infimum is taken across $(\rho, \mathcal{A}, \mathcal{B}) \in C_b(\Lambda, \mathcal{X})^*$.

We remark that switching from $\varphi \in C_c^\infty(\Lambda)$ (with $\varphi(T, \cdot) = 0$) to $\varphi \in \mathrm{BV}_T([0, T]; C_b^2(\mathbb{R}^d))$ does not change the value of the supremum, which will be formalised by the notion of viscosity solutions later in Definition 2.5.7.

Proof. As shown in Lemma A.1 of I. Guo, Loeper, and Shiyi Wang, 2022, we can compute Ψ^* by restricting the domain of the convex conjugate from $\Phi^* : C_b(\Lambda; \mathcal{X})^* \to \mathbb{R} \cup \{+\infty\}$ to $\mathcal{M}(\Lambda; \mathcal{X})$. Since ρ is given by (2.2.4), we have $\rho \in \mathcal{M}_+(\Lambda; \mathbb{R})$; moreover our definition of $\mathcal{A}$ and $\mathcal{B}$ give that $\mathcal{A}, \mathcal{B} \ll \rho$, and therefore $(\mathcal{A}, \mathcal{B}) \in \mathcal{M}_+(\Lambda; \mathbb{R}^d \times \mathbb{S}^d)$, so $(\rho, \mathcal{A}, \mathcal{B}) \in \mathcal{M}_+(\Lambda; \mathcal{X})$. Thus,

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

$$= \int_\Lambda \sup_{a,b} \left(a \cdot \frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\rho} + b : \frac{\mathrm{d}\mathcal{B}}{\mathrm{d}\rho} - F^*(a,b) \right) \mathrm{d}\rho$$
$$= \int_\Lambda F\left(\frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\rho}, \frac{\mathrm{d}\mathcal{B}}{\mathrm{d}\rho} \right) \mathrm{d}\rho. \tag{2.5.21}$$

Since the construction of $(\rho, \mathcal{A}, \mathcal{B})$ give that the restriction of $(\rho, \mathcal{A}, \mathcal{B})$ to $\mathcal{M}(\Lambda; \mathcal{X})$ is equivalent to the restriction of $(\rho, \mathcal{A}, \mathcal{B})$ to $\mathcal{M}_+(\Lambda, \mathcal{X})$ with $\mathcal{A}, \mathcal{B} \ll \rho$, we have

$$\Phi^*(\rho, \mathcal{A}, \mathcal{B}) = \begin{cases} \int_\Lambda F\left(\frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\rho}, \frac{\mathrm{d}\mathcal{B}}{\mathrm{d}\rho} \right) \mathrm{d}\rho & \text{if } (\rho, \mathcal{A}, \mathcal{B}) \in \mathcal{M}(\Lambda, \mathcal{X}), \\ +\infty & \text{otherwise.} \end{cases} \tag{2.5.22}$$

We now compute $\Psi^* : C_b(\Lambda; \mathcal{X})^* \to \mathbb{R} \cup \{+\infty\}$, we first note that if this is restricted to $(\rho, \mathcal{A}, \mathcal{B}) \in \mathcal{M}(\Lambda, \mathcal{X})$, then

$$\begin{aligned} \Psi^*(\rho, \mathcal{A}, \mathcal{B}) &= \sup_{(\gamma,a,b) \in C_b(\Lambda;\mathcal{X})} \left\{ \langle (\gamma, a, b), (\rho, \mathcal{A}, \mathcal{B}) \rangle - \int_{\mathbb{R}^d} \varphi(0,x) \mathrm{d}\mu_0 + \sum_{i=1}^n \lambda_i u_i \right\} \\ &= \sup_{\varphi,\lambda} \Bigg\{ \left\langle \left(x_r \varphi - \partial_t \varphi - \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i}, -\nabla \varphi, -\frac{1}{2} \nabla^2 \varphi \right), (\rho, \mathcal{A}, \mathcal{B}) \right\rangle \\ &\qquad - \int_{\mathbb{R}^d} \varphi(0,x) \mathrm{d}\mu_0 + \sum_{i=1}^n \lambda_i u_i \Bigg\} \\ &= \sup_{\varphi,\lambda} \Bigg\{ \left\{ \int_\Lambda x_r \varphi \mathrm{d}\rho - \partial_t \varphi \mathrm{d}\rho - \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i} \mathrm{d}\rho - \nabla_x \varphi \mathrm{d}\mathcal{A} - \frac{1}{2} \nabla_x^2 \varphi : \mathrm{d}\mathcal{B} \right\} \\ &\qquad - \int_{\mathbb{R}^d} \varphi(0,x) \mathrm{d}\mu_0 + \sum_{i=1}^n \lambda_i u_i \Bigg\}. \end{aligned} \tag{2.5.23}$$

Thus, since Φ^* is independent of (φ, λ) we can simply add (2.5.22) and (2.5.23) together to get the argument of the infimum in Problem 2.5.4, so

$$V = \inf_{(\rho,\mathcal{A},\mathcal{B}) \in \mathcal{M}(\Lambda;\mathcal{X})} (\Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B})). \tag{2.5.24}$$

Now, we apply Lemma A.2 in I. Guo, Loeper, and Shiyi Wang, 2022 and we get that

$$V = \inf_{(\rho,\mathcal{A},\mathcal{B}) \in \mathcal{M}(\Lambda;\mathcal{X})} \{\Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B})\} = \inf_{(\rho,\mathcal{A},\mathcal{B}) \in C_b(\Lambda;\mathcal{X})^*} \{\Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B})\}. \tag{2.5.25}$$

□

Now we apply the Fenchel-Rockerfellar duality theorem as formulated in Villani, 2003, Theorem 1.9. We first note that as the constraints in the functionals Φ and Ψ are affine, the functionals are clearly convex in (γ, a, b). We now check the conditions of the theorem at the point $(0, O^{d\times 1}, O^{d\times d})$ which is represented by $(\varphi, \lambda) = (e^{tx_r}, O^{n\times 1})$ where O refers to the zero matrix of appropriate dimension. Since F is non-negative:

$$F^*(O^{d\times 1}) = -\inf_{(\alpha,\beta)\in\mathbb{R}^d\times\mathbb{S}_+^d} F(\alpha,\beta) \leqslant 0. \tag{2.5.26}$$

Therefore, we have $\Phi(0, O^{d\times 1}, O^{d\times d}) = 0$ and $(0, O^{d\times 1}, O^{d\times d})$ is a point of continuity of Φ since F^* is continuous. Moreover, since $\varphi(t,x) = e^{tx_r}$ and μ_0 is a probability measure:

$$\Psi(0, O^{d\times 1}, O^{d\times d}) = \int_{\mathbb{R}^d} \varphi(0,x)\,\mathrm{d}\mu_0 = 1. \tag{2.5.27}$$

Thus, as Ψ is finite and Φ is finite and continuous at $(0, O^{d\times 1}, O^{d\times d})$ and both take values in $(-\infty, +\infty]$ we may apply the Fenchel-Rockerfellar duality theorem and obtain that:

$$\begin{aligned}
&\inf_{(\gamma,a,b)\in C_b(\Lambda;\mathcal{X})} \{\Phi(-\gamma,-a,-b) + \Psi(\gamma,a,b)\} \\
&\qquad = \sup_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda;\mathcal{X})^*} \{-\Phi^*(-\rho,-\mathcal{A},-\mathcal{B}) - \Psi^*(-\rho,-\mathcal{A},-\mathcal{B})\} \\
&\qquad = \sup_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda;\mathcal{X})^*} \{-\Phi^*(\rho,\mathcal{A},\mathcal{B}) - \Psi^*(\rho,\mathcal{A},\mathcal{B})\} \\
&\qquad = -\inf_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda;\mathcal{X})^*} \{\Phi^*(\rho,\mathcal{A},\mathcal{B}) + \Psi^*(\rho,\mathcal{A},\mathcal{B})\}.
\end{aligned} \tag{2.5.28}$$

Thus rearranging we obtain

$$
\begin{aligned}
V &= \inf_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda;\mathcal{X})^*} \{\Phi^*(\rho,\mathcal{A},\mathcal{B}) + \Psi^*(\rho,\mathcal{A},\mathcal{B})\} \\
&= \sup_{(\gamma,a,b)\in C_b(\Lambda;\mathcal{X})} \{-\Phi(-\gamma,-a,-b) - \Psi(\gamma,a,b)\} \\
&= \sup_{(\gamma,a,b)\in C_b(\Lambda;\mathcal{X})} \Bigg\{ \sum_{i=1}^n \lambda_i u_i - \int_{\mathbb{R}^d} \varphi(0,x)\,\mathrm{d}\mu_0 \;:\; -\gamma + F^*(-a,-b) \leqslant 0, \\
&\qquad\qquad (\gamma,a,b) \text{ is represented by } (\varphi,\lambda) \Bigg\} \\
&= \sup_{(\lambda,\varphi)\in\mathbb{R}^n\times \mathrm{BV}_T([0,T];C_b^2(\mathbb{R}^d))} \Bigg\{ \sum_{i=1}^n \lambda_i u_i - \int_{\mathbb{R}^d} \varphi(0,x)\,\mathrm{d}\mu_0 \;:\; \partial_t\varphi - x_r\varphi + \sum_{i=1}^n \lambda_i G_i(x)\delta_{\tau_i} \\
&\qquad\qquad + F^*\left(\nabla_x\varphi, \frac{1}{2}\nabla_x^2\varphi\right) \leqslant 0 \Bigg\}.
\end{aligned}
\tag{2.5.29}
$$

Now, to obtain equality in the HJB equation constraint and thus the HJB equation (2.5.2) and dual formulation in Theorem 2.5.1, we adapt the classical notion of viscosity solutions from Lions, 1983 to include the required jump discontinuities, in analogy to I. Guo, Loeper, and Shiyi Wang, 2022. First define disjoint intervals $I_k \coloneqq [\tau_{k-1}, \tau_k)$ with $\tau_0 = 0$, with $\bigcup_{k=1}^n I_k = [0,T)$.

Definition 2.5.7 (Viscosity Solution)**.** *For any $\lambda \in \mathbb{R}^n$, we say $\varphi \in \mathrm{BV}_T([0,T];C_b(\mathbb{R}^d))$ is a viscosity subsolution (supersolution) of* (2.5.2) *if $\varphi|_{I_k\times\mathbb{R}^d} \in C_b(I_k;C_b(\mathbb{R}^d))$ is a classical (continuous) viscosity subsolution (supersolution) of* (2.5.2) *in $I_k \times \mathbb{R}^d$ and for all $k = 1,\dots,n$ has jump discontinuities:*

$$
\varphi(t,x) = \varphi(t^-,x) - \sum_{i=1}^n \lambda_i G_i(x) \mathbb{1}_{\{t=\tau_i\}}. \tag{2.5.30}
$$

With terminal condition $\varphi(T,\cdot) = 0$. In addition, $\varphi \in \mathrm{BV}_T([0,T];C_b(\mathbb{R}^d))$ is a viscosity solution of (2.5.2) *if it is both a viscosity subsolution and viscosity supersolution of* (2.5.2)*.*

The expression for V in (2.5.29) involved supersolutions to (2.5.2) and the first step is to show that we can restrict to viscosity solutions, as stated in the first part of Theorem 2.5.1. For this we follow the proof of I. Guo, Loeper, and Shiyi Wang, 2022, Proposition 3.5.

I. Guo, Loeper, and Shiyi Wang, 2022, Remark 3.9 provides a comparison principle, which is the classical comparison principle of viscosity solutions applied to φ on $I_k \times \mathbb{R}^d$ for each k. Using this, one can deduce existence and uniqueness of solutions to (2.5.2) via Crandall, Ishii, and Lions, 1992. The comparison principle also implies that V in (2.5.1) is smaller than the supremum over viscosity solutions. We then use the smoothing argument from Bouchard, Loeper, and Zou, 2017, which shows that any viscosity solution of (2.5.2) can be approached by smooth supersolutions. This, together with (2.5.29) shows that V is larger than the supremum over viscosity solutions. This allows us to conclude the proof of the first point of Theorem 2.5.1. We now seek to obtain the form of the optimal (α, β) for the second part of Theorem 2.5.1.

Lemma 2.5.8. *If the supremum in Theorem 2.5.1 is attained for some λ^* with $\varphi^* \in \mathrm{BV}_T([0,T]; C_b^2(\mathbb{R}^d))$ solving the corresponding HJB equation, and $(\rho^*, \alpha^*, \beta^*)$ being the optimal solution of Problem 2.4.1, then (α^*, β^*) is given by:*

$$(\alpha_t^*, \beta_t^*) = \nabla F^* \left(\nabla_x \varphi^*(t, \cdot), \frac{1}{2} \nabla_x^2 \varphi^*(t, \cdot) \right), \qquad \mathrm{d}\rho^* \text{ - almost everywhere.} \tag{2.5.31}$$

Remark 2.5.9. *As stated in Remark 2.5.3, the supremum in (2.5.1) may not be attained for some (λ^*, φ^*). However, Lemma 2.5.8 can be improved using I. Guo and Loeper, 2021, Proposition 3.12 to show that the ε-optimisers of Theorem 2.5.1 can give convergence to the optimal model in probability on Ω. Namely if $\mathbb{P}^*$ is an optimal probability measure for Problem 2.4.1 with characteristics (α^*, β^*), and $(\lambda^n, \varphi^n) \in \mathbb{R}^n \times \mathrm{BV}_T([0,T]; C_b^2(\mathbb{R}^d))$ is some optimising sequence of (2.5.1), then we have:*

$$\partial_t \varphi^n - x_r \varphi^n + \sum_{i=1}^n \lambda_i^n G_i(x) \delta_{\tau_i} + F^* \left(\nabla_x \varphi^n, \frac{1}{2} \nabla_x^2 \varphi^n \right) \overset{\mathrm{d}\mathbb{P}^*}{\to} 0, \tag{2.5.32}$$

$$\nabla F^* \left(\nabla \varphi^n, \frac{1}{2} \nabla^2 \varphi^n \right) \overset{\mathrm{d}\mathbb{P}^*}{\to} (\alpha^*, \beta^*). \tag{2.5.33}$$

In particular, this motivates our numerical approach in Chapter 3 where the dual problem is solved by using gradient methods to build an approximating sequence.

Remark 2.5.10. *Here we need to assume that $\varphi \in C_b^2$ in space to make sense of (2.5.31).*

In view of the a priori estimates of Loeper and Quirós, 2018, under some convexity condition on F^*, interior regularity holds for viscosity solutions the HJB equations, thus yielding the C^2_{loc} in space regularity needed to write (2.5.31)*, provided F^* fulfills the conditions of Loeper and Quirós, 2018.*

Proof. Let (α^*, β^*) be the optimal solution of Problem 2.4.1, then $(\rho^*, \rho^*\alpha^*, \rho^*\beta^*)$ also achieves the infimum in Problem 2.5.4. Assume that λ^* is the optimal solution solving (2.5.1) with corresponding solution to (2.5.2) φ^*. Then, λ^* achieves the supremum in Problem 2.5.4, so with our optimal solution we may rewrite Problem 2.5.4 as:

$$V = \int_\Lambda \left(F(\alpha^*, \beta^*) - \partial_t \varphi^* - \nabla_x \varphi^* \cdot \alpha^* - \frac{1}{2} \nabla_x^2 \varphi^* : \beta^* + x_r \varphi^* - \sum_{i=1}^n \lambda^* G_i(x) \delta_{\tau_i} \right) \mathrm{d}\rho^* - \int_{\mathbb{R}^d} \varphi^* \, \mathrm{d}\mu_0 + \sum_{i=1}^n \lambda_i^* u_i. \tag{2.5.34}$$

Since (φ^*, λ^*) are optimal, we have from Theorem 2.5.1 that

$$V = \sum_{i=1}^n \lambda_i^* u_i - \int_{\mathbb{R}^d} \varphi^* \, \mathrm{d}\mu_0. \tag{2.5.35}$$

Therefore, (2.5.34) is equivalent to

$$\begin{aligned} 0 &= \int_\Lambda \left(F(\alpha^*, \beta^*) - \partial_t \varphi^* - \nabla_x \varphi^* \cdot \alpha^* - \frac{1}{2} \nabla_x^2 \varphi^* : \beta^* + x_r \varphi^* - \sum_{i=1}^n \lambda^* G_i(x) \delta_{\tau_i} \right) \mathrm{d}\rho^* \\ &= \int_\Lambda \left(F(\alpha^*, \beta^*) + F^* \left(\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^* \right) - \nabla_x \varphi^* \cdot \alpha^* - \frac{1}{2} \nabla_x^2 \varphi^* : \beta^* \right) \mathrm{d}\rho^*. \end{aligned} \tag{2.5.36}$$

Now define $(\bar{\alpha}, \bar{\beta})$ as:

$$(\bar{\alpha}, \bar{\beta}) = \nabla F^* \left(\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^* \right). \tag{2.5.37}$$

Note that

$$(\bar{\alpha}, \bar{\beta}) = \nabla F^* \left(\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^* \right) = \underset{(a,b) \in \mathbb{R}^d \times \mathbb{S}^d_+}{\arg\max} \left(a \cdot \nabla_x \varphi^* + b : \frac{1}{2} \nabla_x^2 \varphi^* - F(a, b) \right). \tag{2.5.38}$$

Thus,

$$F^* \left(\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^* \right) = \nabla_x \varphi^* \cdot \bar{\alpha} + \frac{1}{2} \nabla_x^2 \varphi^* : \bar{\beta} - F(\bar{\alpha}, \bar{\beta}). \tag{2.5.39}$$

Since F is convex, its Legendre transform is an involution (see Rockafellar, 1970, Corol-

lary 12.2.1), so taking the double Legendre transform, we have

$$(A, B) \coloneqq \nabla F(\bar{\alpha}, \bar{\beta}) = \nabla F^{**}(\bar{\alpha}, \bar{\beta}) = \underset{(a,b)\in\mathbb{R}^d\times\mathbb{S}^d_+}{\arg\max} (a \cdot \bar{\alpha} + b : \bar{\beta} - F^*(a, b)). \tag{2.5.40}$$

Therefore,

$$F^{**}(\bar{\alpha}, \bar{\beta}) = A \cdot \bar{\alpha} + B : \bar{\beta} - F^*(A, B) = A \cdot \bar{\alpha} + B : \bar{\beta} - \max_{(x,y)\in\mathbb{R}^d\times\mathbb{S}^d_+} (A \cdot x + B : y - F(x, y)). \tag{2.5.41}$$

Since $F(\bar{\alpha}, \bar{\beta}) = F^{**}(\bar{\alpha}, \bar{\beta})$, from (2.5.38) we have that for the affine terms to cancel, we need $(A, B) = \left(\nabla_x \varphi^*, \frac{1}{2}\nabla_x^2 \varphi^*\right)$. So, substituting into (2.5.36) we have:

$$0 = \int_\Lambda \left(F(\alpha^*, \beta^*) - F(\bar{\alpha}, \bar{\beta}) - \nabla_x \varphi^* \cdot (\alpha^* - \bar{\alpha}) - \frac{1}{2}\nabla_x^2 \varphi^* : (\beta^* - \bar{\beta}) \right) \mathrm{d}\rho^*. \tag{2.5.42}$$

Since $F(\alpha, \beta)$ is assumed to be strongly convex in (α, β), we have from (2.1.10) that for some constant $C > 0$:

$$F(\alpha^*, \beta^*) - F(\bar{\alpha}, \bar{\beta}) \geqslant \langle \nabla F(\bar{\alpha}, \bar{\beta}), (\alpha^* - \bar{\alpha}, \beta^* - \bar{\beta}) \rangle + C \left(||\alpha^* - \bar{\alpha}||^2 + ||\beta^* - \bar{\beta}||^2 \right). \tag{2.5.43}$$

Applying this inequality to (2.5.42) and noting that $\nabla F(\bar{\alpha}, \bar{\beta}) = \left(\nabla_x \varphi^*, \frac{1}{2}\nabla_x^2 \varphi^*\right)$ gives us

$$0 \geqslant \int_\Lambda C \left(||\alpha^* - \bar{\alpha}||^2 + ||\beta^* - \bar{\beta}||^2 \right) \mathrm{d}\rho^* \geqslant 0. \tag{2.5.44}$$

Therefore we have $(\alpha^*, \beta^*) = (\bar{\alpha}, \bar{\beta})$ up to $\mathrm{d}\rho^*$ almost everywhere. □

Chapter 3

SOT Calibration with Stochastic Interest Rates: Numerics

Equipped with the duality and dimension reduction results of Chapter 2, we provide various numerical implementations in different contexts of model calibration. Specifically, we seek to calibrate a stock price local volatility model with volatility coefficient depending on time, the underlying, and the short rate process; driven by a Brownian motion which can be correlated with the randomness driving the rates process. Different approaches can be used in this task, either opting for a *sequential calibration*: the interest rates are calibrated (or given) first, and then a hybrid equity/rates model is calibrated to products exhibiting a dependency to both rates and equity dynamics, or, going for a *joint calibration* where the full covariance process is calibrated at once using rates and equity market prices.

This Chapter is structured as follows: first we give a generic numerical method to solve the two dimensional SOT calibration problem from which we construct our calibration methods. Then we introduce the idea of a "sequential calibration" and provide numerical results on simulated data. We then relax the assumptions on the model characteristics, and develop a joint calibration method tested on both simulated and market data. Some observations in the results of joint calibration, namely the correlation surface and the significantly slower calibration speed, then lead us to test different cost functions and, more importantly, develop a much faster method of computing the calibrated surfaces. This allows us to develop a third

method conceptually in-between the two called "full sequential" calibration. This is more in the spirit of sequential calibration, but without the strong assumptions on the correlation. We compare all three on simulated data with a correct interest rate reference, and then with an incorrect interest rate reference (e.g. if there were errors in the prior calibration of the interest rate). We also demonstrate full sequential calibration applied to the Heston model calibration of I. Guo, Loeper, and Shiyi Wang, 2022 on simulated data. Finally, we extend the time horizon and verify whether applying joint calibration only to call options generates "jointly dependent" surfaces, since over a shorter time horizon we observe that our volatility surfaces generally exhibit weak dependence on one of the state variables. Highlights of this chapter are published as preprints in Joseph, Loeper, and Obłój, 2023a; Joseph, Loeper, and Obłój, 2023b.

3.1 Generic Numerical Method

Our numerical approach focuses on solving the dual problem. Specifically we adapt the methods presented in I. Guo, Loeper, and Shiyi Wang, 2022; I. Guo, Loeper, Obłój, et al., 2022a. Given a fixed $\lambda \in \mathbb{R}^n$, we first want to solve (3.3.4). We use a discretisation on a uniform spatial grid of $[\log S_{\min}, \log S_{\max}] \times [r_{\min}, r_{\max}]$ for the log-stock and interest rates[1]. Recall that we have

$$F^*\left(\nabla_x \varphi, \frac{1}{2}\nabla_x^2 \varphi\right) = \alpha^* \cdot \nabla_x \varphi + \frac{1}{2}\beta^* : \nabla_x^2 \varphi - F\left(\alpha^*, \beta^*\right), \tag{3.1.1}$$

with $(\alpha^*, \beta^*) = \nabla F^*\left(\nabla_x \varphi, \frac{1}{2}\nabla_x^2 \varphi\right)$. It is therefore easy to see that the HJB equation (2.5.2) is a fully nonlinear parabolic PDE. However if (α^*, β^*) are fixed, then it is a linear PDE which we will refer to as the "linearised HJB equation", which we discretise using an implicit finite differences scheme with central differences in the spatial derivatives.

In order to solve the nonlinear PDE, we will use a policy iteration method as in I. Guo, Loeper, and Shiyi Wang, 2022; I. Guo, Loeper, Obłój, et al., 2022a to solve it (see also Peter A Forsyth and Labahn, 2007; J. Wang and Peter A Forsyth, 2008; Bokanowski, Maroso,

[1] In all of the subsequent numerical examples we will use a 100×100 spatial grid and partition the time interval into days, so that $\mathrm{d}t = \frac{1}{365}$.

and Hasnaa Zidani, 2009; Peter A Forsyth and Vetzal, 2011; Reisinger and Witte, 2012; K. Ma and P. Forsyth, 2017). Since outside of the maturities τ_i our solution is continuous in time, we take the initial approximation $\varphi_{\text{init}}(t_n, \cdot) = \varphi(t_{n+1}, \cdot)$, and if at maturity, i.e. for some $i = 1, \ldots, n$, $t_n = \tau_i$, we add $\lambda_j G_j(x)$ for all j such that $\tau_j = \tau_i$ to $\varphi(t_{n+1}, \cdot)$ and obtain our initialisation. With a previous approximation, $\varphi_{\text{old}}(t_n, \cdot)$, the idea behind our iteration is to find optimisers $(\alpha^*_{t_n}, \beta^*_{t_n})$ using (2.5.31) evaluated at $\varphi_{\text{old}}(t_n, \cdot)$, and then solve the linearised HJB equation implicitly backwards in time over $\{t_{n+1}, t_n\}$ to find $\varphi_{\text{new}}(t_n, \cdot)$. We then update $\varphi_{\text{old}}(t_n, \cdot)$ with $\varphi_{\text{new}}(t_n, \cdot)$, and iterate until some specified tolerance ε_2 is reached such that $||\varphi_{\text{old}}(t_n, \cdot) - \varphi_{\text{new}}(t_n, \cdot)||_\infty < \varepsilon_2$, at which point our solution $\varphi(t_n, \cdot)$ is given by $\varphi_{\text{new}}(t_n, \cdot)$.

The main result of Barles and P. Souganidis, 1991 guarantees convergence to the viscosity solution if the numerical scheme is stable, consistent, and monotone (provided there is a comparison principle in the viscosity solution). Explicit methods such as Bonnans and Housnaa Zidani, 2003; Debrabant and Jakobsen, 2013, while easier to implement — since the formula for (α^*, β^*) is then given directly by (2.5.31) at each timestep and thus we entirely avoid the necessity of policy iteration — must satisfy a CFL condition for stability and therefore have restrictions on the size of timestep. As the discretisation used in the linearised HJB equation within the policy iteration method is fully implicit, we do not have to restrict to very small timesteps for stability. A sufficient condition for monotonicity is that β^* is diagonally dominant (see Bonnans and Housnaa Zidani, 2003, Lemma 3.1 or Barles and Jakobsen, 2002). This is difficult for us to guarantee without imposing high lower bounds on β^*_{11} and β^*_{22}, however we observe that in practice the policy iteration converges.

We also mention other methods for the numerical approximation of viscosity solutions to PDEs. First order upwind schemes have been shown to converge to the viscosity solution – see P. E. Souganidis, 1985; Barles and P. Souganidis, 1991; Oberman, 2006. They have been developed to solve an HJB equation arising in control theory in the context of finite differences in Shouyang Wang, Gao, and Teo, 2000, and a finite volume approach in Song Wang, Jennings, and Teo, 2003; Richardson and Song Wang, 2006; second order schemes have also been studied in Jean-David, Luo, and Zhao, 2010 in the context of the Eikonal equation. An efficient method approximating the nonlinear discretisation was developed in

Witte and Reisinger, 2011; Witte and Reisinger, 2012.

We choose a boundary far away enough such that the boundary conditions do not have a significant effect on the HJB equation solution within the region corresponding to the strikes of the calibrating options, and our boundary conditions are such that the second derivative of φ does not change with time between each maturity. That is, for a subsequence of the calibrating option maturity times $(\tau_{i_k})_{k=1,\dots,m}$ such that for $k = 1, \dots, m$ all τ_{i_k} are distinct, and with $\tau_{i_0} = 0$,

$$\partial^2_{zz}\varphi(t,z,r) = \partial^2_{zz}\varphi(\tau_i,z,r), \quad \text{for } z \in \{z_{\min}, z_{\max}\},\ r \in [r_{\min}, r_{\max}],\ t \in (\tau_{i_{k-1}}, \tau_{i_k}], \tag{3.1.2}$$

$$\partial^2_{rr}\varphi(t,z,r) = \partial^2_{rr}\varphi(\tau_i,z,r), \quad \text{for } z \in [z_{\min}, z_{\max}],\ r \in \{r_{\min}, r_{\max}\},\ t \in (\tau_{i_{k-1}}, \tau_{i_k}]. \tag{3.1.3}$$

We started with this boundary condition and compared it with constant Dirichlet boundary conditions $\varphi(t,x) = C_{i_k}$, for some constant $C_{i_k} \in \mathbb{R}$, all $t \in (\tau_{i_{k-1}}, \tau_{i_k}]$, and x on the boundary of our rectangular domain, and observed no significant change.

Finally, as call option payoffs are not differentiable at the strike, we observed numerical instabilities when adding the jump discontinuities and calculating $\partial^2_{rr}\varphi$ and $\partial^2_{zz}\varphi$. To overcome this, we instead took the approximation (3.1.4) for the call and cap payoffs with some parameter $\varepsilon \ll K$. As the numerical instabilities were particularly problematic when a strike was close to a grid point, we took $\varepsilon > \delta z$ to ensure the smoothing had an effect, but also on the same order of magnitude so that the approximation remains valid. When computing the model prices in (3.1.8), we then instead use the exact call payoff since the instability is only observed in the nonlinear HJB equation.

$$(S_T - K)^+ \approx \frac{S_T - K}{2}\left(\tanh\left(\frac{S_T - K}{\varepsilon}\right) + 1\right) \tag{3.1.4}$$

Having solved the HJB equation for a fixed λ, we turn our attention to solving the dual problem in Theorem 2.5.1 and finding the optimal λ^*. First, we observe that we can speed up the optimisation routine by providing a formula for the gradients. This formula is obtained in the same way as I. Guo, Loeper, Obłój, et al., 2022a, but with the discounting appearing as a result of the $-r\varphi$ term in the HJB equation and the Feynman-Kac formula

(see Klebaner, 2012, Theorem 6.8).

Lemma 3.1.1. *Suppose Problem 2.4.1 is admissible, and define the dual objective function as*

$$L(\lambda) = \lambda \cdot u - \varphi(0, X_0). \tag{3.1.5}$$

Then the gradients of the dual objective function are given by

$$\partial_{\lambda_i} L(\lambda) = u_i - \mathbb{E}^{\mathbb{P}} \left[e^{-\int_0^{\tau_i} r_s \, \mathrm{d}s} G_i(X_{\tau_i}) \right]. \tag{3.1.6}$$

Proof. We have for each $i = 1, \ldots, n$ that $\partial_{\lambda_i} L(\lambda) = u_i - \partial_{\lambda_i} \varphi^{\lambda}(0, X_0)$. Also, by Lemma 2.5.8, we have $\nabla F^* \left(\partial_x \varphi^{\lambda}, \frac{1}{2} \nabla_x^2 \varphi^{\lambda} \right) = (\alpha^{\lambda}, \beta^{\lambda})$. Thus, by differentiating the HJB equation (2.5.2), we obtain with $\psi^{\lambda} = \partial_{\lambda_i} \varphi^{\lambda}$:

$$\begin{cases} \partial_t \psi^{\lambda} - x_r \psi^{\lambda} + \alpha^{\lambda} \cdot \nabla_x \psi^{\lambda} + \frac{1}{2} \beta^{\lambda} : \nabla_x^2 \psi^{\lambda} = -G_i(x) \delta_{\tau_i}, & (t, x) \in [0, T) \times \mathbb{R}^2 \\ \psi^{\lambda}(T, x) = 0, & x \in \mathbb{R}^2. \end{cases} \tag{3.1.7}$$

Since solving (3.1.7) on $[0, T]$ with terminal condition $\psi^{\lambda}(T, \cdot) = 0$ is equivalent to solving (3.1.7) on $[0, \tau_i]$ with terminal condition $\psi^{\lambda}(\tau_i, x) = G_i(x)$, we may directly apply the Feynman-Kac formula and obtain (3.1.6). □

Remark 3.1.2. *This result holds in any dimension, and in higher dimensions the Feynman-Kac representation is useful since the expectation can be evaluated using Monte Carlo methods in parallel. For our two dimensional problem, we directly solve the pricing PDE associated with the expectation in* (3.1.6) *via finite differences since it is more accurate.*

Remark 3.1.3 (Complexity of the method)**.** *It is well known that the solution of a linear system of equations requires $O(n^3)$ operations. If we have n_i mesh points in dimension i, then this means that the total number of equations we solve in the implicit finite differences method is $O(\Pi_{i=1}^{d} n_i^3)$. Thus, our PDE approach suffers from the curse of dimensionality as the number of operations grows exponentially with the dimension, so the numerical method here is only appropriate for $d < 3$. With M timesteps and a maximum number of policy iterations of N, we have that the worst case number of operations for solving the HJB*

equation is $O(n_1^3 n_2^3)$ performed in sequence MN times. In comparison, with a total of L constraints, the complexity of solving all of the linear pricing PDEs is $O(Ln_1^3 n_2^3)$, performed in sequence M times. The pricing PDEs can be solved in parallel, so the most costly step is the solution of the HJB equation.

Our initialisation is $\lambda = 0$ since if the reference model is already calibrated, then that will immediately return $||\nabla_\lambda L(\lambda)||_\infty < \varepsilon_1$. Given a guess λ, we solve the HJB equation (2.5.2). From the solution of (2.5.2), we compute the diffusion coefficients $(\alpha^*(t,x), \beta^*(t,x))$ from (2.5.31). For each instrument we compute the model price by solving the linear pricing PDEs for $(t,x) \in [0,\mathcal{T}) \times \mathbb{R}^2$:

$$\begin{cases} \partial_t \psi(t,x) + \alpha^*(t,x) \cdot \nabla_x \psi(t,x) + \frac{1}{2}\beta^*(t,x) : \nabla_x^2 \psi(t,x) - r\psi(t,x) = 0, \\ \psi(\mathcal{T},x) = \mathcal{G}(x). \end{cases} \tag{3.1.8}$$

The boundary conditions used here are the same as the HJB equation boundary conditions (3.1.2)-(3.1.3). We can then compute the gradient $\nabla_\lambda L(\lambda)$ corresponding to the difference between the model prices and the market prices, and finally use the L-BFGS algorithm to update λ.

3.1.1 Renormalisation and Reference Model Iteration

The first renormalisation we make is $\lambda \mapsto \frac{\lambda}{\mathcal{V}_{\text{mkt}}}$ where $\mathcal{V}_{\text{mkt}}$ are the market Vegas of the options computed from their Black-Scholes implied volatility. This converts pricing errors into implied volatility errors since the vega represents how much the option price will change as the volatility changes by 1%. This approximation is to first order and can be seen via a Taylor expansion of the price with respect to the volatility parameter. As we wish to fit the implied volatility, this rescaling seemed to help the gradient method update λ appropriately and ensure that the exit criteria of $||\nabla L(\lambda)||_\infty < \varepsilon_1$ was reached with a good implied volatility fit.

The second rescaling we make is $r_t \mapsto Rr_t$ in order to bring the short rate to the same order as the log-stock, where we choose $R = 100$. This is due to the fact that β_{22} is small since $r \approx 0.05$, and β^* must be diagonally dominant to guarantee monotonicity in the

Algorithm 3: Policy iteration algorithm.

Data: Input an initial λ and market prices u_i.
Result: Calibrated model prices, optimal drift and diffusion

while $||\nabla_\lambda L(\lambda)||_\infty > \varepsilon_1$ **do**
  `/* Solve the HJB equation backwards in time */`
  for $k = N-1, \ldots, 0$ **do**
    `/* Terminal Conditions - adding` λ `multiplied by the payoff */`
    if $t_{k+1} = \tau_i$ *for some* $i = 1, \ldots, n$ **then**
      $\varphi_{t_{k+1}} \leftarrow \varphi_{t_{k+1}} + \sum_{i=1}^{n} \lambda_i G_i \mathbb{1}_{\{t_{k+1}=\tau_i\}}$
    end
    `/* Policy iteration to approximate the optimal characteristics */`
    $\varphi_{t_k}^{\text{new}} \leftarrow \varphi_{t_{k+1}}$ `// Approximate using previous time step`
    while $||\varphi_{t_k}^{\text{new}} - \varphi_{t_k}^{\text{old}}|| > \varepsilon_2$ **do**
      $\varphi_{t_k}^{\text{old}} \leftarrow \varphi_{t_k}^{\text{new}}$ `// Store the old value of` φ
      Approximate (α^*, β^*) with $\varphi_{t_k}^{\text{old}}$ using (2.5.31). `// Use old values to approximate optimal characteristics`
      Plug (α^*, β^*) into (2.5.2) to remove the supremum and solve using one step of an implicit finite difference method, and set the solution to $\varphi_{t_k}^{\text{new}}$.
    end
    $\varphi_{t_k} \leftarrow \varphi_{t_k}^{\text{new}}$ `// Save the solution once the` φ `has converged to the optimal solution`
  end
  `/* Computing the model prices and gradients */`
  Compute the model prices by solving the pricing PDE (3.1.8) using the ADI method.
  Compute the gradients (3.1.6).
  Use the L-BFGS algorithm to update λ.
end

numerical scheme. The choice of $R = 100$ is arbitrary, but seemed to work well for the purposes of the proof of concept numerical experiments and we note that other approaches could also be pursued, e.g., working with rescaled $\frac{S_t}{S_0}$, $\frac{r_t}{r_0}$ quantities.

If Algorithm 3 is directly applied, then often it will output a model with spiky drift and diffusion surfaces which is undesirable. This is a result of our cost function penalising deviations from a reference model, which results in larger local changes to the surfaces, but with the rest of the overall surface being closer to the reference model. Moreover, instabilities from the $\partial^2_{zz}\varphi$ and $\partial^2_{rr}\varphi$ terms in β_{11} and β_{22} at maturity at the strikes of our options when adding in the jump discontinuities are unavoidable. In addition, sometimes the L-BFGS algorithm will make very small steps, thus requiring a large number of iterations and resulting in the calibration taking too long as is a common issue in first order gradient

methods. To avoid all of this and improve the convergence to our calibrated model, we apply a "reference model iteration" technique as in I. Guo, Loeper, Obłój, et al., 2022a. That is, once our optimisation routine finishes, we apply an interpolation and smoothing technique to the output model drift and diffusion terms, and then set those terms to our new reference model and re-run the calibration.

Our original SOT calibration problem only optimises over $\mathbb{P}$ given a fixed reference $\overline{\mathbb{P}}$, however the choice of $\overline{\mathbb{P}}$ will certainly dictate the features of $\mathbb{P}^*$. Therefore, this "reference model iteration" can be viewed as solving a meta-problem where we modify the choice of $\overline{\mathbb{P}}$ to improve $\mathbb{P}^*$. This is because for a fixed reference model, the calibration routine will always converge to some calibrated model since the problem is convex, however this calibrated model might be spiky or take too long to reach. By smoothing a calibrated model (or a model obtained after a certain number of iterations), the perturbations in the volatility surface are spread out and the smoothed model is closer to a calibrated model than the original reference guess. Since changing the reference model changes the cost function, the next epoch of the meta-algorithm will have a different optimiser, which we have observed empirically to be smoother than the previous calibrated model. The final output model is not smoothed since this would no longer be a calibrated model. This not only speeds up overall convergence to a useful calibrated model, but also decreases sharp peaks that can arise in our calibrated local volatility surfaces otherwise. This "reference model iteration" need not converge to a fixed reference model as it is only used as a tool to generate a smoother volatility surface. Thus the exit criteria for this can be defined by the modeller depending on how smooth they would like the local volatility surface while taking into account the extra computational time due to the additional epochs required.

3.1.2 Description of subsequent numerical results presented

The rest of this Chapter will cover numerical results in different calibration scenarios detailing the construction of the calibration problems. In all examples, the key plots to demonstrate the accuracy of calibration are the implied volatility curves. The model prices for these plots are computed for a superset of strikes containing those in the calibration constraints. In the simulated data examples, these can then be superimposed on the generating

model prices to show that the SOT method also recovers non-listed strikes in the region $[K_{\min}, K_{\max}]$. Further, in the simulated data examples we provide a table showing the exact price and IV values obtained in the generating and calibrated models to further show the calibration accuracy. This is omitted in the market data examples due to the terms of use of the data from the provider. For replication purposes, we also include a table of all parameters used in the calibration. At the end of each subsection, we also produce the local volatility surfaces at the maturity dates and in-between to show the dynamic properties of the calibrated model. We additionally provide the reference (and generating models in the simulated case), to show the dependence of the calibrated model on the reference.

3.2 Sequential Calibration

We start with a setting in which a model for the short rate is fixed and has already been calibrated. We refer to this as a "sequential calibration" problem. Our aim is then to calibrate a local volatility model for the stock price, where the local volatility function can depend on the short rate. Specifically, we take a given pre-calibrated Hull-White model for the interest rate, see Hull and A. White, 1990; Hull and A. White, 1994, and local volatility dynamics for the log-price:

$$\begin{aligned} \mathrm{d}Z_t &= \left(r_t - \frac{1}{2}\sigma^2(t, Z_t, r_t)\right)\mathrm{d}t + \sigma(t, Z_t, r_t)\,\mathrm{d}W_t^1, \\ \mathrm{d}r_t &= (b(t) - ar_t)\,\mathrm{d}t + \sigma_r\,\mathrm{d}W_t^2, \\ \mathrm{d}\langle W^1, W^2\rangle_t &= \xi(t, Z_t, r_t)\,\mathrm{d}t, \end{aligned} \tag{3.2.1}$$

where $a, \sigma_r > 0$ are constants and $b(\cdot)$ is a function of time, calibrated so that the dynamics of $(r_t)_{0\leqslant t\leqslant T}$ match the market data (e.g., suitable interest rates caps and floors). Both a and σ_r being positive constants is not a particularly restrictive constraint as remarked in Brigo and Mercurio, 2007; Hull and A. White, 1995. Note that $t \mapsto b(t)$ will therefore need to be calibrated to fit the term structure of interest rates seen in the market. Our aim now is to calibrate $\sigma(t, Z_t, r_t)$ and $\xi(t, Z_t, r_t)$ using our OT-methodology. In order to calibrate the local volatility and correlation, we will want to find a cost function that forces $\alpha_t^{\mathbb{P}}$ and $\beta_t^{\mathbb{P}}$ to

take the form above. As discussed in Section 2.1, we achieve this by using a functional form for F as long as $(\alpha, \beta) \in \Gamma$ for some convex set Γ, with $F = +\infty$ otherwise. The set Γ will enforce that at least β is positive semidefinite and symmetric, and $\beta_{22} = \sigma_r^2$, which defines a set that is convex in β. With no further constraints on β_{12}, we obtain a non-explicit form of the Legendre transform of the cost function, F^*, which in general will require a numerical solver at each (t, x). To reduce the computational difficulty of the numerical solution, as in I. Guo, Loeper, and Shiyi Wang, 2022, we first restrict the correlation to the following form, which still keeps the set Γ convex:

$$\xi(t, Z_t, r_t) = \frac{\sigma_r(t)}{\sigma(t, Z_t, r_t)} \xi_{\text{ref}}(t, Z_t, r_t), \quad \text{for } t \in [0, T], \tag{3.2.2}$$

where $\xi_{\text{ref}}(t, Z_t, r_t) \in \mathbb{R}^2$ here is a fixed reference function. We then set:

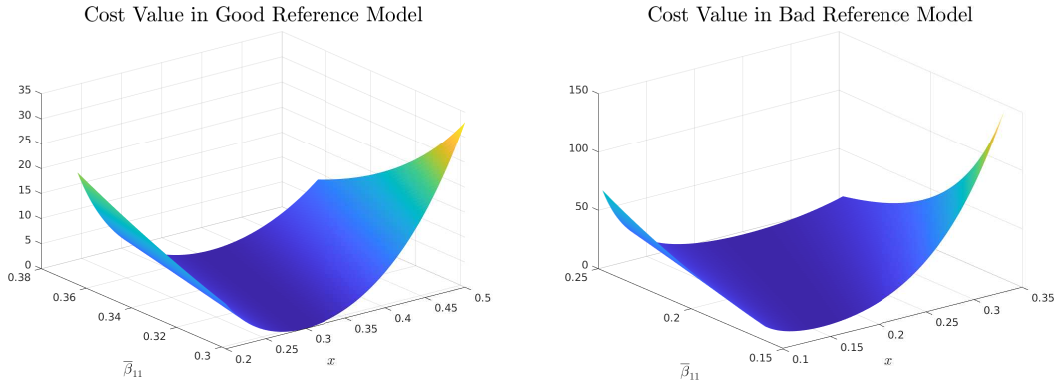
$$\Gamma(t, Z_t, r_t) = \left\{ (\alpha, \beta) \in \mathbb{R}^2 \times \mathbb{S}^2 \ : \ \alpha_1 = r_t - \frac{1}{2}\beta_{11},\ \alpha_2 = (b(t) - a r_t),\ \beta_{12} = \beta_{21} = \xi_{\text{ref}} \sigma_r^2,\ \beta_{22} = \sigma_r^2 \right\}, \tag{3.2.3}$$

where one would change the set Γ suitably if a different short rate model was fixed initially. We also require the inequality $\xi_{\text{ref}}(t, Z_t, r_t)^2 \sigma_r^2 \leqslant \sigma^2(t, Z_t, r_t)$ to keep $\xi(t, Z_t, r_t) \in [-1, 1]$ as a correlation also. Since r_t is on a much lower scale to Z_t, we have that $\sigma_r^2 \ll \sigma^2$, so this condition is not financially restrictive. To enforce this condition, we will define a convex function $H : \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ with a parameter $p > 2$:

$$H(x, \bar{x}, s) \coloneqq \begin{cases} (p-1)\left(\frac{x-s}{\bar{x}-s}\right)^{1+p} + (p+1)\left(\frac{x-s}{\bar{x}-s}\right)^{1-p} - 2p, & \text{if } x, \bar{x} > s, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.2.4}$$

The coefficients of each term ensure that $H(x, \bar{x}, s)$ is minimised over x at $x = \bar{x}$ with $\min_x H(x, \bar{x}, s) = 0$. We fix a reference local volatility function $\bar{\sigma}^2 = \bar{\sigma}^2(t, Z_t, r_t)$ that represents the desired model. Our aim is to find a calibrated model which does not deviate too much from the reference one. To achieve this we set

$$F(t, Z, r, \alpha, \beta) = \begin{cases} H(\beta_{11}, \bar{\sigma}^2, \xi_{\text{ref}}^2 \sigma_r^2), & \text{if } (\alpha, \beta) \in \Gamma(t, Z_t, r_t), \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.2.5}$$



(a) (b)

Figure 3.1: The function $(x, \overline{\beta}_{11}) \mapsto H(x, \overline{\beta}_{11}, \xi^2_{\text{ref}}\sigma^2_r)$ with the reference models given in Table 3.1

Proposition 3.2.1. *The function F defined in (3.2.5) satisfies the assumptions in Section 2.1.*

Proof. For notational ease, define $\bar{x} := \overline{\sigma}^2$ and $s := \xi^2_{\text{ref}}\sigma^2_r$. Since the set $\Gamma(t, Z_t, r_t)$ is convex, and the constraint $x, \bar{x} > s$ is affine, we only need to check the conditions on the set where F is finite. The function $x \mapsto H(x, \overline{x}, s)$ is clearly lower semicontinuous for $x, \bar{x} > s$. Moreover, p-coercivity follows by taking the exponent p in (2.1.11) to have the same value as in (3.2.5). For $x \leqslant \bar{x}$, since the left hand side of (2.1.11) is decreasing to a minimum at $x = \bar{x}$ and the right hand side is increasing, by taking $C > \bar{x}^p$, the inequality is satisfied. For $x > \bar{x}$, it suffices to note that for x sufficiently large, the first term in H will dominate $|x|^p$, thus we can choose a C large enough so that (2.1.11) is satisfied. To check that F satisfies the strong convexity requirement, we only need to check that $\frac{\partial^2}{\partial x^2} H(x, \bar{x}, s) \geqslant M > 0$ (see Nesterov, 2018, Theorem 2.1.11). First, note that since $x, \bar{x} > s$, and $p > 2$:

$$\frac{\partial^2}{\partial x^2} H(x, \bar{x}, s) = p(p^2 - 1)\frac{(x - s)^{p-1}}{(\bar{x} - s)^{p+1}} + p(p^2 - 1)\frac{(x - s)^{-p-1}}{(\bar{x} - s)^{p+1}} > 0. \tag{3.2.6}$$

We obtain the stationary point $x^* = s + (\bar{x} - s)\left(\frac{p+1}{p-1}\right)^{\frac{1}{2p}} > s$ by solving $\frac{\partial^3}{\partial x^3} H(x, \bar{x}, s) = 0$.

This is a local minima since the fourth derivative is given by

$$\frac{\partial^4}{\partial x^4}H(x,\bar{x},s) = p(p^2-1)(p-1)(p-2)\frac{(x-s)^{p-3}}{(\bar{x}-s)^{p+1}} + p(p^2-1)(p+1)(p+2)\frac{(x-s)^{-p-3}}{(\bar{x}-s)^{1-p}} > 0, \tag{3.2.7}$$

since $x, \bar{x} > s$ and $p > 2$. Then by setting $M = \frac{\partial^2}{\partial x^2}H(x^*, \bar{x}, s)$ where $\bar{x} = \overline{\sigma}^2$ and $s = \xi_{\text{ref}}^2\sigma_r^2$, we have strong convexity. □

This cost function will ensure that we retain the Hull-White model in the interest rate, while also matching the market prices for the call options by calibrating the volatility of the stock. Additionally, we wish to enforce that the matrix β from our model characteristics is positive definite and that ξ_{ref} remains a correlation function, and we achieve this by setting $s = \xi_{\text{ref}}^2\sigma_r^2$ as an argument of H in the definition of F. Applying Theorem 2.5.1, we have the following dual formulation with the given cost function $F(\alpha, \beta)$:

Problem 3.2.2 (Hull-White Dual Formulation)**.**

$$V = \sup_{\lambda} \lambda \cdot u - \varphi^{\lambda}(0, Z_0, r_0). \tag{3.2.8}$$

Where $\varphi^{\lambda} = \varphi(t,z,r)$ solves the HJB equation

$$\sum_{i=1}^{n}\lambda_i(\exp(z)-K_i)^+\delta_{\tau_i} + \partial_t\varphi + \sup_{\beta_{11}}\Bigg(\left(r - \frac{1}{2}\beta_{11}\right)\partial_z\varphi + (b(t)-ar)\partial_r\varphi + \frac{1}{2}\beta_{11}\partial_{zz}^2\varphi + \bar{\xi}\sigma_r^2\partial_{zr}^2\varphi + \frac{1}{2}\sigma_r^2\partial_{rr}^2\varphi - r\varphi - H(\beta_{11}, \bar{\sigma}^2, \bar{\xi}^2\sigma_r^2)\Bigg) = 0. \tag{3.2.9}$$

fir $(t,z,r) \in [0,T) \times \mathbb{R}^2$ with terminal condition $\varphi(T,\cdot) = 0$.

Lemma 3.2.3 (Analytic Formula for the Optimal Characteristic)**.** *The optimal characteristic in the HJB equation (3.2.9), β_{11}^* is given by:*

$$\beta_{11}^*(t,z,r) = \bar{\xi}^2\sigma_r^2 + \left(\frac{(\bar{\sigma}^2 - \bar{\xi}^2\sigma_r^2)^p(\varphi_{zz}-\varphi_z)}{4(p^2-1)} + \frac{1}{2}\sqrt{\left(\frac{(\bar{\sigma}^2 - \bar{\xi}^2\sigma_r^2)^p(\varphi_{zz}-\varphi_z)}{2(p^2-1)}\right)^2 + 4(\bar{\sigma}^2 - \bar{\xi}^2\sigma_r^2)^{2p}}\right)^{\frac{1}{p}}. \tag{3.2.10}$$

Proof. By differentiating the argument of the supremum in (3.2.9) with respect to β_{11}, we notice that solving the supremum over β_{11} in (3.2.9) is equivalent to solving the equation

for x:

$$\frac{1}{2}(\partial_{zz}^2\varphi - \partial_z\varphi) = \partial_x H(x, \bar{\sigma}^2, \bar{\xi}^2\sigma_r^2). \tag{3.2.11}$$

Computing the right hand term and rearranging, we have

$$\frac{\partial}{\partial x}H(x, \bar{\sigma}^2, \bar{\xi}^2\sigma_r^2) = (p^2-1)\left(\left(\frac{x-\bar{\xi}^2\sigma_r^2}{\bar{\sigma}^2-\bar{\xi}^2\sigma_r^2}\right)^p - \left(\frac{x-\bar{\xi}^2\sigma_r^2}{\bar{\sigma}^2-\bar{\xi}^2\sigma_r^2}\right)^{-p}\right) \tag{3.2.12}$$

We again rearrange and arrive at the quadratic in $(x-\bar{\xi}^2\sigma_r^2)^p$:

$$(x-\bar{\xi}^2\sigma_r^2)^{2p} - (x-\bar{\xi}^2\sigma_r^2)^p(\bar{\sigma}^2-\bar{\xi}^2\sigma_r^2)^p\frac{\varphi_{zz}-\varphi_z}{2(p^2-1)} - (\bar{\sigma}^2-\bar{\xi}^2\sigma_r^2)^{2p} = 0 \tag{3.2.13}$$

Thus solving this and taking the positive root since $x > \bar{\xi}^2\sigma_r^2$ gives us the desired answer. □

Once the optimal β_{11}^* is obtained, the full model dynamics are specified and we can compute the model price $\psi(0,z,r)$ of an instrument, with payoff G and maturity $\tau \leqslant T$, by solving the standard pricing PDE for $(t,z,r) \in [0,\tau) \times \mathbb{R}^2$:

$$\begin{cases} \partial_t\psi + \left(r - \frac{1}{2}\beta_{11}^*\right)\partial_z\psi + (b(t)-ar)\partial_r\psi + \frac{1}{2}\beta_{11}^*\partial_{zz}^2\psi + \bar{\xi}\sigma_r^2\partial_{zr}^2\psi + \sigma_r^2\partial_{rr}^2\psi - r\psi = 0, \\ \psi(\tau,z,r) = G(z,r). \end{cases} \tag{3.2.14}$$

If we denote $\mathbb{P}^*$ the measure corresponding to the dynamics of the selected model, then via the Feynman-Kac formula, $\mathbb{E}^{\mathbb{P}^*}\left[e^{-\int_0^\tau r_s\,\mathrm{d}s}G(Z_\tau, r_\tau)\middle| Z_0 = z, r_0 = r\right] = \psi(0,z,r)$.

3.2.1 Numerical Results

We present numerical results showcasing the performance of our proposed calibration method. We use simulated data to investigate the advantages and drawbacks of our method, and in particular its dependence on the reference model $\bar{\sigma}(t,x)$ in (3.2.5). Here, the reference is taken to be given by the CEV model (see J. Cox and Ross, 1976; J. Cox, 1996), where in log-stock the volatility is given by $\bar{\sigma}_{\mathrm{CEV}}(t,x) = \bar{\sigma}x^{\gamma-1}$, where $\bar{\sigma} > 0$ and $\gamma \geqslant 0$ are constants. We solve a pricing PDE to compute the generating model prices and consider the following instruments: calls on the underlying with an expiration of 60 days; calls on the underlying with an expiration of 120 days. We keep the interest rate model parameters

the same throughout since that is assumed to be given.

We present two numerical examples, the "good" reference model where the initial reference model is parametrically close to the generating model, and the "bad" reference model where it is not parametrically close to the generating model and in particular has a correlation with a different sign. Note that in the extreme case when the generating model and the reference model are the same, the calibration procedure will stop instantly and recover the generating model. The parameters for all the models are summarised in Table 3.1.

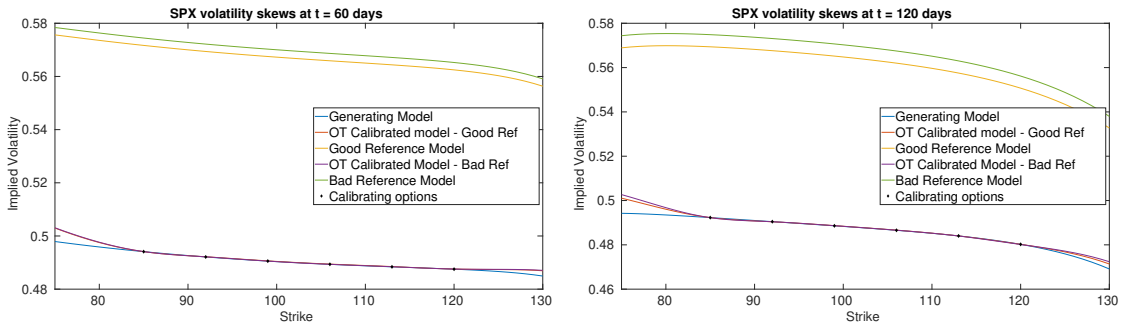
Parameter	Value	Interpretation
Z_0	$\log(92)$	Initial log-underlying price
r_0	0.025×100	Initial short rate scaled by $R = 100$
ε_1	1×10^{-4}	Tolerance for the difference in model and market implied volatility
ε_2	1×10^{-8}	Tolerance for the policy iteration approximation of the optimal characteristics
p	4	Exponent in the cost function
σ	0.78	Volatility scaling of generating CEV model
γ	0.9	Power law in generating CEV model
$b(t)$	$ar_0 + \frac{\sigma_r^2}{2a}(1 - e^{-2at})$	Initial term structure of Hull-White generating model
a	0.4	Speed of mean reversion of Hull-White generating model
σ_r	0.05	Volatility of Hull-White generating model
ξ	-0.6	Instantaneous correlation between short rate and log-stock in generating model
$\overline{\sigma}_{\text{good}}$	0.9	Volatility scaling of the "good" reference CEV model
$\overline{\gamma}_{\text{good}}$	0.9	Power law in the "good" reference CEV model
$\overline{\xi}_{\text{good}}$	-0.4	Instantaneous correlation between short rate and log-stock in the "good" reference model
$\overline{\sigma}_{\text{bad}}$	1.2	Volatility scaling of the "bad" reference CEV model
$\overline{\gamma}_{\text{bad}}$	0.78	Power law in the "bad" reference CEV model
$\overline{\xi}_{\text{bad}}$	0.4	Instantaneous correlation between short rate and log-stock in the "bad" reference model

Table 3.1: Parameter values for the simulated data example.

Our numerical methods converged for both the "good" and "bad" reference model cases, calibrating all the call options to a tolerance of 10^{-4} in implied volatility, with the calibrated model implied volatility replicating the prices of the generating model. The generating and calibrated model implied volatility skews are indistinguishable for both maturities within the range $K = [85, 120]$ in which our options' strikes were taken, indicating that our op-

timal transport model replicates the observed simulated data both at and in-between the calibrating option strikes. The dependence on the reference model is only observed outside of this range, which is to be expected. The plots are given in Figure 3.2 with the exact results summarised in Table 3.2. The reference model iteration came at a significant computational cost, as it essentially repeats the calibration with a different reference to obtain a smoother calibrated surface, however both ran on a standard notebook laptop in a matter of hours. Both reference models were smoothed around 10 times before changes in the volatility surface plots were no longer noticeable.

3.2.1.1 Implied Volatility Plots



(a) (b)

Figure 3.2: Implied volatility skews under the generating model, reference model and OT-calibrated models with both reference models and across two maturities.

3.2.1.2 Plots of surfaces of model characteristics

To better illustrate the features of the OT-calibrated model, we present plots of the surfaces of the model characteristics – see Figures 3.3 and 3.5. Broadly speaking, as discussed above, the OT-calibrated model is close to the generating one in the region specified by the data and close to the reference one otherwise. In addition, the surfaces we obtain for the correlation ξ are highly dependent on the reference model, as one may expect from the the convexity adjustment in (3.2.2). The choice of reference value will therefore mean that ξ is always close to the reference model in this case with some fluctuations obtained through the change in the value β_{11}. Consequently, the modeller's (or trader's) insight in specifying a correct correlation (not least its sign) are, relatively speaking, more important as the market data

		Generating Model		Calibrated Model: Good Reference		Calibrated Model: Bad Reference	
Option Type	Strike	Price	IV	Price	IV	Price	IV
SPX call options $t = 60$ days	85	11.3666	0.4941	11.3666	0.4941	11.3668	0.4941
	92	7.5389	0.4921	7.5398	0.4922	7.5396	0.4922
	99	4.7538	0.4906	4.7549	0.4906	4.7537	0.4905
	106	2.8616	0.4893	2.8625	0.4894	2.8613	0.4893
	113	1.6523	0.4884	1.6532	0.4885	1.6526	0.4884
	120	0.9189	0.4875	0.9192	0.4876	0.9192	0.4876
SPX call options $t = 120$ days	85	14.2787	0.4923	14.2787	0.4923	14.2780	0.4923
	92	10.7017	0.4905	10.7007	0.4904	10.7009	0.4904
	99	7.8563	0.4886	7.8580	0.4887	7.8575	0.4886
	106	5.6560	0.4866	5.6575	0.4866	5.6568	0.4866
	113	3.9917	0.4840	3.9918	0.4840	3.9910	0.4840
	120	2.7493	0.4802	2.7495	0.4802	2.7483	0.4802

Table 3.2: Table of the generating and calibrated model prices and implied volatilities.

alone will not necessarily help to correct a widely wrong reference guess.

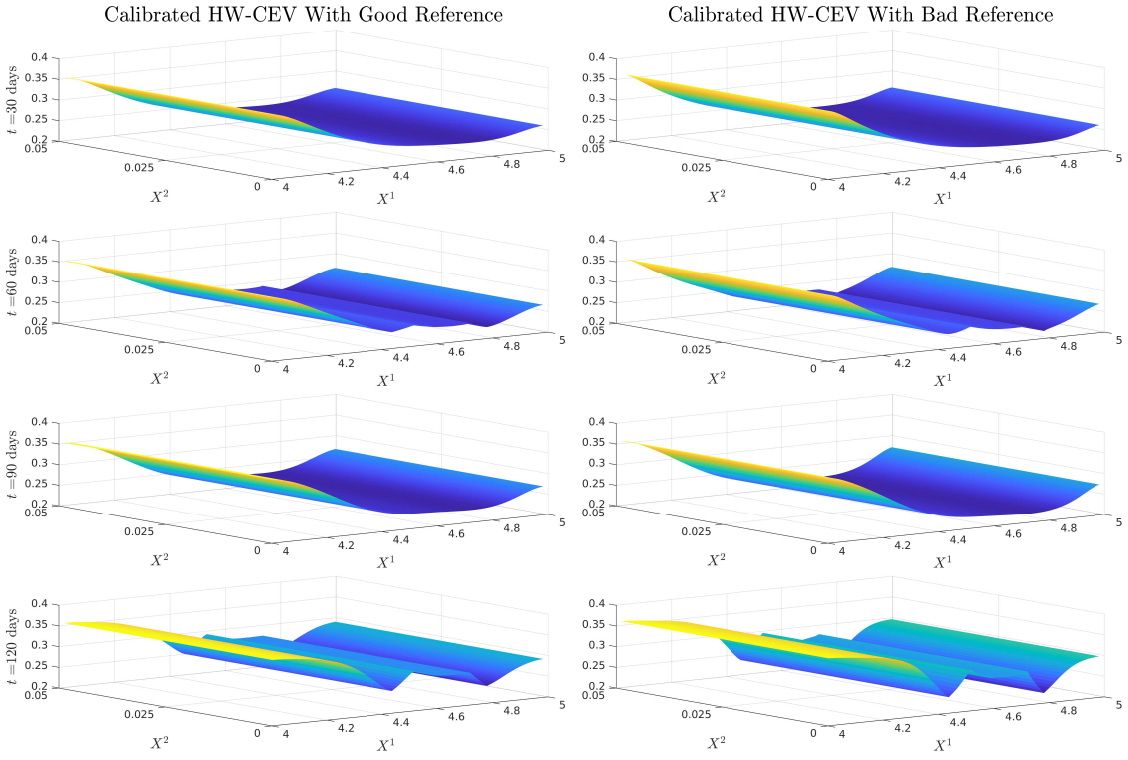
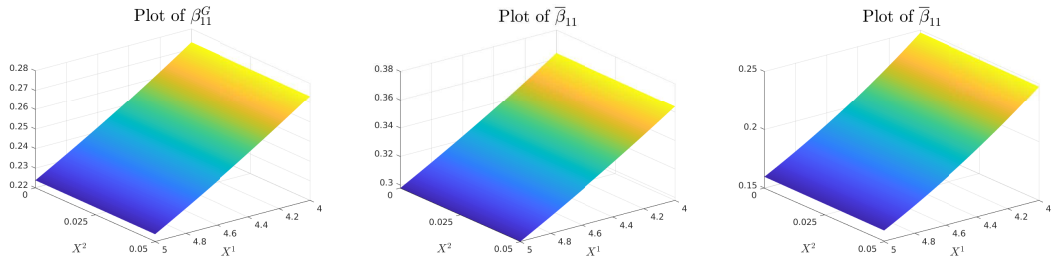


Figure 3.3: Plots of β_{11} at $t = 30$, 60, 90, and 120 days for the calibrated model.



(a) (b) (c)

Figure 3.4: Plots of generating β_{11} (a), good reference model $\overline{\beta}_{11}$ (b), and the bad reference model $\overline{\beta}_{11}$ (c).

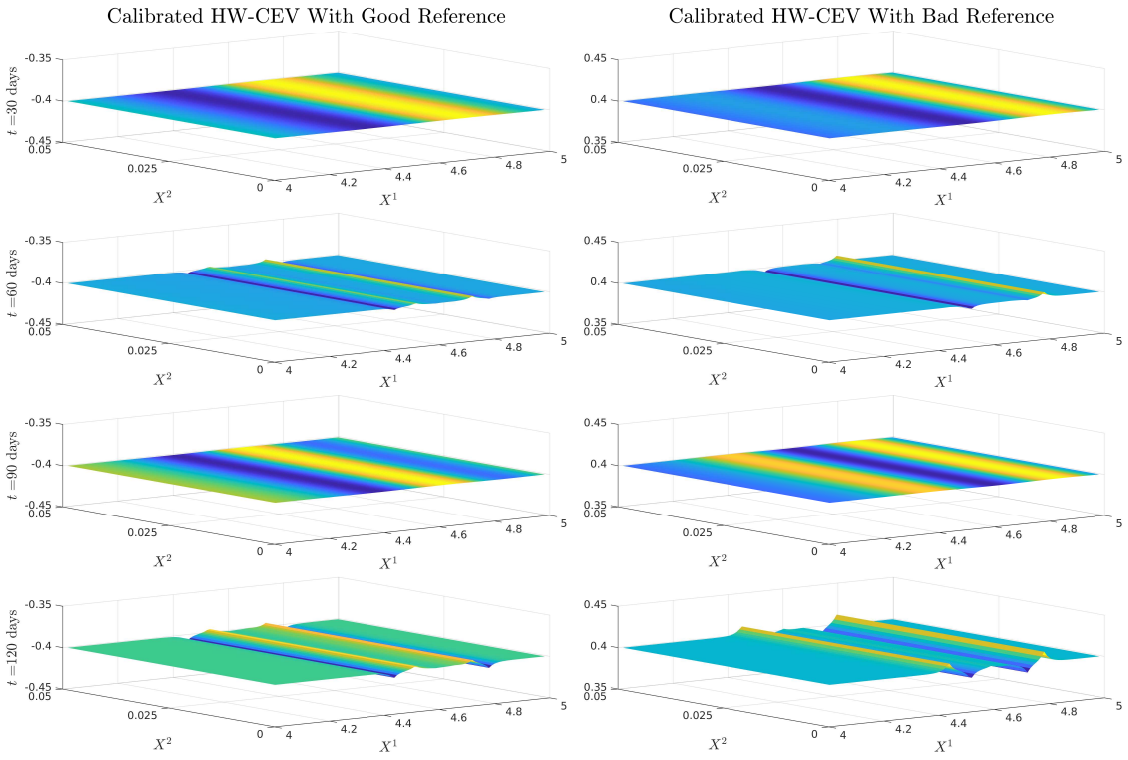


Figure 3.5: Plots of ξ at $t = 30$, 60, 90, and 120 days for the calibrated model. The generating parameter is given by $\xi = -0.6$, the good reference parameter is given by $\xi_{\text{good}} = -0.4$, and the bad reference parameter is given by $\xi_{\text{bad}} = 0.4$.

3.3 Joint Calibration

We now focus on jointly calibrating the log-stock of the S&P 500 with the short rate. Our method offers exact joint calibration of the stock price's process and the stochastic short rate processes to market prices of, respectively, European options, European calls and caps. Our

SOT reformulation can be seen as a projection on to the space of exactly calibrated models: we start with a given reference model and the cost functional which we seek to penalize deviations from that reference. While enjoying the accuracy of exactly calibrating to chosen product prices, the penalization away from the reference model offers several advantages. First, it convexifies the problem giving a unique solution (for a fixed reference). Second, this convexity is known to give regularity to the solution (see Loeper, 2018, Loeper and Quirós, 2018). Third, it gives the user some control about choosing a meaningful reference model from which one tries not to depart too much. To construct our cost function, we first define the convex set $\Gamma(t,x)$ by

$$\Gamma(t,x) \coloneqq \left\{(\alpha,\beta) \in \mathbb{R}^2 \times \mathbb{S}^2_+ \ : \ \alpha_1 = x_2 - \frac{1}{2}\beta_{11}\right\}. \tag{3.3.1}$$

We take the initial measure, μ_0 to be the dirac measure of the today's stock price and interest rate $X_0 \in \mathbb{R}^2$ which is known. Then, using our reference model $(\overline{\alpha},\overline{\beta}) \in \mathbb{R}^2 \times \mathbb{S}^2_+$ we define the cost function $F(\alpha,\beta)$ as:

$$F(\alpha,\beta) = \begin{cases} ||\alpha - \overline{\alpha}||_2^2 + ||\beta - \overline{\beta}||^2_{\text{Fro}}, & \text{if } (\alpha,\beta) \in \Gamma(t,x), \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.3.2}$$

Using this cost function, applying Theorem 2.5.1, we have that our dual problem is given by

$$V = \sup_{\lambda,\varphi} \lambda \cdot u - \varphi(0,X_0). \tag{3.3.3}$$

Where the supremum is taken over $\lambda \in \mathbb{R}^n$ and φ as a viscosity solution to the HJB equation

$$\begin{aligned} \partial_t\varphi + \sup_{\alpha_2\in\mathbb{R},\beta\in\mathbb{S}^2_+} \Bigg\{ \left(x_2 - \frac{1}{2}\beta_{11}\right)\partial_{x_1}\varphi + \alpha_2\partial_{x_2}\varphi + \frac{1}{2}\beta_{11}\partial^2_{x_1x_1}\varphi + \frac{1}{2}\beta_{22}\partial^2_{x_2x_2}\varphi + \beta_{12}\partial^2_{x_1x_2}\varphi \\ - ||\alpha - \overline{\alpha}||_2^2 - ||\beta - \overline{\beta}||^2_{\text{Fro}} \Bigg\} - x_2\varphi + \sum_{i=1}^n \lambda_i G_i(x)\delta_{\tau_i} = 0, \end{aligned} \tag{3.3.4}$$

for $(t,x) \in [0,T] \times \mathbb{R}^2$ with terminal condition $\varphi(T,\cdot) = 0$. To apply Lemma 2.5.8, we must compute the gradient of the Legendre-Fenchel transform of our cost function (3.3.2).

Lemma 3.3.1 (Gradient of Legendre Transform of Cost Function)**.** *For $x_2 \in \mathbb{R}$, $a \in \mathbb{R}^2$, $b \in \mathbb{S}^2$, $\overline{\alpha} \in \mathbb{R}^2$, $\overline{\beta} \in \mathbb{S}^2$, define*

$$A = \overline{\alpha}_2 + \frac{1}{2}a_2 \tag{3.3.5}$$

$$B = \frac{4}{5}\overline{\beta}_{11} + \frac{2}{5}b_{11} - \frac{1}{5}a_1 + \frac{2}{5}x_2 - \frac{2}{5}\overline{\alpha}_1 \tag{3.3.6}$$

$$C = \overline{\beta}_{12} + \frac{1}{2}b_{12} \tag{3.3.7}$$

$$D = \overline{\beta}_{22} + \frac{1}{2}b_{22} \tag{3.3.8}$$

Then the gradient of the Legendre-Fenchel transform of the cost function is given by

$$\nabla F^*(a,b) = \begin{cases} \left(\begin{bmatrix} x_2 - \frac{1}{2}B \\ A \end{bmatrix}, \begin{bmatrix} B & C \\ C & D \end{bmatrix} \right) & \text{if } \begin{bmatrix} B & C \\ C & D \end{bmatrix} \in \mathbb{S}^2_+ \\ \left(\begin{bmatrix} x_2 \\ A \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right) & \text{if } BD \geqslant C^2 \text{ and } \frac{\sqrt{5}}{2}B + D < 0 \\ \left(\begin{bmatrix} x_2 - \frac{1}{2}\tilde{\beta}_{11} \\ A \end{bmatrix}, \begin{bmatrix} \tilde{\beta}_{11} & \tilde{\beta}_{12} \\ \tilde{\beta}_{12} & \tilde{\beta}_{22} \end{bmatrix} \right) & \text{otherwise} \end{cases} \tag{3.3.9}$$

Where $\tilde{\beta}_{11} = \frac{2}{\sqrt{5}}x' + \frac{2}{\sqrt{5}}z'$, $\tilde{\beta}_{12} = y'$, $\tilde{\beta}_{22} = -x' + z'$, with the parameters given by

$$x' = \overline{x}'\frac{\overline{z}' + \zeta}{\overline{z}' + 2\zeta}, \qquad y' = \overline{y}'\frac{\overline{z}' + \zeta}{\overline{z}' + \frac{2+\sqrt{5}}{2}}, \qquad z' = \overline{z}' + \zeta, \tag{3.3.10}$$

$$(\overline{x}', \overline{y}', \overline{z}') = \left(\frac{\sqrt{5}}{4}B - \frac{1}{2}D, C, \frac{\sqrt{5}}{4}B + \frac{1}{2}D \right), \tag{3.3.11}$$

and where ζ is a solution of the following quartic equation that gives the cost function the lowest value and satisfies $\frac{\sqrt{5}}{4}B + \frac{1}{2}D + \zeta \geqslant 0$

$$\overline{x}'^2 \left(\overline{z}' + \frac{2+\sqrt{5}}{2}\zeta \right)^2 + \frac{\sqrt{5}}{2}\overline{y}'^2(\overline{z}' + 2\zeta)^2 - (\overline{z}' + 2\zeta)^2 \left(\overline{z}' + \frac{2+\sqrt{5}}{2}\zeta \right)^2 = 0 \tag{3.3.12}$$

Proof. First, note that since α_1 is fixed from the construction of Γ and β is required to be

symmetric and positive semidefinite, the Legendre-Fenchel transform of our function F is given by

$$F^*(a,b) = \sup_{\substack{\alpha_2\in\mathbb{R}\\ \beta\in\mathbb{S}^2_+}} \left\{\alpha\cdot a + \beta\cdot b - ||\alpha-\overline{\alpha}||_2^2 - ||\beta-\overline{\beta}||_{\text{Fro}}^2\right\}. \tag{3.3.13}$$

We also have that the gradient is given by the maximisers of the above supremum, that is

$$\nabla F^*(a,b) = \operatorname*{arg\,max}_{\substack{\alpha\in\mathbb{R}^2\\ \beta\in\mathbb{S}^2\\ (\alpha,\beta)\in\Gamma(t,x)}} \left\{\alpha\cdot a + \beta\cdot b - ||\alpha-\overline{\alpha}||_2^2 - ||\beta-\overline{\beta}||_{\text{Fro}}^2\right\}. \tag{3.3.14}$$

The condition $(\alpha,\beta)\in\Gamma(t,x)$ requires $\alpha_1 = x_2 - \frac{1}{2}\beta_{11}$ and $\beta\in\mathbb{S}^2_+$. Thus, we have

$$\begin{aligned}
\nabla F^*(a,b) &= \operatorname*{arg\,max}_{\substack{\alpha_2\in\mathbb{R}\\ \beta\in\mathbb{S}^2_+}}\Bigg\{a_1\left(x_2-\frac{1}{2}\beta_{11}\right) + a_2\alpha_2 + b_{11}\beta_{11} + 2b_{12}\beta_{12} + b_{22}\beta_{22} \\
&\qquad\qquad - \sum_{i=1}^{2}(\alpha_i-\overline{\alpha}_i)^2 - \sum_{i,j=1}^{2}(\beta_{ij}-\overline{\beta}_{ij})^2\Bigg\} \\
&= \operatorname*{arg\,max}_{\substack{\alpha_2\in\mathbb{R}\\ \beta\in\mathbb{S}^2_+}}\Bigg\{-\frac{1}{2}a_1\beta_{11} + a_2\alpha_2 + b_{11}\beta_{11} + 2b_{12}\beta_{12} + b_{22}\beta_{22} + \beta_{11}x_2 - \frac{1}{4}\beta_{11}^2 - \overline{\alpha}_1\beta_{11} \\
&\qquad\qquad - \alpha_2^2 + 2\alpha_2\overline{\alpha}_2 - \beta_{11}^2 + 2\beta_{11}\overline{\beta}_{11} - 2\beta_{12}^2 + 4\beta_{12}\overline{\beta}_{12} - \beta_{22}^2 + 2\beta_{22}\overline{\beta}_{22}\Bigg\} \\
&= \operatorname*{arg\,min}_{\substack{\alpha_2\in\mathbb{R}\\ \beta\in\mathbb{S}^2_+}}\left\{(\alpha_2-A)^2 + \frac{5}{4}(\beta_{11}-B)^2 + 2(\beta_{12}-C)^2 + (\beta_{22}-D)^2\right\}.
\end{aligned} \tag{3.3.15}$$

Where

$$A = \overline{\alpha}_2 + \frac{1}{2}a_2, \quad B = \frac{4}{5}\overline{\beta}_{11} + \frac{2}{5}b_{11} - \frac{1}{5}a_1 + \frac{2}{5}x_2 - \frac{2}{5}\overline{\alpha}_1, \quad C = \overline{\beta}_{12} + \frac{1}{2}b_{12}, \quad D = \overline{\beta}_{22} + \frac{1}{2}b_{22}. \tag{3.3.16}$$

Clearly, the minimiser for α_2 is given by $\alpha_2^* = \overline{\alpha}_2 + a_2$. We reformulate the minimisation problem for $\beta\in\mathbb{S}^2_+$ as

$$(\beta_{11}^*,\beta_{12}^*,\beta_{22}^*) = \operatorname*{arg\,min}_{(x,y,z)\in\mathbb{R}^{\geqslant 0}\times\mathbb{R}\times\mathbb{R}^{\geqslant 0}}\left\{\frac{5}{4}(x-B)^2 + 2(y-C)^2 + (z-D)^2 \;:\; xz\geqslant y^2\right\} \tag{3.3.17}$$

We first perform the rescaling $x \mapsto \frac{\sqrt{5}}{2}x = \tilde{x}$, which gives

$$\frac{5}{4}(x-B)^2 \mapsto (\tilde{x}-\tilde{B})^2, \text{ where } \tilde{B} = \frac{\sqrt{5}}{2}B, \text{ and } \frac{2}{\sqrt{5}}\tilde{x}z \geqslant y^2. \tag{3.3.18}$$

We then perform the coordinate rotation around the y-axis clockwise through an angle of $45°$, which is given by the linear transformation

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ y \\ z \end{bmatrix}. \tag{3.3.19}$$

We also note that the inverse of the composition of these transformations is given by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & \frac{2}{\sqrt{5}} \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}. \tag{3.3.20}$$

We then reformulate the minimisation problem in the new (x', y', z') coordinates as:

$$\underset{(x',y',z')\in W}{\arg\min} \left\{(x'-\overline{x}')^2 + (y'-\overline{y}')^2 + (z'-\overline{z}')^2\right\}, \tag{3.3.21}$$

where $(\overline{x}', \overline{y}', \overline{z}') = \left(\frac{1}{2}\tilde{B} - \frac{1}{2}D, C, \frac{1}{2}\tilde{B} + \frac{1}{2}D\right)$, and W is the convex elliptical cone

$$W = \left\{(x', y', z') \in \mathbb{R}^3 \ : \ z' \geqslant 0, \ z'^2 \geqslant \frac{\sqrt{5}}{2}y'^2 + x'^2\right\} \tag{3.3.22}$$

Thus, this minimisation problem has become minimising the Euclidean distance from $(\overline{x}', \overline{y}', \overline{z}')$ to W in (x', y', z') coordinates, which presents three cases to us:

1. $(\overline{x}', \overline{y}', \overline{z}') \in W$, then the solution to (3.3.21) is attained at $(x', y', z') = (\overline{x}', \overline{y}', \overline{z}')$.

2. The constraint $\overline{z}'^2 \geqslant \frac{\sqrt{5}}{2}\overline{y}'^2 + \overline{x}'^2$ is satisfied, but $\overline{z}' < 0$. Then we must have that the solution lies on $z' = 0$, which therefore gives that $x' = y' = 0$.

3. Neither constraint is satisfied by $(\overline{x}', \overline{y}', \overline{z}')$, so the solution, (x', y', z'), lies on the

boundary, $\partial W = \left\{(x', y', z') \in \mathbb{R}^3 \ : \ z' \geqslant 0, x'^2 + \frac{\sqrt{5}}{2} y'^2 = z'^2\right\}$.

We now solve the third case, by finding the outward pointing normal vector of the surface ∂W at the point (x', y', z') solving (3.3.21), we have

$$\begin{bmatrix} \overline{x}' \\ \overline{y}' \\ \overline{z}' \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} + \zeta \begin{bmatrix} \frac{x'}{\sqrt{x'^2 + \frac{\sqrt{5}}{2} y'^2}} \\ \frac{\frac{\sqrt{5}}{2} y'}{\sqrt{x'^2 + \frac{\sqrt{5}}{2} y'^2}} \\ -1 \end{bmatrix}. \tag{3.3.23}$$

From (3.3.23) and the equation of the boundary ∂W, we obtain the following equations for (x', y', z'):

$$z' = \overline{z}' + \zeta, \qquad x' = \overline{x}' \frac{\overline{z}' + \zeta}{\overline{z}' + 2\zeta}, \qquad y' = \overline{y}' \frac{\overline{z}' + \zeta}{\overline{z}' + \frac{2+\sqrt{5}}{2}\zeta} \tag{3.3.24}$$

Thus, substituting these into the boundary equation for ∂W, we obtain that ζ satisfies

$$\overline{x}'^2 \left(\overline{z}' + \frac{2+\sqrt{5}}{2}\zeta\right)^2 + \frac{\sqrt{5}}{2}\overline{y}'^2(\overline{z}' + 2\zeta)^2 - (\overline{z}' + 2\zeta)^2 \left(\overline{z}' + \frac{2+\sqrt{5}}{2}\zeta\right)^2 = 0 \tag{3.3.25}$$

Thus, to find the solution of (3.3.21), we pick the value of ζ solving (3.3.25) such that $\overline{z}' + \zeta \geqslant 0$ and that ζ gives the smallest value in the objective function of (3.3.21), then use (3.3.24) to obtain the solution in $(x', y'z')$ coordinates, and finally apply (3.3.20) to obtain the result. □

3.3.1 Numerical Results: Simulated Data

We outline the reference model choice which we wish to use in our calibration. We consider the following instruments: calls on the underlying stock with an expiration of 60 days; caplets on the short rate with a maturity of 60 days and a fixed notional amount of \$1,000; calls on the underlying stock with an expiration of 120 days; caplets on the short rate with a maturity of 120 days and a fixed notional amount of \$1,000.

Our reference model follows (3.2.1). With $X_t^1 = \log(S_t)$ and $X_t^2 = r_t$, the characteristics

of our reference model with reference parameters $(\overline{\gamma}, \overline{\sigma}, \overline{b}, \overline{a}, \overline{\rho}, \overline{\sigma}_r)$ are therefore given by:

$$(\overline{\alpha}(t, X_t), \overline{\beta}(t, X_t)) = \left(\begin{bmatrix} X_t^2 - \frac{1}{2}\overline{\sigma}^2 \exp(X_t^1)^{2(\overline{\gamma}-1)} \\ \overline{b}(t) - \overline{a} X_t^2 \end{bmatrix}, \begin{bmatrix} \overline{\sigma}^2 \exp(X_t^1)^{2(\overline{\gamma}-1)} & \bar{\rho}\bar{\sigma}\bar{\sigma}_r \exp(X_t^1)^{\overline{\gamma}-1} \\ \bar{\rho}\bar{\sigma}\bar{\sigma}_r \exp(X_t^1)^{\overline{\gamma}-1} & \bar{\sigma}_r^2 \end{bmatrix} \right). \tag{3.3.26}$$

We test our calibration method on simulated data created using a generating model that is from the same family of models as our reference model, but with different parameters: we choose a "good" reference model that is parametrically close to our generating model, and a "bad" reference model that is parametrically further away, and in particular has the wrong sign for the correlation. A numerical issue was that often the volatility surfaces we obtained would have some very small (but still positive values) for β_{11} and β_{22}. Thus, when we compute the calibrated correlation,

$$\rho_{\text{cal}} = \frac{\beta_{12}}{\sqrt{\beta_{11}\beta_{22}}}, \tag{3.3.27}$$

this would give us very sharp spikes on the calibrated ρ_{cal} from points in our β_{11} or β_{22} surfaces that are close to zero. Thus, we added the following bounds:

$$\beta_{11}(t, x) \geqslant \delta_1 \qquad \beta_{22}(t, x) \geqslant \delta_2, \qquad \text{for all } (t, x) \in [0, T] \times \mathbb{R}^2. \tag{3.3.28}$$

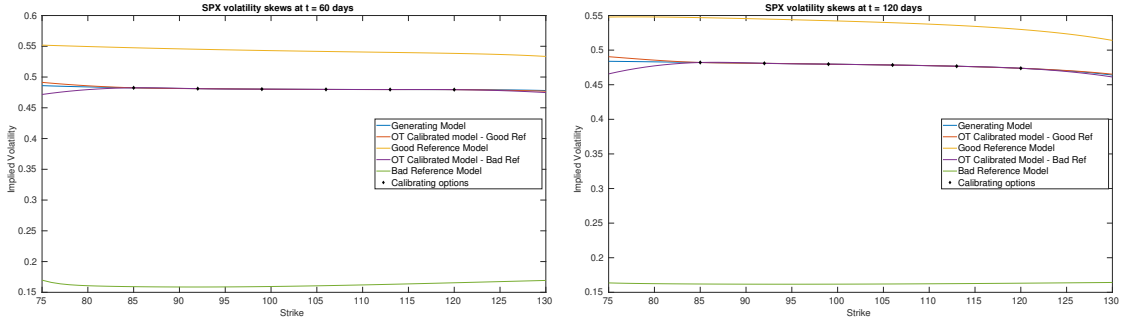
It is difficult to directly adapt the result of Lemma 3.3.1 as jointly enforcing the lower bounds and the positive semidefinite constraint creates a number of subcases that become computationally difficult to solve (see Section 3.3.4 for a discussion on this). Since this classification of cases needs to be applied at each gridpoint separately on each step of the policy iteration algorithm, we found it was more efficient in practice to break from the iterations and smooth whenever either bound in (3.3.28) was violated. This helped us generate more stable correlation surfaces and more realistic volatility surfaces, at the expense of more reference model smoothing iterations being required.

Hull-White CEV Model		
Parameter	Value	Interpretation
X_0^1	$\log(92)$	Initial log-stock price
X_0^2	0.025×100	Initial short rate scaled by $R = 100$

ε_1	1×10^{-4}	Tolerance for the difference in scaled model and market implied volatility
ε_2	1×10^{-12}	Tolerance for the policy iteration approximation of the optimal characteristics
σ	0.60	Volatility scaling of generating CEV model
γ	0.95	Power law in generating CEV model
$b(t)$	$aX_0^2 + \frac{\sigma_r^2}{2a}(1 - e^{-2at})$	Initial term structure of Hull-White generating model
a	0.05	Speed of mean reversion of Hull-White generating model
σ_r	0.04	Volatility of Hull-White generating model
ρ	-0.4	Instantaneous correlation between short rate and log-stock in generating model
$\overline{\sigma}_{\text{good}}$	0.90	Volatility scaling of "good" reference CEV model
$\overline{\gamma}_{\text{good}}$	0.89	Power law in "good" reference CEV model
$\overline{b}_{\text{good}}(t)$	$\overline{a}X_0^2 + \frac{\overline{\sigma}_r^2}{2\overline{a}}(1 - e^{-2\overline{a}t})$	Initial term structure of Hull-White "good" reference model (with "good" reference parameters used in the definition)
$\overline{a}_{\text{good}}$	0.03	Speed of mean reversion of Hull-White "good" reference model
$(\overline{\sigma}_r)_{\text{good}}$	0.02	Volatility of Hull-White "good" reference model
$\overline{\rho}_{\text{good}}$	-0.20	Instantaneous correlation between short rate and log-stock in "good" reference model
δ_1	0.075	Lower bound for β_{11} (the calibrated asset volatility) when calibrating with the "good" reference model
δ_2	0.00	Lower bound for β_{11} (the calibrated scaled interest rate volatility) when calibrating with the "good" reference model
$\overline{\sigma}_{\text{bad}}$	0.20	Volatility scaling of "bad" reference CEV model
$\overline{\gamma}_{\text{bad}}$	0.95	Power law in "bad" reference CEV model
$\overline{b}_{\text{bad}}(t)$	$\overline{a}X_0^2 + \frac{\overline{\sigma}_r^2}{2\overline{a}}(1 - e^{-2\overline{a}t})$	Initial term structure of Hull-White "bad" reference model (with "bad" reference parameters used in the definition)
$\overline{a}_{\text{bad}}$	0.03	Speed of mean reversion of Hull-White "bad" reference model
$(\overline{\sigma}_r)_{\text{bad}}$	0.05	Volatility of Hull-White "bad" reference model
$\overline{\rho}_{\text{bad}}$	0.40	Instantaneous correlation between short rate and log-stock in "bad" reference model
δ_1	0.075	Lower bound for β_{11} (the calibrated asset volatility) when calibrating with the "bad" reference model
δ_2	0.60	Lower bound for β_{22} (the calibrated scaled interest rate volatility) when calibrating with the "bad" reference model

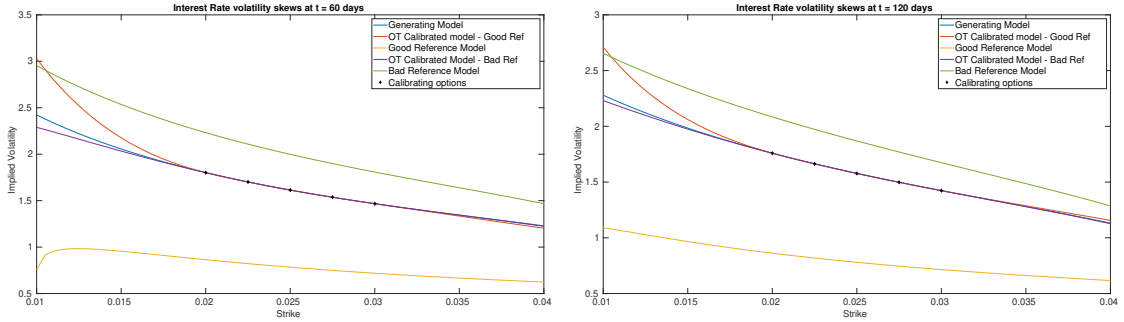
Table 3.3: Parameter values for the HW-CEV simulated data example.

3.3.1.1 Implied Volatility Plots



(a) (b)

Figure 3.6: Implied volatility skews of the calibrating SPX options under the generating model, both reference models and both calibrated HW-CEV models.



(a) (b)

Figure 3.7: Implied volatility skews of the calibrating Short Rate options under the generating model, both reference models and both calibrated HW-CEV models.

Figures 3.6 and 3.7 demonstrate an exact implied volatility fit for all of our calibrating options starting from either a good or a bad reference, see Table 3.4 the degree of accuracy obtained. The surface plots in Section 3.3.1.2 show similar properties to those in the sequential calibration case of Section 3.2.1.2, namely that inside the region of strikes our volatility is close to the generating model, and outside it sticks towards the reference model. The key difference here is in the correlation plot of Figure 3.10 where in comparison with Figure 3.5 we have much less dependency on the reference model, although knowing at least the sign of the correlation is still required.

		Generating Model		Calibrated Model: Good Reference		Calibrated Model: Bad Reference	
Option Type	Strike	Price	IV	Price	IV	Price	IV
SPX call options $t = 60$ days	85	11.2142	0.4825	11.2144	0.4825	11.2139	0.4825
	92	7.3755	0.4811	7.3753	0.4811	7.3752	0.4811
	99	4.6051	0.4803	4.6049	0.4803	4.6051	0.4803
	106	2.7426	0.4799	2.7422	0.4798	2.7428	0.4799
	113	1.5667	0.4797	1.5664	0.4796	1.5665	0.4797
	120	0.8624	0.4795	0.8621	0.4795	0.8624	0.4795
Interest rate caplets $t = 60$ days	0.0200	1.5460	1.8015	1.5459	1.8015	1.5459	1.8015
	0.0225	1.3018	1.7011	1.3019	1.7012	1.3018	1.7011
	0.0250	1.0820	1.6137	1.0820	1.6137	1.0821	1.6137
	0.0275	0.8869	1.5365	0.8869	1.5364	0.8870	1.5366
	0.0300	0.7161	1.4673	0.7162	1.4673	0.7161	1.4672
SPX call options $t = 120$ days	85	14.0842	0.4821	14.0839	0.4821	14.0841	0.4821
	92	10.5020	0.4809	10.5026	0.4809	10.5022	0.4809
	99	7.6696	0.4797	7.6690	0.4797	7.6694	0.4797
	106	5.4943	0.4785	5.4940	0.4785	5.4936	0.4784
	113	3.8607	0.4767	3.8598	0.4767	3.8595	0.4767
	120	2.6499	0.4738	2.6496	0.4738	2.6498	0.4738
Interest rate caplets $t = 120$ days	0.0200	3.8312	1.7587	3.8313	1.7588	3.8313	1.7587
	0.0225	3.3763	1.6638	3.3762	1.6638	3.3764	1.6639
	0.0250	2.9491	1.5776	2.9491	1.5776	2.9490	1.5775
	0.0275	2.5500	1.4980	2.5499	1.4980	2.5501	1.4980
	0.0300	2.1787	1.4233	2.1789	1.4234	2.1788	1.4233

Table 3.4: Table of the generating and calibrated model prices and implied volatilities.

3.3.1.2 Plots of surfaces of model characteristics

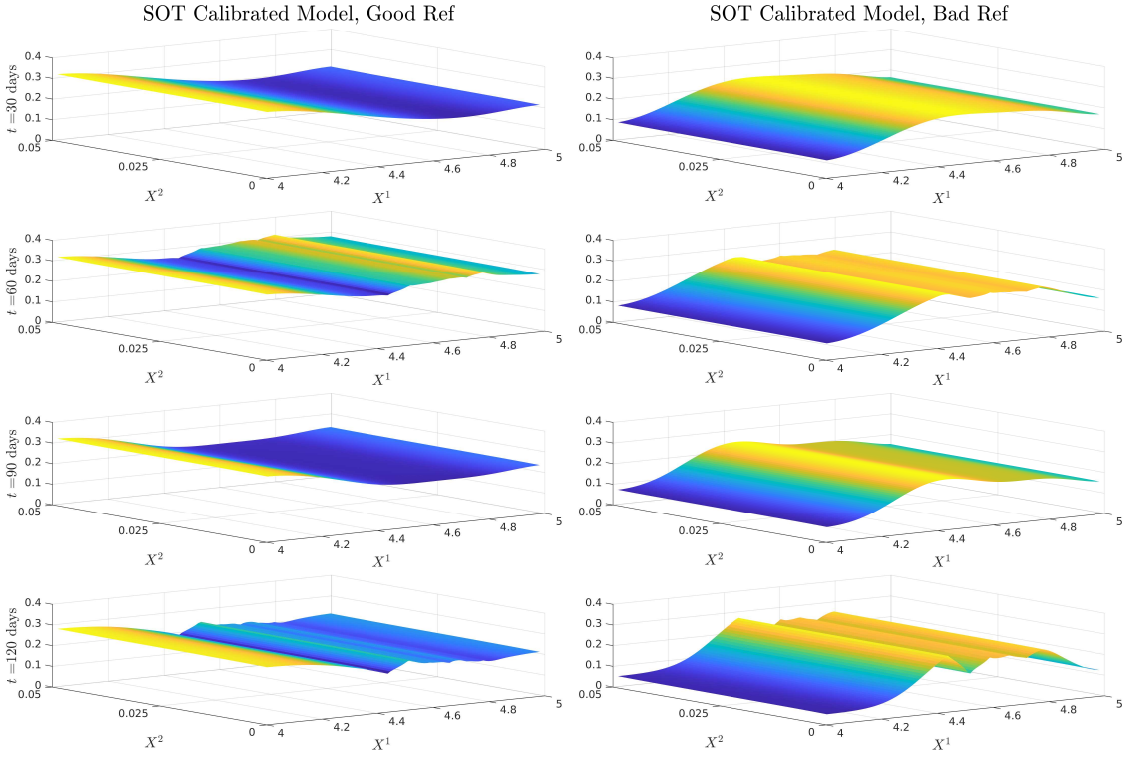
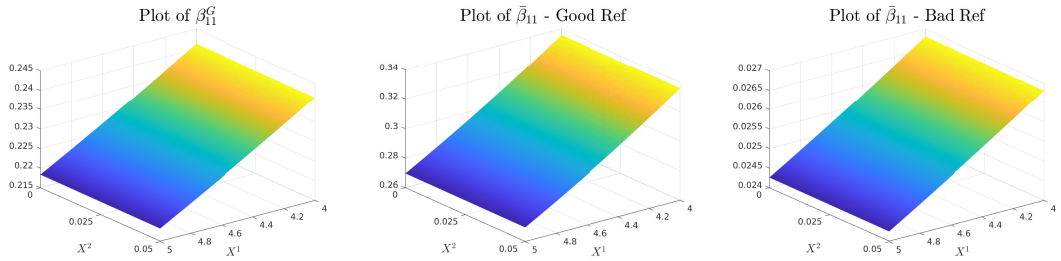


Figure 3.8: Plots of SOT calibrated β_{11} at $t = 30$, 60, 90 and 120 days.



(a) (b) (c)

Figure 3.9: Plots of β_{11} in the generating model (a), $\overline{\beta}_{11}$ in the good reference model (b), and $\overline{\beta}_{11}$ in the bad reference model. Note that these generating and reference values are time homogeneous.

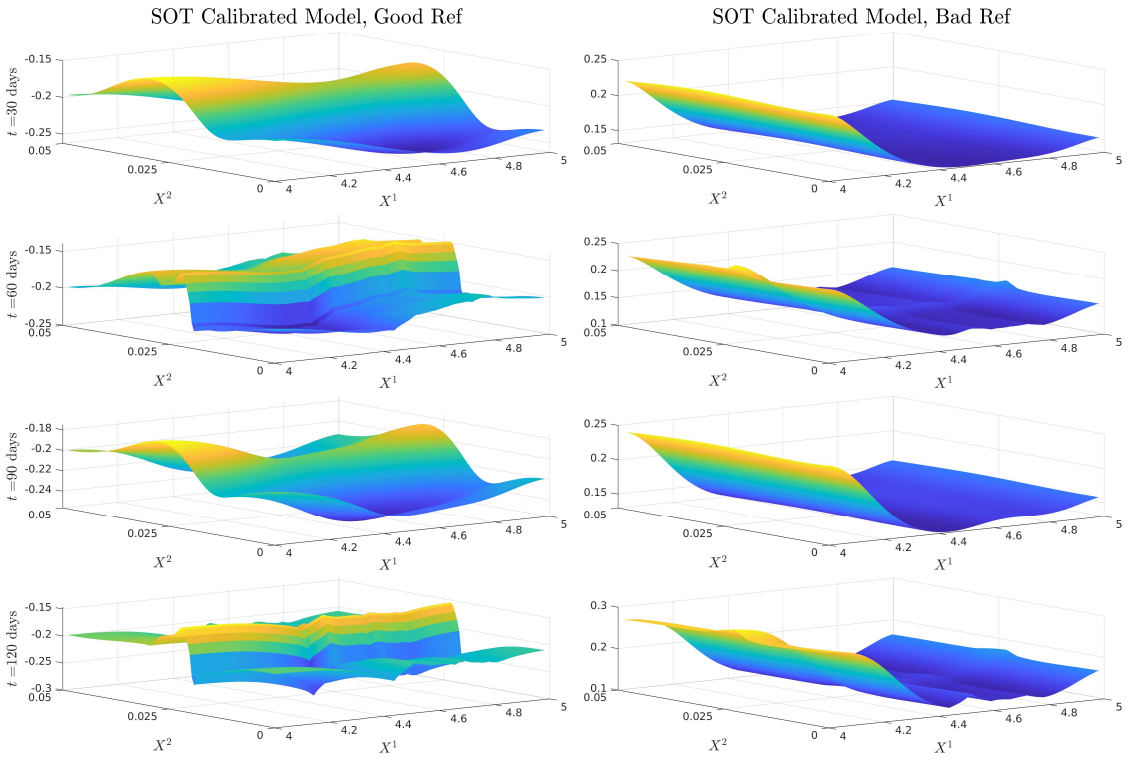


Figure 3.10: Plots of SOT calibrated $\rho = \frac{\beta_{12}}{\sqrt{\beta_{11}\beta_{22}}}$ at $t = 30$, 60, 90 and 120 days. Note the scaling is undone in β_{12} and β_{22}. Here, the generating parameter is given by $\rho = -0.4$, the good reference parameter is given by $\rho_{\text{good}} = -0.2$, the bad reference parameter is given by $\rho_{\text{bad}} = 0.4$.

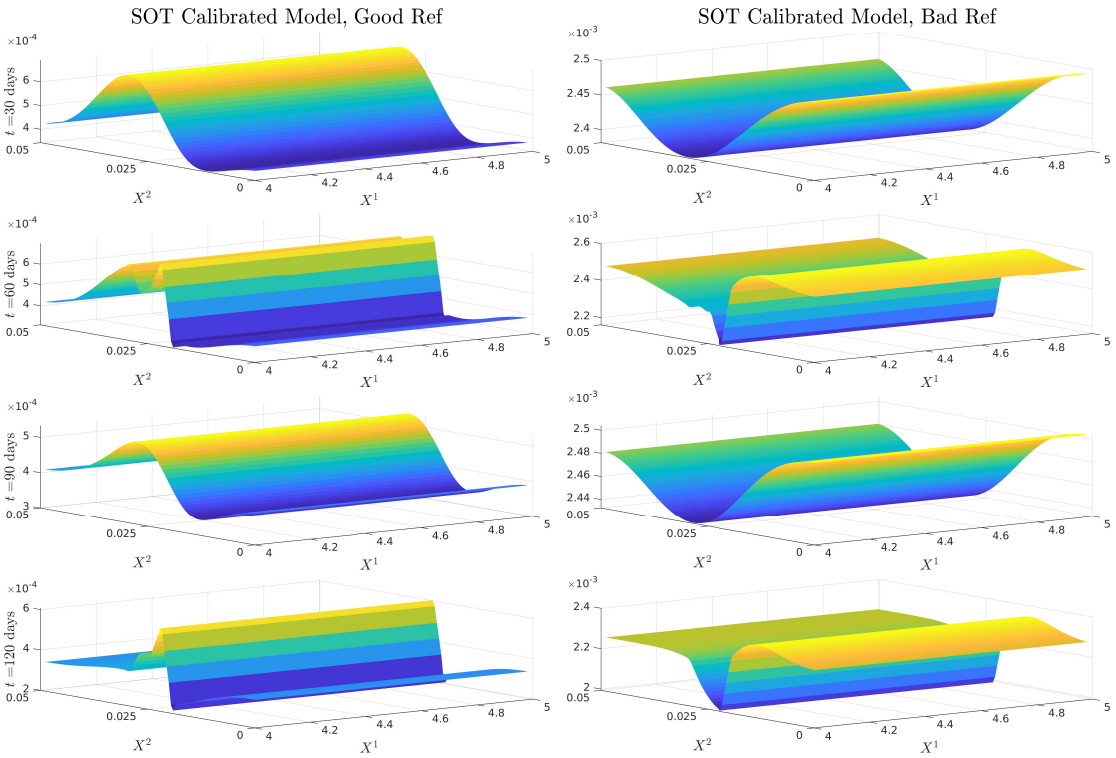


Figure 3.11: Plots of SOT calibrated β_{22} at $t = 30$, 60, 90 and 120 days. Note that the interest rate scaling has been undone in these plots so this is the volatility of the interest rate. The generating parameter is $\beta_{22} = 1.6\times10^{-3}$, the good reference parameter is $\overline{\beta}_{22} = 4\times10^{-4}$, and the bad reference parameter is $\overline{\beta}_{22} = 2.5 \times 10^{-3}$.

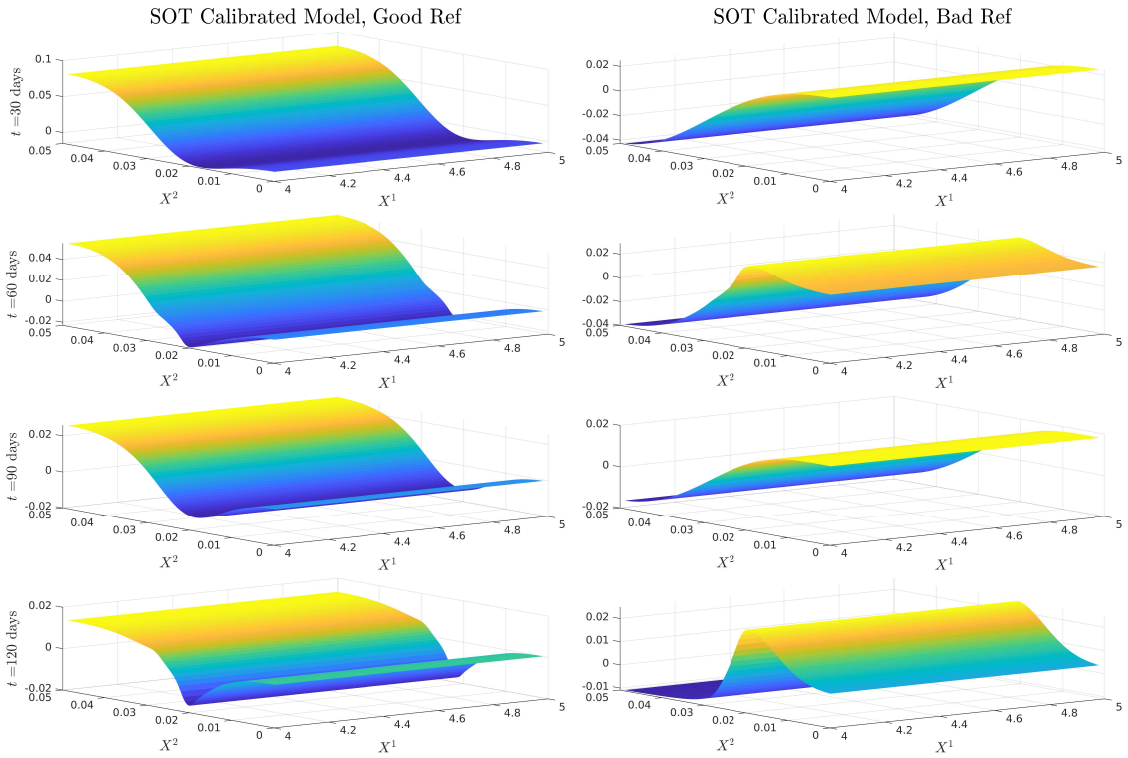


Figure 3.12: Plots of SOT calibrated α_2 at $t = 30$, 60, 90 and 120 days.

3.3.2 Numerical Results: Market Data

We now test our calibration procedure on market data.[2] We used the one month LIBOR as a proxy for the short rate, and obtained the implied volatility and prices of the following options: calls on the SPX with expiry 19/08/2022; caps on the one month LIBOR with notional \$10,000,000 and expiry 23/08/2022; calls on the SPX with expiry 18/11/2022; caps on the one month LIBOR with notional \$10,000,000 and expiry 23/11/2022. We have a total of $n = 32$ options with payoff functions $G_i(X_{\tau_i}) = (\exp(X^1_{\tau_i}) - K_i)^+$ for the calls and $G_i(X_{\tau_i}) = N\tau_i(X^2_{\tau_i} - K_i)^+$ for the caps, where N is the notional value.

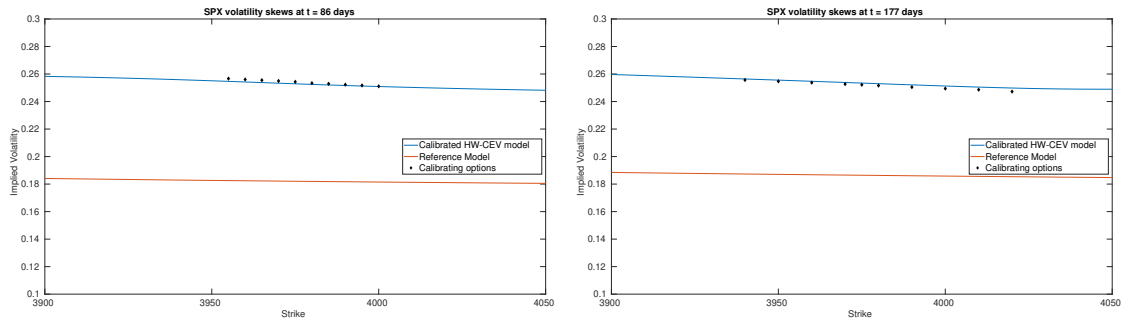
The main difference here is that we no longer have a prior distribution to compare our reference model with. We use the same models as in Section 3.3.1 and list the reference model parameters in a table. The numerical method is largely unchanged from Section 3.3.1, with the main difference being that the tolerance must be higher since the calibration takes far longer than in the simulated data case.

[2]Data obtained from a Bloomberg terminal at the Saïd Business school in Oxford on 23/05/2022.

Hull-White CEV Model		
Parameter	Value	Interpretation
X_0^1	$\log(3973.75)$	Initial log-stock price
X_0^2	0.02543×100	Initial short rate scaled by $R = 100$
ε_1	5×10^{-3}	Tolerance for the difference in scaled model and market implied volatility
ε_2	1×10^{-12}	Tolerance for the policy iteration approximation of the optimal characteristics
$\overline{\sigma}$	0.40	Volatility scaling of reference CEV model
$\overline{\gamma}$	0.90	Power law in reference CEV model
$\overline{b}(t)$	$\overline{a}X_0^2 + \frac{\overline{\sigma}_r^2}{2\overline{a}}(1 - e^{-2\overline{a}t})$	Initial term structure of Hull-White reference model
$\overline{a}$	0.005	Speed of mean reversion of Hull-White reference model
$\overline{\sigma}_r$	0.025	Volatility of Hull-White reference model
$\overline{\rho}$	-0.20	Instantaneous correlation between short rate and log-stock in reference model
δ_1	0.05	Lower bound for β_{11}, the calibrated SPX volatility
δ_2	4	Lower bound for β_{22}, the calibrated volatility of the scaled short rate

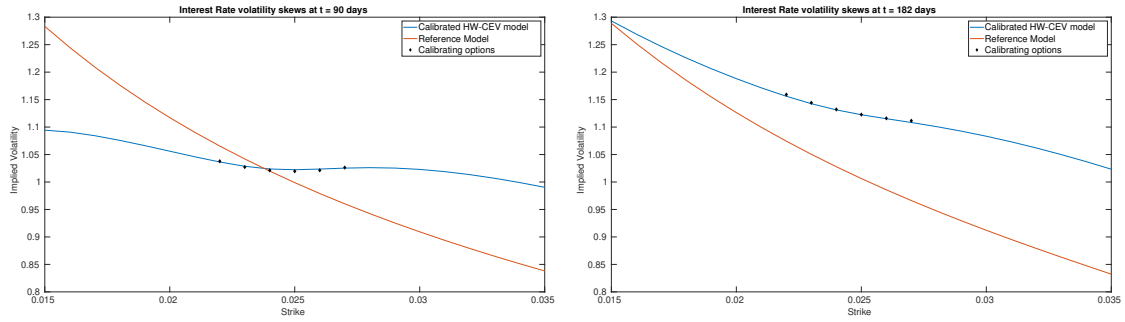
Table 3.5: Parameter values for the market data example.

3.3.2.1 Implied Volatility Plots



(a) (b)

Figure 3.13: Implied volatility skews of the SPX calibrating options under the reference model and calibrated HW-CEV model.



(a) (b)

Figure 3.14: Implied volatility skews of the Short Rate calibrating options under the reference model and calibrated HW-CEV model.

Figures 3.13 and 3.14 demonstrate a good fit of the market implied volatilities from the OT calibrated model. Since in our experiments it took far more gradient steps in the market data case than in the simulated data case, we relaxed the L^∞ exit criteria with a higher tolerance value of 5×10^{-3} in the market case. This can be seen by the small mismatch of the calibration constraints in Figures 3.14 and 3.14. This slower gradient iteration also meant that we took fewer smoothed reference iterations since each epoch took much longer due to the larger number of gradient steps required, as can be seen by the spikier characteristic surfaces in Section 3.3.2.2. We detail how to overcome these computational limitations to obtain a better calibrated model in Section 3.3.4.

3.3.2.2 Plots of surfaces of model characteristics

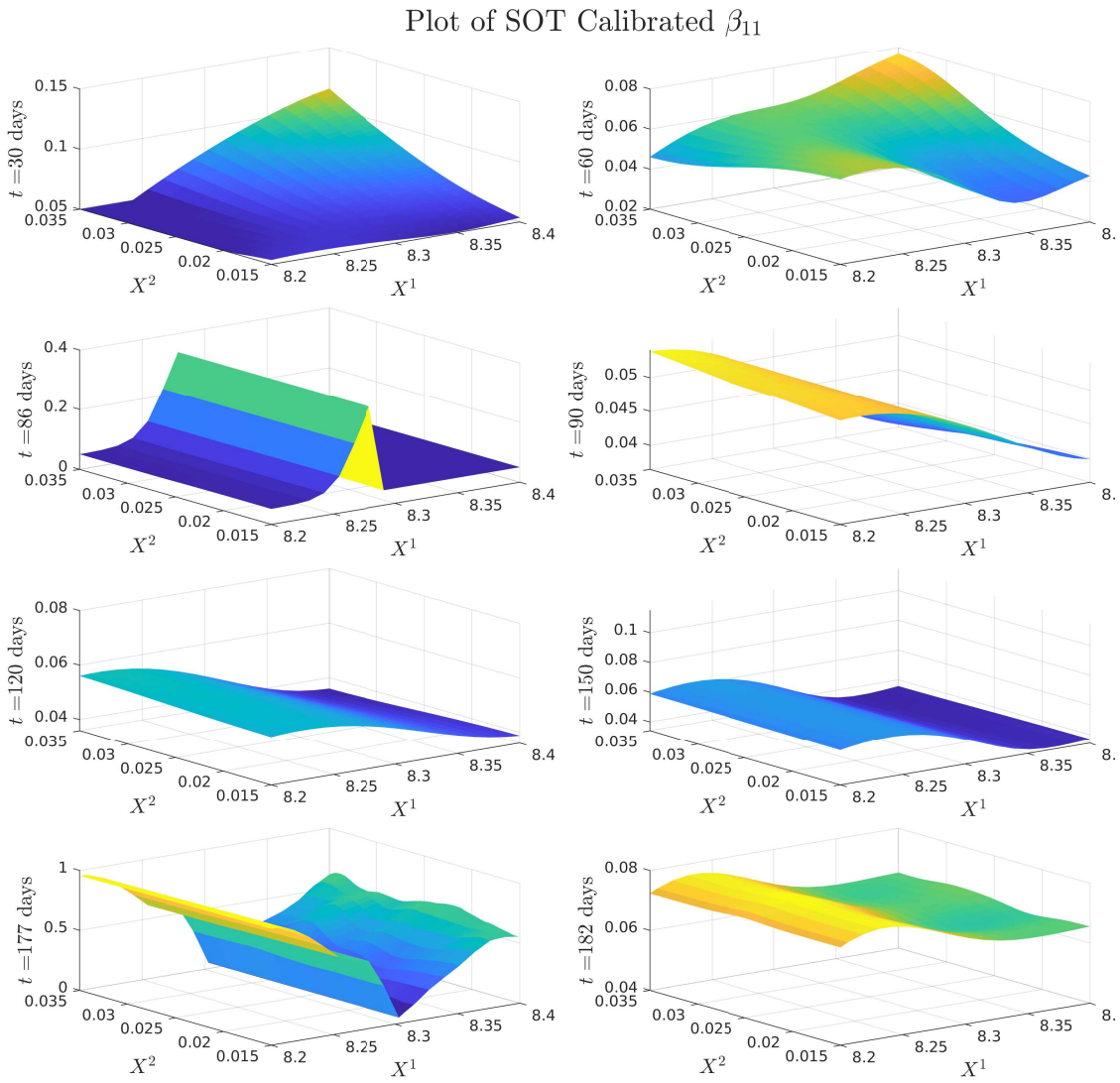


Figure 3.15: Plot of SOT calibrated β_{11} at $t = 30$, 60, 86, 90, 120, 150, 177, and 182 days.

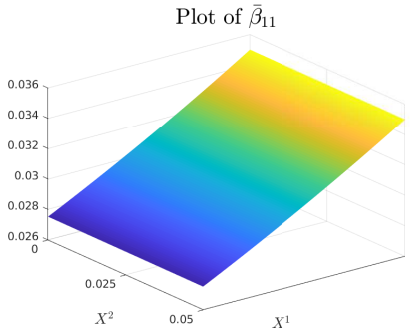


Figure 3.16: Plot of reference parameter $\overline{\beta}_{11}$. Note that this is time homogeneous.

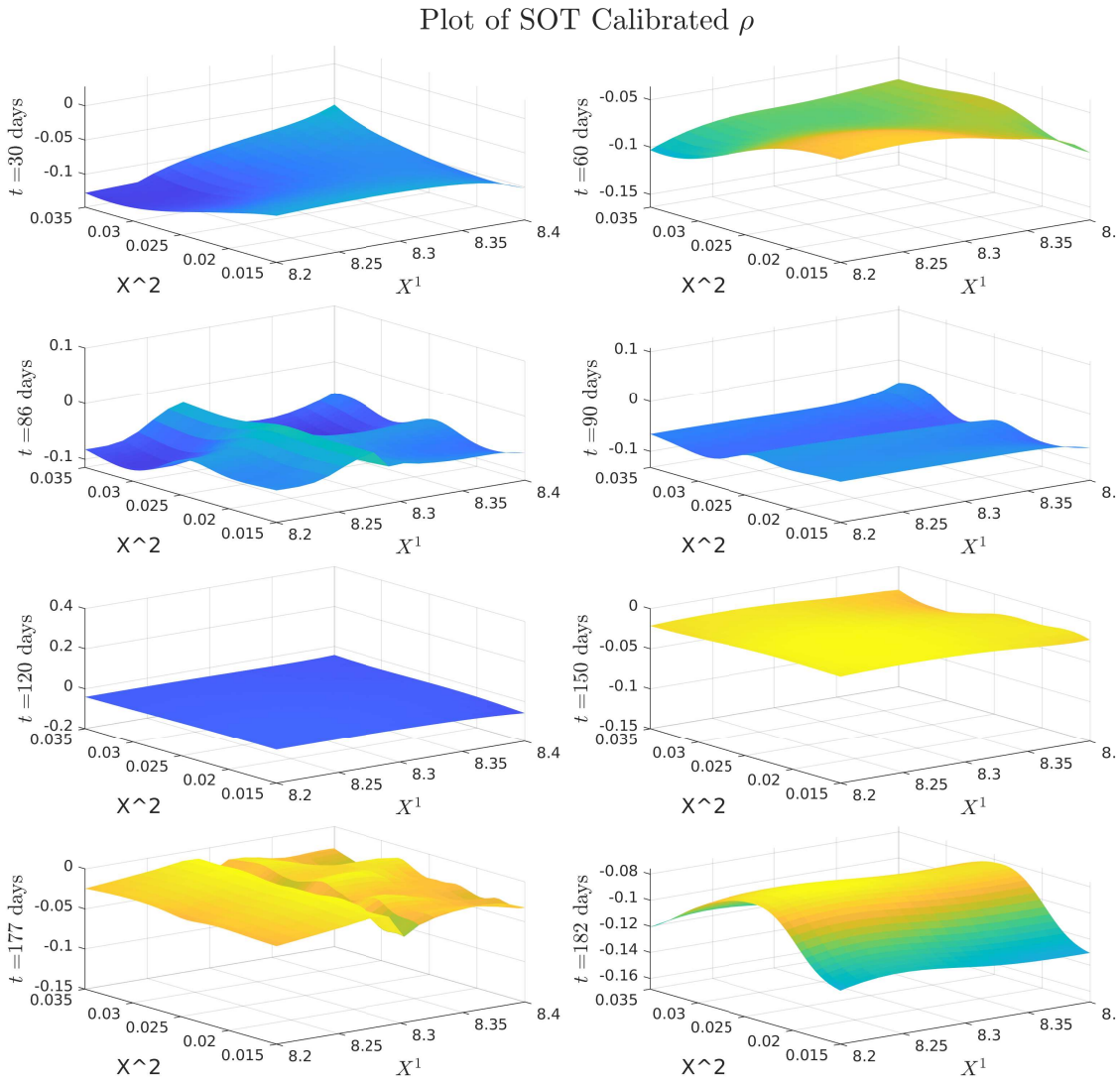


Figure 3.17: Plot of (a) calibrated $\rho = \frac{\beta_{12}}{\sqrt{\beta_{11}\beta_{22}}}$ at $t = 30$, 60, 86, 90, 120, 150, 177, and 182 days. Note that the scaling in β_{12} and β_{22} was undone in this plot. The reference model parameter used was $\overline{\rho} = -0.2$.

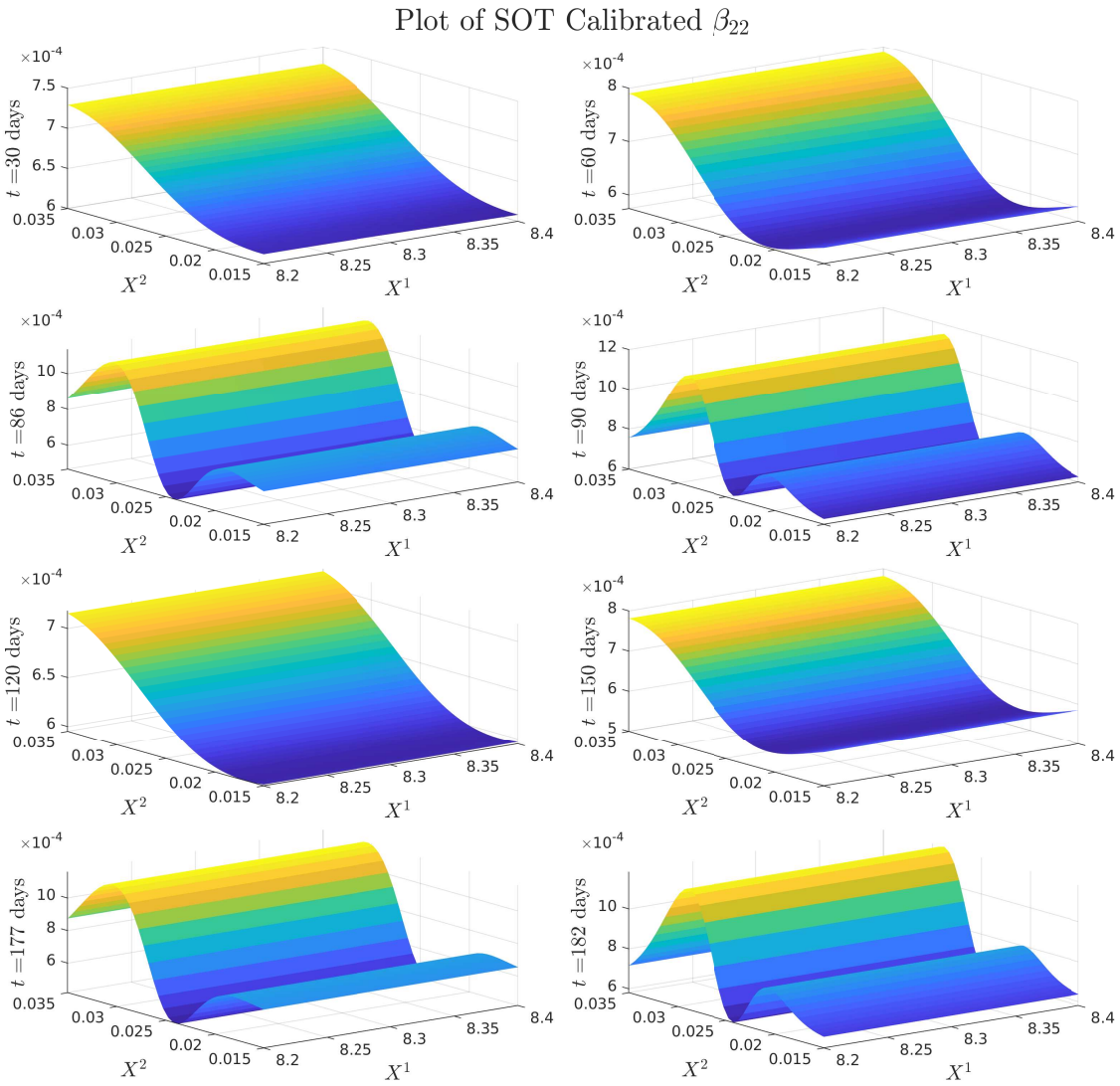


Figure 3.18: Plot of SOT calibrated β_{22} at $t = 30$, 60, 86, 90, 120, 150, 177, and 182 days. Note that the scaling applied to β_{22} has been undone in this plot. The reference parameter used was $\overline{\beta}_{22} = 6.25 \times 10^{-4}$.

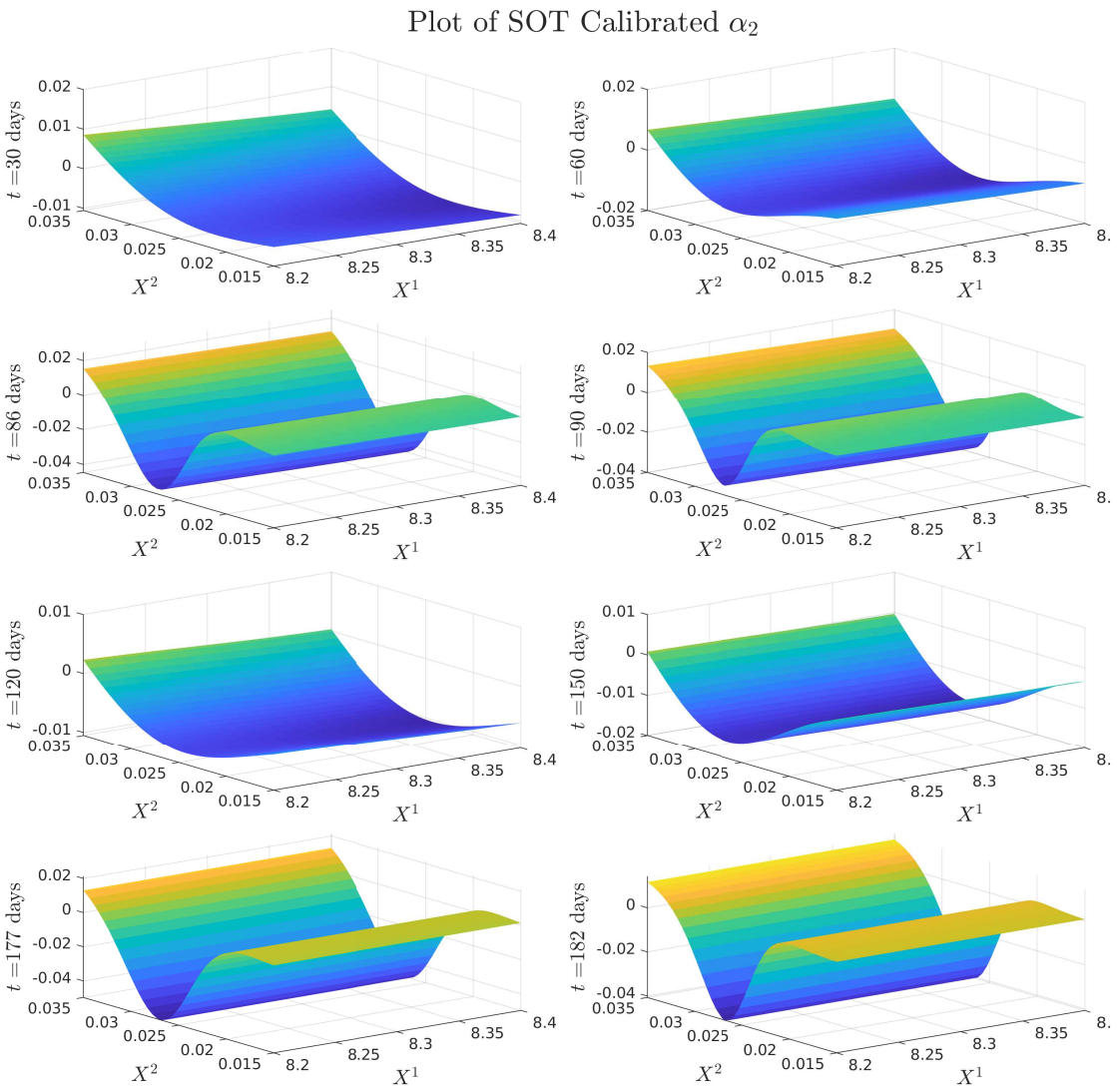


Figure 3.19: Plot of SOT calibrated α_2 at $t = 30$, 60, 86, 90, 120, 150, 177, and 182 days.

3.3.3 A Reweighted Cost Function to Reduce Dependence on the Reference Model Correlation

Thus far in both of our joint and sequential calibration results, we have seen a heavy dependence on the reference correlation $\overline{\rho}$ in our calibrated surface plots. We conjectured this was due to the fact that $\beta \in \mathbb{S}^2_+$, so

$$||\beta - \overline{\beta}||^2_{\text{Fro}} = (\beta^2_{11} - \overline{\beta}_{11})^2 + 2(\beta_{12} - \overline{\beta}_{12})^2 + (\beta_{22} - \overline{\beta}_{22})^2, \tag{3.3.29}$$

and thus there is a larger penalisation of β_{12} away from the reference model than for the other characteristics. We instead took the same set Γ as in (3.3.1) and defined our cost function to be:

$$F(\alpha,\beta) = \begin{cases} ||\alpha - \overline{\alpha}||_2^2 + (\beta_{11} - \overline{\beta}_{11})^2 + \kappa(\beta_{12} - \overline{\beta}_{12})^2 + (\beta_{22} - \overline{\beta}_{22})^2, & \text{if } (\alpha,\beta) \in \Gamma, \\ +\infty & \text{otherwise,} \end{cases} \tag{3.3.30}$$

where $\kappa > 0$. We use the same parametric reference model as (3.2.1), and the generating and "good reference" parameters in Table 3.3. The adaptation of Lemma 3.3.1 to obtain the optimisers for the HJB equation with the cost function (3.3.30) is trivial. The choice of κ should determine how strong the penalisation away from the reference $\overline{\beta}_{12}$ is. We found in practice that the convergence significantly worsens as κ is lowered, with the calibration failing to converge at all for values of κ below one. We obtained convergence for $\kappa = 1$ albeit extremely slowly, taking c.44 hours. For brevity, we only show the main plot of interest (Figure 3.20), which compares ρ between the two approaches. We see that, while there is a difference between the two, both correlations remain centred around $\overline{\rho}$, and thus we have not significantly reduced the strong dependence on the reference model.

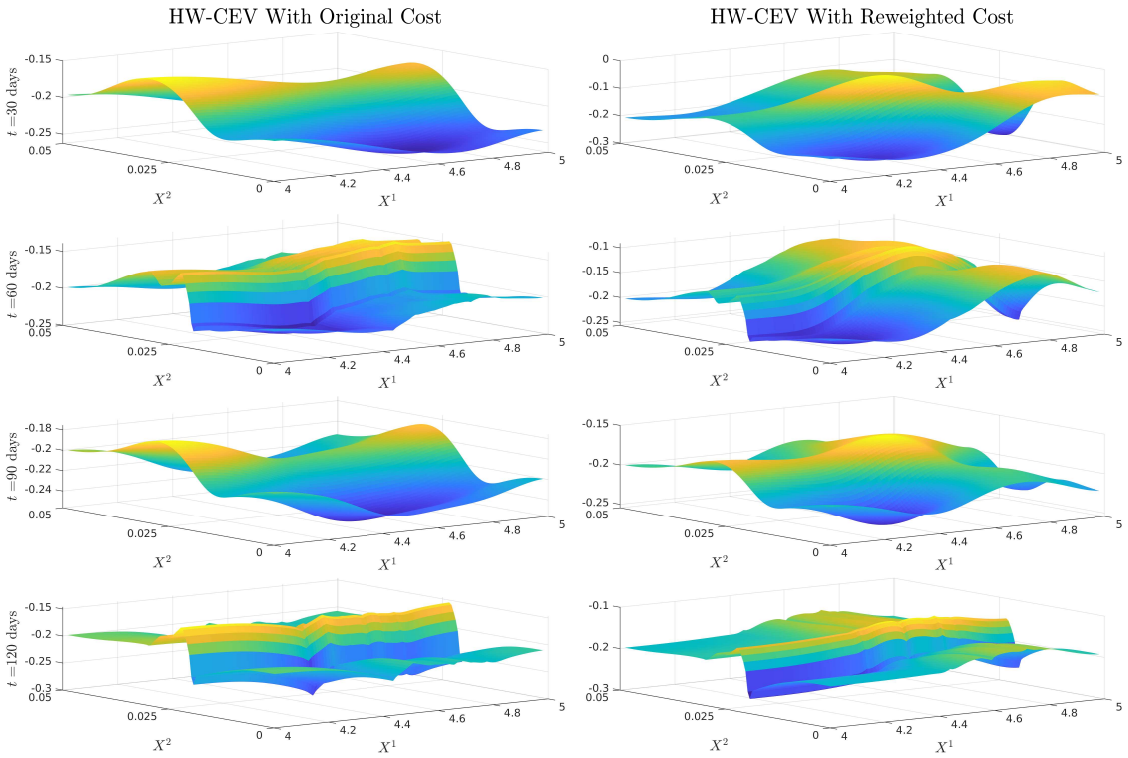


Figure 3.20: Plots of SOT calibrated $\rho = \frac{\beta_{12}}{\sqrt{\beta_{11}\beta_{22}}}$ calibrated to the original and reweighted cost functions. Note that our generating parameter is given by $\rho = -0.4$ and the reference parameter is given by $\overline{\rho} = -0.2$.

3.3.4 A Faster Joint Calibration Method with Bounds on β_{11}^* and β_{22}^*, and a "Parametrically Calibrated" Reference Model

Throughout our analysis, we have aimed to non-parametrically calibrate a model while penalising deviations away from a reference model. Here, we demonstrate a use case for our joint calibration method, which is reducing errors obtained in a parametric method. We choose a parametric model and use a least squares approach to give us a reasonable first "parametric calibration" from which we apply our non-parametric SOT methods to improve the accuracy of calibration. While Lemma 3.3.1 offers an exact solution for the joint calibration problem, computing the optimal surfaces proves slow in practice. In particular, the evaluation of (3.3.25) can be slow and leads to numerical errors in our computation. Moreover, Lemma 3.3.1 gives us no control over the values of β_{11} and β_{22} aside from positivity, and therefore we can obtain small (but still positive) value for the volatility, or extremely large values at some gridpoints. We conjecture that these arise from the fact that the com-

putation in Lemma 3.3.1 defines β_{11} in terms of $\partial^2_{xx}\varphi$ and β_{22} in terms of $\partial^2_{rr}\varphi$, which can lead to instabilities when the options are at the money and close together. This motivates introducing the following bounds on β_{11} and β_{22} to help obtain more realistic surfaces:

$$\delta^l_{11} \leqslant \beta_{11}(t,x) \leqslant \delta^u_{11} \qquad \delta^l_{22} \leqslant \beta_{22}(t,x) \leqslant \delta^u_{22}, \qquad \text{for all } (t,x) \in [0,T] \times \mathbb{R}^2. \tag{3.3.31}$$

This assumption is desirable from our perspective to remove large spikes from the calibrated volatility surface, and has its roots in the uncertain volatility model (see Lyons, 1995; Avellaneda, Levy, and Parás, 1995; Avellaneda, Friedman, et al., 1997; Karoui, Jeanblanc-Picquè, and Shreve, 1998). These bounds can be functions of (t,x), although we only use constant bounds in our numerical examples for simplicity. The values chosen for these bounds in our examples can be found in Table 3.7. The drift and diffusion then take values in the following admissible (convex) set, which depends on the interest state variable:

$$\Gamma_{\text{joint}}(r) \coloneqq \left\{ (\alpha,\beta) \in \mathbb{R}^2 \times \mathbb{S}^2_+ \ : \ \alpha_1 = r - \frac{1}{2}\beta_{11},\ \delta^l_{11} \leqslant \beta_{11} \leqslant \delta^u_{11},\ \delta^l_{22} \leqslant \beta_{22} \leqslant \delta^u_{22} \right\}. \tag{3.3.32}$$

Define $||\cdot||_{\text{Fro}}$ as the Frobenius norm, which is given by $||M||_{\text{Fro}} = \sqrt{\sum_{i,j} M^2_{i,j}}$. We consider the following cost function F to assess locally – with the drift and diffusion coefficients evaluated at $(t,x) = (t,z,r)$ – the deviation of our candidate model from our reference model.

$$F(\alpha,\beta) = \begin{cases} ||\alpha - \overline{\alpha}||^2_2 + ||\beta - \overline{\beta}||^2_{\text{Fro}}, & \text{if } (\alpha,\beta) \in \Gamma_{\text{joint}}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.3.33}$$

With the bounds on β_{11}, β_{22}, we take the Legendre-Fenchel transform of F to be

$$F^*(a,b) \coloneqq \sup_{\substack{\alpha \in \mathbb{R}^2, \beta \in \mathbb{S}^2_+ \\ \delta^l_{11} \leqslant \beta_{11} \leqslant \delta^u_{11} \\ \delta^l_{22} \leqslant \beta_{22} \leqslant \delta^u_{22}}} \{\alpha \cdot a + \beta : b - F(\alpha,\beta)\}. \tag{3.3.34}$$

We consider the same reference model as in (3.2.1), but with a Vasicek short rate model (see Vasicek, 1977) instead for ease with the "parametric calibration"

$$\mathrm{d}r_t = a(b - r_t)\,\mathrm{d}t + \sigma_r\,\mathrm{d}W_t^2, \tag{3.3.35}$$

Under the reference model, the drift coefficient $\overline{\alpha}(t, X_t)$ and the diffusion coefficient $\overline{\beta}(t, X_t) = \frac{d\langle X\rangle_t}{dt}$ of the process X are thus given by:

$$(\overline{\alpha}(t, X_t), \overline{\beta}(t, X_t)) = \left(\begin{bmatrix} r_t - \frac{1}{2}\overline{\sigma}^2 \exp(Z_t)^{2(\overline{\gamma}-1)} \\ \overline{a}\left(\overline{b} - r_t\right) \end{bmatrix}, \begin{bmatrix} \overline{\sigma}^2 \exp(Z_t)^{2(\overline{\gamma}-1)} & \bar{\rho}\bar{\sigma}\bar{\sigma}_r \exp(Z_t)^{\overline{\gamma}-1} \\ \bar{\rho}\bar{\sigma}\bar{\sigma}_r \exp(Z_t)^{\overline{\gamma}-1} & \bar{\sigma}_r^2 \end{bmatrix} \right). \tag{3.3.36}$$

Recall that in practice, we will only consider call options on S and caps on interest rates. Given these constraints, we obtain the reference model (3.3.36) via a least squares parametric calibration. Specifically, let $\tilde{u}(\overline{\sigma}, \overline{\sigma}_r, \overline{\gamma}, \overline{\rho}, \overline{a}, \overline{b}) \in \mathbb{R}^n$ denote the prices of the n options attained by our model (3.3.36) with parameters $(\overline{\sigma}, \overline{\sigma}_r, \overline{\gamma}, \overline{\rho}, \overline{a}, \overline{b})$. The reference model is obtained by numerically solving:

$$\min_{\overline{\sigma}, \overline{\sigma}_r, \overline{\gamma}, \overline{\rho}, \overline{a}, \overline{b}} \frac{1}{n} \sum_{i=1}^{n} \left(\tilde{u}_i(\overline{\sigma}, \overline{\sigma}_r, \overline{\gamma}, \overline{\rho}, \overline{a}, \overline{b}) - u_i \right)^2. \tag{3.3.37}$$

Where the minimisation is taken over $\overline{\sigma}, \overline{\sigma}_r, \overline{a}, \overline{b} > 0$ and $\overline{\rho} \in [-1, 1]$. To minimise this, we simply compute the price of the options given the parameters $(\overline{\sigma}, \overline{\sigma}_r, \overline{\gamma}, \overline{\rho}, \overline{a}, \overline{b})$ using the Feynman-Kac formula, then numerically compute the gradients with respect to the parameters using a central difference approximation, and then apply the L-BFGS algorithm of Liu and Nocedal, 1989. We also rescaled the model and market prices by their Black-Scholes vegas to convert pricing errors into implied volatility errors. We ran the optimisation algorithm to a first order error of 1×10^{-3}, which corresponded to an error of implied volatilities less than 10^{-2} in the SPX call options, and of 10^{-1} in the short rate cap options. This whole procedure took just under 50 minutes on a laptop.

Parameter	Initial Value	Parametrically Calibrated Value	Interpretation
$\overline{\sigma}$	0.4	0.4115	Volatility scaling of the CEV model
$\overline{\gamma}$	0.9	0.9362	Power law in the CEV model
$\overline{\rho}$	-0.2	-0.2037	Instantaneous correlation between short rate and log-stock
$\overline{\sigma}_r$	0.03	0.0232	Volatility of the Vasicek model
$\overline{a}$	0.5	0.0156	Speed of mean reversion in the Vasicek model
$\overline{b}$	0.03	0.2852	Mean to which Vasicek model reverts

Table 3.6: Reference Model Parameters

3.3.4.1 Analytic Formula for (3.3.34)

The bounds in (3.3.31) prove problematic to ad hoc penalise as in Section 3.3.1. Moreover, Lemma 3.3.1 cannot directly be applied and must be modified to provide a global optimiser. For illustrative purposes, we detail that procedure and offer an alternative since it is computationally slow. With our cost function, $F(\alpha,\beta) = ||\alpha-\overline{\alpha}||^2 + ||\beta-\overline{\beta}||^2$, and the constraints that $\delta_{11}^l \leqslant \beta_{11} \leqslant \delta_{11}^u$, $\delta_{22}^l \leqslant \beta_{22} \leqslant \delta_{22}^u$, $\beta_{12}^2 \leqslant \beta_{11}\beta_{22}$, our optimiser β^* is given by

$$\beta^* = \text{argmax}_{\substack{\beta_{12}^2 \leqslant 2\beta_{11}\beta_{22} \\ \delta_{11}^l \leqslant \beta_{11} \leqslant \delta_{11}^u \\ \delta_{22}^l \leqslant \beta_{22} \leqslant \delta_{22}^u}} \left\{ (\beta_{11}-\tilde{\beta}_{11})^2 + (\beta_{12}-\tilde{\beta}_{12})^2 + (\beta_{22}-\tilde{\beta}_{22})^2 \right\}, \tag{3.3.38}$$

where

$$\tilde{\beta}_{11} = \overline{\beta}_{11} + \frac{1}{5}(\partial_{zz}^2\varphi - \partial_z\varphi), \tag{3.3.39}$$

$$\tilde{\beta}_{12} = \frac{1}{\sqrt{2}}(\overline{\beta}_{12} + \frac{1}{4}\partial_{zr}^2\varphi), \tag{3.3.40}$$

$$\tilde{\beta}_{22} = \overline{\beta}_{22} + \frac{1}{4}\partial_{rr}^2\varphi. \tag{3.3.41}$$

With no constraints on β, it is obvious that the optimiser is given by $\beta^* = (\tilde{\beta}_{11}, \tilde{\beta}_{12}, \tilde{\beta}_{22})$. Thus, our constrained problem is seen as minimising the distance to $(\tilde{\beta}_{11}, \tilde{\beta}_{12}, \tilde{\beta}_{22})$. As a demonstration, we now plot the constraints as surfaces in terms of $(\beta_{11}, \beta_{12}, \beta_{22})$. For illustrative purposes, we take larger boundary values for δ_{11}^u and δ_{22}^u, and a larger region than are used in the calibration problem, but the concept remains the same.

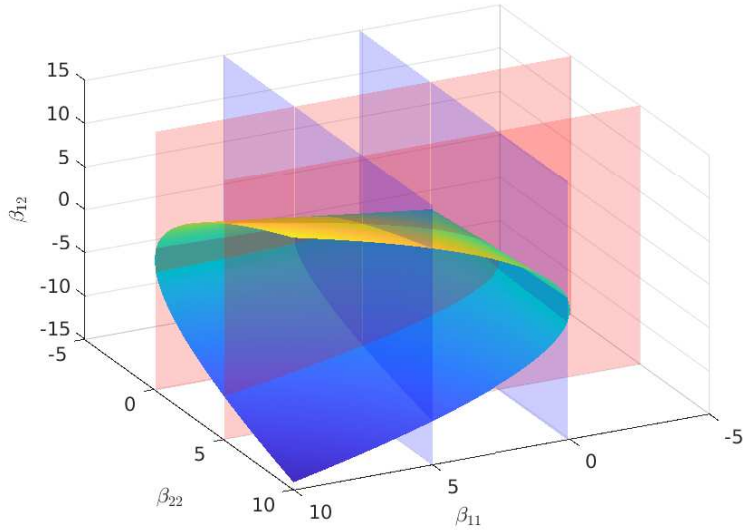


Figure 3.21: Plot of boundaries of constraints

We can see that when $\tilde{\beta}_{12}^2 \leqslant 2\tilde{\beta}_{11}\tilde{\beta}_{22}$, the vertical projections onto the planes give the correct answer. When $\tilde{\beta}_{12}^2 > 2\tilde{\beta}_{11}\tilde{\beta}_{22}$, the closest point on the surface $z^2 = 2xy$ to $(\tilde{\beta}_{11}, \tilde{\beta}_{12}, \tilde{\beta}_{22})$ is given by the orthogonal projection. However, we also must enforce the constraints $\delta_{11}^l \leqslant \beta_{11} \leqslant \delta_{11}^u$ and $\delta_{22}^l \leqslant \beta_{22} \leqslant \delta_{22}^u$, so we need to check when the orthogonal projection "overshoots" the admissible region.

Case 1: $\tilde{\beta}_{11} > \delta_{11}^u$ **and** $\tilde{\beta}_{22} > \delta_{22}^u$

Since the surface $z^2 = 2xy$ is increasing in both x and y, the orthogonal projection will give us a point further away from our admissible region. Therefore, the vertical projections will give us the closest point within the admissible region.

Case 2: $\tilde{\beta}_{11} > \delta_{11}^u$ **and** $\tilde{\beta}_{22} < \delta_{22}^l$

The orthogonal projection will project onto a point further from δ_{11}^u, and again the vertical projections give us the closest point within the admissible region.

Case 3: $\tilde{\beta}_{11} < \delta_{11}^l$ **and** $\tilde{\beta}_{22} > \delta_{22}^u$

The orthogonal projection will project onto a point further from δ_{22}^u, and again the vertical projections give us the closest point within the admissible region.

Case 4: $\tilde{\beta}_{11} < \delta_{11}^l$ **and** $\tilde{\beta}_{22} < \delta_{22}^l$

If no orthogonal projection onto the surface $z^2 = 2xy$ exists, then the closest point is given

by the vertical projections. If an orthogonal projection exists within the admissible region, then the closest point is given by that. However, if the orthogonal projection exists outside the region, then the closest point is given by the vertical projection onto the admissible region from the orthogonal projection onto the surface $z^2 = 2xy$.

Case 5: $\tilde{\beta}_{11} < \delta^l_{11}$ **and** $\delta^l_{22} \leqslant \tilde{\beta}_{22} \leqslant \delta^u_{22}$

$\beta^*_{11} = \delta^l_{11}$ since going any further along the β_{11} axis will increase the distance from $(\tilde{\beta}_{11}, \tilde{\beta}_{12}, \tilde{\beta}_{22})$. Therefore, the optimiser lies along the curve $z^2 = 2\delta^l_{11} y$.

Case 6: $\tilde{\beta}_{11} > \delta^u_{11}$ **and** $\delta^l_{22} \leqslant \tilde{\beta}_{22} \leqslant \delta^u_{22}$

$\beta^*_{11} = \delta^u_{11}$ since going any further along the β_{11} axis will increase the distance from $(\tilde{\beta}_{11}, \tilde{\beta}_{12}, \tilde{\beta}_{22})$. Therefore, the optimiser lies along the curve $z^2 = 2\delta^u_{11} y$.

Case 7: $\tilde{\beta}_{22} < \delta^l_{22}$ **and** $\delta^l_{11} \leqslant \tilde{\beta}_{11} \leqslant \delta^u_{11}$

$\beta^*_{22} = \delta^l_{22}$ since going any further along the β_{22} axis will increase the distance from $(\tilde{\beta}_{11}, \tilde{\beta}_{12}, \tilde{\beta}_{22})$. Therefore, the optimiser lies along the curve $z^2 = 2\delta^l_{22} x$.

Case 8: $\tilde{\beta}_{22} > \delta^u_{22}$ **and** $\delta^l_{11} \leqslant \tilde{\beta}_{11} \leqslant \delta^u_{11}$

$\beta^*_{22} = \delta^u_{22}$ since going any further along the β_{22} axis will increase the distance from $(\tilde{\beta}_{11}, \tilde{\beta}_{12}, \tilde{\beta}_{22})$. Therefore, the optimiser lies along the curve $z^2 = 2\delta^u_{22} x$.

Case 9: $\delta^l_{11} \leqslant \tilde{\beta}_{11} \leqslant \delta^u_{11}$ **and** $\delta^l_{22} \leqslant \tilde{\beta}_{22} \leqslant \delta^u_{22}$

This is the same as in Case 4, except that an orthogonal projection will always exist. If the orthogonal projection onto $z^2 = 2xy$ lies in the admissible region, then the optimisers are given by that. Otherwise, the closest point is given on the boundary nearest to the orthogonal projection.

Discussion on Case 1-9

This entire procedure is computationally slow. The computation of Case 5-8 relies on solving a cubic equation, and finding the minimiser among all of its roots. Case 4 and Case 9 require computing an orthogonal projection, finding which region it lies in and then applying the procedure for Case 5-8 to the nearest boundary if it overshoots. For different values of λ tried in the L-BGFS optimisation algorithm, these cases can become more common which therefore slows down our calibration algorithm. Moreover, there is no expectation of continuity on the surface with these different cases, and numerical errors in computing the orthogonal projections and roots of the cubic can also give sharp and unrealistic spikes.

If these are a rare occurrence in our calibrated surfaces however, we can instead apply vertical projections as a penalisation away from these more pathological surfaces since the vertical projection gives a "wrong" surface and introduces calibration error. In all three calibration routines where only the vertical projections are used, the resulting calibrated surface had precisely zero cases on the boundary of the constraint and therefore the approximation and the analytic answer coincide. We remark that there were cases of $\beta_{11}^* = \delta_{11}^{l,u}$ or $\beta_{22}^* = \delta_{22}^{l,u}$, however since $\beta_{12}^* \neq \pm\sqrt{\beta_{11}^*\beta_{22}^*}$, the vertical projections onto the admissible region is in fact the analytic solution here.

Procedure for computing the minimiser in Case 5-8

Assume $\tilde{\beta}_{11} < \delta_{11}^l$. Then $\beta_{11}^* = \delta_{11}^l$ since going any further along the β_{11} axis will increase the distance from $(\tilde{\beta}_{11}, \tilde{\beta}_{12}, \tilde{\beta}_{22})$. Therefore, the optimiser lies along the curve $z^2 = 2\delta_{11}^l y$. Therefore, we minimise:

$$f(y) = (y - \tilde{\beta}_{22})^2 + \left(\sqrt{2\delta_{11}^l y} - \tilde{\beta}_{12}\right)^2, \tag{3.3.42}$$

subject to $y \in [\delta_{22}^l, \delta_{22}^u]$. We have

$$f'(y) = 2(y - \tilde{\beta}_{22}) + \frac{\sqrt{2\delta_{11}^l}\left(\sqrt{2\delta_{11}^l y - \tilde{\beta}_{12}}\right)}{\sqrt{y}}. \tag{3.3.43}$$

Since $f''(y) > 0$, this gives that the minimum is attained as the positive solution to the cubic

$$2w^3 + 2(\delta_{11}^l - \tilde{\beta}_{12})w - \tilde{\beta}_{12} = 0. \tag{3.3.44}$$

If no positive solution exists for $w \in [\sqrt{\delta_{22}^l}, \sqrt{\delta_{22}^u}]$, then w is taken to be the boundary point with the lowest value for f. If multiple admissible roots exist, then take the root minimising $f(w^2)$. Since the problem is symmetric, the same method can be obtained in the other cases with precisely one of the constraints on $\tilde{\beta}_{11}$ or $\tilde{\beta}_{22}$ holding.

3.3.4.2 An Alternative Approach

In the step 9 of Algorithm 3 we have to Approximate (α^*, β^*) from $\varphi_{t_k}^{\text{old}}$. This is a key step and it needs to be done efficiently, and in our numerical experiments the analytic method of Lemma 3.3.1 proved far slower than the alternative which we now describe.

Corrias, 1996 provides a numerical scheme to evaluate the Legendre-Fenchel transform in $O(N^d \log N)$ time for N discretisation points and d dimensions. We remark as well that Lucet, 1997 provides a faster version in $O(N^d)$ time, however it is difficult to separate the β terms while maintaining positive semi-definiteness. Since (3.3.34) requires the evaluation of the spatial derivatives at all points of the grid, we cannot directly apply the methods to fully take advantage of the faster computational speed. However, we use the observation in equation (2.1) of Corrias, 1996, which allows us to decompose the maximisation component-wise, and apply the idea of evaluating the transform on bounded intervals for computational ease.

Lemma 3.3.2 (Approximating the Optimal Coefficients)**.** *Define* (α^*, β^*) *by:*

$$\alpha_1^*(t,z,r) = r - \frac{1}{2}\beta_{11}^*(t,z,r), \tag{3.3.45}$$

$$\alpha_2^*(t,z,r) = \overline{\alpha}_2(t,z,r) + \frac{1}{2}\partial_r\varphi(t,z,r), \tag{3.3.46}$$

$$\beta_{11}^*(t,z,r) = \begin{cases} \overline{\beta}_{11}(t,z,r) + \frac{1}{5}(\partial_{zz}^2\varphi(t,z,r) - \partial_z\varphi(t,z,r)), & \text{when } \overline{\beta}_{11} + \frac{1}{5}(\partial_{zz}^2\varphi - \partial_z\varphi) \in [\delta_{11}^l, \delta_{11}^u], \\ \delta_{11}^l, & \text{when } \overline{\beta}_{11} + \frac{1}{5}(\partial_{zz}^2\varphi - \partial_z\varphi) < \delta_{11}^l, \\ \delta_{11}^u, & \text{when } \overline{\beta}_{11} + \frac{1}{5}(\partial_{zz}^2\varphi - \partial_z\varphi) > \delta_{11}^u, \end{cases} \tag{3.3.47}$$

$$\beta_{22}^*(t,z,r) = \begin{cases} \overline{\beta}_{22}(t,z,r) + \frac{1}{4}\partial_{rr}^2\varphi(t,z,r), & \text{when } \overline{\beta}_{22} + \frac{1}{4}\partial_{rr}^2\varphi \in [\delta_{22}^l, \delta_{22}^u], \\ \delta_{22}^l, & \text{when } \overline{\beta}_{22} + \frac{1}{4}\partial_{rr}^2\varphi < \delta_{22}^l, \\ \delta_{22}^u, & \text{when } \overline{\beta}_{22} + \frac{1}{4}\partial_{rr}^2\varphi > \delta_{22}^u, \end{cases} \tag{3.3.48}$$

$$\beta_{12}^*(t,z,r) = \begin{cases} \overline{\beta}_{12}(t,z,r) + \frac{1}{4}\partial_{zr}^2\varphi(t,z,r), & \text{when } \overline{\beta}_{12} + \frac{1}{4}\partial_{zr}^2\varphi \in \left[-\sqrt{\beta_{11}^*\beta_{22}^*}, \sqrt{\beta_{11}^*\beta_{22}^*}\right], \\ -\sqrt{\beta_{11}^*\beta_{22}^*}, & \text{when } \overline{\beta}_{12} + \frac{1}{4}\partial_{zr}^2\varphi < -\sqrt{\beta_{11}^*\beta_{22}^*}, \\ \sqrt{\beta_{11}^*\beta_{22}^*}, & \text{when } \overline{\beta}_{12} + \frac{1}{4}\partial_{zr}^2\varphi > \sqrt{\beta_{11}^*\beta_{22}^*}, \end{cases} \tag{3.3.49}$$

and $\beta^*_{21} = \beta^*_{12}$. *The matrix* β^* *is positive semidefinite and whenever*

$$-\sqrt{\beta^*_{11}\beta^*_{22}} < \beta^*_{12} < \sqrt{\beta^*_{11}\beta^*_{22}} \tag{3.3.50}$$

holds, then (α^*, β^*) *are the optimisers in* (3.3.34).

The proof proceeds by solving (3.3.34) sequentially: we first solve for β_{11} and β_{22} subject to their bounds. This is easy to do as the expression is quadratic in each variable. We then solve for β_{12} and use it to enforce the condition of positive semi-definiteness on the matrix β. If (3.3.50) holds then the positive semi-definitiveness condition is not binding and the procedure returns the optimizer. Otherwise, we set $\beta^*_{12} = \pm\sqrt{\beta^*_{11}\beta^*_{22}}$ to ensure β^* is positive semi-definite but this means our approximation may differ from a global search over all positive semi-definite matrices β. However, in all of our numerical experiments, (3.3.50) in fact holds everywhere so that, in practice, Lemma 3.3.2 provides the optimisers in (3.3.34) but at a fraction of the computational cost involved in solving (3.3.34) analytically as a constrained optimization.

Proof of Lemma 3.3.2. Since Γ_{joint} defined in (3.3.32) enforces that $\alpha^*_1 = r - \frac{1}{2}\beta^*_{11}$, we immediately have the first equality given an optimiser β^*_{11}. We remark that any reference model $(\overline{\alpha}, \overline{\beta})$ will also follow the same constraint. Additionally, $\beta \in \mathbb{S}^2_+$ is enforced by the set Γ_{joint}. Therefore, since our cost function is given by (3.3.33), its Legendre-Fenchel transform $F^* : \mathbb{R}^2 \times \mathbb{S}^2 \to \mathbb{R}$ is given by:

$$\begin{aligned} F^*(a,b) = \sup_{\substack{\alpha_2\in\mathbb{R},\\ \beta\in\mathbb{S}^2_+,\\ \beta_{11}\in[\delta^l_{11},\delta^u_{11}],\\ \beta_{22}\in[\delta^l_{22},\delta^u_{22}]}} \Bigg\{ & a_1\left(r - \frac{1}{2}\beta_{11}\right) + a_2\alpha_2 + b_{11}\beta_{11} + 2b_{12}\beta_{12} + b_{22}\beta_{22} - \frac{1}{4}(\beta_{11} - \overline{\beta}_{11})^2 \\ & - (\alpha_2 - \overline{\alpha}_2)^2 - (\beta_{11} - \overline{\beta}_{11})^2 - 2(\beta_{12} - \overline{\beta}_{12})^2 - (\beta_{22} - \overline{\beta}_{22})^2 \Bigg\} \end{aligned} \tag{3.3.51}$$

The gradient $\nabla F^*(a,b)$ is given by the maximisers of (3.3.51), and by rearranging the α_2

term, we obtain the second equality:

$$\alpha_2^* = \arg\max_{\alpha_2 \in \mathbb{R}} \left\{ a_2\alpha_2 - \alpha_2^2 + 2\alpha_2\overline{\alpha}_2 - \overline{\alpha}_2^2 \right\} = \arg\min_{\alpha_2 \in \mathbb{R}} \left\{ \left(\alpha_2 - \left(\overline{\alpha}_2 + \frac{1}{2} a_2 \right) \right)^2 \right\} = \overline{\alpha}_2 + \frac{1}{2} a_2. \tag{3.3.52}$$

Now, similarly rearranging (3.3.51), we have

$$[\beta_{12}^*, \beta_{11}^*, \beta_{22}^*] = \left[\arg\min_{\beta_{12}} \left\{ \left(\beta_{12} - \left(\overline{\beta}_{12} + \frac{1}{2} b_{12} \right) \right)^2 \right\}, \arg\min_{\substack{\beta_{11} \\ \beta_{11} \in [\delta_{11}^l, \delta_{11}^u]}} \left\{ \left(\beta_{11} - \left(\overline{\beta}_{11} + \frac{1}{5} \left(2b_{11} - a_1 \right) \right) \right)^2 \right\}, \right.$$
$$\left. \arg\min_{\substack{\beta_{22} \\ \beta_{22} \in [\delta_{22}^l, \delta_{22}^u]}} \left\{ \left(\beta_{22} - \left(\overline{\beta}_{22} + \frac{1}{2} b_{22} \right) \right)^2 \right\} \right], \text{ subject to } \beta^* \in \operatorname{int}(\mathbb{S}_+^2). \tag{3.3.53}$$

We solve the minimisations of β_{11} and β_{22} inside the β_{12} minimisation problem by taking:

$$\beta_{11}^* = \begin{cases} \overline{\beta}_{11} + \frac{1}{5}(2b_{11} - a_1), & \text{if } \overline{\beta}_{11} + \frac{1}{5}(2b_{11} - a_1) \in [\delta_{11}^l, \delta_{11}^u] \\ \delta_{11}^u, & \text{if } \overline{\beta}_{11} + \frac{1}{5}(2b_{11} - a_1) > \delta_{11}^u \\ \delta_{11}^l, & \text{if } \overline{\beta}_{11} + \frac{1}{5}(2b_{11} - a_1) < \delta_{11}^l \end{cases} \tag{3.3.54}$$

$$\beta_{22}^* = \begin{cases} \overline{\beta}_{22} + \frac{1}{2} b_{22}, & \text{if } \overline{\beta}_{22} + \frac{1}{2} b_{22} \in [\delta_{22}^l, \delta_{22}^u], \\ \delta_{22}^u, & \text{if } \overline{\beta}_{22} + \frac{1}{2} b_{22} > \delta_{22}^u, \\ \delta_{22}^l, & \text{if } \overline{\beta}_{22} + \frac{1}{2} b_{22} < \delta_{22}^l. \end{cases} \tag{3.3.55}$$

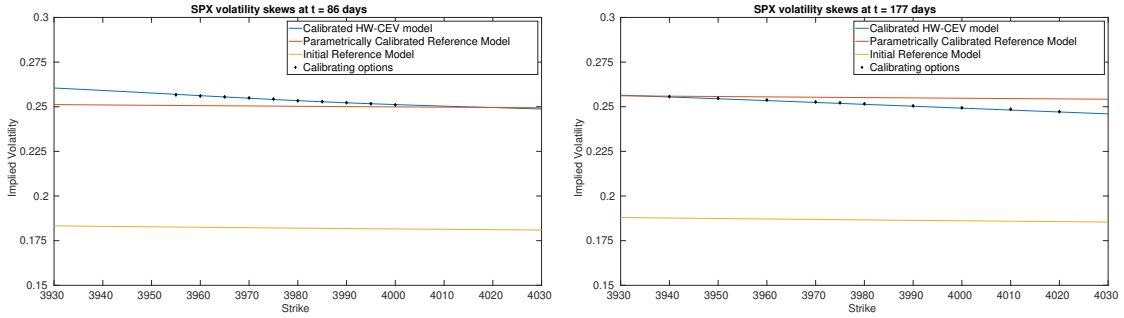
Given β_{11}^* and β_{22}^* as above, we now may rewrite the outside constraint as $\beta_{12} \in [-\sqrt{\beta_{11}^*\beta_{22}^*}, \sqrt{\beta_{11}^*, \beta_{22}^*}]$ and thus obtain:

$$\beta_{12}^* = \begin{cases} \overline{\beta}_{12} + \frac{1}{2} b_{12}, & \text{if } \overline{\beta}_{12} + \frac{1}{2} b_{12} \in [-\sqrt{\beta_{11}^*\beta_{22}^*}, \sqrt{\beta_{11}^*\beta_{22}^*}], \\ \sqrt{\beta_{11}^*\beta_{22}^*}, & \text{if } \overline{\beta}_{12} + \frac{1}{2} b_{12} > \sqrt{\beta_{11}^*\beta_{22}^*}, \\ -\sqrt{\beta_{11}^*\beta_{22}^*}, & \text{if } \overline{\beta}_{12} + \frac{1}{2} b_{12} < -\sqrt{\beta_{11}^*\beta_{22}^*}. \end{cases} \tag{3.3.56}$$

In particular, in the first case, the condition $\beta \in \mathbb{S}_+^2$ is automatically satisfied and the procedure returns the optimizer. By taking $a = \nabla_x \varphi$ and $b = \frac{1}{2}\nabla_x^2 \varphi$, we conclude the proof. □

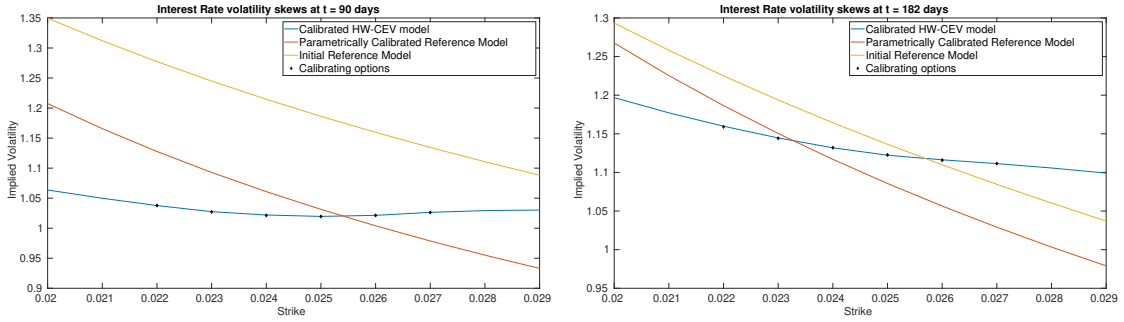
3.3.4.3 Market Data Example with Parametrically Calibrated Reference Model

We test our faster joint calibration procedure the same market data as in Section 3.3.



(a) (b)

Figure 3.22: Implied volatility skews of the SPX calibrating options under the reference model and calibrated HW-CEV model.



(a) (b)

Figure 3.23: Implied volatility skews of the Short Rate calibrating options under the reference model and calibrated HW-CEV model.

We display the optimisation parameters, and the bounds for β_{11} and β_{22} in Table 3.7 below. It is difficult to choose the bounds since too relaxed a bound will result in large values for β_{11} and β_{22} and also result in numerical instabilities for β_{12}; whereas too tight a bound has been observed to severely slow down the calibration. Even with these bounds chosen, we still saw some instabilities in β_{11} on the maturity dates of the calls, and in β_{22} on the maturity dates of the Caps. Imposing the bounds reduced these instabilities significantly, but did not remove them entirely. We see in Figures 3.22 and 3.23 that we obtain an exact fit of the market implied volatilities, with less observable calibration error than in the original joint calibration method (see Figures 3.13 and 3.14). Additionally, the characteristic

surfaces obtained (see Section 3.3.4.4) avoid the large values seen in the original method (see Section 3.3.2.2). As has been the case in prior experiments, we see that the calibrated ρ still has a large dependence on the reference model, so for practical purposes a good prior reference guess is required.

Parameter	Value	Interpretation
ε_1	1×10^{-3}	Tolerance for the difference in model and observed IV
ε_2	1×10^{-8}	Tolerance in the policy iteration for φ
δ_{11}^l	0.01	Lower bound for β_{11}
δ_{11}^u	0.5	Upper bound for β_{11}
δ_{22}^l	2×10^{-4}	Lower bound for β_{22}
δ_{22}^u	1×10^{-3}	Upper bound for β_{22}

Table 3.7: Optimisation Parameters and Bounds

3.3.4.4 Plots of Drift and Diffusion Surfaces

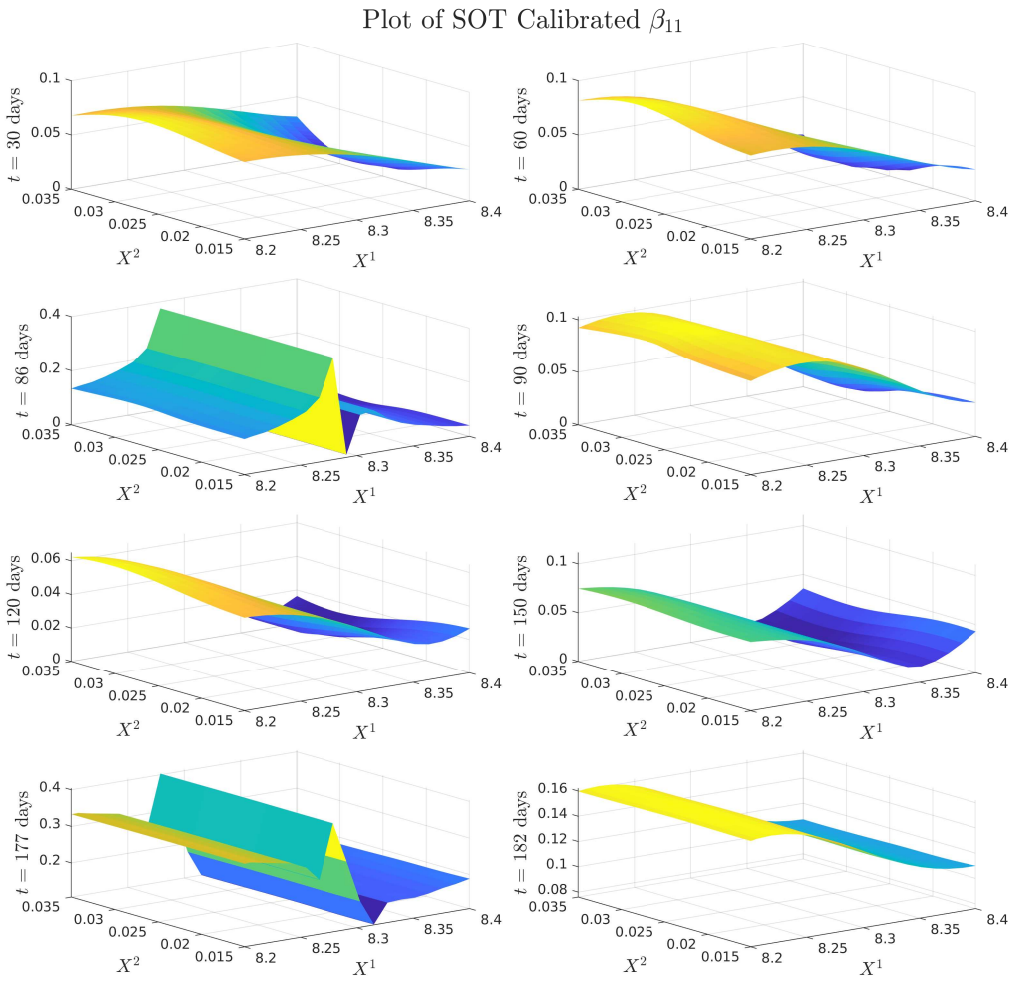


Figure 3.24: Plots of SOT-calibrated β_{11}, the volatility of the log-stock.

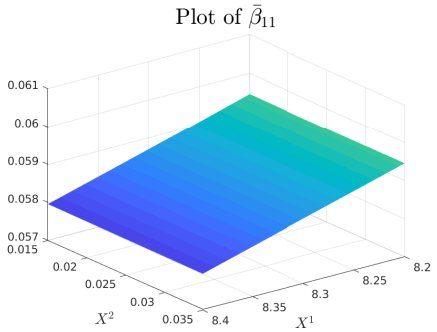


Figure 3.25: Plot of parametrically calibrated reference $\overline{\beta}_{11}$. Note that this is taken to be time homogeneous.

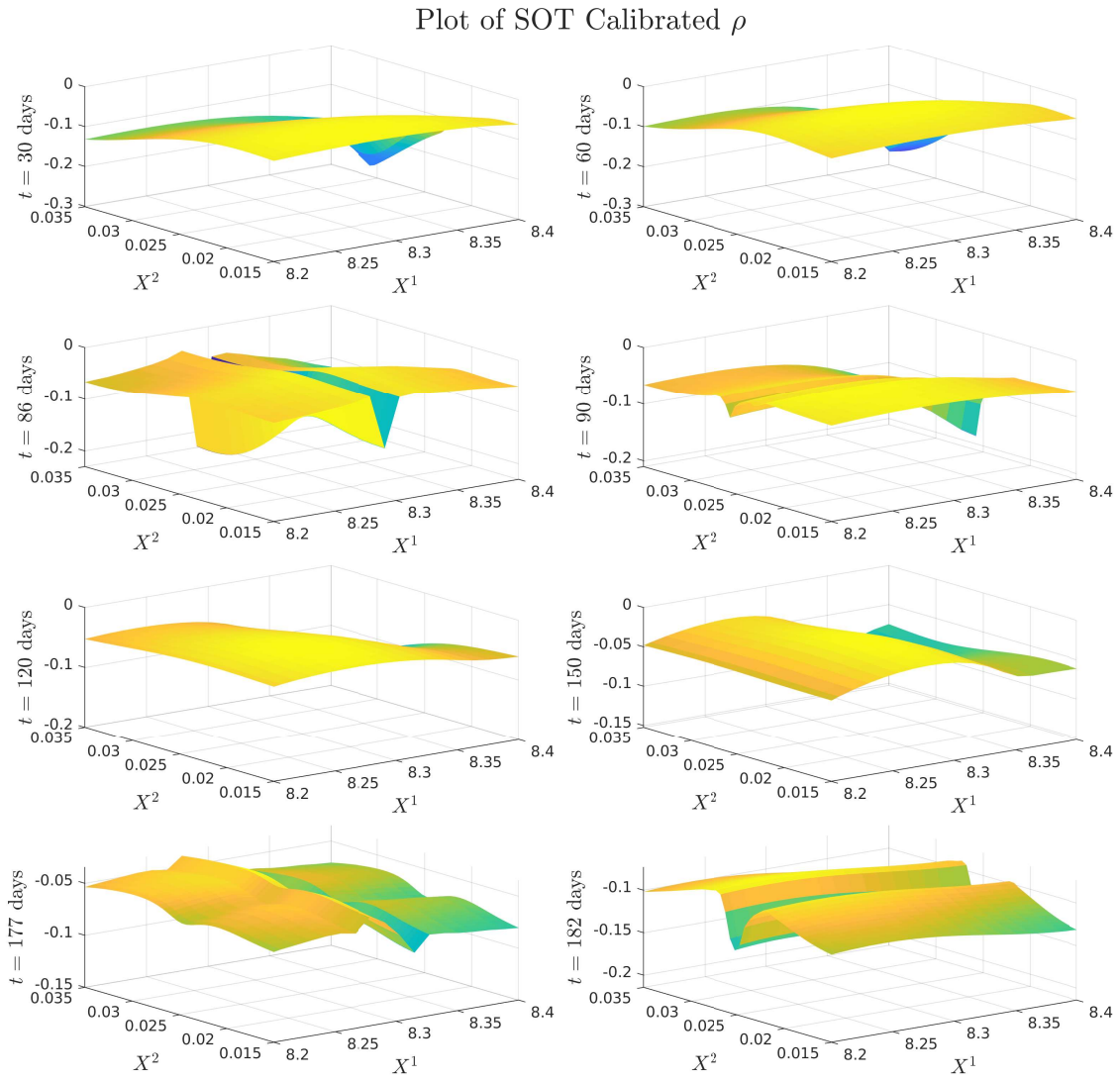


Figure 3.26: Plots of SOT calibrated correlation $\rho = \frac{\beta_{12}}{\sqrt{\beta_{11}\beta_{22}}}$. The β_{12} and β_{22} used here have the scaling undone in order to obtain the actual correlation between $\log(S_t)$ and r_t. The parametrically calibrated reference model takes constant value $\overline{\rho} = -0.2037$.

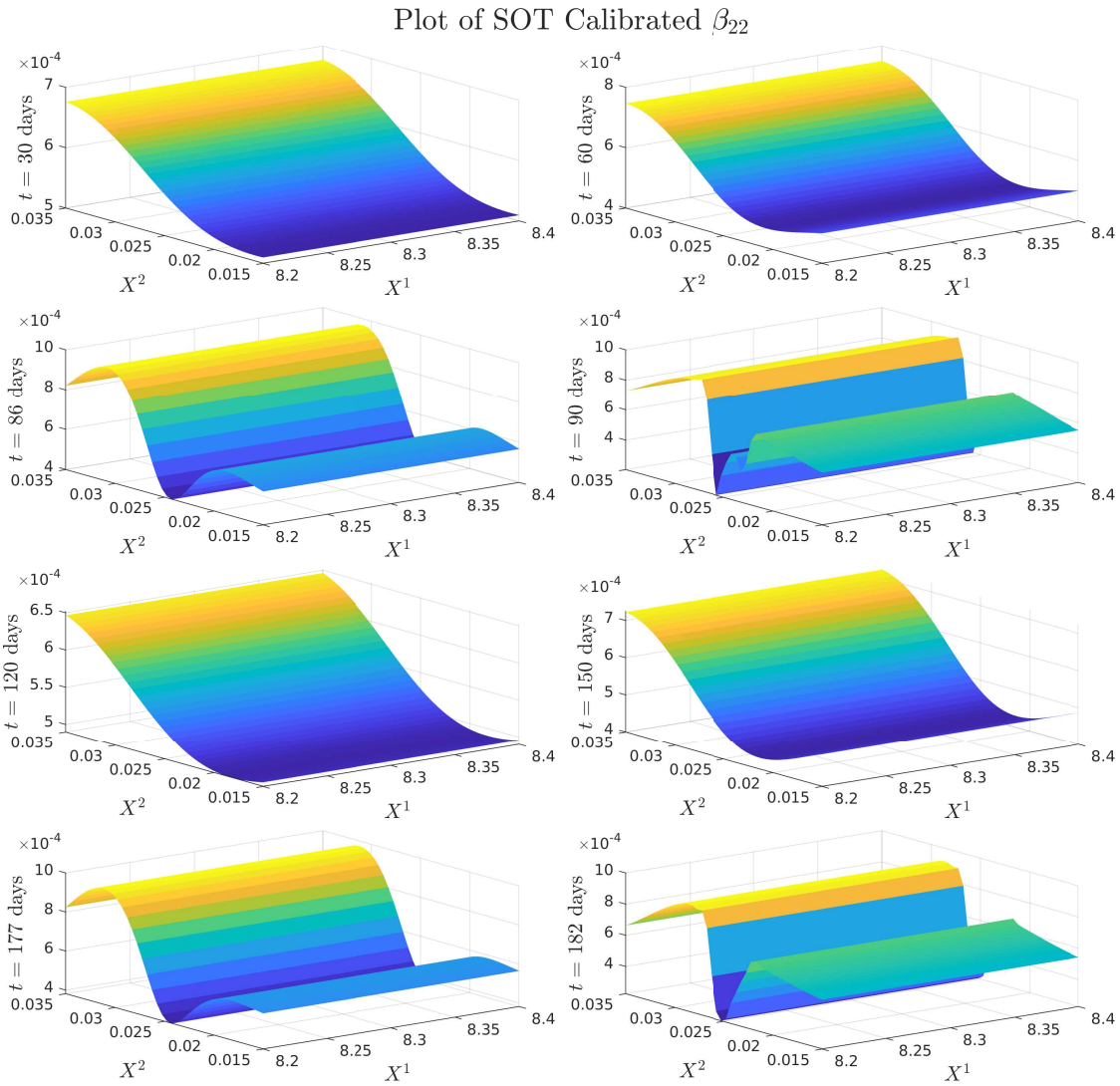


Figure 3.27: Plots of SOT calibrated β_{22}, the volatility of the short rate. Note that the β_{22} used have the scaling undone to represent the actual volatility of the short rate. The parametrically calibrated reference model takes constant value $\overline{\beta}_{22} = 5.3753 \times 10^{-4}$.

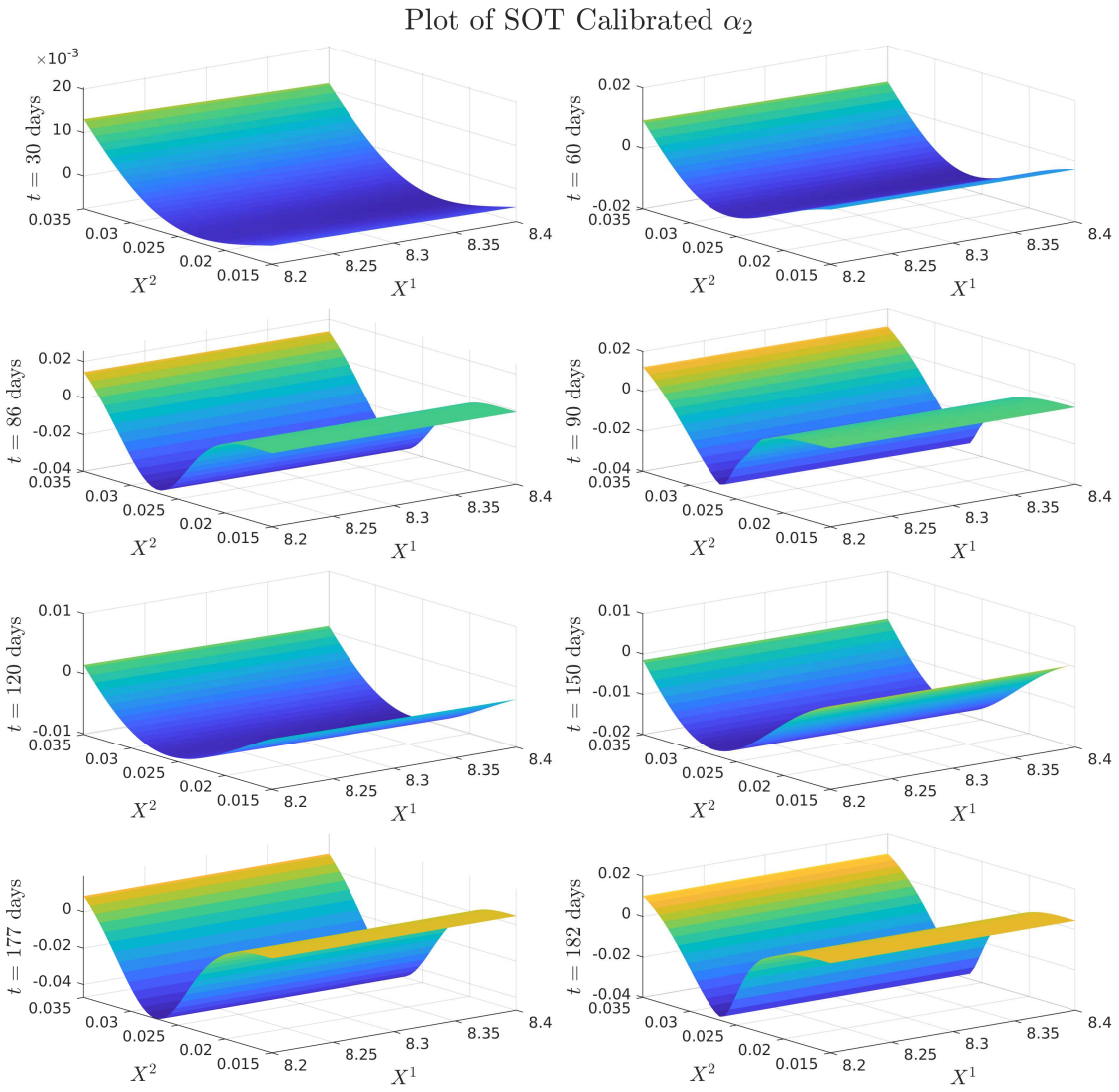


Figure 3.28: Plots of SOT calibrated α_2. Note that the α_2 used have the scaling undone to represent the actual drift of the short rate.

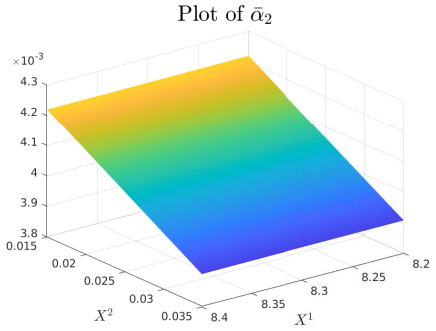


Figure 3.29: Plot of parametrically calibrated reference $\overline{\alpha}_2$. Note that this is taken to be time homogeneous.

3.4 Full Sequential Calibration and a Comparison of all Three Methods

In Section 3.3.4 we developed an efficient joint calibration method for European options on a stock and on interest rates. In Section 3.2, we developed a sequential approach in which the interest rates model was calibrated first and frozen for the subsequent calibration of a local volatility model for the stock prices. This was done under rigid restrictions on the structure of the correlation via (3.2.2) – it was conjectured this would speed up the numerics. We want to now compare the two approaches numerically, as well as introduce and compare a third approach which also proceeds sequentially but relaxes the assumptions on the correlation.

We test the methods on simulated data. We use the generating model as the reference model for the short rate, and our generating and reference models are of the form of the Hull-White-CEV models in (3.2.1). In the two cases where the interest rate is frozen, the surface plots of α_2 and β_{22} will remain the same as the reference model, however α_2 and β_{22} may be perturbed in the joint calibration case so we show all surface plots. The difference in the calibration approach for all three methods is what process(es) are being calibrated, or equivalently the cost function employed, which then also leads to different formulae for the optimal coefficients. We now define the final method, "full sequential" calibration and provide numerical results comparing the three approaches.

3.4.1 "Full Sequential" Calibration

We introduce the notion of "full sequential" calibration, in which the correlation is treated as a free parameter rather than prescribed via (3.2.2). Our set Γ now becomes:

$$\Gamma_{\text{full seq}}(t,r) = \left\{ (\alpha,\beta) \in \mathbb{R}^2 \times \mathbb{S}^2_+ \ : \ \alpha_1 = r - \frac{1}{2}\beta_{11},\ \alpha_2 = (b(t) - ar),\ \beta_{22} = \sigma_r^2 \right\}. \quad (3.4.1)$$

We use the same cost function as in joint calibration, however since we have from the definition of $\Gamma_{\text{full seq}}$ that $\alpha_2 = \overline{\alpha}_2$ and $\beta_{22} = \overline{\beta}_{22}$, and moreover that $\alpha_1 - \overline{\alpha}_1 = \frac{1}{2}(\overline{\beta}_{11} - \beta_{11})$,

our cost function will simplify to:

$$F_{\text{full seq}}(t,r,\alpha,\beta) = \begin{cases} \frac{5}{4}(\beta_{11}-\overline{\beta}_{11})^2 + 2(\beta_{12}-\overline{\beta}_{12})^2, & \text{if } (\alpha,\beta) \in \Gamma_{\text{full seq}}(t,r), \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.4.2}$$

Problem 3.4.1 (Full Sequential Dual Formulation)**.**

$$V = \sup_{\lambda} \lambda \cdot u - \varphi^{\lambda}(0, Z_0, r_0). \tag{3.4.3}$$

Where $\varphi^{\lambda} = \varphi(t,z,r)$ *solves the HJB equation for* $(t,z,r) \in [0,T) \times \mathbb{R}^2$ *with* $\varphi(T,z,r) = 0$ *for all* $(z,r) \in \mathbb{R}^2$*):*

$$\sum_{i=1}^{n} \lambda_i G_i(x)\delta_{\tau_i} + \partial_t\varphi + \sup_{\substack{\beta_{11},\beta_{12},\\ \beta\in\mathbb{S}_+^2}} \Bigg(\left(r - \frac{1}{2}\beta_{11}\right)\partial_z\varphi + (b(t)-ar)\partial_r\varphi + \frac{1}{2}\beta_{11}\partial^2_{zz}\varphi + \beta_{12}\partial^2_{zr}\varphi + \frac{1}{2}\sigma_r^2\partial^2_{rr}\varphi - r\varphi - F_{\textit{full seq}}(t,r,\alpha,\beta)\Bigg) = 0. \tag{3.4.4}$$

In a similar way to Lemma 3.3.2, we obtain an approximation of the optimal β_{11} and β_{12} by first optimising over β_{11}, and then over β_{12} applying the positive semidefinite constraint in this variable given β^*_{11}.

Lemma 3.4.2. *Let* $\beta^*_{22} = \sigma_r^2$ *and define* β^*_{11} *and* $\beta^*_{12} = \beta^*_{21}$ *by:*

$$\beta^*_{11} = \begin{cases} \overline{\beta}_{11} + \frac{1}{5}(\partial^2_{zz}\varphi - \partial_z\varphi), & \textit{when } \overline{\beta}_{11} + \frac{1}{5}(\partial^2_{zz}\varphi - \partial_z\varphi) \in [\delta^l_{11}, \delta^u_{11}] \\ \delta^l_{11}, & \textit{when } \overline{\beta}_{11} + \frac{1}{5}(\partial^2_{zz}\varphi - \partial_z\varphi) < \delta^l_{11}, \\ \delta^u_{11}, & \textit{when } \overline{\beta}_{11} + \frac{1}{5}(\partial^2_{zz}\varphi - \partial_z\varphi) > \delta^u_{11}, \end{cases} \tag{3.4.5}$$

$$\beta^*_{12} = \begin{cases} \overline{\beta}_{12} + \frac{1}{4}\partial^2_{zr}\varphi, & \textit{when } \overline{\beta}_{12} + \frac{1}{4}\partial^2_{zr}\varphi \in \left[-\sigma_r\sqrt{\beta^*_{11}}, \sigma_r\sqrt{\beta^*_{11}}\right], \\ -\sigma_r\sqrt{\beta^*_{11}}, & \textit{when } \overline{\beta}_{12} + \frac{1}{4}\partial^2_{zr}\varphi < -\sigma_r\sqrt{\beta^*_{11}}, \\ \sigma_r\sqrt{\beta^*_{11}}, & \textit{when } \overline{\beta}_{12} + \frac{1}{4}\partial^2_{zr}\varphi > \sigma_r\sqrt{\beta^*_{11}}. \end{cases} \tag{3.4.6}$$

Then β^* is a positive semi-definite matrix and whenever

$$-\sigma_r\sqrt{\beta^*_{11}} < \beta^*_{12} < \sigma_r\sqrt{\beta^*_{11}} \tag{3.4.7}$$

then β^* is the optimizer in (3.4.4).

3.4.2 Numerical Results

We provide a table of parameters used in the generating and reference models for our problem.

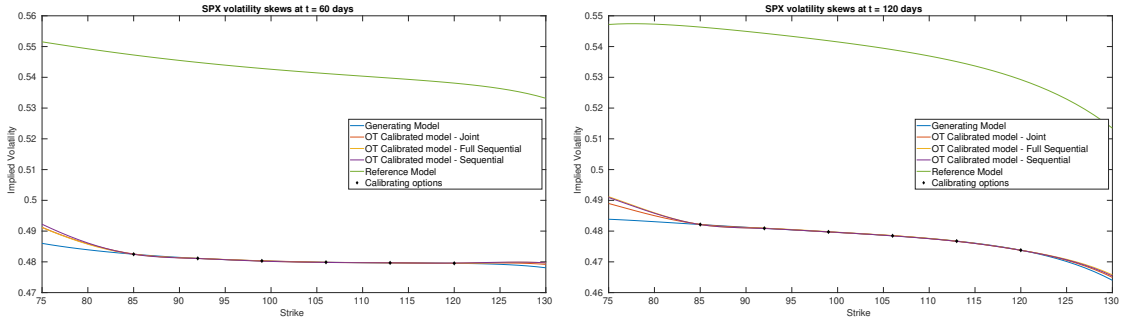
Hull-White CEV Model		
Parameter	Value	Interpretation
X_0^1	$\log(92)$	Initial log-stock price
X_0^2	0.025×100	Initial short rate scaled by $R = 100$
ε_1	1×10^{-4}	Tolerance for the difference in scaled model and market implied volatility
ε_2	1×10^{-12}	Tolerance for the policy iteration approximation of the optimal characteristics
δ_{11}^l	0.05	Lower bound of β_{11} in full sequential and joint calibration
δ_{11}^u	1	Upper bound of β_{11} in full sequential and joint calibration
p	4	Exponent in sequential calibration cost function
σ	0.60	Volatility scaling of generating CEV model
γ	0.95	Power law in generating CEV model
$b(t)$	$aX_0^2 + \frac{\sigma_r^2}{2a}(1 - e^{-2at})$	Initial term structure of Hull-White generating model
a	0.05	Speed of mean reversion of Hull-White generating model
σ_r	0.04	Volatility of Hull-White generating model
ρ	-0.40	Instantaneous correlation between short rate and log-stock in generating model
$\overline{\sigma}$	0.90	Volatility scaling of reference CEV model
$\overline{\gamma}$	0.89	Power law in reference CEV model
$\overline{b}(t)$	$\overline{a}X_0^2 + \frac{\overline{\sigma}_r^2}{2\overline{a}}(1 - e^{-2\overline{a}t})$	Initial term structure of Hull-White reference model
$\overline{a}$	0.05	Speed of mean reversion of Hull-White reference model
$\overline{\sigma}_r$	0.04	Volatility of Hull-White reference model
$\overline{\rho}$	-0.20	Instantaneous correlation between short rate and log-stock in reference model

Table 3.8: Parameter values of generating and reference models used in all three methods.

		Generating Model		Calibrated Model: Sequential		Calibrated Model: Full Sequential		Calibrated Model: Joint	
Option Type	Strike	Price	IV	Price	IV	Price	IV	Price	IV
SPX call options $t = 60$ days	85	11.2142	0.4825	11.2139	0.4825	11.2139	0.4825	11.2152	0.4826
	92	7.3755	0.4811	7.3757	0.4811	7.3749	0.4811	7.3750	0.4811
	99	4.6051	0.4803	4.6045	0.4803	4.6038	0.4802	4.6056	0.4804
	106	2.7426	0.4799	2.7420	0.4798	2.7419	0.4798	2.7426	0.4799
	113	1.5667	0.4797	1.5665	0.4797	1.5667	0.4797	1.5673	0.4797
	120	0.8624	0.4795	0.8623	0.4795	0.8629	0.4796	0.8626	0.4796
SPX call options $t = 120$ days	85	14.0842	0.4821	14.0839	0.4821	14.0838	0.4821	14.0857	0.4822
	92	10.502	0.4809	10.5021	0.4809	10.5007	0.4808	10.5027	0.4809
	99	7.6696	0.4797	7.6712	0.4798	7.6699	0.4798	7.6691	0.4797
	106	5.4943	0.4785	5.4952	0.4785	5.4936	0.4784	5.4952	0.4785
	113	3.8607	0.4767	3.8609	0.4768	3.8597	0.4767	3.8594	0.4767
	120	2.6499	0.4738	2.6498	0.4738	2.6495	0.4738	2.6499	0.4738

Table 3.9: Table of the generating and calibrated model prices and implied volatilities.

3.4.2.1 Implied Volatility Plots



(a) (b)

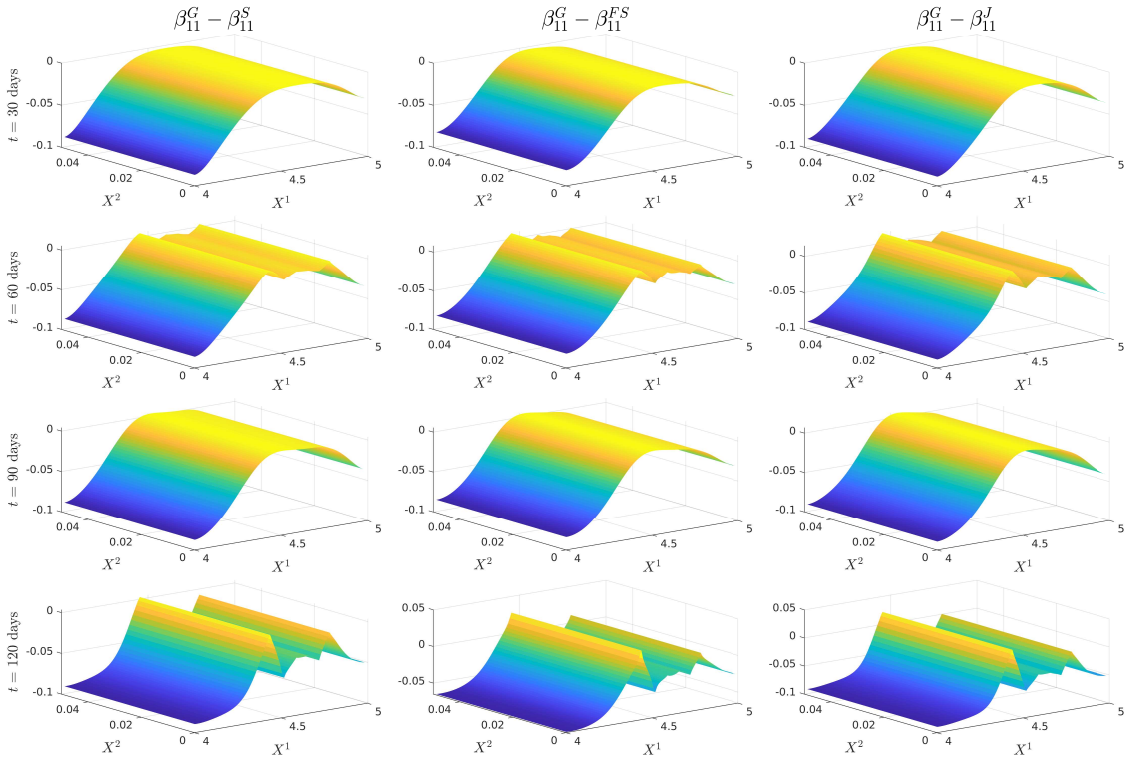
Figure 3.30: Implied volatility skews of the SPX calibrating options under the reference model and calibrated HW-CEV model.

3.4.2.2 Plots of the Volatility and Correlation surfaces

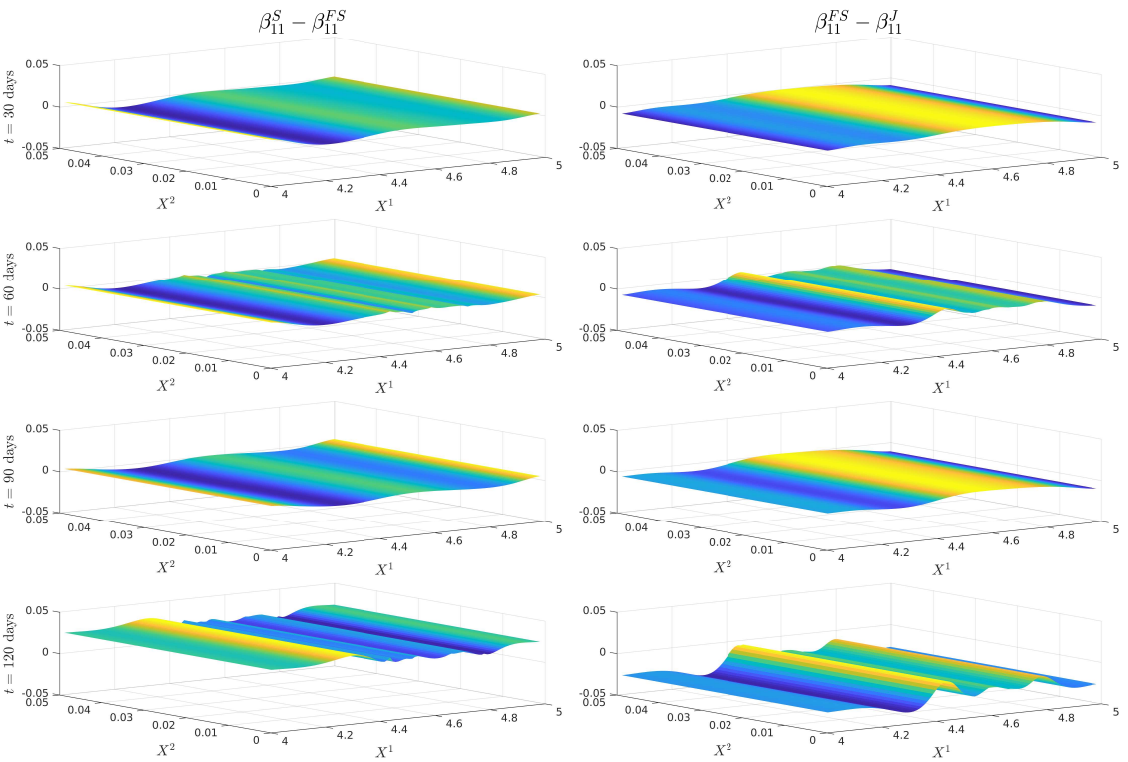
For notational ease in the comparison, denote via a superscript S, FS, J, G the characteristic under the SOT sequential calibrated model, SOT full sequential calibrated model, SOT jointly calibrated model and generating model respectively. In order to provide a meaningful comparison of the three methods, we show plots of the difference in the surfaces for all five characteristics, and how all of them differ from the generating model. We omit all plots for α_1 since in all three cases it is entirely specified by β_{11}. We remark that the formulae for β_{11}^{FS} and β_{11}^{J} given in Lemma 3.3.2 and Lemma 3.4.2 are upon a first glance the same, and both differ from the formula for β_{11}^{S} given in (3.2.10). However, owing to the different cost

functions and the fact that the joint problem allows for perturbations in the interest rate, whereas the full sequential problem does not, there is no expectation of the optimisers φ^*_{FS} and φ^*_J being the same since the global optimisers of both problems may differ. We expect the correlation coefficient ρ^S to be almost constant and close to the reference model, owing to the assumption on the correlation in (3.2.2), whereas we expect some deviation away from the reference model for ρ^{FS} and ρ^J since we relax the assumptions on the correlation. The short rate volatility β_{22} and short rate drift α_2 are assumed to be a priori correct due to the matching generating and reference models. Note that the coefficients $\alpha_2^S, \alpha_2^{FS}, \beta_{22}^S, \beta_{22}^{FS}$ are all fixed, and equal to these in the generating model, which we recall is also the reference model, i.e., we have[3] $\alpha_2^S = \alpha_2^{FS} = \alpha_2^G$ and $\beta_{22}^S = \beta_{22}^{FS} = \beta_{22}^G$. However, in the joint calibration case we do not fix α_2^J and β_{22}^J and our optimisers given in Lemma 3.3.2 do not guarantee replication of the generating model. Nonetheless, since the cost function penalises deviations from the reference model, we expect that α_2^J and β_{22}^J will be close to the generating model with some small perturbations away from α_2^G and β_{22}^G.

[3]We confirmed this numerically in our results.

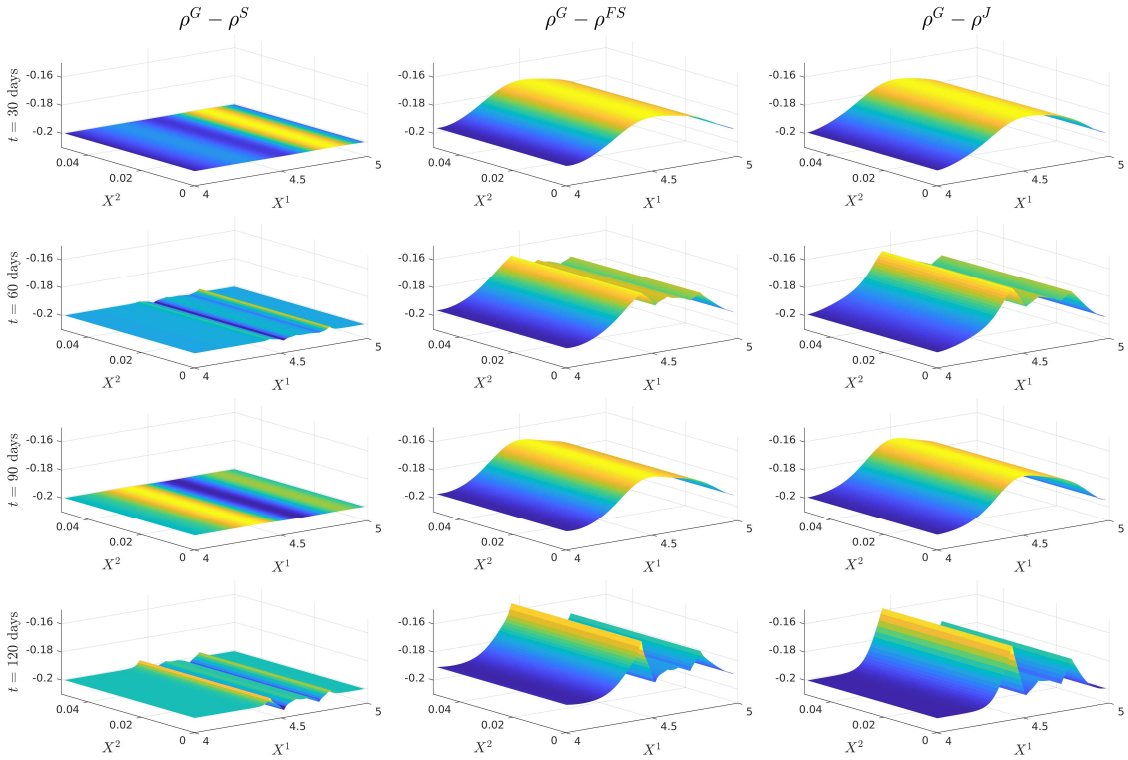


(a)

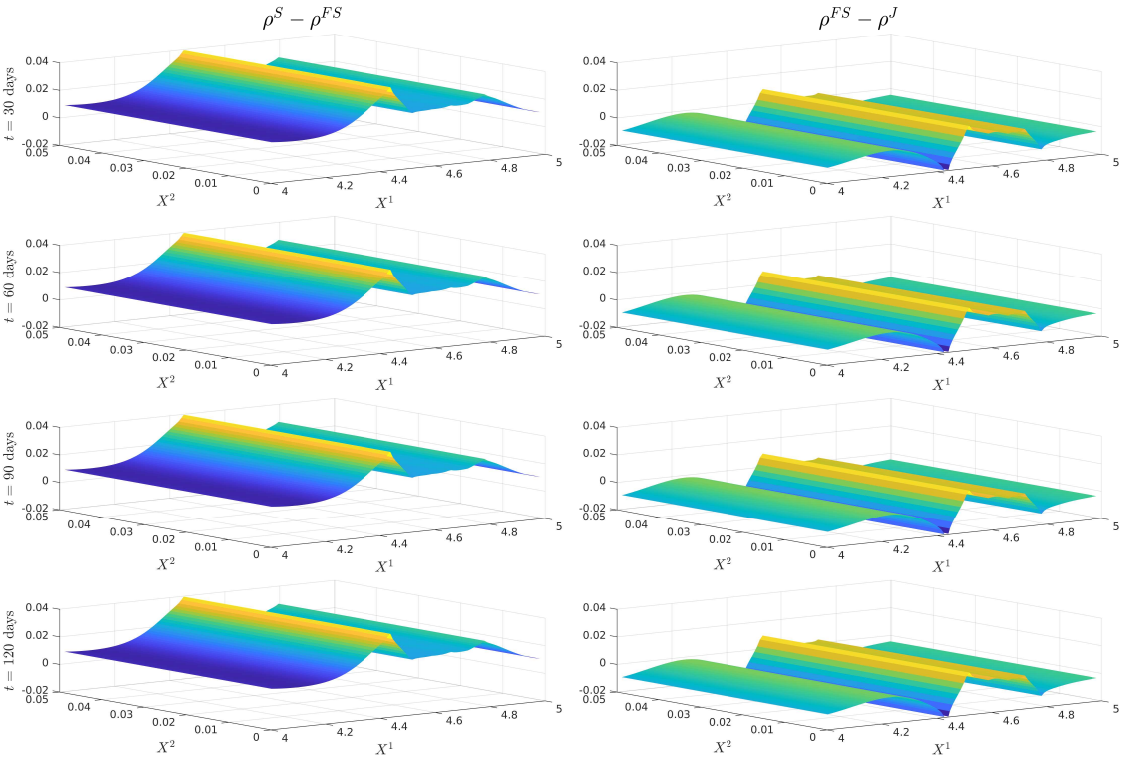


(b)

Figure 3.31: Plots of the difference between the generating model β_{11} and SOT calibrated β_{11} in the three methods (a), and plots of the difference between calibrated β_{11} in the three methods (b).



(a)



(b)

Figure 3.32: Plots of the difference between the generating model $\rho = \frac{\beta_{12}}{\sqrt{\beta_{11}\beta_{22}}}$ and SOT calibrated ρ in the three methods (a), and plots of the difference between calibrated ρ in the three methods (b). Note that we undid the scaling in β_{12} and β_{22} to recover the correlation between the log-stock and the short rate. 118

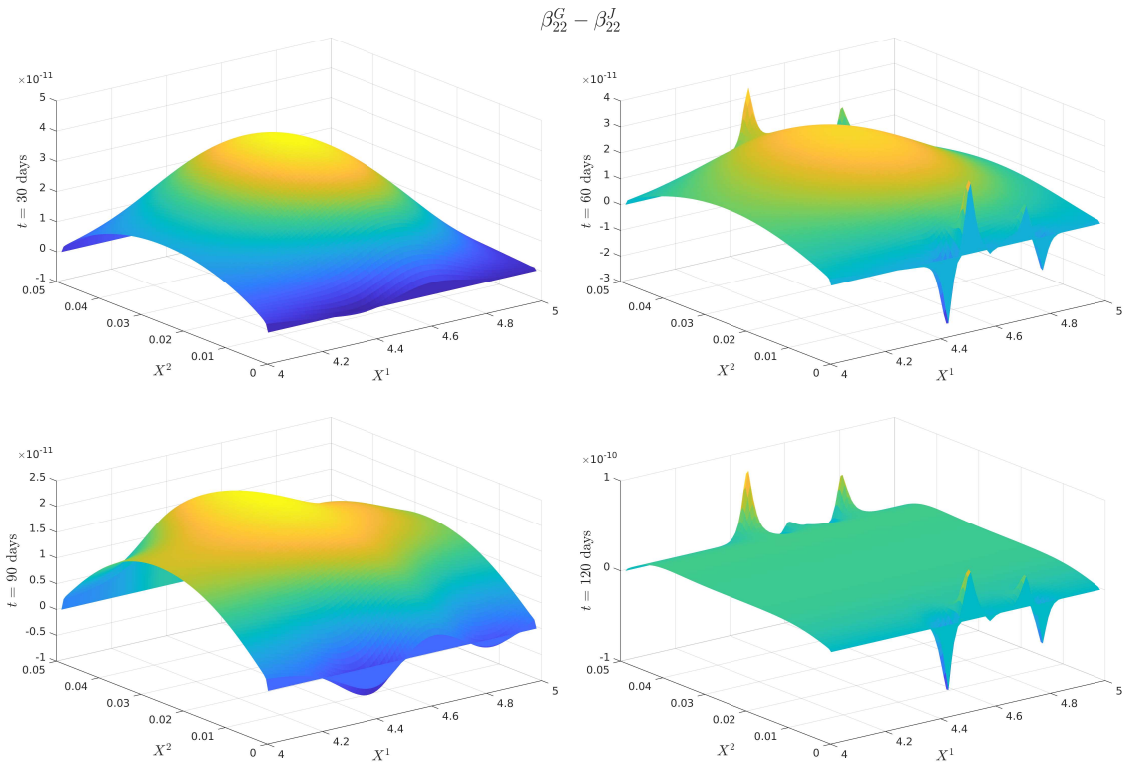


Figure 3.33: Plots of the difference between the generating model β_{22} and SOT jointly calibrated β_{22}. Note that we undid the scaling in β_{22} to recover the volatility of the short rate.

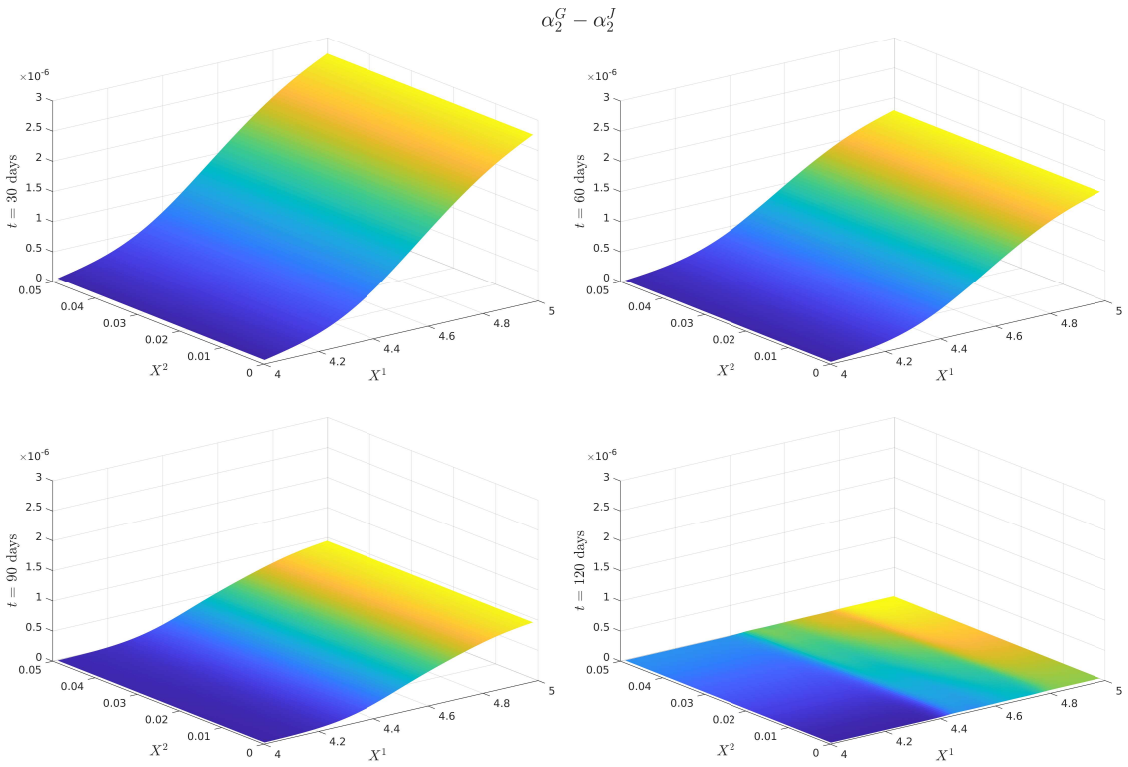
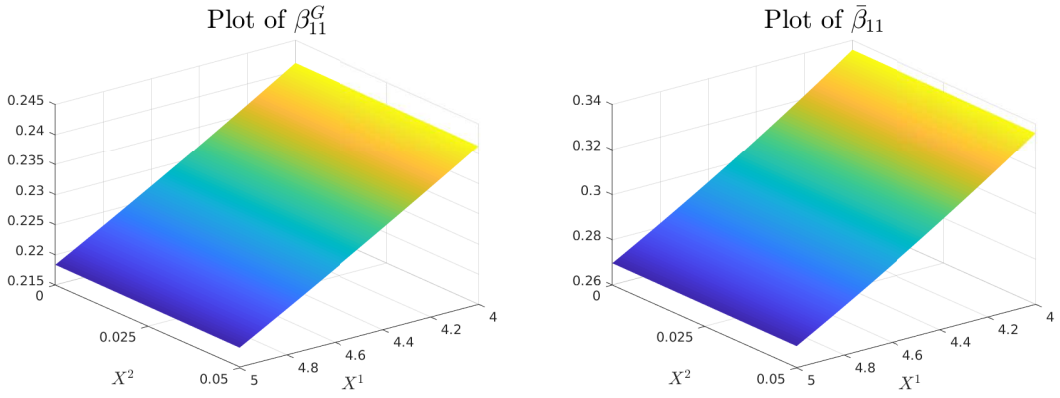


Figure 3.34: Plots of the difference between the generating model α_2 and SOT jointly calibrated α_2. Note that we undid the scaling in α_2 to recover the drift of the short rate.

We only show the plots of β_{11} and ρ since α_1 is entirely determined by β_{11} and Figure 3.31 and Figure 3.34 demonstrate that all three cases are extremely close or exactly equal to the generating model for β_{22} and α_2.



(a) (b)

Figure 3.35: Plots of the generating model log-stock volatility β_{11}^G (a), and reference model log-stock volatility $\overline{\beta}_{11}$ (b).

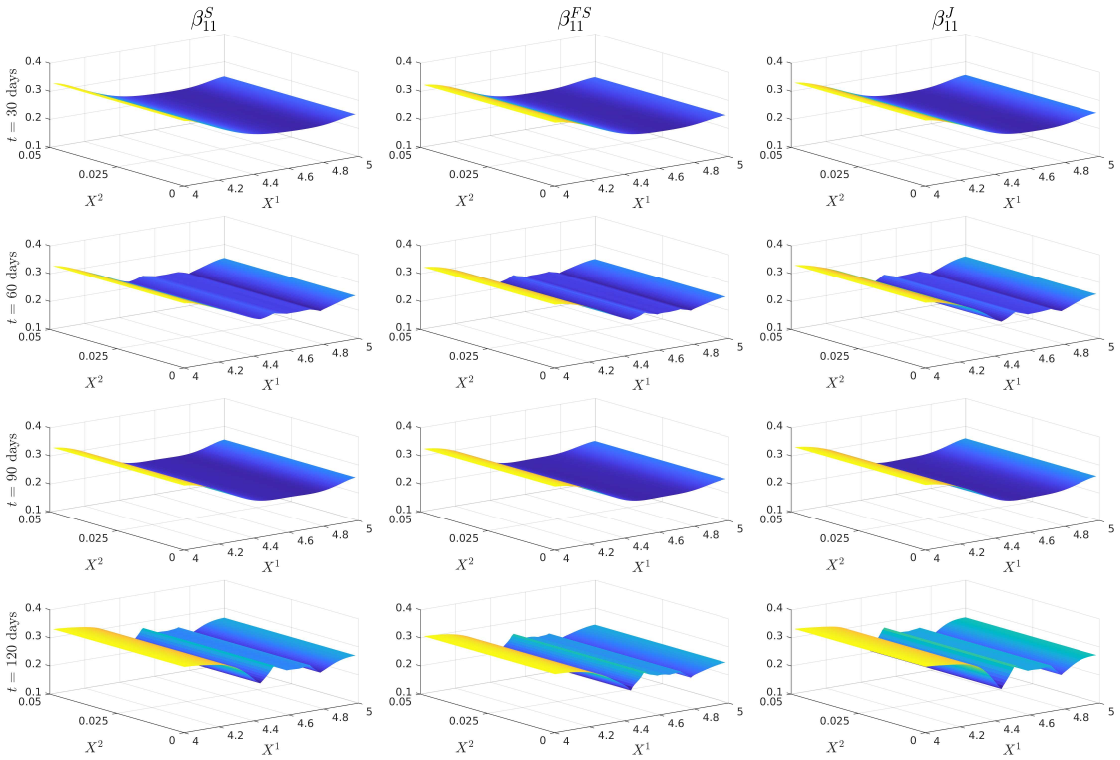


Figure 3.36: Plots of calibrated β_{11} in all three cases.

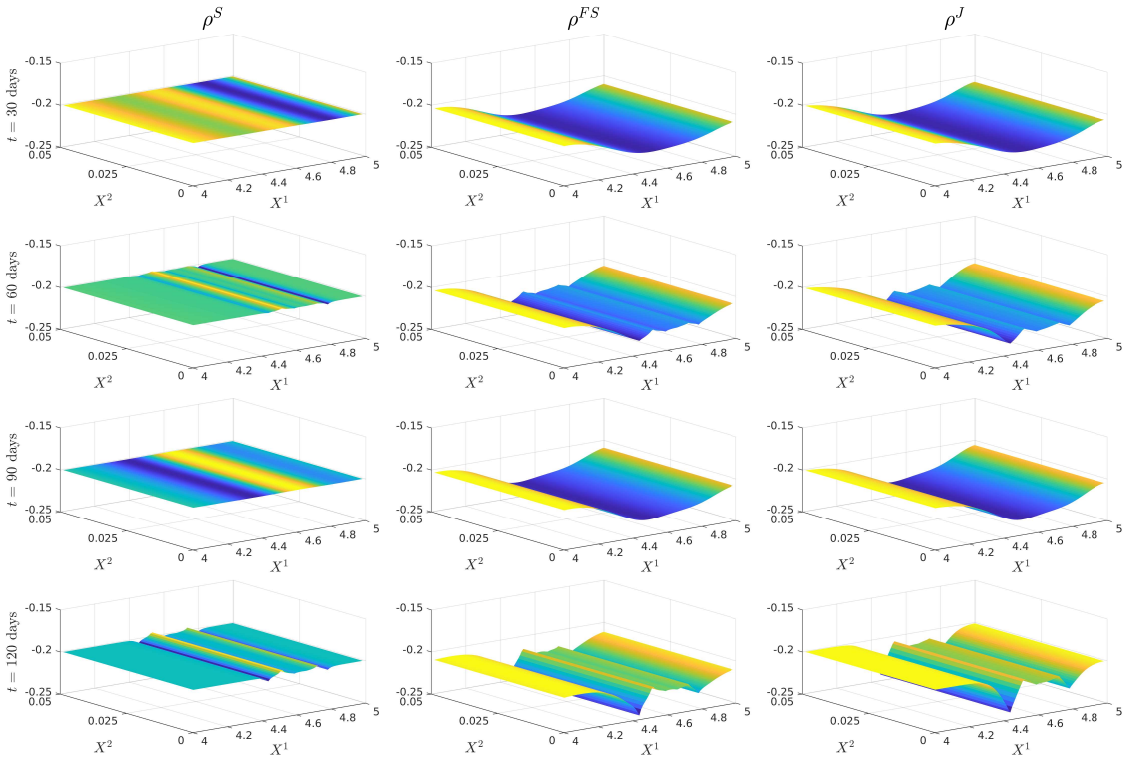


Figure 3.37: Plots of calibrated $\rho = \frac{\beta_{12}}{\sqrt{\beta_{11}\beta_{22}}}$ in all three cases, where the scaling has been removed from β_{12} and β_{22}. The generating model correlation $\rho^G = -0.4$ and the reference model correlation $\overline{\rho} = -0.2$.

3.4.3 Discussion of the results

Figure 3.31 (a) demonstrates that $\beta_{11}^S, \beta_{11}^{FS}, \beta_{11}^J$ all are near the generating model within the region of our strikes and then deviate in the same manner towards the reference model outside the region of calibrating call option strikes. This is expected behaviour since in all three cases, the cost functions penalise deviations away from a reference model volatility while enforcing that the observed call option prices are replicated. A remarkable result is that Figure 3.31 (b) shows that $\beta_{11}^S, \beta_{11}^{FS}, \beta_{11}^J$ are all close at $t = 30, 60, 90$ days despite being defined differently. We do however see some expected perturbations on those dates and a larger perturbation at $t = 120$ days which is also observed in the larger fluctuations seen in Figure 3.31 (a). Figure 3.32 (a) demonstrates strong dependence on the reference model: $\rho^S, \rho^{FS}, \rho^J$ all differ from ρ^G by approximately -0.2, however ρ^{FS} and ρ^J do perturb slightly towards to generating model in the region of strikes. Since our calibrating option has a payoff function only explicitly depending on the log-stock, this was expected behaviour

since ρ^{FS} and ρ^{J} are defined via $\partial^2_{zr}\varphi$. As remarked in Section 3.2, the extremely strong dependence on the reference model for ρ^{S} arises mainly from the assumption on the form of the correlation given in (3.2.2). The differences observed in Figure 3.32 (b) arise from a fundamentally different handling of ρ^{S} and that the volatility of the short rate is fixed so that ρ^{FS} will be different to ρ^{J}. Nonetheless, since all three depend strongly on the reference model, these fluctuations are small.

All three methods converged to a good accuracy with a maximum error in implied volatility of 10^{-4}, as shown in Table 3.9 and Figure 3.30, where in all three cases, we applied Algorithm 3 with the smoothed reference model iteration. A minimum of 10 smoothing iterations were applied in all three cases to generate smoother surfaces. Since each smoothing iteration terminates either when calibration error is achieved or when a threshold of function evaluations were attained, in practice around 20 were required for the sequential calibration case to converge, as opposed to full sequential and joint calibration which both converged on the 10^{th} smoothing iteration. The fastest in terms of computational time was the full sequential approach, taking 2:43:19, then the sequential approach, taking 3:19:10 and the slowest was the joint calibration approach taking 3:40:00. Each epoch of a smoothed reference iteration was the fastest in sequential calibration. Joint calibration being the slowest was to be expected since it still computes the optimisers α_2^* and β_{22}^* at each point, which is more expensive that simply fixing them as in the other two methods.

3.4.4 Comparison of the Three Methods with a Wrong Stochastic Interest Rate

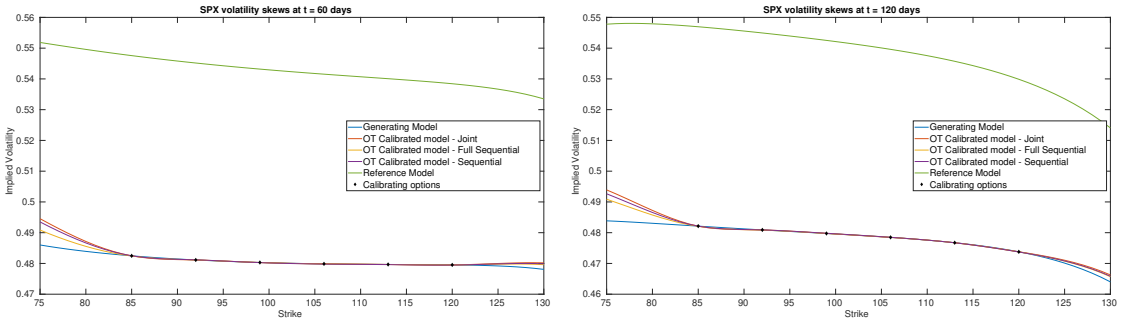
So far in our comparisons, we have assumed that α_2 and β_{22} are known and calibrated to perfectly replicate the term structure in the market. In practice, this will not always be the case, and therefore we compare the performance of all three methods when the reference α_2 and β_{22} are incorrect. We calibrate to the same call options as in Table 3.9, and use the same type of reference model with the parameters the same as in Table 3.8 except for $\overline{a} = 0.1$, $\overline{\sigma}_r = 0.02$ and $\overline{b}(t)$ defined in the same way but with the new parameters. Since the model characteristics α_2 and β_{22} remain fixed as the reference model values in Sequential and Full Sequential calibration, those SOT model characteristics will remain incorrect, however

we expect some perturbations in the joint calibration case away from the reference model since those are not fixed by the calibration method. Our only calibration instruments are call options since the Sequential and Full Sequential methods do not allow for calibration to interest rate derivatives within this framework, and therefore we do not expect for our surfaces to be near to the generating interest rate surface.

		Generating Model		Calibrated Model: Sequential		Calibrated Model: Full Sequential		Calibrated Model: Joint	
Option Type	Strike	Price	IV	Price	IV	Price	IV	Price	IV
SPX call options $t = 60$ days	85	11.2142	0.4825	11.2134	0.4824	11.2144	0.4825	11.2141	0.4825
	92	7.3755	0.4811	7.3757	0.4811	7.3761	0.4812	7.3754	0.4811
	99	4.6051	0.4803	4.6049	0.4803	4.6045	0.4803	4.6048	0.4803
	106	2.7426	0.4799	2.7422	0.4798	2.7417	0.4798	2.7428	0.4799
	113	1.5667	0.4797	1.5665	0.4797	1.5662	0.4796	1.5666	0.4797
	120	0.8624	0.4795	0.8623	0.4795	0.8621	0.4795	0.8623	0.4795
SPX call options $t = 60$ days	85	14.0842	0.4821	14.0829	0.4821	14.0841	0.4821	14.084	0.4821
	92	10.5020	0.4809	10.5003	0.4808	10.5017	0.4809	10.5019	0.4809
	99	7.6696	0.4797	7.6704	0.4798	7.6712	0.4798	7.6700	0.4798
	106	5.4943	0.4785	5.4947	0.4785	5.4948	0.4785	5.4938	0.4785
	113	3.8607	0.4767	3.8601	0.4767	3.8601	0.4767	3.8606	0.4767
	120	2.6499	0.4738	2.6493	0.4737	2.6491	0.4737	2.6502	0.4738

Table 3.10: Table of the generating and calibrated model prices and implied volatilities.

3.4.4.1 Implied Volatility Plots



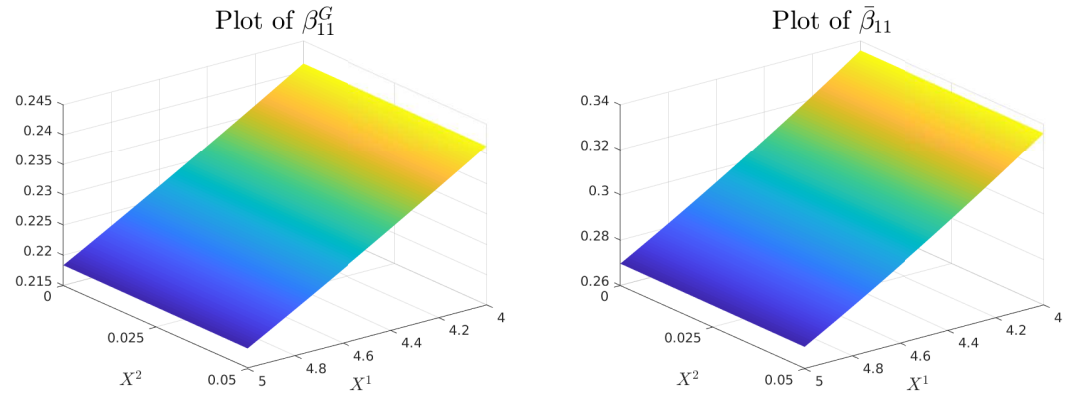
(a) (b)

Figure 3.38: Implied volatility skews of the SPX calibrating options under the reference model and calibrated HW-CEV model.

Figure 3.38 demonstrates that we get an exact fit in all three methods, even under an incorrect fixed stochastic rate, which demonstrates the robustness of the method to errors in computing the interest rate. In Section 3.4.4.2, we only show the plots of β_{11} and ρ since α_1 is entirely determined by β_{11}. We also omit the difference plots similar to those

shown in Section 3.4.2.2, since the generating plots are included and the surfaces alone can demonstrate that the calibrated surfaces of interest are all extremely close to the generating model. We also confirmed numerically that all three cases are extremely close or exactly equal to the reference model for β_{22} and α_2.

3.4.4.2 Plots of the drift and diffusion surfaces



(a) (b)

Figure 3.39: Plots of the generating model log-stock volatility β_{11}^G (a), and reference model log-stock volatility $\overline{\beta}_{11}$ (b).

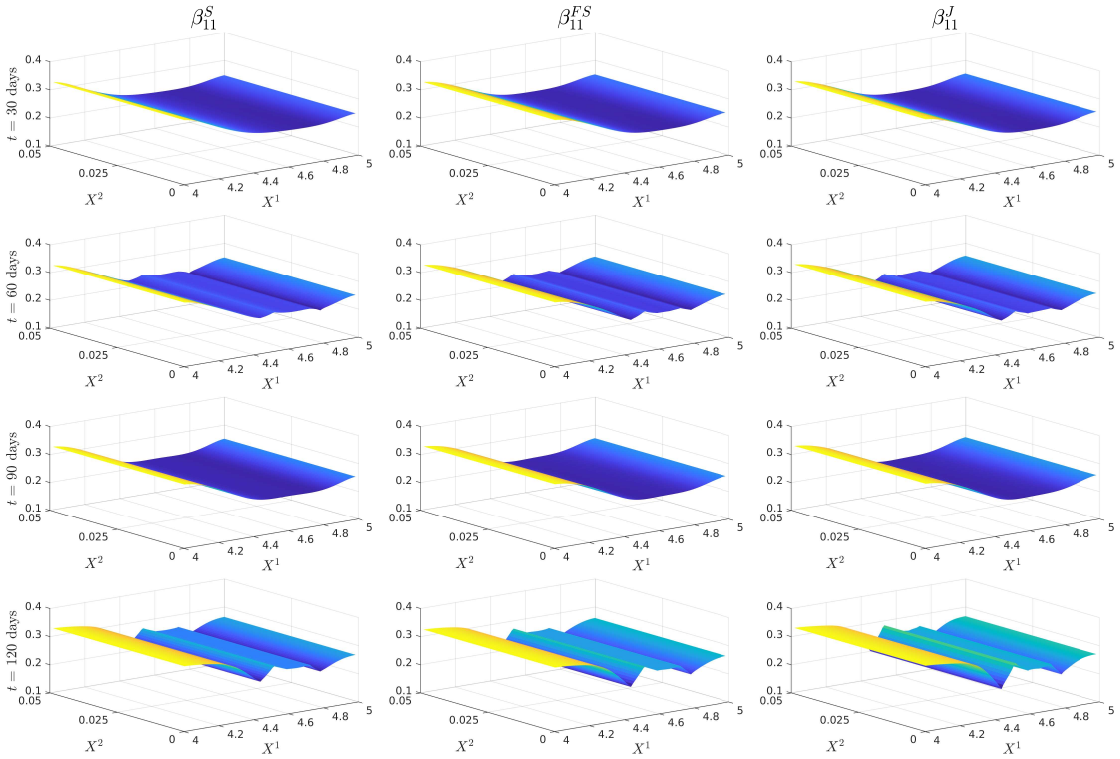


Figure 3.40: Plots of calibrated β_{11} in all three cases.

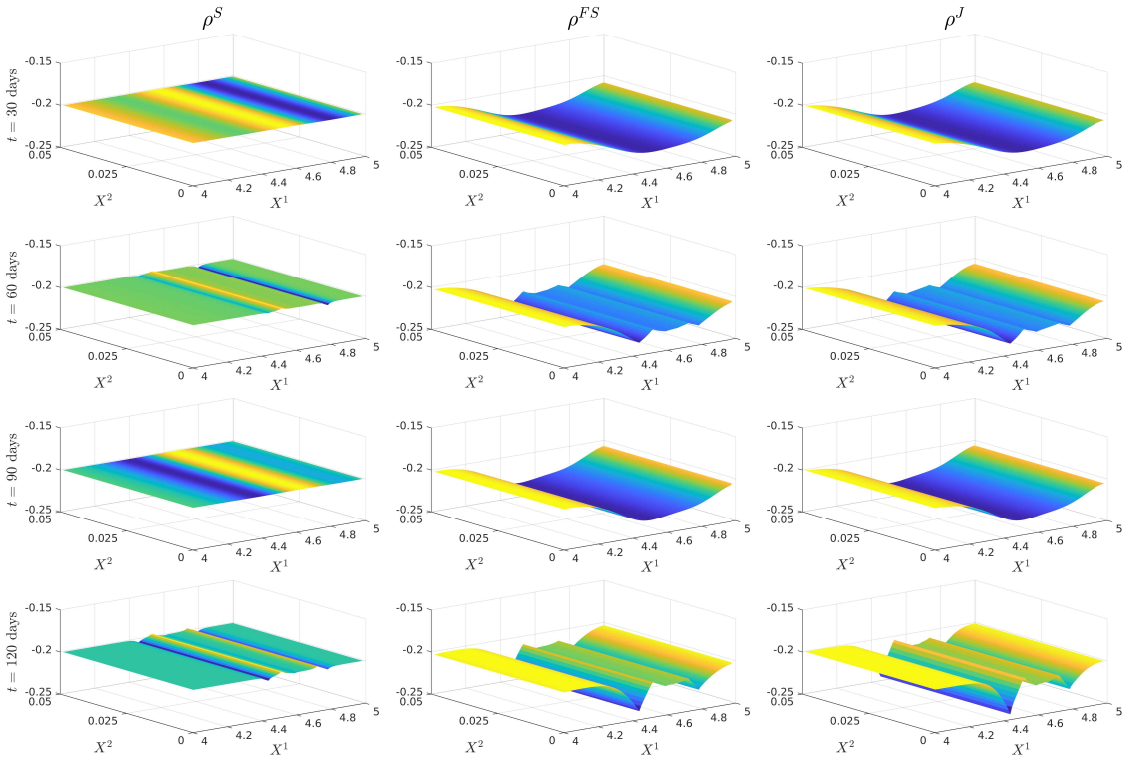


Figure 3.41: Plots of calibrated $\rho = \frac{\beta_{12}}{\sqrt{\beta_{11}\beta_{22}}}$ in all three cases, where the scaling has been removed from β_{12} and β_{22}. The generating model correlation $\rho^G = -0.4$ and the reference model correlation $\overline{\rho} = -0.2$.

3.5 Full Sequential Calibration in the Heston Model

We observe there that the method of Section 3.4.1 can be applied to the local-stochatic volatility calibration setting of I. Guo, Loeper, and Shiyi Wang, 2022 to relax the constraint on the correlation imposed in that paper. In this setting, we assume our interest rate is zero and that our reference model is given by a Heston model, derived in Heston, 1993. The state variables are therefore our log-stock and a correlated stochastic term for the volatility, with dynamics given by:

$$\mathrm{d}Z_t = -\frac{1}{2}V_t\,\mathrm{d}t + \sqrt{V_t}\,\mathrm{d}W_t^1, \tag{3.5.1}$$

$$\mathrm{d}V_t = \kappa(\theta - V_t)\,\mathrm{d}t + \xi\sqrt{V_t}\,\mathrm{d}W_t^2, \tag{3.5.2}$$

$$\mathrm{d}\langle W^1, W^2\rangle_t = \rho\,\mathrm{d}t, \tag{3.5.3}$$

where $\kappa, \theta, \xi > 0$ and $\rho \in [-1, 1]$. We assume that the state variable V_t is known, so the model characteristics that we want to calibrate are the volatility of the log-stock and the correlation between the state variables. As before, the instruments that we calibrate this model to are European call options on the chosen stock. We now define our cost function in a similar manner to Section 3.4.1 by defining the convex set Γ_{LSV}:

$$\Gamma_{\text{LSV}}(v) = \left\{ (\alpha, \beta) \in \mathbb{R}^2 \times \mathbb{S}_+^2 \ : \ \alpha_1 = -\frac{1}{2}\beta_{11},\ \alpha_2 = \kappa(\theta - v),\ \beta_{22} = \xi^2 v,\ \delta_{11}^l \leqslant \beta_{11} \leqslant \delta_{11}^u \right\}. \tag{3.5.4}$$

$$F_{\text{LSV}}(v, \alpha, \beta) = \begin{cases} \frac{5}{4}(\beta_{11} - \overline{\beta}_{11})^2 + 2(\beta_{12} - \overline{\beta}_{12})^2, & \text{if } (\alpha, \beta) \in \Gamma_{\text{LSV}}(v), \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.5.5}$$

Since we have no stochastic discount factor, we no longer need the sub-probability measure approach of Joseph, Loeper, and Obłój, 2023a, and therefore can directly apply the duality result of I. Guo, Loeper, and Shiyi Wang, 2022, Proposition 3.7. With our cost function defined in (3.5.5), we therefore obtain the following dual formulation:

Problem 3.5.1 (Full Sequential Dual Formulation)**.**

$$V = \sup_{\lambda} \lambda \cdot u - \varphi^{\lambda}(0, Z_0, V_0). \tag{3.5.6}$$

Where $\varphi^{\lambda} = \varphi(t, z, v)$ solves the HJB equation:

$$\sum_{i=1}^{n} \lambda_i G_i(x) \delta_{\tau_i} + \partial_t \varphi + \sup_{\substack{\beta_{11}, \beta_{12}, \\ \beta \in \mathbb{S}_+^2}} \Big(-\frac{1}{2}\beta_{11}\partial_z \varphi + \kappa(\theta - v)\partial_v \varphi + \frac{1}{2}\beta_{11}\partial_{zz}^2 \varphi + \beta_{12}\partial_{zv}^2 \varphi + \frac{1}{2}\xi^2 v \partial_{vv}^2 \varphi - F_{\text{LSV}}(v, \alpha, \beta) \Big) = 0, \quad (t, z, v) \in [0, T] \times \mathbb{R}^2. \tag{3.5.7}$$

Similar to Lemma 3.4.2, we obtain the following approximation for β_{11}^* and β_{12}^*:

Lemma 3.5.2. *Let* $\beta^*_{22} = \xi^2 v$ *and define* β^*_{11} *and* $\beta^*_{12} = \beta^*_{21}$ *by*

$$\beta^*_{11}(t,z,v) = \begin{cases} \overline{\beta}_{11}(t,z,v) + \frac{1}{5}(\partial^2_{zz}\varphi(t,z,v) - \partial_z\varphi(t,z,v)), & \text{when } \overline{\beta}_{11} + \frac{1}{5}(\partial^2_{zz}\varphi - \partial_z\varphi) \in [\delta^l_{11}, \delta^u_{11}] \\ \delta^l_{11}, & \text{when } \overline{\beta}_{11} + \frac{1}{5}(\partial^2_{zz}\varphi - \partial_z\varphi) < \delta^l_{11}, \\ \delta^u_{11}, & \text{when } \overline{\beta}_{11} + \frac{1}{5}(\partial^2_{zz}\varphi - \partial_z\varphi) > \delta^u_{11}, \end{cases} \tag{3.5.8}$$

$$\beta^*_{12}(t,z,v) = \begin{cases} \overline{\beta}_{12}(t,z,v) + \frac{1}{4}\partial^2_{zv}\varphi(t,z,v), & \text{when } \overline{\beta}_{12} + \frac{1}{4}\partial^2_{zv}\varphi \in \left[-\xi\sqrt{v\beta^*_{11}}, \xi\sqrt{v\beta^*_{11}}\right], \\ -\xi\sqrt{v\beta^*_{11}}, & \text{when } \overline{\beta}_{12} + \frac{1}{4}\partial^2_{zv}\varphi < -\xi\sqrt{v\beta^*_{11}}, \\ \xi\sqrt{v\beta^*_{11}}, & \text{when } \overline{\beta}_{12} + \frac{1}{4}\partial^2_{zv}\varphi > \xi\sqrt{v\beta^*_{11}}. \end{cases} \tag{3.5.9}$$

Then β^* *is a positive semi-definite matrix and whenever*

$$-\xi\sqrt{v\beta^*_{11}} < \beta^*_{12} < \xi\sqrt{v\beta^*_{11}} \tag{3.5.10}$$

then β^* *is the optimizer in* (3.5.7).

We simulate the call option prices from the generating model and test it against two reference models: a "good" reference model and a "bad" reference model. The parameters used in Γ_{LSV} are those of the reference model.

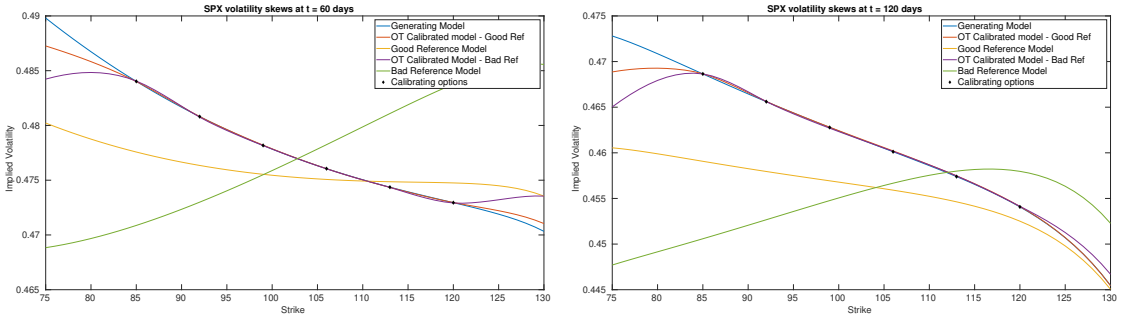
Heston Model		
Parameter	Value	Interpretation
X_0^1	log(92)	Initial log-stock price
X_0^2	0.25	Initial volatility
ε_1	1×10^{-4}	Tolerance for the difference in scaled model and market implied volatility
ε_2	1×10^{-12}	Tolerance for the policy iteration approximation of the optimal characteristics
κ	1	Speed of volatility mean reversion in the generating model
θ	0.05	Long-term mean of the volatility in the generating model
ξ	0.2	Volatility scaling of the volatility in the generating model
ρ	-0.4	Instantaneous correlation between the log-stock and volatility in the generating model
$\overline{\kappa}_{\text{good}}$	1.5	Speed of volatility mean reversion in the reference model
$\overline{\theta}_{\text{good}}$	0.07	Long-term mean of the volatility in the reference model
$\overline{\xi}_{\text{good}}$	0.15	Volatility scaling of the volatility in the reference model
$\overline{\rho}_{\text{good}}$	-0.2	Instantaneous correlation between the log-stock and volatility in the reference model
$\overline{\kappa}_{\text{bad}}$	2	Speed of volatility mean reversion in the reference model
$\overline{\theta}_{\text{bad}}$	0.09	Long-term mean of the volatility in the reference model
$\overline{\xi}_{\text{bad}}$	0.3	Volatility scaling of the volatility in the reference model
$\overline{\rho}_{\text{bad}}$	0.2	Instantaneous correlation between the log-stock and volatility in the reference model

Table 3.11: Parameter values of generating and reference models used in the LSV calibration.

		Generating Model		Calibrated Model: Good Reference		Calibrated Model: Bad Reference	
Option Type	Strike	Price	IV	Price	IV	Price	IV
SPX call options $t = 60$ days	85	11.0144	0.4840	11.0143	0.4840	11.0148	0.4841
	92	7.1928	0.4808	7.1928	0.4808	7.1920	0.4808
	99	4.4416	0.4782	4.4419	0.4782	4.4414	0.4782
	106	2.6037	0.4761	2.6036	0.4761	2.6034	0.4760
	113	1.4564	0.4744	1.4562	0.4743	1.4562	0.4743
	120	0.7813	0.4729	0.7814	0.4730	0.7811	0.4729
SPX call options $t = 120$ days	85	13.4256	0.4686	13.4260	0.4687	13.4256	0.4686
	92	9.8367	0.4656	9.8372	0.4656	9.8370	0.4656
	99	7.0267	0.4628	7.0273	0.4628	7.0250	0.4627
	106	4.9025	0.4601	4.9043	0.4602	4.9034	0.4602
	113	3.3437	0.4574	3.3450	0.4575	3.3451	0.4575
	120	2.2243	0.4541	2.2249	0.4541	2.2252	0.4541

Table 3.12: Table of the generating and calibrated model prices and implied volatilities.

3.5.1 Implied volatility Plots



(a) (b)

Figure 3.42: Implied volatility plots for the generating model, both SOT calibrated models and both reference models.

We remark that this approach can recover the implied volatility within the range of the strikes even when the reference model has completely the wrong shape of implied volatility. We now display the surfaces of SOT calibrated β_{11} and ρ. We notice that while the generating and reference models in β_{11} are the same, our calibrated model is perturbed from the reference. We also notice that our calibrated correlation ρ is close to the reference model correlation parameter in both cases, with deviations towards the generating model when we are in the range of our strikes and when V_t is near zero. This strong dependence is also seen in our previous calibration approaches, and thus a good a priori estimate of ρ would be needed in practice.

3.5.2 Plots of the local-stochastic volatility and correlation

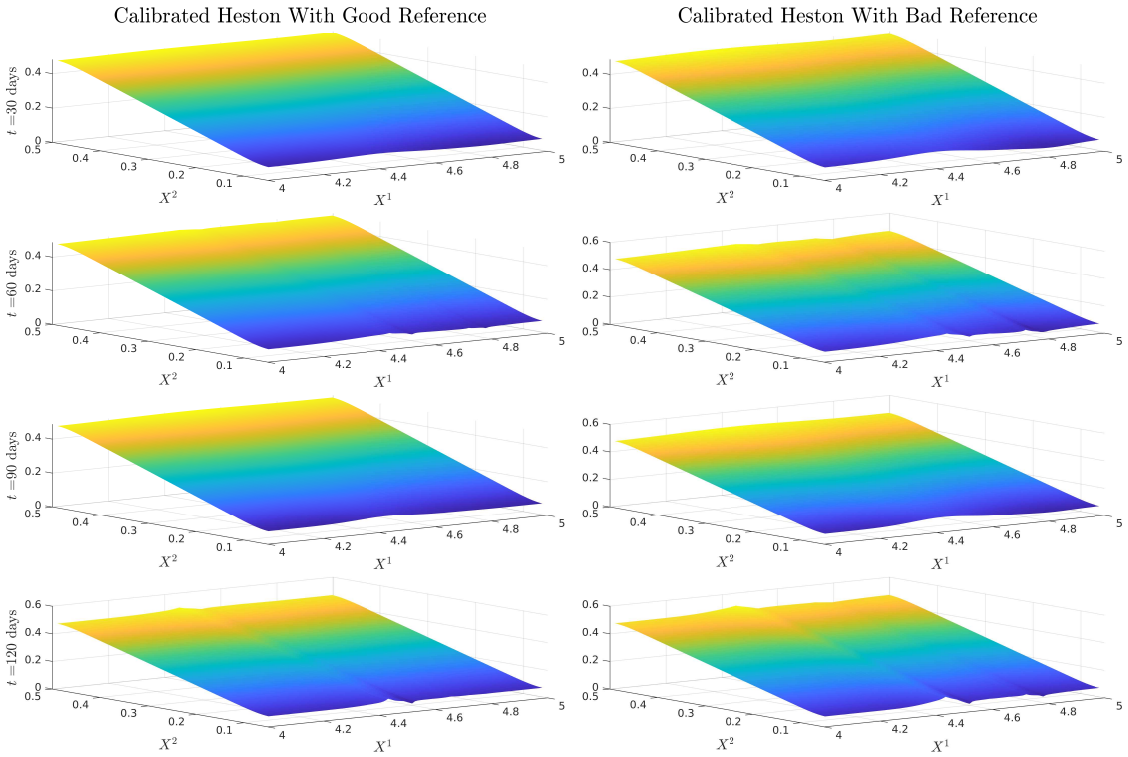


Figure 3.43: Plots of SOT calibrated β_{11} under good and bad reference models. Note that the generating and reference values for β_{11} are given by X^2.

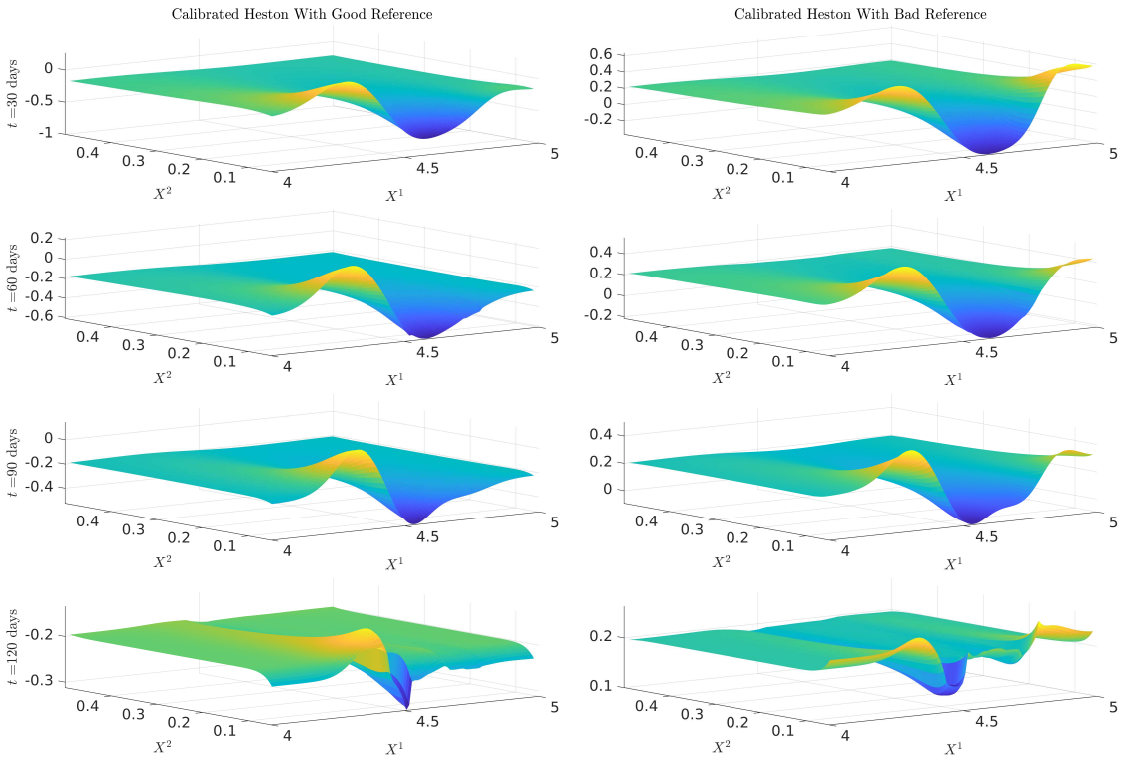


Figure 3.44: Plots of SOT calibrated ρ under good and bad reference models. Note that $\rho = -0.4$ in the generating model, $\overline{\rho}_{\text{good}} = -0.2$ and $\overline{\rho}_{\text{bad}} = 0.2$.

3.6 Joint Calibration Method Only With Call Options Over a Long Time Horizon

Our calibrated surfaces thus far have in general depended on the X^1 variable when only calibrating to call options. Since the interest rates are not identically zero, any product must be jointly dependent on an underlying asset and the interest rate through the stochastic discount factor, and therefore we should expect to see more interest rate dependence as the time horizon increases. Now, we take a time horizon of 5 years, and consider the joint calibration problem on simulated data with maturities at $t = 1, 2, 3, 4, 5$ years. There are far more options to calibrate in this setup, so we instead take the stopping tolerance to be 1×10^{-3} in the rescaled model prices. We also use the faster approach with bounds described in Section 3.3.4 since there are far more options to calibrate to. The parameters used are the same as in Table 3.8 but with $\overline{a} = 0.03$, $\overline{\sigma}_r = 0.02$, and the bounds $\delta_{22}^l = 5 \times 10^{-5}$ and $\delta_{22}^u = 1 \times 10^{-3}$.

3.6.1 Implied Volatility Plots

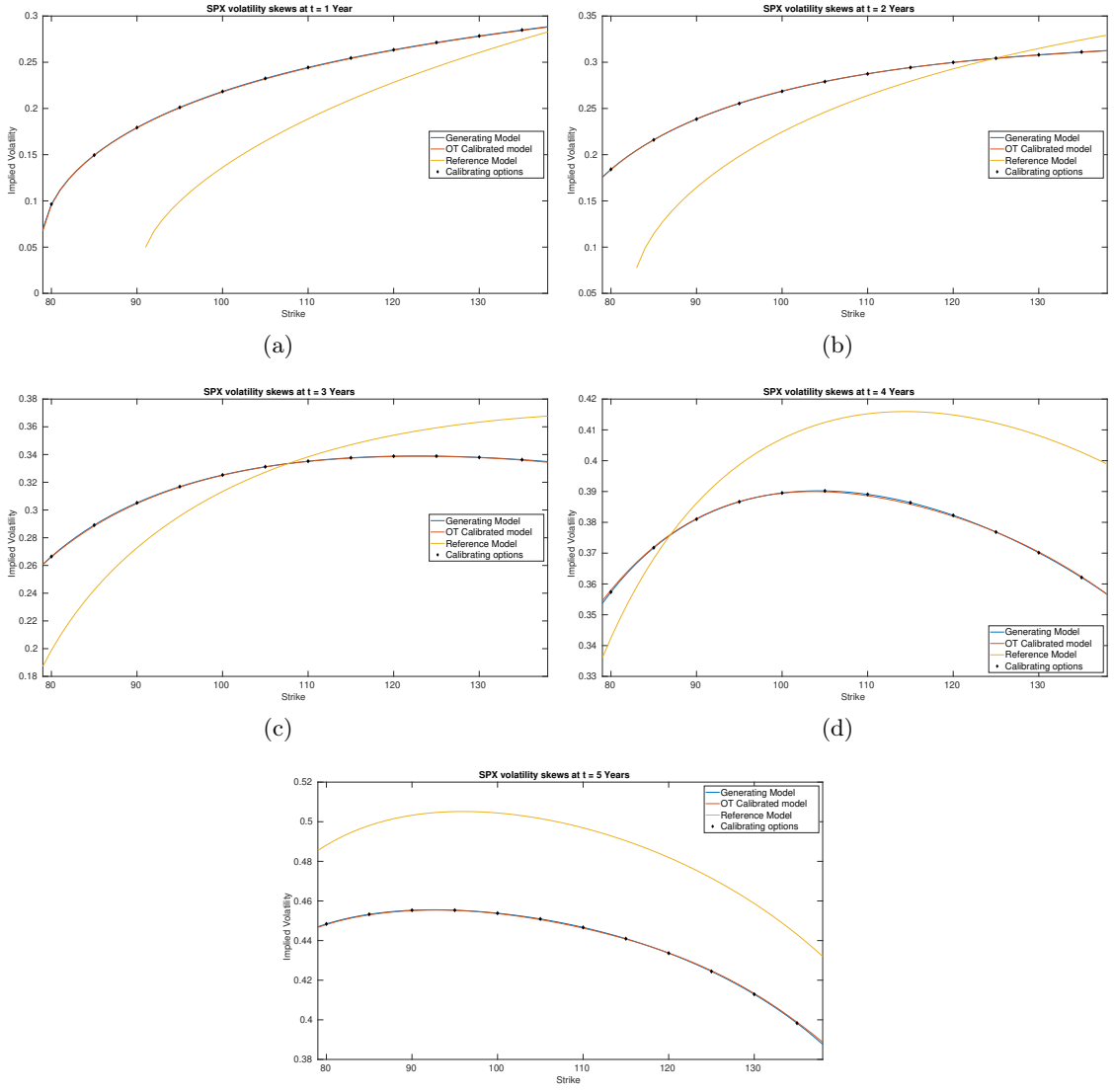


Figure 3.45: Implied volatility skews of the SPX calibrating options under the reference model and calibrated HW-CEV model.

3.6.2 Plots of Drift and Diffusion Surfaces

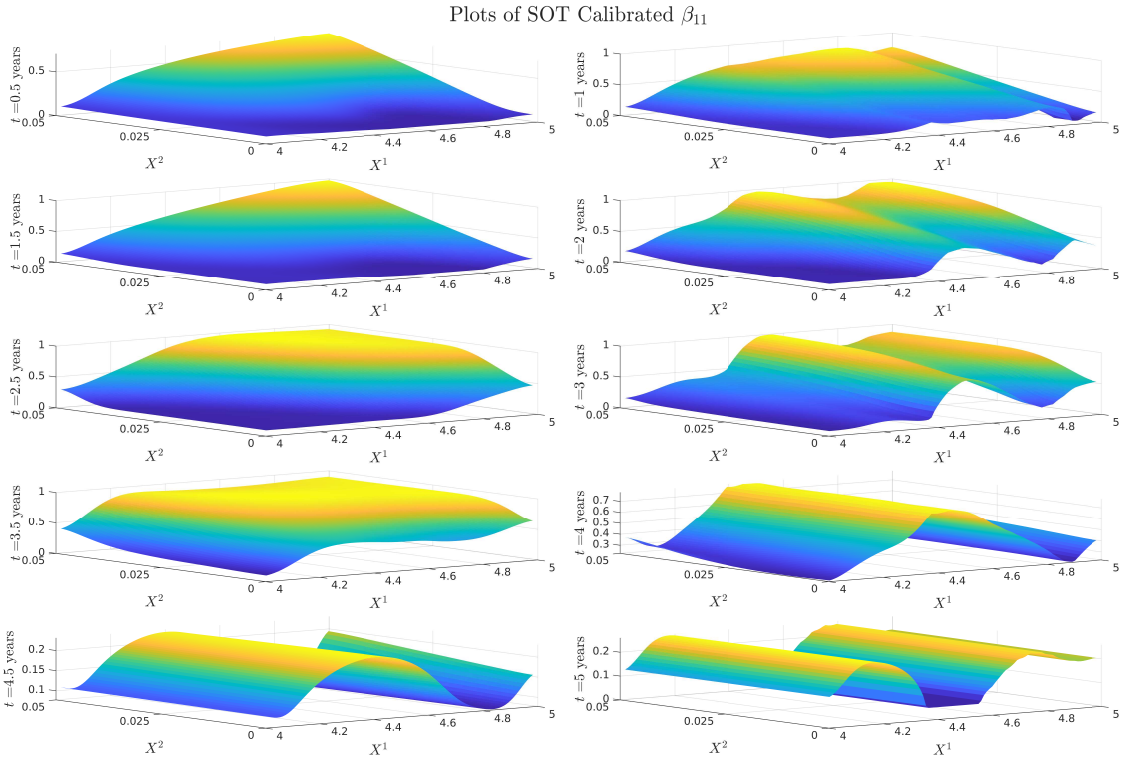
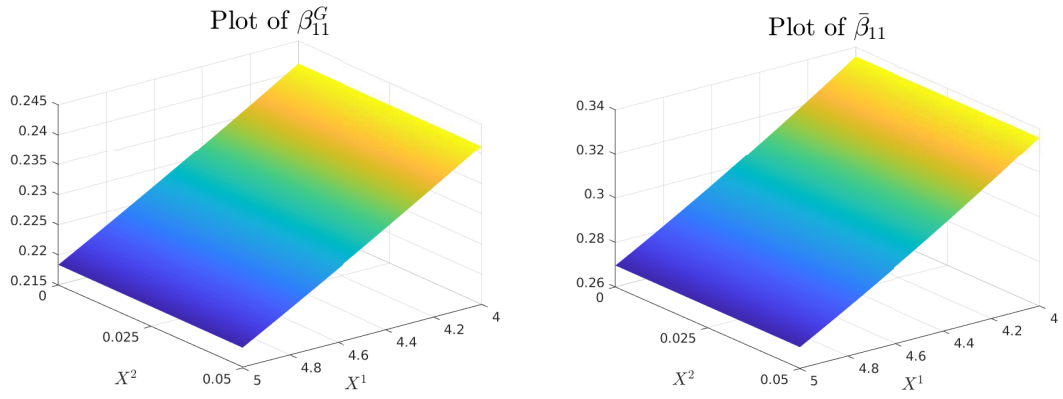


Figure 3.46: Plots of SOT-calibrated β_{11}, the volatility of the log-stock.



(a) (b)

Figure 3.47: Plots of the generating model log-stock volatility β_{11}^G (a), and reference model log-stock volatility $\overline{\beta}_{11}$ (b).

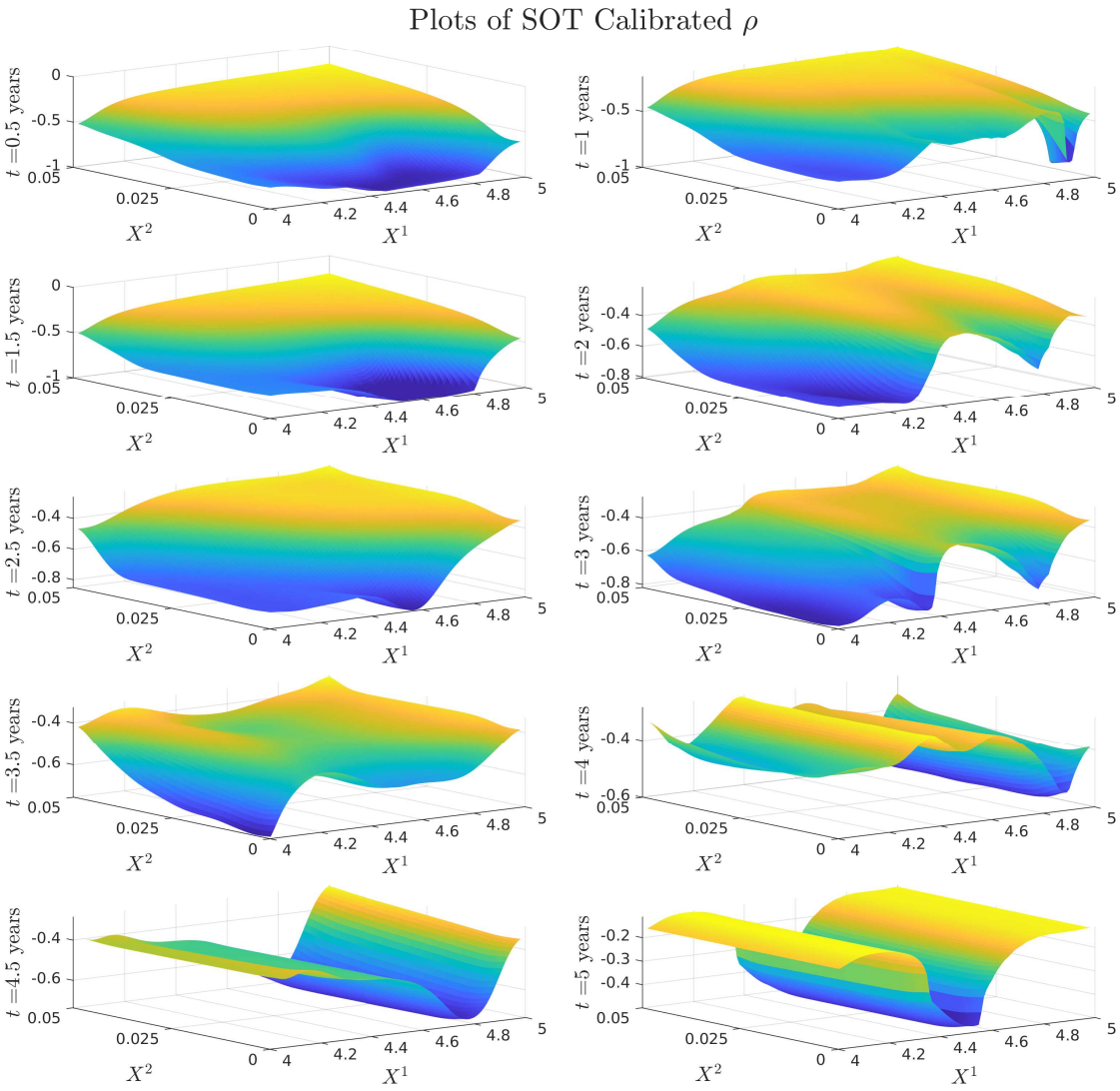


Figure 3.48: Plots of SOT calibrated correlation $\rho = \frac{\beta_{12}}{\sqrt{\beta_{11}\beta_{22}}}$. The β_{12} and β_{22} used here have the scaling undone in order to obtain the actual correlation between $\log(S_t)$ and r_t. The generating parameter used was $\rho = -0.4$ and the reference parameter was $\overline{\rho} = -0.2$.

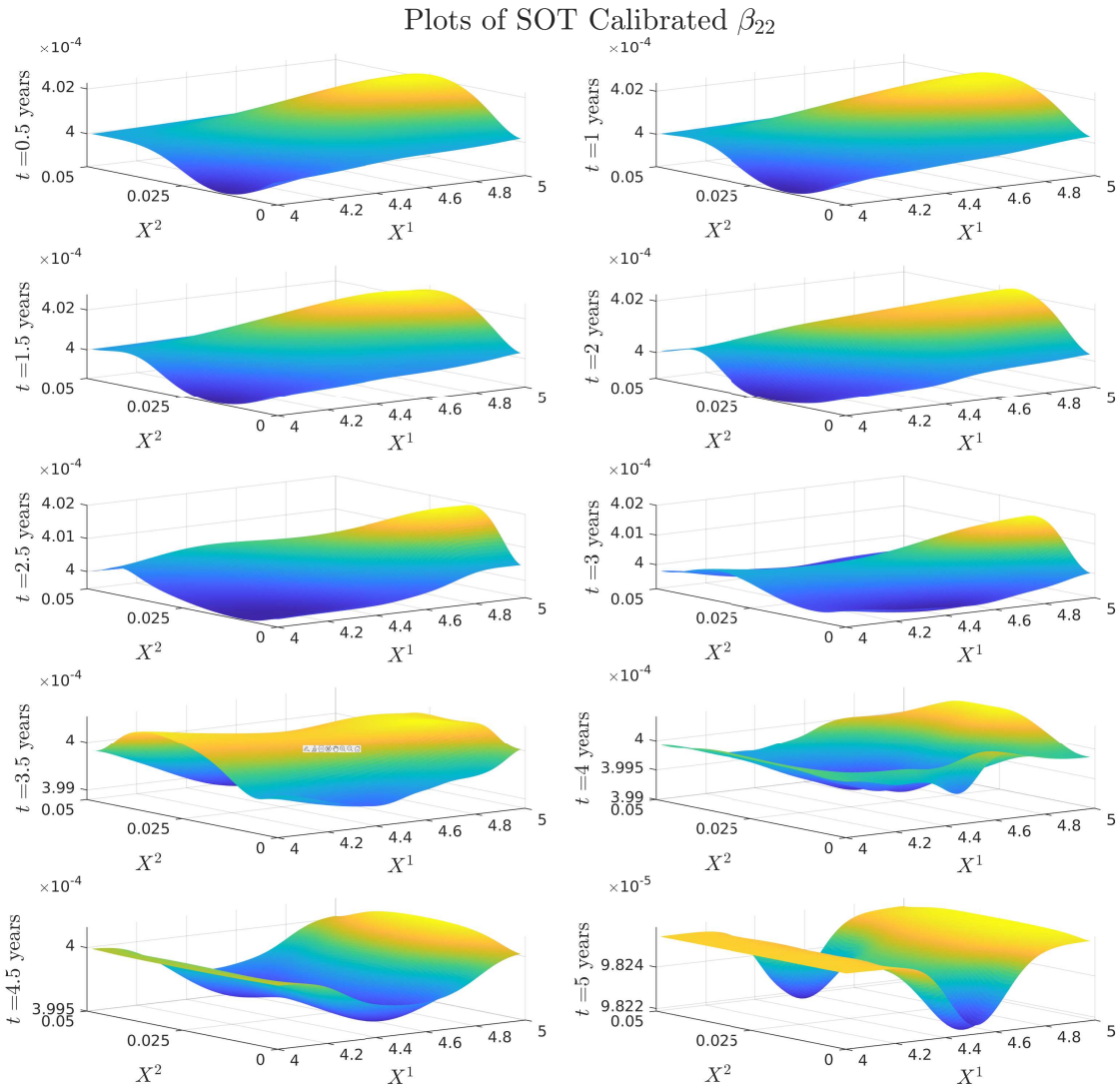


Figure 3.49: Plots of SOT calibrated β_{22}, the volatility of the short rate. Note that the β_{22} used have the scaling undone to represent the actual volatility of the short rate. The generating parameter used was $\beta_{22}^G = 1.6 \times 10^{-3}$ and the reference parameter was $\overline{\beta}_{22} = 4 \times 10^{-4}$.

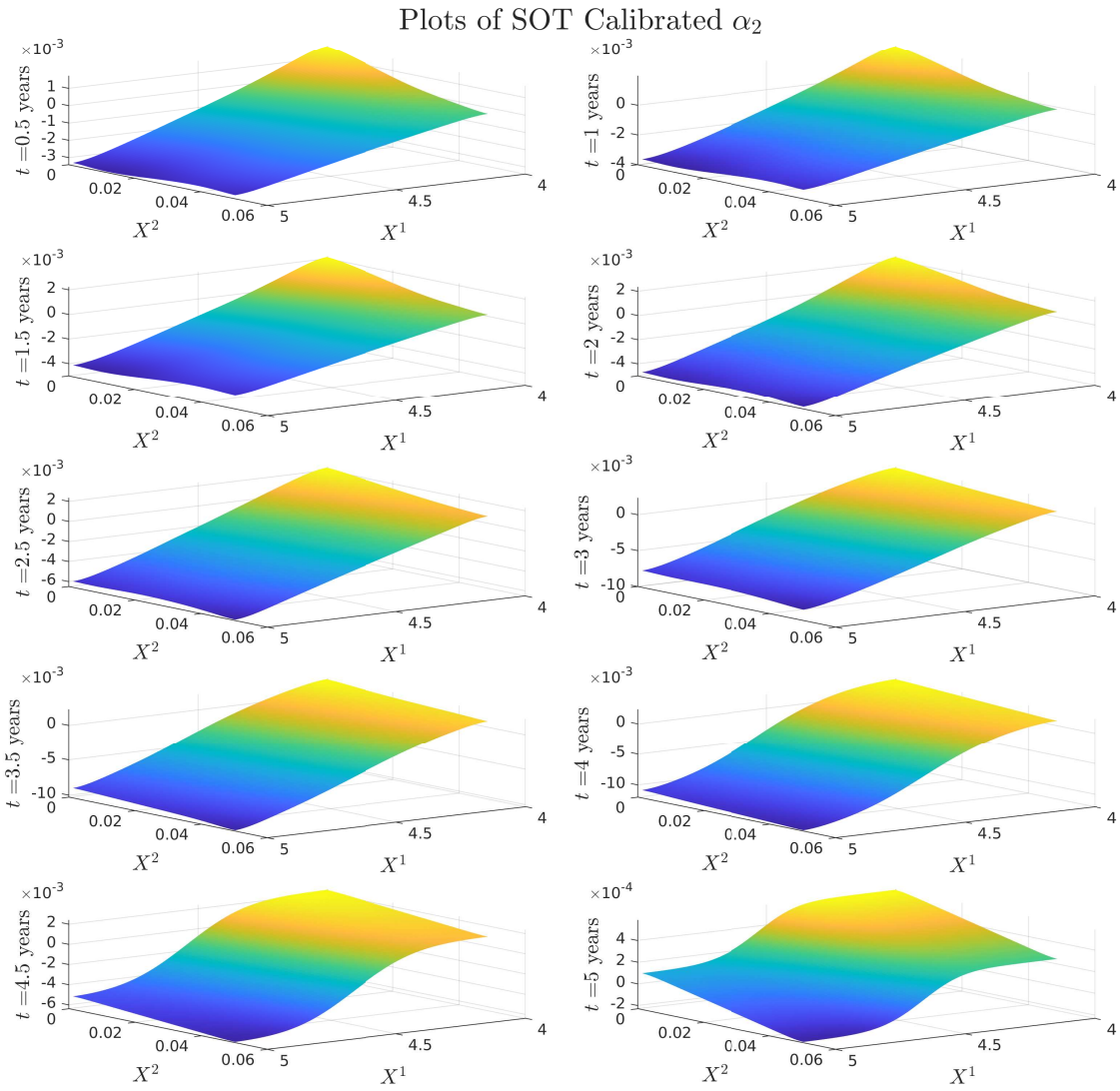


Figure 3.50: Plots of SOT calibrated α_2. Note that the α_2 used have the scaling undone to represent the actual drift of the short rate.

3.6.3 Discussion of the results

We can see that we get an exact fit of the call options in Figure 3.45 across all five maturities. The surfaces in Section 3.6.2 clearly no longer have the homogeneity seen in the previous experiments. It is clear from all of the surfaces that this inhomogeneity propogates through time, thus with a longer time horizon, one can obtain jointly dependent volatilities for the equity and the rate. From Figures 3.46-3.50, we see dependence on both the short rate and the log-stock at most of the displayed times. Figure 3.48 also displays much less of a dependence on the reference model, with much larger perturbations downwards towards

the generating model. Figures 3.49-3.50 demonstrate much larger perturbations on the calibrated surfaces than seen in Section 3.4 and Section 3.4.4 where the time horizon was much shorter.

Chapter 4

Measure Preserving Martingale Sinkhorn Algorithm

We are interested in describing Markov martingales with prescribed marginal distributions. This problem is of fundamental importance in mathematical finance where it corresponds to fitting models to market data. The best known solution, which had a tremendous impact on the financial industry, was given by Dupire, 1994. However, this requires knowing marginal distributions at all times requiring elaborate interpolation techniques and suffering from serious numerical stability challenges, see Bain, Mariapragassam, and Reisinger, 2021. Instead, we want to build a diffusion process with prescribed marginal distributions at finitely many times. This problem is also intimately related to the classical Skorokhod embedding problem (SEP), see Obłój, 2004, and has recently been re-cast using techniques of optimal transport theory.

In this Chapter, we focus on the martingale Benamou-Brenier (mBB) problem, i.e., among martingales with prescribed initial and terminal distributions we want to select the one which is closest to Brownian motion, or the constant volatility martingale. This problem was introduced and studied by Backhoff-Veraguas, Beiglböck, Huesmann, et al., 2020 and its solution is given by the so-called Bass martingale, going back to Bass, 1983 and his solution to the SEP. It is also known as the *stretched Brownian motion (sBM)*, and has recently been studied by Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023; Backhoff-Veraguas,

Schachermayer, and Tschiderer, 2023. Importantly, Conze and Henry-Labordere, 2021 provided a fixed-point-like iteration scheme for computing the sBM. Its convergence is a topic of a very recent study in Acciaio, Marini, and Pammer, 2023.

Our main contribution is two-fold. First, we link the mBB problem with the literature on semimartingale optimal transport problems. We use the duality results in these works to represent sBM. This mirrors closely the duality results presented in Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023 but provides a different point of view. In particular, if we want to generalise the setup to the case where full marginals are not fixed but only constrained (e.g., by a finite number of option prices) this approach is likely to offer more flexibility. Second, we provide a numerical scheme to compute the sBM which mirrors the famous Sinkhorn algorithm used to solve the entropic optimal transport problem, or the Schrödinger problem. Our scheme, dubbed the *measure preserving martingale Sinkhorn* (MPMS) offers a new vantage point on the iterative scheme of Conze and Henry-Labordere, 2021. The MPMS iterations we derive are also reminiscent of the "back-and-forth method" of Jacobs and Léger, 2020, which provides a fast iterative method to solve a standard optimal transport problem. Our MPMS scheme displays fast convergence in practice and it offers promising approaches to the higher dimensional setups. This chapter is published in Joseph, Loeper, and Obłój, 2023c.

4.1 Martingale Benamou-Brenier problem and its duality

We denote $\mathcal{P}_2(\mathbb{R})$ the space of probability measures on $\mathbb{R}$ with a finite second moment. We work with probability measures on $\mathbb{R}$ – this both simplifies the presentation and is the case of interest for applications in mathematical finance. However, our arguments readily generalise to the case of probability measures on $\mathbb{R}^d$. Throughout, we fix $\mu_0, \mu_T \in \mathcal{P}_2(\mathbb{R})$ which admit a density on $\mathbb{R}$. It will be convenient to identify measures with their densities, so $\mu_0(x)$ will denote the density of μ_0. It will be clear from the context if we work with a measure or with its density, a function. We use the shorthand notation for $\nu \in \mathcal{P}_2(\mathbb{R})$ that $\mathrm{d}\nu = \nu(\mathrm{d}x)$. We

also assume that μ_0 and μ_T are in strict convex order, i.e., $\int x \mathrm{d}\mu_0 = \int x \mathrm{d}\mu_T$ and

$$\int (x-K)^+ \,\mathrm{d}\mu_0 < \int (x-K)^+ \,\mathrm{d}\mu_T, \tag{4.1.1}$$

for all K in the interior of the convex hull of the support of μ_T.[1]

We fix a reference volatility value $\bar{\sigma} > 0$ and consider the following martingale Benamou-Brenier (mBB) problem

$$\mathbf{MT}_{\mu_0,\mu_T} = \inf_{\substack{M_0\sim\mu_0, M_T\sim\mu_T \\ M_t = M_0 + \int_0^t \sigma_s \mathrm{d}B_s}} \mathbb{E}\left[\int_0^T (\sigma_t - \bar{\sigma})^2 \mathrm{d}t\right], \tag{mBB}$$

where the optimisation is taken over filtered probability spaces with a Brownian motion $(B_t)_{t\geqslant 0}$. This problem is equivalent to the following problem

$$\mathbf{P}_{\mu_0,\mu_T} = \sup_{\substack{M_0\sim\mu_0, M_T\sim\mu_T \\ M_t = M_0 + \int_0^t \sigma_s \mathrm{d}B_s}} \mathbb{E}\left[\int_0^T |\sigma_t| \mathrm{d}t\right] = \sup_{\substack{M_0\sim\mu_0, M_T\sim\mu_T \\ M_t = M_0 + \int_0^t \sigma_s \mathrm{d}B_s}} \mathbb{E}\left[M_T B_T\right], \tag{4.1.2}$$

in the sense that the two problems share the optimiser and $\mathbf{MT}_{\mu_0,\mu_T} = T\bar{\sigma}^2 + \int x^2 \mathrm{d}\mu_T - \int x^2 \mathrm{d}\mu_0 - 2\bar{\sigma}\mathbf{P}_{\mu_0,\mu_T}$. These problems were studied in detail by Backhoff-Veraguas, Beiglböck, Huesmann, et al., 2020, with further results relevant for the higher-dimensional setup in the recent works of Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023; Backhoff-Veraguas, Schachermayer, and Tschiderer, 2023. In particular, Backhoff-Veraguas, Beiglböck, Huesmann, et al., 2020 show that (mBB) admits a unique optimiser (in distribution) which has a particular representation and is known as the *standard stretched Brownian motion* from μ_0 to μ_T, or the Bass martingale. To describe this process we need to introduce some further notation.

For $\alpha_0 \in \mathcal{P}_2(\mathbb{R})$, we denote $B^\alpha = (B^\alpha_t)_{t\geqslant 0}$ a Brownian motion with a non-trivial starting law $B^\alpha_0 \sim \alpha_0$. We write $B^{\delta_0} = B$, as usual, and for a given B^α let $B_t = B^\alpha_t - B^\alpha_0$ be the associated standard Brownian motion. The heat kernel is denoted

$$\mathbf{R}_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}, \tag{4.1.3}$$

[1] This assumption, in the language of MOT, means that there is just one irreducible component.

and $*$ stands for convolution, so that $B_T^\alpha \sim \alpha_T = \alpha_0 * \mathbf{R}_T$. Finally, $\cdot_\#$ denotes the push-forward operator:

$$F_\# \mu_0 = \mu_T \Leftrightarrow \mu_T = \mu_0 \circ F^{-1} \Leftrightarrow \forall E,\ \mu_T(E) = \mu_0(F^{-1}(E)). \tag{4.1.4}$$

Theorem 4.1.1 (Backhoff-Veraguas, Beiglböck, Huesmann, et al., 2020)**.** *There exists a probability measure $\alpha_0 \in \mathcal{P}_2(\mathbb{R})$ and an increasing function $F : \mathbb{R} \to \mathbb{R}$ such that $F_\#(\alpha_0 * R_T) = \mu_T$. The problem* (mBB) *admits a unique optimiser given by*

$$M_t = \mathbb{E}[F(B_T^\alpha)|\mathcal{F}_t] = F(t, B_t^\alpha), \tag{4.1.5}$$

where

$$\begin{aligned} \partial_t F + \frac{1}{2}\partial_{xx} F = 0, \quad 0 < t < T, x \in \mathbb{R}, \\ F(T, x) = F(x). \end{aligned} \tag{4.1.6}$$

It follows that the optimiser solves

$$\mathrm{d}M_t = F_x(t, B_t^\alpha)\mathrm{d}B_t^\alpha = \sigma(t, M_t)\mathrm{d}B_t, \tag{4.1.7}$$

subject to $M_0 \sim \mu_0$ and $\sigma = F_x \circ F^{-1}$, where the inverse is always taken with respect to the spatial variable.

To gain a different perspective on this process, we note that a class of processes linked to (mBB) was studied via its dual and PDE methods by Huesmann and Trevisan, 2019. More generally, it is a special case of the semimartingale optimal transport studied in Xiaolu Tan and Touzi, 2013; I. Guo, Loeper, and Shiyi Wang, 2019; I. Guo and Loeper, 2021. In its general Markovian formulation, one considers

$$\mathbf{MT}^H_{\mu_0,\mu_T} := \inf_{\substack{M_0 \sim \mu_0, M_T \sim \mu_T \\ M_t = M_0 + \int_0^t \sigma(s,M_s)\mathrm{d}B_s}} \mathbb{E}\left[\int_0^T H(\sigma(t, M_t)^2)\mathrm{d}t\right], \tag{mBB(H))}$$

for a convex function H satisfying suitable regularity assumptions. Then, following the

general result stated in I. Guo and Loeper, 2021, we know that duality holds and the minimizer in (mBB(H)) can be obtained by solving the dual problem

$$\mathbf{MT}^{H}_{\mu_0,\mu_T} = \sup_{\varphi} \left\{ \int_{\mathbb{R}} \varphi(T,x)\,\mathrm{d}\mu_T - \int_{\mathbb{R}} \varphi(0,x)\,\mathrm{d}\mu_0 \right\}, \tag{4.1.8}$$

where the supremum is taken over all super-solutions φ of

$$\partial_t \varphi + H^* \left(\frac{1}{2} \partial_{xx} \varphi \right) \leqslant 0, \tag{4.1.9}$$

with $H^*(p) = \sup_{\sigma\in\mathbb{R}}\{p\sigma - H(\sigma)\}$ the Legendre-Fenchel transform of H. And further, if the supremum is attained by φ then the optimiser is given via

$$\sigma^2(t,x) = H^*_p \left(\frac{1}{2} \partial_{xx} \varphi(t,x) \right). \tag{4.1.10}$$

Note that for the cost function

$$H(\beta) = \frac{1}{2} \left(\sqrt{\beta} - \bar{\sigma} \right)^2, \quad \text{we have } H^*(p) = \bar{\sigma}^2 \frac{p}{1-2p}, \tag{4.1.11}$$

and therefore the HJB equation becomes

$$\partial_t \varphi + \frac{1}{2} \frac{\bar{\sigma}^2 \partial_{xx}\varphi}{1 - \partial_{xx}\varphi} = 0, \tag{4.1.12}$$

which is well studied in its log-normal form in Loeper, 2018, see also Bouchard, Loeper, and Zou, 2016; Bouchard, Loeper, and Zou, 2017; Bouchard, Loeper, Soner, et al., 2019. Assuming that φ is optimal for (4.1.8), then the optimal diffusion in (mBB) is given by

$$\sigma = \frac{\bar{\sigma}}{1 - \partial_{xx}\varphi}. \tag{4.1.13}$$

Note that for φ to solve the HJB equation, we have that $\partial_{xx}\varphi < 1$. This means that the potential $v = \frac{x^2}{2} - \varphi$ is convex. It further satisfies

$$\partial_t v + \frac{1}{2} \bar{\sigma}^2 \left(1 - \frac{1}{\partial_{xx} v} \right) = 0. \tag{4.1.14}$$

Note that the derivative of v, $\xi = \partial_x v$ follows the *linearised* equation

$$\partial_t \xi + \frac{\sigma^2}{2}\partial_{xx}\xi = 0, \quad \text{with } \sigma = \frac{\bar{\sigma}}{\partial_{xx} v} = \frac{\bar{\sigma}}{1 - \partial_{xx}\varphi}. \tag{4.1.15}$$

It follows that if we let M^σ solve $\mathrm{d}M_t^\sigma = \sigma(t, M_t^\sigma)\mathrm{d}B_t$, with σ in (4.1.13), then $Z_t := \xi(t, M_t^\sigma)$ is a martingale. Moreover, we have

$$\mathrm{d}\langle Z\rangle_t = (\partial_x \xi(t, M_t^\sigma))^2 \sigma^2(t, M_t^\sigma)\mathrm{d}t = \bar{\sigma}^2 \mathrm{d}t, \tag{4.1.16}$$

and therefore $\frac{1}{\bar{\sigma}}Z$ is a Brownian motion. Observe that the inverse map of ξ is $\partial_y v^*$. If we put $F(t,y) = \partial_y v^*(t, \bar{\sigma} y)$, then $M_t^\sigma = F(t, \frac{1}{\bar{\sigma}} Z_t)$. Further, from (4.1.14), we see that the Legendre-Fenchel transform v^* satisfies

$$\partial_t v^* + \frac{1}{2}\bar{\sigma}^2(\partial_{yy} v^* - 1) = 0, \tag{4.1.17}$$

which in turn readily implies that F solves the heat equation, which it has to by the martingale property of M^σ, see (4.1.6) above. Note that these functional relations follow simply from φ solving the HJB equation (4.1.12). The marginals are enforced via optimisation over such φ in (4.1.8), which gives $M_0^\sigma \sim \mu_0$ and $M_T^\sigma \sim \mu_T$ at the optimiser. However, since F solves the heat equation, we have $F(0,\cdot) = F(T,\cdot) * \mathbf{R}_T$. Comparing with (4.1.17) we see that

$$v^*(0,\cdot) = \left(v^*(T,\cdot) - \frac{\bar{\sigma}^2 T}{2}\right) * \mathbf{R}_{T\bar{\sigma}}, \tag{4.1.18}$$

$$v(0,\cdot) = (v^*(T, \bar{\sigma}\cdot) * \mathbf{R}_{T\bar{\sigma}})^* + \frac{\bar{\sigma}^2 T}{2}. \tag{4.1.19}$$

We could thus replace the condition that φ solves the HJB equation by the above relation between v, or φ, at times 0 and T. Recall that with H in (4.1.11) we have

$$\int_{\mathbb{R}} \varphi(T,x)\,\mathrm{d}\mu_T - \int_{\mathbb{R}} \varphi(0,x)\,\mathrm{d}\mu_0 = \mathbf{MT}^H_{\mu_0,\mu_T} = \frac{1}{2}\mathbf{MT}_{\mu_0,\mu_T} = \frac{\bar{\sigma}^2 T}{2} + \int_{\mathbb{R}} \frac{x^2}{2}(\mathrm{d}\mu_T - \mathrm{d}\mu_0) - \bar{\sigma}\mathbf{P}_{\mu_0,\mu_T}. \tag{4.1.20}$$

Letting $v(T, \bar{\sigma}^{-1}x) = \psi(x)$, we see that (4.1.8) can be equivalently written as

$$\mathbf{P}_{\mu_0,\mu_T} = \inf_{\psi} \left\{ \int_{\mathbb{R}} \psi \mathrm{d}\mu_T - \int_{\mathbb{R}} (\psi^* * \mathbf{R}_{T\bar{\sigma}})^* \mathrm{d}\mu_0 \right\}, \tag{4.1.21}$$

which in the special case $\bar{\sigma} = 1$ recovers the duality recently obtained by Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023. The densities $\mu(t, x)$ of the marginal distributions M_t^σ satisfy the Fokker-Planck equation

$$\partial_t \mu = \frac{1}{2} \partial_{xx} \left(\sigma^2 \mu \right), \tag{4.1.22}$$

which implicitly encodes the compatibility condition since both $\mu(0, \cdot) = \mu_0(\cdot)$ and $\mu(T, \cdot) = \mu_T(\cdot)$ are fixed. In particular, the measure α_0 and the function F from Theorem 4.1.1, are recovered by taking $\bar{\sigma} = 1$ and

$$\alpha_0(\cdot) = \partial_x v(0, \cdot)_{\#}\mu_0 \quad \text{and} \quad F(y) = F(T, y) = \partial_y v^*(T, y). \tag{4.1.23}$$

As an aside, we provide some concrete examples of this construction in the cases of Brownian and Geometric Brownian Motions with $\overline{\sigma} = 1$ for simplicity.

Example 4.1.2 (Brownian Motion)**.** *Let $T = 1$, $\mu_0 \sim \delta_0$ and $\mu_1 \sim N(0, 1)$. We assume that our process M is given by a Brownian motion, so that $\sigma(t, M_t) \equiv 1$. Then, from* (4.1.15) *we have $\partial_{xx}^2 v = 1$, and a solution of* (4.1.14) *is $v(t, x) = \frac{1}{2}x^2$. The maximiser in the Legendre transform is given by $\partial_y v^*(t, y) = y$, so that $F(t, y) = y$. With B being a standard Brownian motion started from 0, we observe that $F(t, B_t)$ is a Brownian motion for $t \in [0, 1]$ with $F(0, 0) \sim \delta_0$, $F(1, B_1) \sim N(0, 1)$, and F satisfying the heat equation as required.*

Example 4.1.3 (Geometric Brownian Motion)**.** *Let $T = 1$, $\mu_0 \sim \delta_1$ and $\mu_1 \sim$* Lognormal$(0, 1)$*. Assume that the process M is given by a Geometric Brownian, so that $\sigma(t, M_t) = M_t$. Then, from* (4.1.15) *we have $\partial_{xx}^2 v = \frac{1}{x}$, and a solution of* (4.1.14) *is given by $v(t, x) = x \log |x| + x\left(\frac{t}{2} - 1\right) - \frac{t}{2}$. Since $(\partial_x v)^{-1} = \partial_y v^*$, we have $\partial_y v^*(t, y) = \exp\left(y - \frac{t}{2}\right)$, so $F(t, y) = \exp\left(y - \frac{t}{2}\right)$. With B being a standard Brownian motion, we have $F(t, B_t)$ is a Geometric Brownian motion for $t \in [0, 1]$ with $F(0, 0) \sim \delta_1$, $F(1, B_1) \sim$* Lognormal$(0, 1)$*,*

and F satisfying the heat equation as required.

We now explore in more detail these relations and compare with those known for the Sinkhorn system.

4.2 Measure Preserving Martingale Sinkhorn's system (MPMS)

We fix $\bar{\sigma} = 1 = T$ for simplicity of notation. As before, we write $\mu_0 = \mu$ and $\mu_1 = \nu$. We start by recalling classical results in entropic optimal transport (EOT).

4.2.1 Sinkhorn's system

The Schrödinger Problem, a.k.a. EOT problem is stated as finding

$$\inf_{\pi \in \Pi(\mu_0, \mu_1)} \mathbf{KL}(\pi, \mathbf{P}_{\text{ref}}), \tag{4.2.1}$$

where **KL** is the Kullback-Leibler divergence (see Kullback and Leibler, 1951), and we can understand the above either as a problem on the pathspace or equivalently as a problem on $\mathbb{R}^2$. In the first case, $\Pi(\mu_0, \mu_1)$ denotes measures on $C([0,1]; \mathbb{R})$ with marginals μ_0, μ_1 at times $0, 1$ respectively, and $\mathbf{P}_{\text{ref}}$ is the Wiener measure; in the second case, $\Pi(\mu_0, \mu_1)$ denotes probabilty measure on $\mathbb{R} \times \mathbb{R}$ with given marginals μ_0, μ_1, and $\mathbf{P}_{\text{ref}}$ is the standard Gaussian measure on $\mathbb{R} \times \mathbb{R}$. It is well known that there exist f_0, g_1 such that:

$$\mu_0 = (g_1 * \mathbf{R}_1) f_0, \tag{4.2.2}$$

$$\mu_1 = g_1 (\mathbf{R}_1 * f_0), \tag{4.2.3}$$

$$\mu_t = (g_1 * \mathbf{R}_{1-t})(f_0 * \mathbf{R}_t), \tag{4.2.4}$$

The system (4.2.2)-(4.2.3) is known as the *Sinkhorn's system*. It induces an interpolation between μ_0 and μ_1 given by (4.2.4). We refer the reader to Léonard, 2014; Nutz, 2022 for great surveys of Entropic Optimal Transport and the Sinkhorn's system. In a nutshell, $\mathbf{R}_1(x, y) f_0(x) g_1(y)$ is the probability measure on $\mathbb{R} \times \mathbb{R}$ that solves the Schrödinger's prob-

lem. It induces a probability measure on the canonical space $C([0,1];\mathbb{R})$ that is absolutely continuous with respect to the Wiener measure. The *Sinkhorn algorithm* consists in solving iteratively

$$\text{Update } f_0: \qquad \mu_0 = (g_1^n * \mathbf{R}) f_0^{n+1}, \tag{4.2.5}$$

$$\text{Update } g_1: \qquad \mu_1 = g_1^{n+1} (\mathbf{R} * f_0^{n+1}). \tag{4.2.6}$$

Algorithm 4: Sinkhorn algorithm

Input: μ_0, μ_1
Output: f_0, g_1

```
Set g_1^0 = 1
for n = 1 to max_iter do
    Solve backward heat equation for g^n        ∂_t g^n + ½ ∂²_xx g^n = 0, g^n(1) = g_1^n
    Update f_0^n such that f_0^n g^n(0) = μ_0
    Solve forward heat equation for f^n         ∂_t f^n − ½ ∂²_xx f^n, f^n(0) = f_0^n
    Update g_1^{n+1} such that g_1^{n+1} f^n(1) = μ_1
end
return f_0, g_1
```

It can be seen as iterative renormalizations of the kernel $\mathbf{R}_1$ to have the proper marginals μ_0, μ_1. It is shown in Sinkhorn, 1964 that this algorithm converges. Moreover, it yields a continuous dynamic interpolation between measures through the system

$$\partial_t \mu + \nabla_x \cdot (\mu \nabla \varphi) = \frac{1}{2} \Delta \mu, \tag{4.2.7}$$

$$\mu(0) = \mu_0, \tag{4.2.8}$$

$$\partial_t \varphi + \frac{1}{2} |\nabla \varphi|^2 + \frac{1}{2} \Delta \varphi = 0, \tag{4.2.9}$$

$$\varphi(1) = \log(g_1). \tag{4.2.10}$$

The stochastic process associated to this interpolation – with marginals given by μ_t – has unit diffusion and a variable drift.

4.2.2 MPMS or Sinkhorn for the Bass martingale

Inspired by the above classical results, we propose to re-interpret Theorem 4.1.1. Our proposed system is naturally in agreement with the fixed point problem addressed in Conze and Henry-Labordere, 2021, as we explain below. We work again in the setup of Section 4.1 and $\mu_t = \mu(t, \cdot)$ denote the marginals of the optimiser M^σ of (mBB). We write F_1 for the map F in Theorem 4.1.1. Theorem 4.1.1 can be summarised as follows:

$$\mu_0 = (\mathbf{R}_1 * F_1)_{\#}\alpha_0, \tag{4.2.11}$$

$$\mu_1 = F_{1\#}(\mathbf{R}_1 * \alpha_0), \tag{4.2.12}$$

$$\mu_t = (\mathbf{R}_{1-t} * F_1)_{\#}(\mathbf{R}_t * \alpha_0). \tag{4.2.13}$$

This offers a clear analogy to the Sinkhorn system (4.2.2)-(4.2.4) and suggests an iterative scheme to solve for the Bass martingale. The name "measure preserving martingale Sinkhorn" therefore arises since we build a martingale interpolation by iteratively finding the measure preserving maps (i.e. Brenier maps) solving Monge-Ampère, in particular the only difference between MPMS and classical Sinkhorn visually is the pushforward operator as opposed to left and right multiplication. This approach removes the implicit iteration enforcing the martingale constraint in the martingale Sinkhorn algorithm of De March, 2018 (see Algorithm 2). We start with $F_1^0(x) = x$ and

$$\text{Update } \alpha_0: \qquad \mu_0 = (F_1^n * \mathbf{R}_1)_{\#}\alpha_0^{n+1}, \tag{4.2.14}$$

$$\text{Update } F_1: \qquad \mu_1 = (F_1^{n+1})_{\#}(\mathbf{R}_1 * \alpha_0^{n+1}). \tag{4.2.15}$$

It is insightful to comment on the link between MPMS and the fixed-point iteration of Conze and Henry-Labordere, 2021. For a measure ν, let G_ν denote its cumulative distribution function. Recall that for two measures, ν_1 and ν_2, we have for $h = G_{\nu_1}^{-1} \circ G_{\nu_2}$ that $\nu_1 = h_{\#}\nu_2$. It follows that (4.2.11)-(4.2.12) can be re-written as

$$\mathbf{R}_1 * F_1 = G_{\mu_0}^{-1} \circ G_{\alpha_0} \quad \text{and} \quad F_1 = G_{\mu_1}^{-1} \circ G_{\mathbf{R}_1 * \alpha_0}. \tag{4.2.16}$$

Algorithm 5: Measure Preserving Martingale Sinkhorn algorithm

Input: μ_0, μ_1
Output: α_0, F_1
1 Set $F_1^0 = Id$
2 **for** $n = 0$ *to* `max_iter` **do**
3 | Solve backward heat equation for F^n: $\partial_t F^n + \frac{1}{2}\partial_{xx} F^n = 0$, $F^n(1) = F_1^n$
4 | Update α_0^{n+1} such that $F^n(0)_{\#}\alpha_0^{n+1} = \mu_0$
5 | Solve forward heat equation for α^{n+1}: $\partial_t \alpha^{n+1} - \frac{1}{2}\partial_{xx}\alpha^{n+1} = 0$, $\alpha^{n+1}(0) = \alpha_0^{n+1}$
6 | Update F_1^{n+1} such that $(F_1^{n+1})_{\#}\alpha^{n+1}(1) = \mu_1$
7 **end**
8 **return** α_0, F_1

Note also that $G_{\mathbf{R}_1 * \alpha_0} = \mathbf{R}_1 * G_{\alpha_0}$ so that we obtain

$$G_{\alpha_0} = G_{\mu_0} \circ (\mathbf{R}_1 * F_1) = G_{\mu_0} \circ \left(\mathbf{R}_1 * \left(G_{\mu_1}^{-1} \circ (\mathbf{R}_1 * G_{\alpha_0})\right)\right) \tag{4.2.17}$$

which is the fixed point relation for G_{α_0} in Conze and Henry-Labordere, 2021: in this sense the two algorithms are equivalent. Specifically, given a candidate α_0^n, then computing $G_{\mu_1}^{-1} \circ (\mathbf{R}_1 * G_{\alpha_0^n})$ corresponds to lines 5&6 in Algorithm 5, and then applying $G_{\mu_0} \circ (\mathbf{R}_1 * \cdot)$ to the output corresponds to lines 3&4 in Algorithm 5, and yields the updated α_0^{n+1}. One immediate advantage of the representation in Algorithm 5 is that it naturally extends to arbitrary dimensions. Further, we believe this formulation and its analogy to the classical Sinkhorn, will allow for more in depth study of the algorithm and its further relaxations. We now investigate properties of the algorithm, starting by establishing the fact that each iteration of MPMS algorithm increases the objective function of the dual problem (4.1.8). Throughout we we write $f(t, \cdot) = f_t(\cdot)$ for a function $f : [0, T] \times \mathbb{R} \to \mathbb{R}$ as a shorthand.

Theorem 4.2.1. *Let $\varphi_t(x)$ solve* (4.1.12) *and $v_t(x) := \frac{x^2}{2} - \varphi_t(x)$ as before. Let $\psi_t(y) := \frac{y^2}{2} + v_t^*(y)$. Let α_0 be defined by $\partial_x v(0, \cdot)_{\#}\alpha_0 = \mu_0$ and $\alpha_T = \mathbf{R}_T * \alpha_0$. Then any update of $\varphi_T(\cdot)$ (and consequently of φ through* (4.1.12)*) that increases $\int_{\mathbb{R}} \varphi_T(x)\,\mathrm{d}\mu_T - \int_{\mathbb{R}} \psi_T(x)\,\mathrm{d}\alpha_T$ increases the dual objective function $\int_{\mathbb{R}} \varphi_T(x)\,\mathrm{d}\mu_T - \int_{\mathbb{R}} \varphi_0(x)\,\mathrm{d}\mu_0$. Consequently, the MPMS iterations strictly increase the objective function of the dual problem* (4.1.8) *unless the algorithm has reached a fixed point.*

Proof. Starting from $t = 0$, with distribution μ_0, let $\tilde{\mu}_t$ be the law of a diffusion process with

diffusion given by (4.1.13), denoted $\sigma_t(\varphi_t)$. We have that $v_t(x) = \frac{x^2}{2} - \varphi_t(x)$ is convex and at time t, let

$$\alpha_t = \partial_x v_{t\#}\tilde{\mu}_t.$$

As seen before, α is the distribution of a Brownian motion with initial law α_0. Let $\psi_t(y) = \frac{y^2}{2} + v^*(y)$, and note ψ_t solves the heat equation:

$$\partial_t \psi_t + \frac{1}{2}\bar{\sigma}^2 \partial_{yy}\psi_t = \bar{\sigma}^2. \tag{4.2.18}$$

Therefore,

$$\int_{\mathbb{R}} \psi_0 \mathrm{d}\alpha_0 + \bar{\sigma}^2 T = \int_{\mathbb{R}} \psi_T \mathrm{d}\alpha_T \tag{4.2.19}$$

$$\mathrm{argsup}_{f(y)-g(x)\leqslant |y-x|^2/2} \left\{ \int_{\mathbb{R}} f \mathrm{d}\tilde{\mu}_t - \int_{\mathbb{R}} g \mathrm{d}\alpha_t \right\} = \{\varphi_t, \psi_t\}. \tag{4.2.20}$$

One could also say that the potential φ_t solves gives the Bass martingale between μ_0 and $\tilde{\mu}_T$. We now have

$$\int_{\mathbb{R}} \varphi_T \mathrm{d}\mu_T - \int_{\mathbb{R}} \varphi_0 \mathrm{d}\mu_0 = \underbrace{\int_{\mathbb{R}} \varphi_T \mathrm{d}\mu_T - \int_{\mathbb{R}} \psi_T \mathrm{d}\alpha_T}_{\text{(I)}} + \underbrace{\int_{\mathbb{R}} \psi_0 \mathrm{d}\alpha_0 - \int_{\mathbb{R}} \varphi_0 \mathrm{d}\mu_0}_{\text{(II)}} + \bar{\sigma}^2 T. \tag{4.2.21}$$

The value (II) is minimal with respect to the pair $\{\varphi_0, \psi_0\}$ for the pair of measures $\{\alpha_0, \mu_0\}$, so any update of φ improves (II). Therefore any improvement of (I) improves the dual objective function. The MPMS algorithm updates $\varphi \mapsto \varphi^{\mathrm{new}}$ such that

$$\partial_x v_T^{\mathrm{new}}{}_{\#}\mu_T = \alpha_T, \tag{4.2.22}$$

or in other words

$$\int_{\mathbb{R}} \varphi_T^{\mathrm{new}} \mathrm{d}\mu_T - \int_{\mathbb{R}} \psi_T^{\mathrm{new}} \mathrm{d}\alpha_T = \sup_{f(y)-g(x)\leqslant |y-x|^2/2} \left\{ \int_{\mathbb{R}} f \mathrm{d}\mu_T - \int_{\mathbb{R}} g \mathrm{d}\alpha_T \right\}. \tag{4.2.23}$$

Now assume that $\tilde{\mu}_T \neq \mu_T$, otherwise the algorithm terminates. Note that we do not have

to attain the sup in (4.1.8) for the result to hold true, we only need to increase $\int_{\mathbb{R}} f \mathrm{d}\mu_T - \int_{\mathbb{R}} g \mathrm{d}\alpha_T$. Therefore the algorithm increases (I) and when doing so, it updates $\{\varphi_0, \psi_0\}$ in a way that can only increase (II) since the previous pair of dual potentials was optimal (minimal). Therefore,

$$\int_{\mathbb{R}} \varphi_T^{\text{new}} \mathrm{d}\mu_T - \int_{\mathbb{R}} \varphi_0^{\text{new}} \mathrm{d}\mu_0 > \int_{\mathbb{R}} \varphi_T \mathrm{d}\mu_T - \int_{\mathbb{R}} \varphi_0 \mathrm{d}\mu_0. \tag{4.2.24}$$

The algorithm then updates $\alpha_0 \mapsto \alpha_0^{\text{new}} = \partial_x v_0^{\text{new}}{}_{\#}\mu_0$, and therefore the whole curve α_t, but this has no effect on the dual value. □

For completeness, we now provide a short formal proof of the contraction property in Acciaio, Marini, and Pammer, 2023, Theorem 1.2, taking advantage of our PDE formulation.

Proposition 4.2.2 (Contraction Property of Acciaio, Marini, and Pammer, 2023, Theorem 1.2)**.** *The MPMS algorithm satisfies the contraction property* $W^\infty(\alpha_0^{n+1}, \alpha_0^{n+2}) \leqslant W^\infty(\alpha_0^n, \alpha_0^{n+1})$.

Proof. The algorithm follows the updates: $v_0^n \mapsto \alpha_0^n \mapsto \alpha_T^n \mapsto v_T^{n+1} \mapsto v_0^{n+1}$. We then have the following sequence of inequalities:

$$\text{Update } \alpha_0\colon \quad W^\infty(\alpha_0^n, \alpha_0^{n+1}) \leqslant \|\partial_x v_0^n - \partial_x v_0^{n+1}\|_{L^\infty}, \tag{4.2.25}$$

$$\text{Heat equation contraction:} \quad W^\infty(\alpha_T^n, \alpha_T^{n+1}) \leqslant W^\infty(\alpha_0^n, \alpha_0^{n+1}), \tag{4.2.26}$$

$$\text{Update } v\colon \quad \|\partial_x v_T^{n+1} - \partial_x v_T^{n+2}\|_{L^\infty} = W^\infty(\alpha_T^n, \alpha_T^{n+1}), \tag{4.2.27}$$

$$\text{HJB contraction:} \quad \|\partial_x v_0^{n+1} - \partial_x v_0^{n+2}\|_{L^\infty} \leqslant \|\partial_x v_T^{n+1} - \partial_x v_T^{n+2}\|_{L^\infty}. \tag{4.2.28}$$

The first and third point are a direct consequence of the fact that $\partial_x v_t(\cdot)_{\#}\mu = \alpha$, and for the third point that we are in dimension 1. The second point is a well known fact, that can be easily recovered by considering the case of two Dirac masses evolving through the heat flow. For the fourth point, observe that $\xi = \partial_x v$ solves the parabolic homogeneous PDE

$$\partial_t \xi + \frac{\partial_{xx}\xi}{2(\partial_x \xi)^2} = 0. \tag{4.2.29}$$

We claim that $\xi_T \to \xi_0$ is a contraction in L^∞. Indeed, this is a general fact for homogeneous

nonlinear parabolic equations (see, for example Lieberman, 1996). Consider the linearized equation of (4.2.29). It is a parabolic equation of the form

$$\partial_t u + \mathcal{A}(\xi)\partial_{xx} u + \mathcal{B}(\xi)\partial_x u = 0. \tag{4.2.30}$$

It satisfies the maximum principle on bounded solutions, and therefore $\|u_0\|_{L^\infty(\mathbb{R})} \leqslant \|u_T\|_{L^\infty(\mathbb{R})}$ with equality only if u is constant. The claim follows easily, and the fourth point of the argument follows. Finally,

$$W^\infty(\alpha_0^{n+1}, \alpha_0^{n+2}) \leqslant \|\partial_x v_0^{n+1} - \partial_x v_0^{n+2}\|_{L^\infty} \leqslant W^\infty(\alpha_0^n, \alpha_0^{n+1}). \tag{4.2.31}$$

□

We observe that the only step of the proof that does not work in dimension greater than 1 is the third argument.

4.2.3 An aside on MPMS in d dimensions

We start by considering the d-dimensional version of (mBB) as formulated in Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023. Fix a reference volatility $\bar{\sigma} > 0$, and marginals $\mu_0, \mu_T \in \mathcal{P}_2(\mathbb{R}^d)$ in strict convex order. Then, as in one dimension, we have the equivalent formulation:

$$\mathbf{MT}_{\mu_0,\mu_T} = \inf_{\substack{M_0\sim\mu_0, M_T\sim\mu_T \\ M_t = M_0 + \int_0^t \sigma_s \mathrm{d}B_s}} \mathbb{E}\left[\int_0^T |\sigma_t - \bar{\sigma}\mathrm{I}_d|^2_{\mathrm{HS}} \mathrm{d}t\right], \tag{mBBd}$$

where $|\cdot|_{\mathrm{HS}}$ is the Hilbert-Schmidt norm given by $|A|_{\mathrm{HS}} = \sqrt{\mathrm{tr}(A^\intercal A)}$ for a $d \times d$ matrix A, and I_d is the $d \times d$ identity matrix. We also have the equivalent formulation of the (mBBd):

$$\mathbf{P}_{\mu_0,\mu_T} = \sup_{\substack{M_0\sim\mu_0, M_T\sim\mu_T \\ M_t = M_0 + \int_0^t \sigma_s \mathrm{d}B_s}} \mathbb{E}\left[\int_0^T \mathrm{tr}(\sigma_t)\mathrm{d}t\right], \tag{4.2.32}$$

which can be viewed as maximising the covariance of M and Brownian motion. We then consider the more general formulation in a Markovian setting, where we are given a cost

function $H : \mathbb{R}^{d\times d} \to (-\infty, +\infty]$:

$$\mathbf{MT}^{H}_{\mu_0,\mu_T} := \inf_{\substack{M_0\sim\mu_0, M_T\sim\mu_T \\ M_t = M_0 + \int_0^t \sigma(s,M_s)\mathrm{d}B_s}} \mathbb{E}\left[\int_0^T H(\sigma^\intercal(t,M_t)\sigma(t,M_t))\mathrm{d}t\right]. \tag{mBBd(H))}$$

Once again, applying the general duality result of I. Guo and Loeper, 2021, we arrive at the dual formulation for (mBBd(H)):

$$\mathbf{MT}^{H}_{\mu_0,\mu_T} = \sup_{\varphi}\left\{\int_{\mathbb{R}^d}\varphi(T,x)\,\mathrm{d}\mu_T - \int_{\mathbb{R}^d}\varphi(0,x)\,\mathrm{d}\mu_0\right\}, \tag{4.2.33}$$

where the supremum is taken over all super-solutions φ of

$$\partial_t\varphi + H^*\left(\frac{1}{2}\nabla_x^2\varphi\right) \leqslant 0. \tag{4.2.34}$$

It is well known that for a positive semidefinite matrix, A, there exists a unique positive semidefinite B such that $B^\intercal B = A$, which we we denote by $\sqrt{A}$ (see for example Horn and Johnson, 2012, Theorem 7.2.6). Therefore, for $\beta \in \mathbb{R}^{d\times d}$, we take the cost function H to be:

$$H(\beta) = \begin{cases} \frac{1}{2}|\sqrt{\beta} - \bar{\sigma}\mathrm{I}_d|^2_{\mathrm{HS}}, & \text{if } \beta \text{ positive semidefinite,} \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.2.35}$$

Then, we have that the problems (mBBd(H)) and $\mathbf{P}_{\mu_0,\mu_T}$ are equivalent when H is given by (4.2.35), and

$$\int_{\mathbb{R}^d}\varphi(T,x)\,\mathrm{d}\mu_T - \int_{\mathbb{R}^d}\varphi(0,x)\,\mathrm{d}\mu_0 = \mathbf{MT}_{\mu_0,\mu_T} = \frac{d\bar{\sigma}^2T}{2} + \int_{\mathbb{R}^d}\frac{|x|^2}{2}(\mathrm{d}\mu_T - \mathrm{d}\mu_0) - \bar{\sigma}\mathbf{P}_{\mu_0,\mu_T}. \tag{4.2.36}$$

Using the fact that $\beta \in \mathbb{S}^d_+$, and its matrix square root is also positive semidefinite, we compute the Legendre transform to be

$$\begin{aligned} H^*(p) &= \sup_{\beta\in\mathbb{S}^d_+}\left\{\mathrm{tr}(\beta p) - \frac{1}{2}\mathrm{tr}\left(\beta - 2\bar{\sigma}\sqrt{\beta} + \bar{\sigma}^2\mathrm{I}_d\right)\right\} \\ &= \sup_{\sigma\in\mathbb{S}^d_+}\left\{\mathrm{tr}\left(\sigma^2\left(p - \frac{1}{2}\mathrm{I}_d\right) + \bar{\sigma}\sigma - \frac{1}{2}\bar{\sigma}^2\mathrm{I}_d\right)\right\} \end{aligned}$$

$$= \inf_{\sigma \in \mathbb{S}^d_+} \left\{ \operatorname{tr}\left(\sigma^2 \left(\frac{1}{2} \mathrm{I}_d - p \right) - \bar{\sigma}\sigma + \frac{1}{2}\bar{\sigma}^2 \mathrm{I}_d \right) \right\}. \tag{4.2.37}$$

The right hand side is clearly minimised at $\sigma = \bar{\sigma}(\mathrm{I}_d - 2p)^{-1}$, so we have

$$H^*(p) = \operatorname{tr}\left(\bar{\sigma}^2 p (\mathrm{I}_d - 2p)^{-1} \right). \tag{4.2.38}$$

Thus, the HJB equation (4.2.34) becomes:

$$\partial_t \varphi + \frac{\bar{\sigma}^2}{2} \operatorname{tr}\left(\nabla_x^2 \varphi (\mathrm{I}_d - \nabla_x^2 \varphi)^{-1} \right). \tag{4.2.39}$$

If φ is optimal for (4.2.33), then the optimal diffusion in (mBBd) is given by

$$\sigma = \bar{\sigma} \left(\mathrm{I}_d - \nabla_x^2 \varphi \right)^{-1}. \tag{4.2.40}$$

We notice that (4.2.39) has been studied in the context of market impact in Abergel and Loeper, 2017. As in the one-dimensional case, for φ to solve (4.2.39), we must have $\mathrm{I}_d - \nabla_x^2 \varphi \geqslant 0$, and thus the potential $v = \frac{|x|^2}{2} - \varphi$ is convex. Additionally, the potential v satisfies

$$\partial_t v + \frac{\bar{\sigma}^2}{2} (d - \operatorname{tr}\left((\nabla_x^2 v)^{-1} \right) = 0. \tag{4.2.41}$$

We set $\xi = \nabla_x v$, and let J_ξ be its Jacobian matrix. Since trace and differentials commute, we apply the chain rule to J_ξ^{-1}, so that $\partial_{x_i} \operatorname{tr}(J_\xi^{-1}) = \operatorname{tr}(-J_\xi^{-1}(\partial_{x_i} J_\xi) J_\xi^{-1})$, then by the cyclic property of the trace, we obtain for each $i = 1, \ldots, d$:

$$\partial_t \xi^i + \frac{1}{2} \operatorname{tr}\left(\sigma^2 \partial_{x_i} J_\xi \right), \qquad \text{with } \sigma = \bar{\sigma} J_\xi^{-1} = \bar{\sigma} (\nabla_x^2 v)^{-1}. \tag{4.2.42}$$

Therefore, if M^σ solves $\mathrm{d}M_t^\sigma = \sigma(t, M_t^\sigma)\, \mathrm{d}B_t$, with σ defined in (4.2.40), then $Z_t \coloneqq \nabla_x v(t, M_t^\sigma)$ is a martingale, and

$$\mathrm{d}\langle Z \rangle_t = (\nabla_x^2 v(t, M_t^\sigma))^2 \sigma^2(t, M_t^\sigma)\, \mathrm{d}t = \bar{\sigma}^2 \mathrm{I}_d\, \mathrm{d}t, \tag{4.2.43}$$

and therefore $\frac{1}{\bar{\sigma}}Z$ is a Brownian motion. Since the inverse of $\nabla_x v$ is $\nabla_y v^*$, if we set $F(t,y) = \nabla_y v^*(t, \bar{\sigma} y)$, then $M_t^{\sigma} = F(t, \frac{1}{\bar{\sigma}} Z_t)$. Then, by (4.2.39), we have that v^* satisfies

$$\partial_t v^* + \frac{1}{2}\bar{\sigma}^2(\Delta_y v^* - d) = 0, \tag{4.2.44}$$

where Δ_y is the Laplacian, and therefore F solves the heat equation coordinate-wise. The heat kernel in $\mathbb{R}^d$ is given by

$$\mathbf{R}_t^d = \frac{1}{(2\pi t)^{d/2}} e^{-\frac{|x|^2}{2t}}, \tag{4.2.45}$$

so we could equivalently write $F(0,\cdot) = F(T,\cdot) * \mathbf{R}_{T\bar{\sigma}}^d$, where the convolution is performed coordinate-wise. Comparing with (4.2.44), we see that

$$v^*(0,\cdot) = \left(v^*(T,\cdot) - \frac{dT\bar{\sigma}^2}{2}\right) * \mathbf{R}_{T\bar{\sigma}}^d, \tag{4.2.46}$$

$$v(0,\cdot) = \left(v^*(T,\bar{\sigma}\cdot) * \mathbf{R}_{T\bar{\sigma}}^d\right)^* + \frac{dT\bar{\sigma}^2}{2}. \tag{4.2.47}$$

Due to the equivalence in (4.2.36) with H defined in (4.2.35), by setting $v(T, \bar{\sigma}^{-1}x) = \psi(x)$, then (4.2.33) can equivalently be written as

$$\mathbf{P}_{\mu_0,\mu_T} = \inf_{\psi} \left\{ \int_{\mathbb{R}} \psi \, \mathrm{d}\mu_T - \int_{\mathbb{R}} (\psi^* * \mathbf{R}_{T\bar{\sigma}}^d)^* \, \mathrm{d}\mu_0 \right\}, \tag{4.2.48}$$

which, as before, in the special case $\bar{\sigma} = 1$ recovers the duality obtained by Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023. When constructing the MPMS iterations, the only difference is $\mathbf{R}_1^d$ in the updates (4.2.14)-(4.2.15). In particular, we obtain Algorithm 6.

In the context of Theorem 4.2.1, with $\varphi_t(x)$ solving (4.2.39), $v_t = \frac{|x|^2}{2} - \varphi_t(x)$, and $\psi_t(y) = v_t^*(y) + \frac{|y|^2}{2}$, any update of $\varphi_T(\cdot)$ that increases $\int_{\mathbb{R}^d} \varphi_T(x) \mathrm{d}\mu_T - \int_{\mathbb{R}^d} \psi_T(x) \, \mathrm{d}\alpha_T$ also increases the value of the dual objective function in (4.2.33). The proof of this and that the MPMS iterations increase the value of (4.2.33) is identical in d dimensions.

Algorithm 6: Measure Preserving Martingale Sinkhorn algorithm in d dimensions

Input: μ_0, μ_1
Output: $\alpha_0, \tilde{F}$

Set $\tilde{F_i}^0 = Id$
for $n = 0$ *to* `max_iter` **do**
  Solve backward heat equation for each F_i^n: $\partial_t F_i^n + \frac{1}{2}\Delta F_i^n = 0$, $F_i^n(1) = \tilde{F_i}^n$
  Update α_0^{n+1} such that $F^n(0)_\# \alpha_0^{n+1} = \mu_0$
  Solve forward heat equation for α^{n+1}: $\partial_t \alpha^{n+1} - \frac{1}{2}\Delta \alpha^{n+1} = 0$, $\alpha^{n+1}(0) = \alpha_0^{n+1}$
  Update $\tilde{F}^{n+1}$ such that $(\tilde{F}_1^{n+1})_\# \alpha^{n+1}(1) = \mu_1$
end
return $\alpha_0, \tilde{F}$

4.3 Examples and implementation of Algorithm 5

We discuss now the practical implementation of Algorithm 5 and present some examples. This algorithm has already been shown to perform very well in Conze and Henry-Labordere, 2021; this relies to a large degree on the fact that in one dimension, the transport problem, or the Monge-Ampère equation, can be solved explicitly. This is the observation in (4.2.17) which we use for the computation of the pushforwards in Algorithm 5. However, we are also interested in exploring ways to extend this algorithm to other settings, e.g., when the marginal distributions are not entirely fixed or to higher dimensions. One possibility would be to only approximatively solve the transport problem in step 6 in Algorithm 5. This motivated our numerical experiments reported in Section 4.3.2.

4.3.1 Direct MPMS Implementation

In another small departure from the methods in Conze and Henry-Labordere, 2021, instead of using Gauss-Hermite quadrature to compute the convolutions, we use the observation that α and F solve the forward and backward heat equation respectively to construct an implicit finite difference scheme to compute α_1 and F_0 from α_0 and F_1 respectively. If $0 = t_0 < \cdots < t_M = 1$ is a discretisation of the time interval, then we solve the linear systems $AF_{t_k} = F_{t_{k+1}}$ from $F_{t_{k+1}}$ and $\tilde{A}\alpha_{t_k} = \alpha_{t_{k-1}}$. Since $A = \tilde{A}$, we only need to compute $(A^{-1})^M$ once to be able to compute the convolutions. We take a finite interval $[z_{\min}, z_{\max}]$

as our computational domain, and assign the following boundary conditions:

$$\alpha_{t_k}(z_{\min}) = 0, \quad \alpha_{t_k}(z_{\max}) = 1, \quad F_{t_k}(z_{\min}) = z_{\min}, \quad F_{t_k}(z_{\max}) = z_{\max}, \quad \text{for } k = 1, \dots, M. \tag{4.3.1}$$

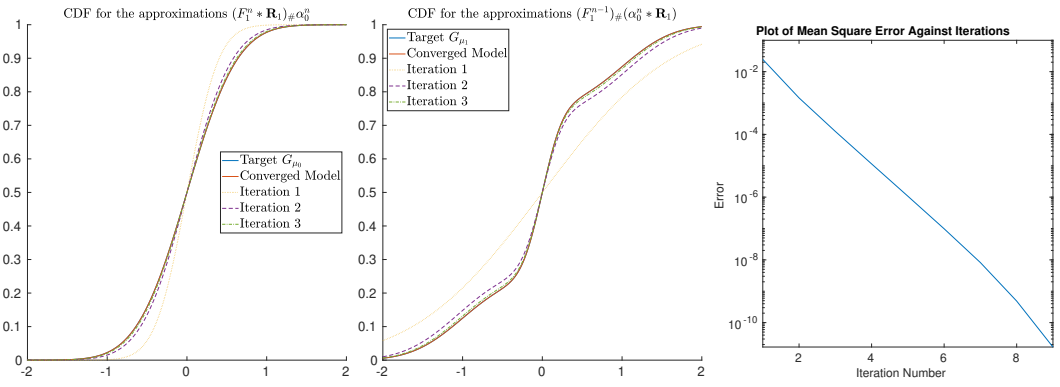
In all of our examples, we measure the error by $\frac{1}{N}\sum_{k=1}^{N}\left(G_{\mu_0}^{-1}(y_k) - G^{-1}_{(F_1^n * \mathbf{R}_1)_{\#}\alpha_0^n}(y_k)\right)^2$ where $0 = y_1 < \cdots < y_N = 1$ is a uniform discretisation of $[0,1]$, so that we are minimising in mean square the horizontal distance between the CDFs.

4.3.1.1 Mixed Gaussian Example

We start with a mixed Gaussian example with our marginals given by

$$\mu_0 = \mathcal{N}(0, 0.5), \quad \mu_1 = \frac{1}{4}\mathcal{N}(-1, 0.25) + \frac{1}{2}\mathcal{N}(0, 0.5) + \frac{1}{4}\mathcal{N}(1, 0.25). \tag{4.3.2}$$

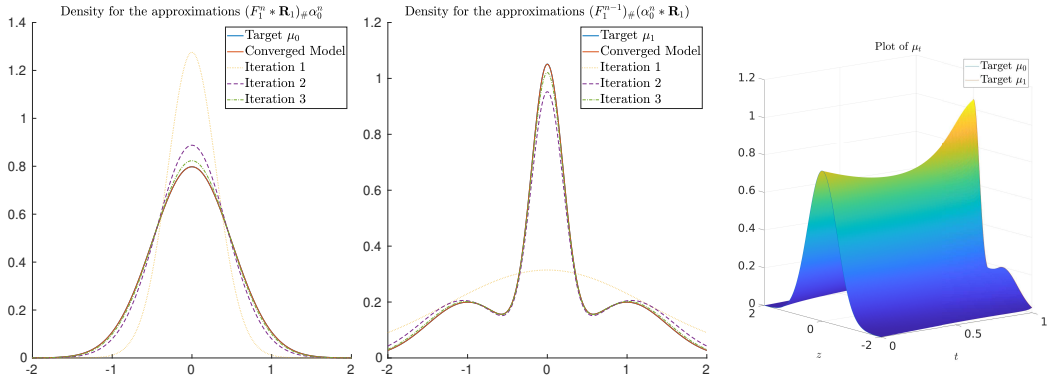
We discretised the interval $[-4, 4]$ into 1000 spatial gridpoints, and the time interval $[0, 1]$ into 50 gridpoints. We ran Algorithm 5 until an error of 1×10^{-10} was reached, which took 9 iterations – equivalent to 0.067 seconds[2].



(a) (b) (c)

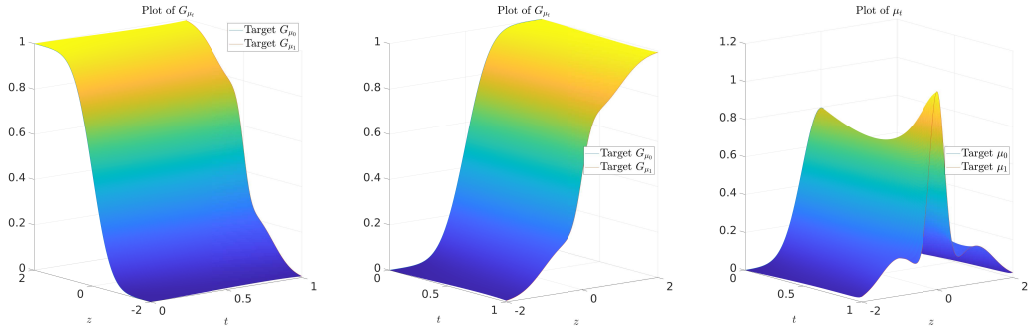
Figure 4.1: Example (4.3.2). (a) Plots of CDF for the approximations of μ_0 after various iterations along with the target CDF, (b) Plots of CDF for the approximations of μ_1 after various iterations along with the target CDF, (c) Plot of mean square error as a function of iteration.

[2]All computational times reported are for a Lenovo Thinkpad Core i5 standard laptop.



(a) (b) (c)

Figure 4.2: Example (4.3.2). (a) Plots of the density for the approximations of μ_0 along with the target density, (b) Plots of the density for the approximations of μ_1 along with the target density (c) Plot of converged density martingale interpolation viewed from $t = 0$.



(a) (b) (c)

Figure 4.3: Example (4.3.2). (a) Plot of CDF martingale interpolation viewed from $t = 0$, (b) Plot of CDF martingale interpolation viewed from $t = 1$, (c) Plot of density martingale interpolation viewed from $t = 1$.

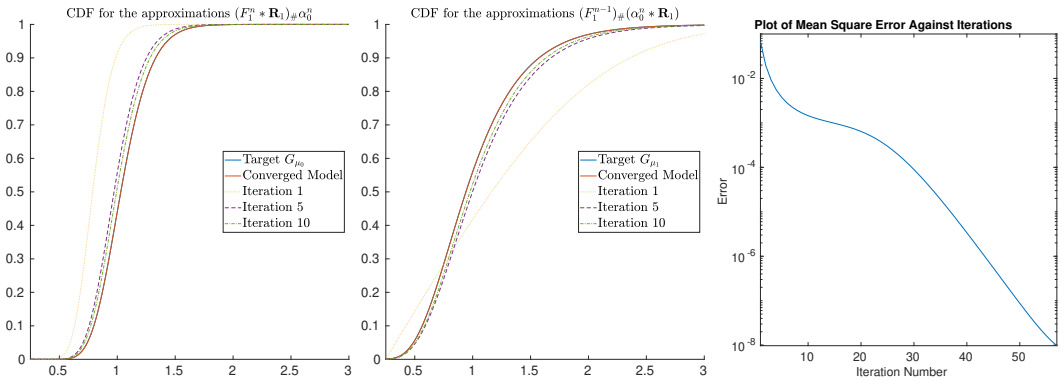
4.3.1.2 Lognormal Example

We next try a lognormal example, where the prescribed marginals are given by

$$\mu_0 = \text{Lognormal}\left(r - \frac{1}{2}\sigma_0^2, \sigma_0\right), \quad \mu_1 = \text{Lognormal}(2r - \sigma_1^2, \sigma_1). \tag{4.3.3}$$

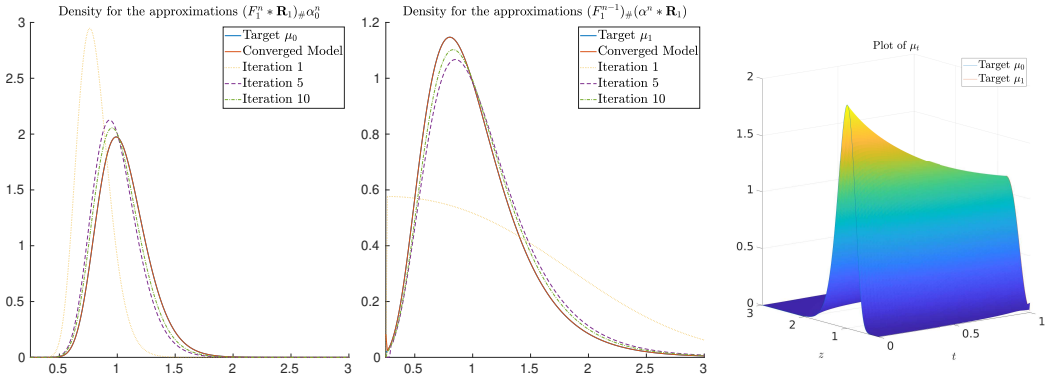
Where $r = 0.05$, $\sigma_0 = 0.2$, and $\sigma_1 = 0.4$, so that this resembles a Geometric Brownian motion with different volatilites. We took the same time discretisation as in the weighted Gaussian example, but our spatial interval was given instead by $[0.25, 7]$, which was discretised into

1000 gridpoints. We ran Algorithm 5 until an error of 1×10^{-8} was attained, which took 57 iterations – equivalent to 0.35 seconds.



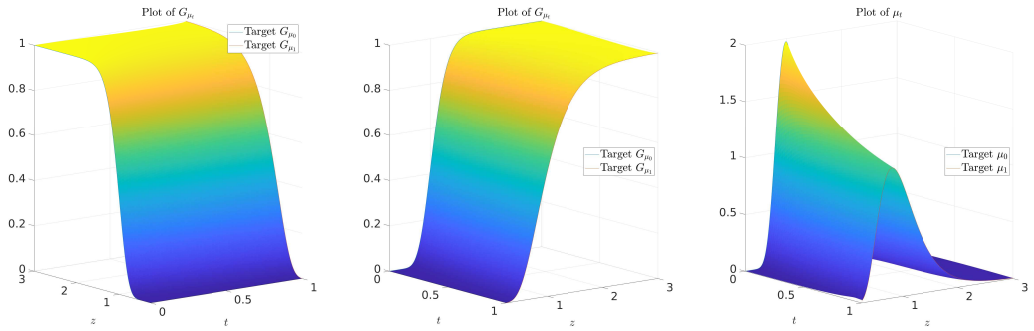
(a) (b) (c)

Figure 4.4: Example (4.3.3). (a) Plots of CDF for the approximations of μ_0 after various iterations along with the target CDF, (b) Plots of CDF for the approximations of μ_1 after various iterations along with the target CDF, (c) Plot of mean square error as a function of iteration.



(a) (b) (c)

Figure 4.5: Example (4.3.3). (a) Plots of the density for the approximations of μ_0 along with the target density, (b) Plots of the density for the approximations of μ_1 along with the target density (c) Plot of converged density martingale interpolation viewed from $t = 0$.



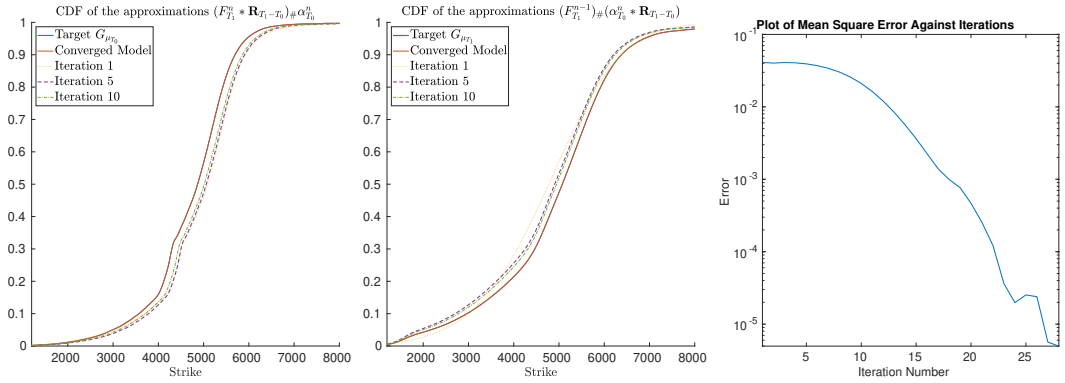
(a) (b) (c)

Figure 4.6: Example (4.3.3). (a) Plot of CDF martingale interpolation viewed from $t = 0$, (b) Plot of CDF martingale interpolation viewed from $t = 1$, (c) Plot of density martingale interpolation viewed from $t = 1$.

4.3.1.3 Market Data Example

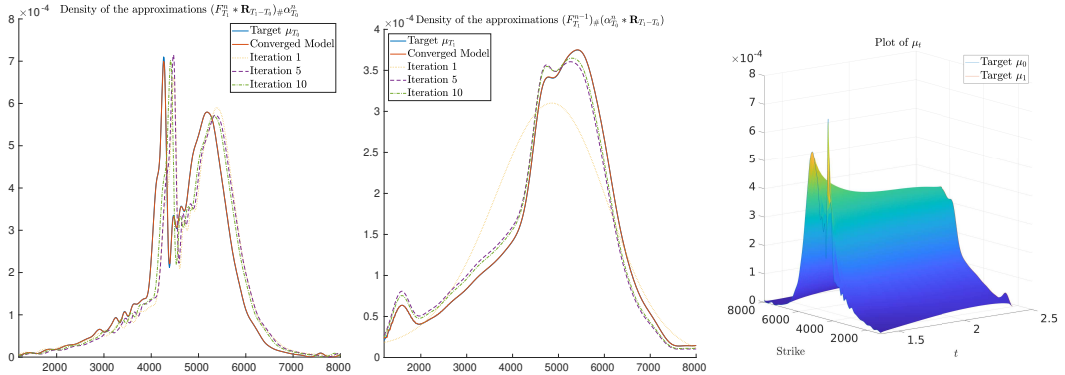
Finally, we benchmark the MPM-Sinkhorn iterations using the Breeden-Litzenberger formula of Breeden and Litzenberger, 1978. Let $C(T, K)$ be the price of a European call option on an underlying with maturity $T > 0$ and strike $K > 0$, then the density of the underlying at time T is given by $\mu_T(K) = \frac{\partial^2 C(T,K)}{\partial K^2}$. In order to compute the second derivative of the call option price, we interpolated the strikes and prices with a spline. Owing to numerical instabilities in computing the second derivative, we also applied a smoothing formula to generate more reasonable densities. We obtained[3] SPX call option data with maturities at 20/12/2024 and 19/12/2025, which we denote T_0 and T_1 respectively. We discretised $[T_0, T_1]$ into 50 gridpoints, and took the spatial interval of $[1200, 8000]$ and discretised it into 1000 gridpoints. We then rescaled the domain and CDFs by 1000. Since the CDFs are numerically computed from interpolated options data, we took a higher tolerance at 5×10^{-6}, and obtained convergence in 28 iterations which took 0.18 seconds.

[3]Data obtained from `https://www.cboe.com/delayed_quotes/spx` on 08/08/2023.



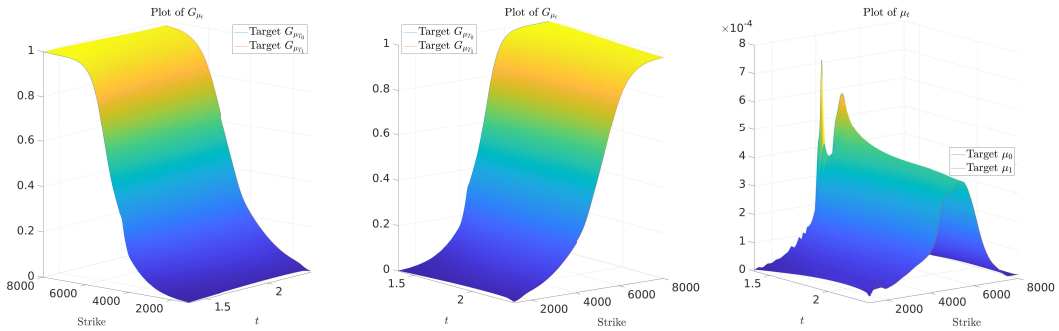
(a) (b) (c)

Figure 4.7: SPX market data example. (a) Plots of CDF for the approximations of μ_0 after various iterations along with the target CDF, (b) Plots of CDF for the approximations of μ_1 after various iterations along with the target CDF, (c) Plot of mean square error as a function of iteration.



(a) (b) (c)

Figure 4.8: SPX market data example. (a) Plots of the density for the approximations of μ_0 along with the target density, (b) Plots of the density for the approximations of μ_1 along with the target density (c) Plot of converged density martingale interpolation viewed from $t = 0$.



(a) (b) (c)

Figure 4.9: SPX market data example. (a) Plot of CDF martingale interpolation viewed from $t = T_0$, (b) Plot of CDF martingale interpolation viewed from $t = T_1$, (c) Plot of density martingale interpolation viewed from $t = T_1$.

4.3.2 An alternative implementation with an approximate transport solution

We now provide an alternative implementation of Algorithm 5, where we deal with the densities directly. We write $f = F'$ and f^n for the derivative of F^n. In the algorithm, we have to twice solve the transport relation $\nu_1 = h_{\#}\nu_2$ with two of ν_1, ν_2, h given. For a given $\nu_1 \in \mathcal{P}_2(\mathbb{R})$ and a (differentiable strictly increasing) h, a change of variables gives $\nu_2(y) = \nu_1(h(y))h'(y)$. In particular, line 4 in Algorithm 5 reads $\alpha_0^n(y) = \mu_0(F^n(0,y))f^n(0,y)$. On the other hand, for a given $\nu_1, \nu_2 \in \mathcal{P}_2(\mathbb{R})$, we have $h = G_{\nu_1}^{-1} \circ G_{\nu_2}$. We thus consider the equivalent implicit equation for h: $\nu_2(y) = \nu_1(h(y))h'(y)$ and use the previous estimate for $h(y)$ to solve for $h'(y)$ and integrate numerically. In summary, our alternative implementation for Algorithm 5 reads as follows:

Algorithm 7: MPMS: density implementation

1 Set $F_1^0 = Id$, $f_1^0 = 1$;

2 **for** $n = 0$ *to* `max_iter` **do**

3 Compute α_0^{n+1}: $\alpha_0^{n+1}(x) = \mu_0\left((F_1^n * \mathbf{R}_1)(x)\right)(f_1^n * \mathbf{R}_1)(x)$;

4 Compute f_1^{n+1} using F_1^n and α_0^n: $f_1^{n+1}(x) = \frac{(\alpha_0^n * \mathbf{R}_1)(x)}{\mu_1(F_1^n(x))}$;

5 Compute F_1^{n+1}(x): $F_1^{n+1}(x) = \int_{-\infty}^{x} f_1^{n+1}(y)dy$.

6 **end**

7 **return** α_0, F_1, f_1

In practice, the above has to be implemented on a discrete grid and has to address the issue of division by zero. We let x_i, $i = 1, \ldots, N$ be a discretisation of the desired interval to apply the convolution on and let $y_i = \frac{1}{m}(x_i - c)$ be a rescaling of the discretisation such that most of the mass of μ_1 is contained in $[y_1, y_N]$, i.e., $\int_{y_1}^{y_N} \mu_1 dy \approx 1$. We also replace f_1^{n+1} with $f_1^{n+1} \wedge C$ to avoid issues around division in line 4 above; in the region of the domain where both $\alpha^{(n)} * \mathbf{R}_1$ and $\mu_1(F_1^n)$ are close to zero, we set $f_1^{n+1} = 0$. Let `conv` denote a numerical implementation of the convolution and $\texttt{conv}(\cdot)_k$ denote the k^{th} element of the vector returned by `conv`. Then, the above algorithm can be implemented as follows:

$$
\begin{aligned}
&F_1^0(x_k) = x_k, \quad f_1^0(x_k) = 1, \quad k = 1, \ldots, N; \\
&\begin{cases}
F_1^n * \mathbf{R}_1(x_k) &= \texttt{conv}(mF_1^n + c, \mathbf{R}_1)_k, \\
f_1^n * \mathbf{R}_1(x_k) &= \texttt{conv}(mf_1^n + c, \mathbf{R}_1)_k, \\
\alpha^n(y_k) &= \frac{1}{m}\left(f_1^n * \mathbf{R}_1(x_k) - c * \mathbf{R}_1(x_k)\right) \mu_0\left(\frac{1}{m}(F_1^n * \mathbf{R}_1(x_k) - c * \mathbf{R}_1(x_k)\right),
\end{cases}
\end{aligned}
\tag{4.3.4}
$$

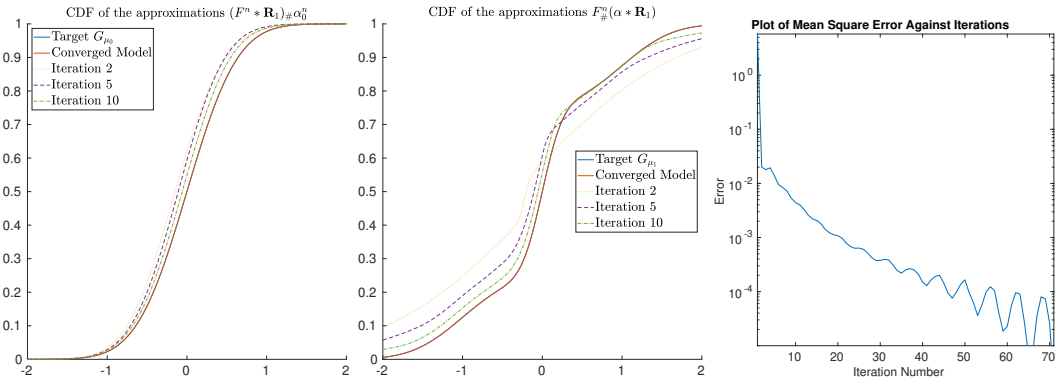
$$
\begin{cases}
\alpha_0^n * \mathbf{R}_1(x_k) &= \texttt{conv}(m\alpha_0^n + c, \mathbf{R}_1)_k \\
f_1^n(y_k) &= \begin{cases}
\left(\frac{1}{m}\left(\alpha_0^n * \mathbf{R}_1(x_k) - c * \mathbf{R}_1(x_k)\right) / \mu_1(F_1^n(y_k))\right) \wedge C, \\
\qquad \text{if } \alpha_0^n * \mathbf{R}_1(x_k) - c * \mathbf{R}_1(x_k) \neq 0 \neq \mu_1(F_1^n(y_k)), \\
0, \\
\qquad \text{if } \alpha_0^n * \mathbf{R}_1(x_k) - c * \mathbf{R}_1(x_k) = 0 = \mu_1(F_1^n(y_k)),
\end{cases} \\
F_1^{n+1}(y_k) &= F_1^n(y_1) + \int_{y_1}^{y_k} f_1^n(y) dy, \quad k = 1, \ldots, N.
\end{cases}
\tag{4.3.5}
$$

Since we only approximately solve $\mu_1 = F_{\#}(\alpha_0 * \mathbf{R}_1)$, we must control the error at both $t = 0$ and $t = 1$. We therefore take the maximum of the mean square error of the inverse CDFs at $t = 0$ and $t = 1$. Since the CDFs and their inverse must both be numerically computed, we had to take lower tolerances in this implementation.

4.3.2.1 Mixed Gaussian Example

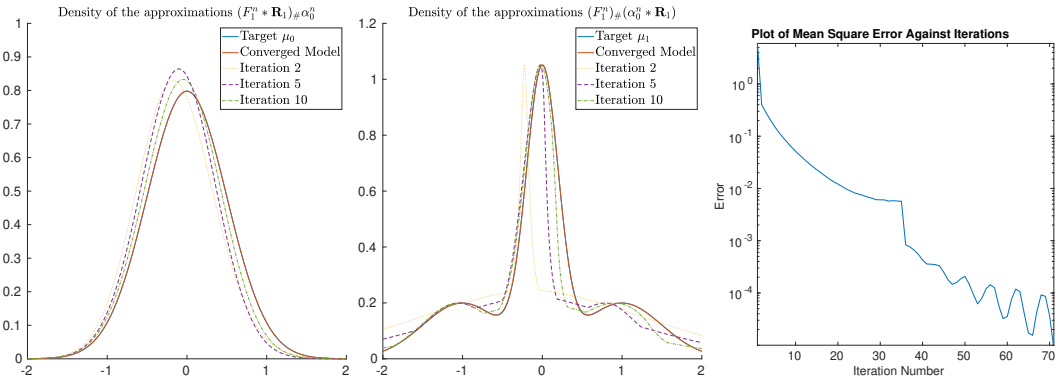
We start with the weighted Gaussian with marginals given by (4.3.2). We discretised the interval $[-4, 4]$ into 1000 spatial gridpoints, and applied the iterations (4.3.4)-(4.3.5) until a

mean square error of 1×10^{-5} was achieved. The upper bound in (4.3.5) was taken to be $C = 1.5$. This took a total of 35 iterations which was equivalent to 0.13 seconds on a laptop.



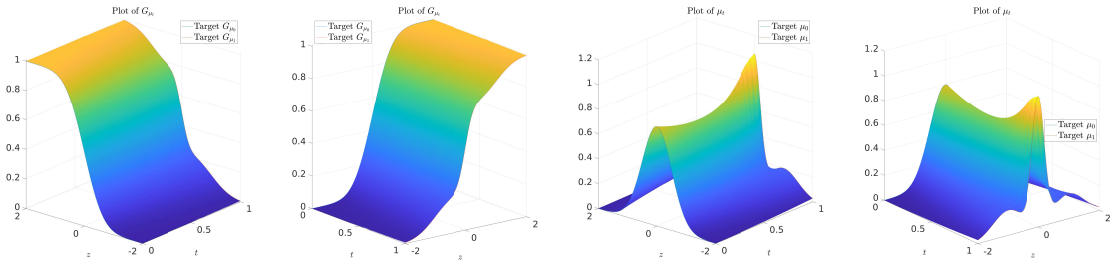
(a) (b) (c)

Figure 4.10: (a) Plots of μ_0 after various iterations along with the target density, (b) Plots of μ_0 after various iterations along with the target density, (c) Plot of mean square errors in μ_0.



(a) (b) (c)

Figure 4.11: (a) Plots of the density of μ_0 after various iterations along with the target density, (b) Plots of the density of μ_1 defined with F at the previous iteration after various iterations along with the target density, (c) Plot of mean square errors in μ_1.

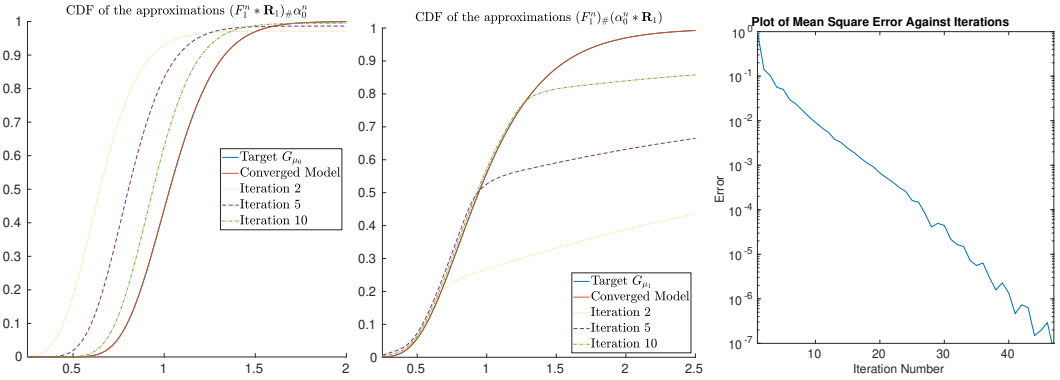


(a) (b) (c) (d)

Figure 4.12: (a) Plot of CDF martingale interpolation viewed from $t = 0$, (b) Plot of CDF martingale interpolation viewed from $t = 1$, (c) Plot of density martingale interpolation viewed from $t = 0$, (d) Plot of density martingale interpolation viewed from $t = 1$.

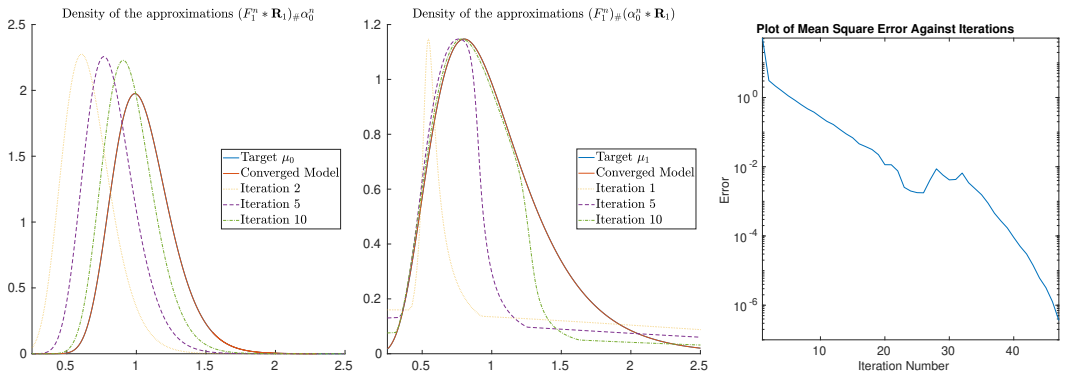
4.3.2.2 Lognormal Example

We next try repeat the lognormal example with marginals given by (4.3.3). We used the domain $[-3.9, 4.1]$ and mapped it onto $[0.2, 5]$ for μ_0 and $[0.25, 10]$ for μ_1 each with 1000 gridpoints. As before, the iterations (4.3.4)-(4.3.5) were repeated until an error of 1×10^{-6} was achieved, this time the upper bound in (4.3.5) was set to $C = 2.4$. This took 47 iterations, which was equivalent to 0.19 seconds on a laptop.



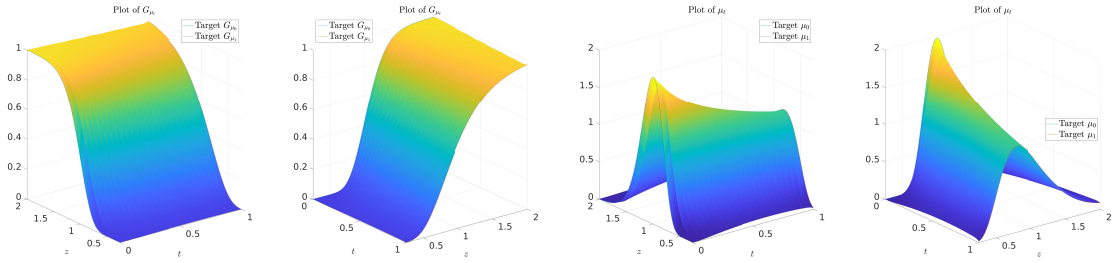
(a) (b) (c)

Figure 4.13: (a) Plots of μ_0 after various iterations along with the target density, (b) Plots of μ_0 after various iterations along with the target density, (c) Plot of mean square errors in μ_0.



(a) (b) (c)

Figure 4.14: (a) Plots of the density of μ_0 after various iterations along with the target density, (b) Plots of the density of μ_1 defined with F at the previous iteration after various iterations along with the target density, (c) Plot of mean square errors in μ_1.

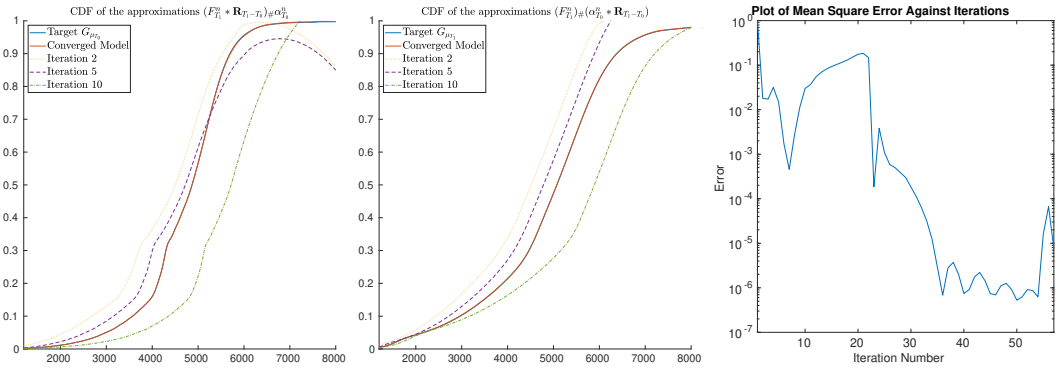


(a) (b) (c) (d)

Figure 4.15: (a) Plot of CDF martingale interpolation viewed from $t = 0$, (b) Plot of CDF martingale interpolation viewed from $t = 1$, (c) Plot of density martingale interpolation viewed from $t = 0$, (d) Plot of density martingale interpolation viewed from $t = 1$.

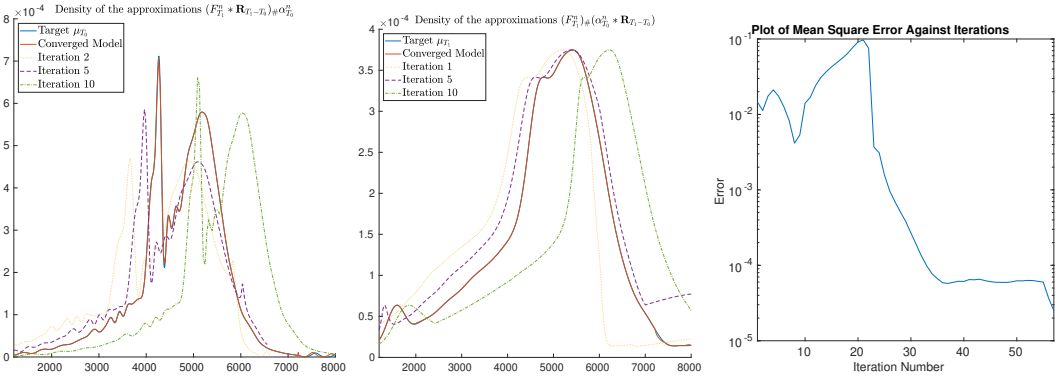
4.3.2.3 Market Data Example

Finally, we repeat the Breeden-Litzenberger example using Algorithm 7. We used the domain $[-10, 10]$, and mapped it onto $[1200, 8000]$ for μ_{T_0} and μ_{T_1}. Due to numerical issues in computing the inverse CDFs, we used the mean square difference between CDFs as the error. We iterated (4.3.4)-(4.3.5) until we an error of 5×10^{-5} was attained, and took the upper bound in (4.3.5) to be $C = 2.7$. This took a total of 57 iterations, for a total computational time of 0.18 seconds. If we wanted to extend the model onto the interval $[0, T_0]$, since we have no marginal constraints, we can use the fact that F solves the heat equation, so that for $t \in [0, T_0]$, we have $F(t, \cdot) = \mathbf{R}_{T_0 - t} * F(T_1, \cdot)$.



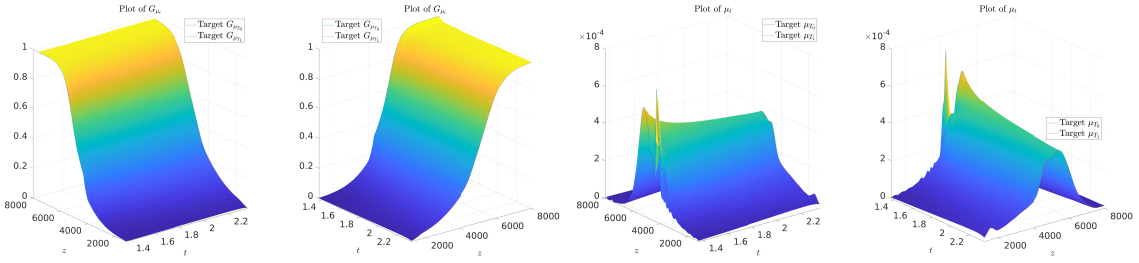
(a) (b) (c)

Figure 4.16: (a) Plots of μ_{T_0} after various iterations along with the target density, (b) Plots of μ_{T_1} after various iterations along with the target density, (c) Plot of mean square errors in μ_{T_0}.



(a) (b) (c)

Figure 4.17: (a) Plots of the density of μ_{T_0} after various iterations along with the target density, (b) Plots of the density of μ_{T_1} defined with F at the previous iteration after various iterations along with the target density, (c) Plot of mean square errors in μ_{T_1}.



(a) (b) (c) (d)

Figure 4.18: (a) Plot of CDF martingale interpolation viewed from $t = 0$, (b) Plot of CDF martingale interpolation viewed from $t = 1$, (c) Plot of density martingale interpolation viewed from $t = 0$, (d) Plot of density martingale interpolation viewed from $t = 1$.

4.4 Geometric Bass Martingale and Local Volatility Construction

We now turn our attention to using the MPMS algorithm to construct local volatility surfaces. We compare the performance in our examples of the MPMS algorithm in the Bass setting described above, and in the Geometric Bass setting introduced in the recent work of Backhoff-Veraguas, Loeper, and Obłój, 2024 (see also Beiglböck, Pammer, and Riess, 2024 for a weak optimal transport construction). We recall some of the key results of Backhoff-Veraguas, Loeper, and Obłój, 2024 and explain modifications to the MPMS implementation used when constructing the local volatility surfaces. Given $\mu_0, \mu_T \in \mathcal{P}_2(\mathbb{R})$ supported on $(0, +\infty)$, and a constant reference volatility $\bar{\sigma} > 0$, the Geometric martingale Benamou-Brenier problem is defined by

$$\mathbf{GMT}_{\mu_0,\mu_T} = \inf_{\substack{M_0\sim\mu_0, M_T\sim\mu_T \\ M_t = M_0 + \int_0^t M_s\sigma_s \mathrm{d}B_s}} \mathbb{E}\left[\int_0^T (\sigma_t - \bar{\sigma})^2 \mathrm{d}t\right]. \tag{GmBB}$$

As in the regular Bass martingale case, (GmBB) is equivalent to

$$\mathbf{GP}_{\mu_0,\mu_T} = \sup_{\substack{M_0\sim\mu_0, M_T\sim\mu_T \\ M_t = M_0 + \int_0^t M_s\sigma_s \mathrm{d}B_s}} \mathbb{E}\left[\int_0^T |\sigma_t| \mathrm{d}t\right]. \tag{4.4.1}$$

For $f \in L^1(\mathrm{d}\mu)$, define the f-reflected measure by

$$f_\dagger \mu = \left(y \mapsto \frac{1}{f(y)}\right)_\# \left(\frac{f(y)}{\int_{\mathbb{R}} f(x)\mathrm{d}\mu}\mathrm{d}\mu\right) \tag{4.4.2}$$

Further, let $\tilde{\mu} = \left(x \mapsto \frac{x}{m}\right)_\# \mu$ for $m = \int_{\mathbb{R}} x\mathrm{d}\mu$, and define $f_\ddagger \mu = f_\dagger \tilde{\mu}$. Writing $\nu_i = \mathrm{Id}_\ddagger \mu_i$, by Backhoff-Veraguas, Loeper, and Obłój, 2024, Theorem 3.1, we have $\mathbf{P}_{\nu_0,\nu_T} = \mathbf{GP}_{\mu_0,\mu_T}$. Therefore, we may directly apply Algorithm 5 with inputs ν_0, ν_T to solve $\mathbf{GP}_{\mu_0,\mu_T}$. Given the optimisers of Algorithm 5, α_0, F_T, we have the relation $\nu_t = (\mathbf{R}_{T-t} * F_T)_\#(\mathbf{R}_t * \alpha_0)$, and moreover by Backhoff-Veraguas, Loeper, and Obłój, 2024, Proposition 6.2, we can express

the volatility in terms of the optimiser F:

$$\sigma_{\text{GBass}}(t,x) = \frac{x}{m\partial_x F^{-1}\left(t, \frac{m}{x}\right)}. \tag{4.4.3}$$

In comparison with the Bass martingale, where we have $\mu_t = (\mathbf{R}_{T-t} * F_T)_{\#}(\mathbf{R}_t * \alpha_0)$, i.e. $M_t = F(t, B_t)$, by Itô's formula we have

$$\mathrm{d}M_t = \partial_x F(t, B_t)\mathrm{d}B_t = \partial_x F\left(t, F^{-1}(t, M_t)\right)\mathrm{d}B_t = M_t\sigma_{\text{Bass}}(t, M_t)\mathrm{d}B_t, \tag{4.4.4}$$

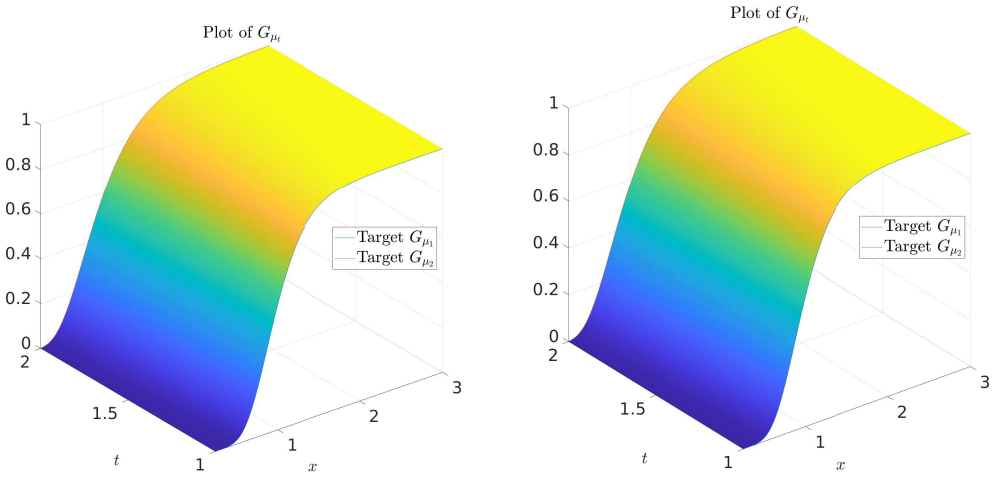
where

$$\sigma_{\text{Bass}}(t,x) = \frac{\partial_x F\left(t, F^{-1}(t,x)\right)}{x} = \frac{1}{x\partial_x F^{-1}(t,x)}. \tag{4.4.5}$$

We now provide some numerical examples comparing the MPMS algorithm in the Bass and Geometric Bass cases. In both, we recover the CDF interpolation and the volatility surfaces σ_{Bass} and σ_{GBass}. In comparison with Section 4.3, we want a finer discretisation of the timesteps to recover a higher resolution volatility surface in time. Our previous approach relied on pre-allocating a 50×1000 to use as a transition matrix in the fully implicit finite differences scheme between $F_T \mapsto F_0$ and $\alpha_0 \mapsto \alpha_T$ respectively. Since computation of the Bass volatility (4.4.5) and Geometric Bass volatility (4.4.3) requires numerical inversion, differentiation, and interpolation, we take 1000 timesteps and instead use the Crank-Nicholson method for higher order accuracy in time. However, pre-allocating a 1000×1000 matrix inversion and 1000 matrix multiplications will take a large amount of time, and lead to numerical inaccuracies, so we therefore instead solve the system of linear equations via the Crank-Nicholson method at each timestep. This slows the algorithm down in exchange for a much higher resolution of the volatility surface in time. We use the same boundary conditions as before in (4.3.1), however to reduce their effect on the solution in the region where most of the mass of μ_0 and μ_T is concentrated, we take a larger computational domain. We use the same mean square error as before: for $0 = y_1 < \cdots < y_N = 1$ a uniform discretisation of $[0,1]$, $\texttt{error} = \frac{1}{N}\sum_{k=1}^{N}\left(G^{-1}_{\mu_0}(y_k) - G^{-1}_{(F_T^n * \mathbf{R}_T)_{\#}\alpha_0^n}(y_k)\right)^2$. Note that since $\mu \mapsto \mathrm{Id}_{\ddagger}\mu$ is an involution, in the Geometric Bass case we compute the error by performing the map $(F_T^n * \mathbf{R}_T)_{\#}\alpha_0^n \mapsto \mathrm{Id}_{\ddagger}\left((F_T^n * \mathbf{R}_T)_{\#}\alpha_0^n\right)$ to compute the model μ_0.

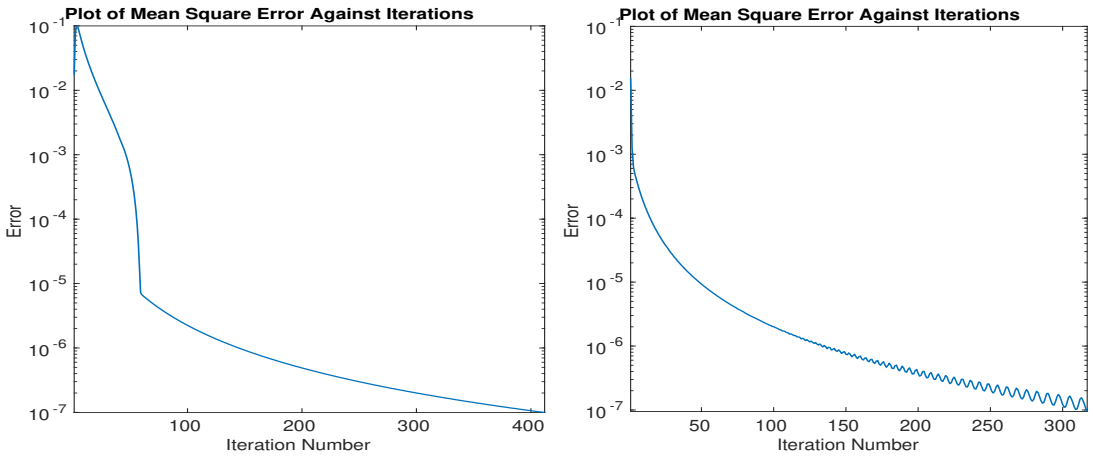
4.4.1 Geometric Brownian Motion Example

Our first example is one where the marginals follow a Geometric Brownian Motion, with $M_0 = 1$, $\bar{\sigma}(t) = 0.3\sqrt{t}$, between times $T_1 = 1$, and $T_2 = 2$. Our marginals are therefore given by $\mu_t = \text{Lognormal}\left(-\frac{1}{2}\bar{\sigma}(t)^2, \bar{\sigma}(t)\right)$ for $t \in \{1, 2\}$, and our volatility function is given by $\sigma(t, x) \equiv 0.3$. In both the Bass and Geometric Bass cases, we took our spatial domain to be $[-2, 10]$, and iterated until an error of 1×10^{-7} was achieved. In the Bass martingale case this took 412 iterations for a total time of 92.4 seconds, corresponding to the observed mean square error in volatility being 1.99×10^{-6}. In the Geometric Bass martingale case, this took 317 iterations for a total time of 74.3 seconds. We verified that in this case (4.4.3) and (4.4.5) recovered the volatility surface $\sigma \equiv 0.3$ to an observed mean square error in local volatility being 1.05×10^{-6}.



(a) (b)

Figure 4.19: Geometric Brownian Motion Example. (a) Plot of CDF Bass martingale interpolation, (b) Plot of CDF Geometric Bass martingale interpolation.



(a) (b)

Figure 4.20: Geometric Brownian Motion Example. (a) Plot of mean square error as a function of iteration for the Bass martingale case, (b) Plot of mean square error as a function of iteration for the Geometric Bass martingale case.

We generated the model prices for 121 different strikes in the range $[0.4, 1.6]$ for call options along with their implied volatilities. The errors are reported in the table below. Since our MPMS method calibrates the CDFs of the underlying at discrete times, we computed the model prices using numerical integration with respect to the model CDF.

Maturity	Call Price Errors		Implied Volatility Errors	
	Max Bass	Mean Bass	Max Bass	Mean Bass
$T = 1$	2.96×10^{-4}	1.41×10^{-4}	2.01×10^{-3}	5.65×10^{-4}
$T = 2$	1.46×10^{-5}	9.38×10^{-6}	1.46×10^{-4}	3.44×10^{-5}

Table 4.1: Pricing and IV errors in Geometric Brownian motion example: Bass Case

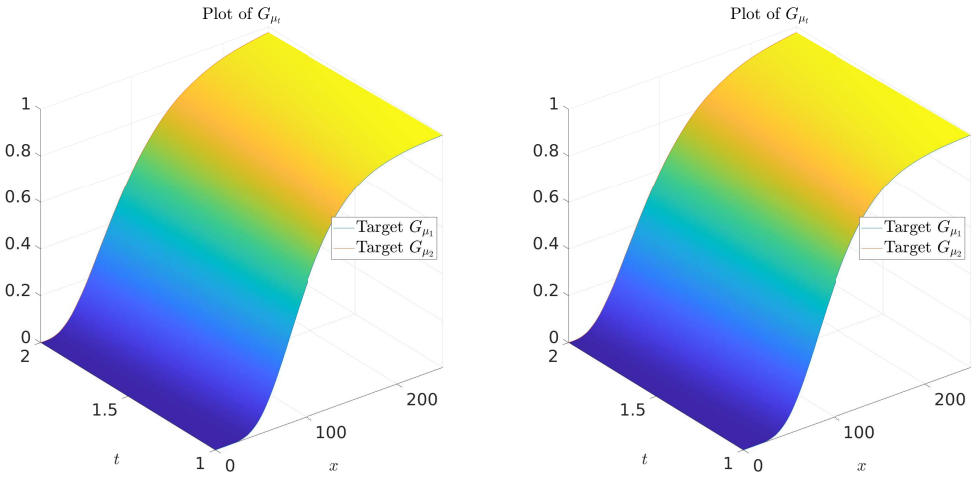
Maturity	Call Price Errors		Implied Volatility Errors	
	Max GBass	Mean GBass	Max GBass	Mean GBass
$T = 1$	2.80×10^{-4}	1.01×10^{-4}	1.93×10^{-3}	4.55×10^{-4}
$T = 2$	6.22×10^{-5}	4.57×10^{-5}	2.07×10^{-4}	1.20×10^{-4}

Table 4.2: Pricing and IV errors in Geometric Brownian motion example: Geometric Bass Case

4.4.2 CEV Model Example

Our next example is a simulated data example, under the CEV model of J. Cox, 1996. Our volatility function is given by $\sigma_{\text{CEV}}(t, x) = \bar{\sigma} x^{\gamma}$, for parameters $\bar{\sigma} = 1$ and $\gamma = -0.2$. We take an initial value of $S_0 = 100$, and perform the rescaling $x \mapsto \frac{x}{S_0}$ for numerical stability in

our PDE solution. In the Bass case, we ran the MPMS algorithm until an error of 1×10^{-7} was obtained, which took 309 iterations for a total time of 73.3 seconds; this corresponded to a mean square error in volatility of 8.74×10^{-6}. In the Geometric Bass case, we found that constructing the reflected measure introduced numerical errors and so we could not attain the same mean square error as in the Bass case. We ran the MPMS algorithm for 500 iterations, and achieved a mean square error of 1.26×10^{-5}, for a total computational time of 115.7 seconds. We verified again here that (4.4.3) and (4.4.5) recovered $\sigma_{\text{CEV}}(t, x)$ to an observed mean square error in local volatility of 4.59×10^{-5}.



(a) (b)

Figure 4.21: CEV model example. (a) Plot of CDF Bass martingale interpolation, (b) Plot of CDF Geometric Bass martingale interpolation.

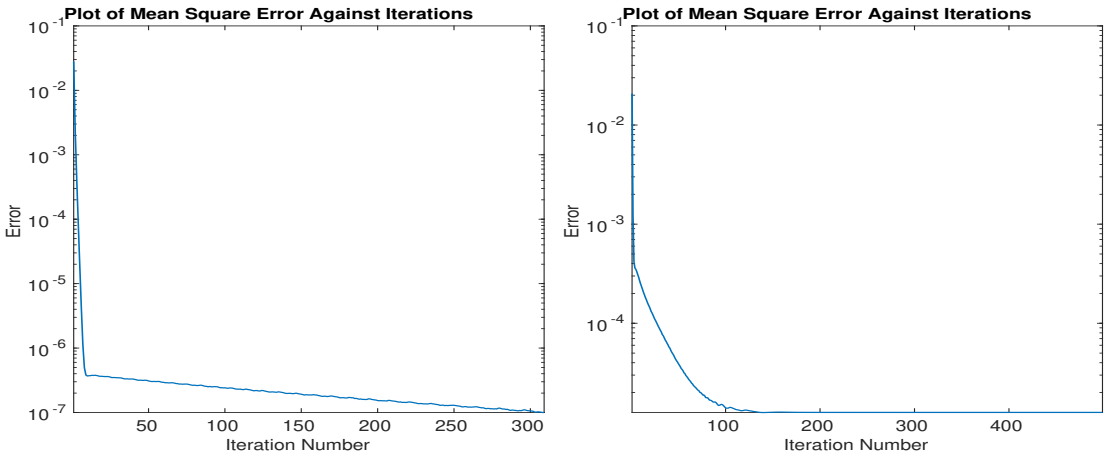


Figure 4.22: CEV model example. (a) Plot of mean square error as a function of iteration for the Bass martingale case, (b) Plot of mean square error as a function of iteration for the Geometric Bass martingale case.

We generated the model prices for 121 different strikes in the range $[40, 160]$ for call options along with their implied volatilities. The errors are reported in the table below. As before, we use numerical integration with respect to the model CDF to compute the option prices.

Maturity	Call Price Errors		Implied Volatility Errors	
	Max Bass	Mean Bass	Max Bass	Mean Bass
$T = 1$	2.14×10^{-4}	1.06×10^{-4}	9.31×10^{-4}	1.06×10^{-4}
$T = 2$	1.02×10^{-5}	4.99×10^{-6}	4.76×10^{-5}	1.37×10^{-5}

Table 4.3: Pricing and IV errors in CEV example: Bass Case

Maturity	Call Price Errors		Implied Volatility Errors	
	Max GBass	Mean GBass	Max GBass	Mean GBass
$T = 1$	2.08×10^{-3}	9.31×10^{-4}	5.28×10^{-5}	3.15×10^{-5}
$T = 2$	8.36×10^{-3}	6.85×10^{-3}	7.34×10^{-4}	1.92×10^{-4}

Table 4.4: Pricing and IV errors in CEV example: Geometric Bass Case

4.4.3 Market Data Examples

We close this Chapter by comparing local volatility construction on market data call option implied volatilities provided by BNP Paribas. In our previous simulated examples, the generating model was a martingale and therefore we could directly apply the MPMS

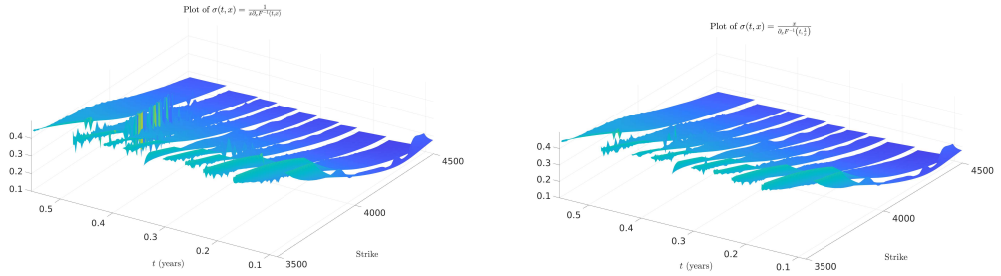
algorithm; however when dealing with market data here we must incorporate the discount and dividends. To overcome this issue, we instead consider call options on the forwards $F_{t,T} = \frac{S_t D_t}{D_T} - d_t$, where D_t is the discount factor at time t, and d_t are the cash dividends. We make the assumption that the dividends are issued at T_i^+ for each maturity, and thus consider the independent problems on $[T_i^+, T_{i+1}]$. For each maturity, we calculate the Black-Scholes price with respect to the forward: $C(T,K) = \mathrm{BS}(F_{0,T}, T, K, \sigma_{\mathrm{Imp}})$, where σ_{Imp} is the observed implied volatility. For each T_i, we then construct the model CDFs using the Breeden-Litzenberger formula, $\partial_K C(T_i, K) + 1 = G_{\mu_{T_i}}$. Next, we compute the "implied dividends" by solving $\int_{\mathbb{R}} F_{T_i^+, T_{i+1}}(x)\,\mathrm{d}\mu_{T_i} = \int_{\mathbb{R}} x\,\mathrm{d}\mu_{T_{i+1}}$. We can therefore define the function $g^i(x) = x\frac{D_{T_i}}{D_{T_{i+1}}} - d_{T_i}$, so that we have $F_{T_i,T_{i+1}} \sim g^i_{\#}\mu_i$, and so on each interval $[T_i^+, T_{i+1}]$, we apply the MPMS algorithm to the marginals $(g^i_{\#}\mu_{T_i}, \mu_{T_{i+1}})$. We additionally extrapolated the CDF obtained via linear interpolation: at strike $K = \$0$, the call price is equal to the forward; at strike $K = \$10,000$ (which is significantly higher than the maximum strike with liquidity), we set the CDF to be equal to 1.

When computing the price surfaces between maturity, we used numerical integration with respect to the Bass (or geometric Bass) interpolated CDFs. To compute the implied volatility surfaces, we need a continuum of forwards, so in the absence of other information, we made the assumption that the short rate was piecewise constant in-between maturities, and thus computing the forward curve was trivial for $t \in [T_i^+, T_{i+1}]$.

Case 1: Maturities of 27/11/2023, 04/12/2023, 29/12/2023, 19/01/2024, 31/01/2024, 16/02/2024, 29/02/2024, 15/03/2024, 28/03/2024, 19/04/2024, and 17/05/2024, as observed on 27/10/2023.

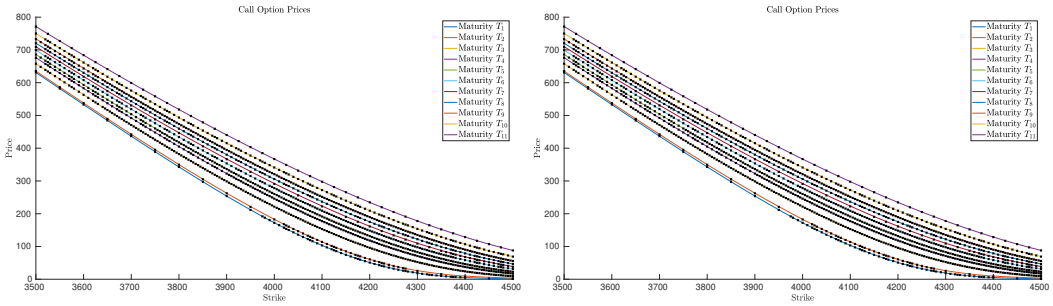
Case 2: Maturities of 04/06/2021, 18/06/2021, 30/06/2021, 16/07/2021, 30/07/2021, 20/08/2021, 31/08/2021, 17/09/2021, 30/09/2021, 1915/10/2021, and 19/11/2021, as observed on 27/04/2021.

As in Section 4.4.2, we perform the rescaling $S_T \mapsto \frac{S_T}{S_0}$, and our computational domain is $[-0.5, 3] \times [T_k, T_{k+1}]$ for $k = 1, \ldots, n-1$ where $T_1, \ldots, T_n$ are the maturities of the options, with 1001 spatial gridpoints and 101 time gridpoints. We ran the MPMS algorithm either to a tolerance of 1×10^{-10} or for 2000 iterations, which took a total of just under 5 minutes in all cases using the parallel computing toolbox in matlab on a laptop.



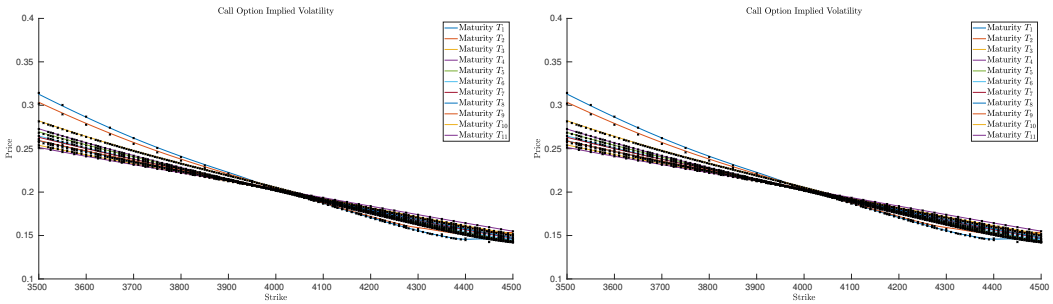
(a) (b)

Figure 4.23: Market data case 1. (a) Bass local volatility, (b) Geometric Bass local volatility.



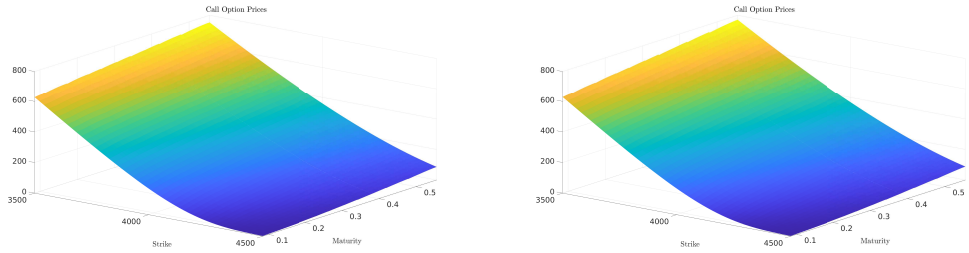
(a) (b)

Figure 4.24: Market data case 1. (a) Call prices in Bass case, (b) Call prices in Geometric Bass case.



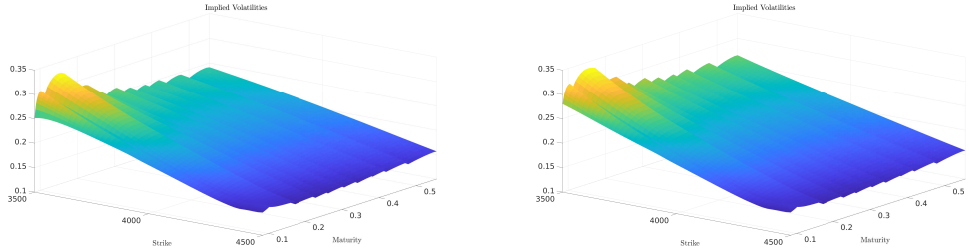
(a) (b)

Figure 4.25: Market data case 1. (a) Implied volatilities in Bass case, (b) Implied volatilities in Geometric Bass case.



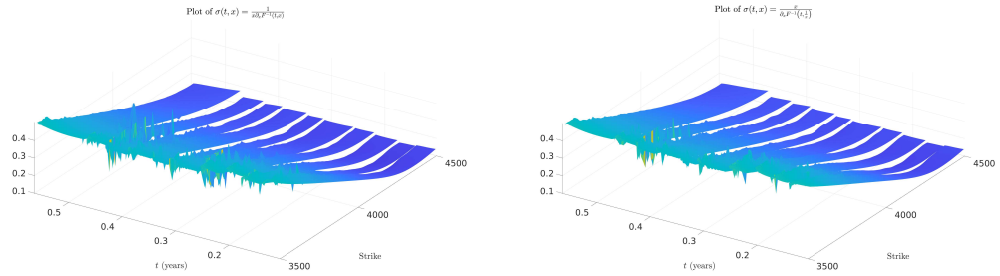
(a) (b)

Figure 4.26: Market data case 1. (a) Call price surface in Bass case, (b) Call price surface in Geometric Bass case.



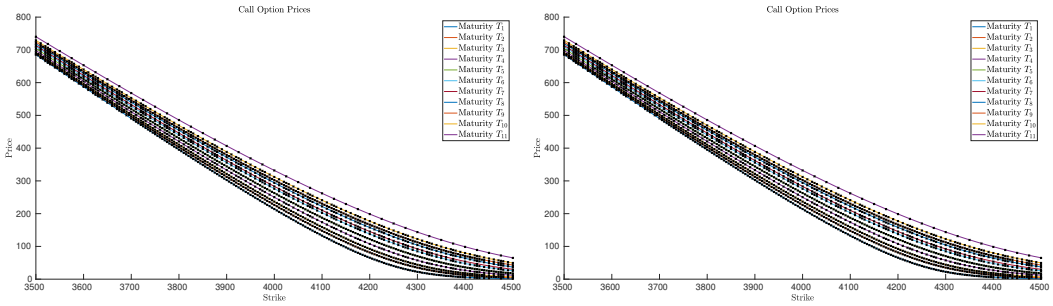
(a) (b)

Figure 4.27: Market data case 1. (a) Implied volatility surface in Bass case, (b) Implied volatility surface in Geometric Bass case.



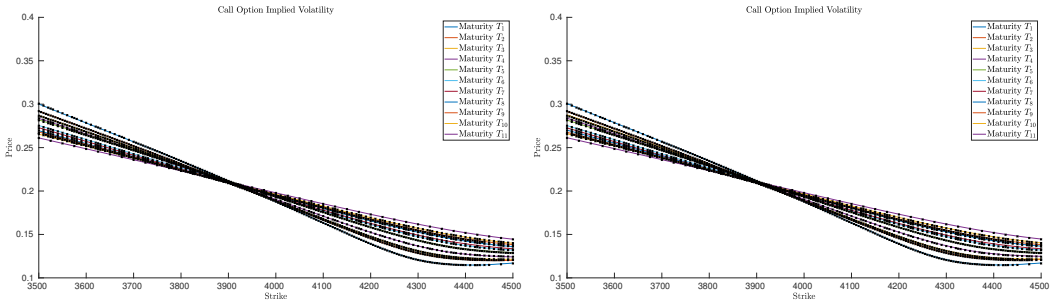
(a) (b)

Figure 4.28: Market data case 2. (a) Bass local volatility, (b) Geometric Bass local volatility.



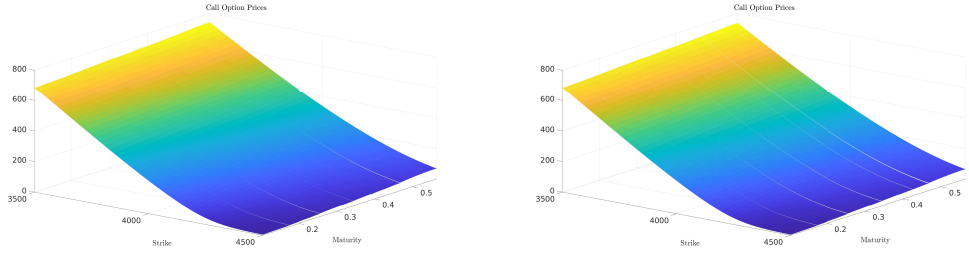
(a) (b)

Figure 4.29: Market data case 2. (a) Call prices in Bass case, (b) Call prices in Geometric Bass case.



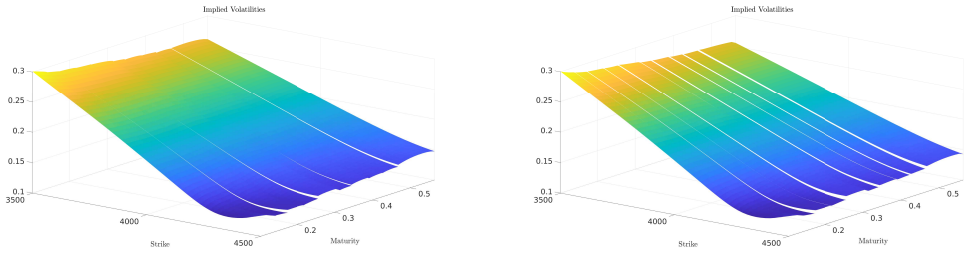
(a) (b)

Figure 4.30: Market data case 2. (a) Implied volatilities in Bass case, (b) Implied volatilities in Geometric Bass case.



(a) (b)

Figure 4.31: Market data case 2. (a) Call price surface in Bass case, (b) Call price surface in Geometric Bass case.



(a) (b)

Figure 4.32: Market data case 2. (a) Implied volatility surface in Bass case, (b) Implied volatility surface in Geometric Bass case.

4.4.4 Discussion on the Results

In Section 4.4.1 and Section 4.4.2, we rapidly obtain in both cases the correct local volatility surface, and thus the correct model prices and implied volatilities. In our market data comparisons of Section 4.4.3, we observed significantly more stable volatility surfaces in the Bass case, along with a more realistic continuous time model with the call prices. In part, this is due to the fact that we have to first compute $\nu_i = \mathrm{Id}_{\ddagger}\mu_i$, which requires numerical interpolation in practice and run the MPMS algorithm on the ν_i. The volatility surface instabilities in both cases arise from the fact that computing $\partial_x F^{-1}$ requires numerical inversion and differentiation, which can accumulate numerical instabilities and errors. The more pronounced instabilities in σ_{GBass}, in comparison with σ_{Bass}, evidently are due to the fact that $\partial_x F^{-1}$ then has to be evaluated at $\frac{m}{x}$, accumulating further numerical instabilities. These instabilities (in both cases) appear at maturities, since in-between maturities $F(t,\cdot)$ is computed using the heat equation which has a smoothing effect. In both cases, the volatility surfaces are piecewise continuous in time, however this is expected as a feature of the model since the surfaces are constructed independently on each $[T_i^+, T_{i+1}]$, and we performed the dividend shifting via $\mu_{T_i^+} = g^i_{\#}\mu_{T_i}$ at maturity. This can also be seen in the piecewise continuous call price surfaces, as we are in reality constructing a diffusion on the forward curves. Finally, we notice that we obtain a good accuracy in calibration to the market prices and implied volatilities in both cases.

Chapter 5

Conclusion

In Chapter 2, we developed a non-parametric optimal-transport driven method to calibrate a local volatility model in a stochastic interest rate environment. While standard mimicking results from Brunick and Shreve, 2013 give the existence of a Markov model, the model increases in dimension due to the presence of the stochastic discount factor $Y_t = \exp\left(-\int_0^t r_s \, \mathrm{d}s\right)$ along with the state $X = (\mathbf{S}_t, r_t)$. We derived via Lemma 2.2.8 a transformation into sub-probability measures, i.e. the aforementioned "discounted densities", that reduces the dimension to that of the state, X. Moreover, the discounted superposition principle of Theorem 2.2.7 shows that we lose no generality in a European option calibration setting by considering Markov models with coefficients just depending on X, as opposed to (X, Y). The discounted density transformation is generic to calibration with any European options with payoff functions jointly on the underlying and the interest rate, that are separable with the discount factor. Within our Markov setting, we derive a generic duality result via standard techniques in Theorem 2.5.1 allowing us to numerically solve the problem via standard PDE techniques.

The setup allows a modeller to flexibly impose their prior beliefs on the calibration procedure with a reference model representing their best approximation of the market dynamics, and bounds penalising extreme volatility values. These bounds could also depend on time, log-stock and stochastic rate to find a calibrated model bounded between two parametric volatility surfaces, in analogy with the uncertain volatility model of Avellaneda, Friedman, et al., 1997. As seen in Chapter 3, the duality result we generated is applicable in a variety of

different calibration methods since the cost function itself determines the style of calibration required. The sequential calibration method of Section 3.2 gives a fast method to calibrate a local volatility model to European Call options at the expense of strong assumptions on the correlation, so in practice a good prior estimate of the correlation would be required. The original joint calibration method of Section 3.3 gives an exact calibration to both stock and interest rate derivatives, at the expense of computational speed from the analytic solution to the optimisers. While the joint calibration had far less restrictive assumptions on the correlation, we found in practice these were heavily influenced by the reference model. Indeed, the reweighted approach of Section 3.3.3 did dampen the effects somewhat, however the correlation was still heavily dependent on the reference. Thus, a good prior estimate of the correlation is also necessary in the joint calibration, although to less of an extent than in sequential calibration.

The computational speed of joint calibration is significantly worsened when we impose bounds on our volatility surfaces to obtain a more realistic calibrated model, as we discovered the method would often output very spiky surfaces by prioritising larger local changes over smaller global changes to the surface. This motivated our faster joint calibration method of Section 3.3.4, which does address these issues at the expense of guaranteeing that the surfaces are the global optimiser. While these were not the full analytic solution, they coincided with the analytic solution whenever $\rho_{\text{cal}} \neq \pm 1$, which was always the case in our numerical experiments, and the computational performance improvements were significant. This could also have been a priori imposed as a constraint by shrinking the convex cone formed by the constraint $\beta \in \mathbb{S}_+^2$ if the modeller had beliefs on correlation bounds. Our full sequential calibration approach of Section 3.4.1 provided a middle ground between the two methods and is the fastest at calibrating only to call options, regardless of whether the interest rate model is correct. We compared all three approaches in Section 3.4.2 and interestingly found very small differences in the volatility and correlation parameters β_{11}, ρ between the three methods. Section 3.4.4 gives us similar results in the comparison, but also demonstrates that the sequential methods can still recover the correct prices with an incorrect interest rate, showing the robustness of the methods.

One common feature of the volatility surface in our SOT calibrated models is the interest

rate homogeneity of the local volatility β_{11}. Roughly speaking, this is because, in all three cases, the formulae for β_{11} only depend on the z-derivatives of φ and our calibrating options are not jointly dependent. However, over a long time period we observe this not to be the case in Section 3.6. In particular, the dependence arises since over a longer period of time, the discount factor will have more of an effect on the prices of the options, which is reflected in the $-x_r\varphi$ term in the HJB equation. Additionally, we observed that the calibration to significantly more options did not substantially slow down the calibration routine. We also used the full sequential method in Section 3.5 to solve the LSV calibration problem of I. Guo, Loeper, and Shiyi Wang, 2022 with less restrictive assumptions on the correlation, resulting in an exact calibration under both a good and a bad reference model.

The computational time cost of building a continuous time model was significant with restrictive bounds, the choice of the reference model, and the number of smoothing iterations all contributing towards the total computational time. While the duality result is generic in dimension and the features of state variables, the finite difference policy iteration scheme of Algorithm 3 will struggle to generate a continuous time calibrated model in higher dimensions. The main limiting step is the solution of the fully nonlinear HJB equation since the linear pricing PDEs (3.1.8) could instead have been solved via Monte Carlo methods once the optimal characteristics $(\alpha, \beta) \in \mathbb{R}^d \times \mathbb{S}_+^d$ were determined.

In part, this practical difficulty in extending the numerical method of Chapter 3 to higher dimensions inspired our approach in Chapter 4. The core idea being that calibrating a discrete time model and then extending to a continuous time model should, in principle at least, be computationally easier than directly calibrating a continuous time model. In Section 4.1, we derived the Bass martingale using the PDE framework of Xiaolu Tan and Touzi, 2013; Huesmann and Trevisan, 2019 and recovered the duality result of Backhoff-Veraguas, Beiglböck, Schachermayer, et al., 2023. Section 4.2 introduced the measure preserving martingale Sinkhorn algorithm, in analogy with the well-known Sinkhorn algorithm. In Section 4.2.2, we demonstrate a monotonicity property in any dimension, and provide a short proof of the contraction result of Acciaio, Marini, and Pammer, 2023 in one dimension.

We then demonstrate implementations of MPMS in Section 4.3 in a variety of settings with extremely fast convergence in all cases. The direct "CDF" approach of Section 4.3.1

is much easier to implement and more stable than the approximate "density approach" of Section 4.3.2. Nevertheless, it is an interesting result that our approximate method converged in one dimension. In Section 4.4, we also used the MPMS algorithm to solve the Geometric Bass martingale problem, and recovered the generating local volatility surfaces in both the Bass and Geometric Bass cases under Black-Scholes and CEV models.

We additionally provided a fast local volatility calibration with market data via the Breeden-Litzenberger formula, with results shown in the region of liquid strikes. In addition, we provided an empirical comparison of the same options observed on different dates, and found that the local volatilities were consistent in our plots. Outside of the plotted region, and where the density function takes much smaller values, we found that the surface exhibited large instabilities. This is potentially due to the local volatility being a function of $\partial_x F^{-1}$, and since $F^{-1} = G^{-1}_{\mathbf{R}_T * \alpha_0} \circ G_{\mu_T}$, and we have that G_{μ_T} is nearly constant, so $\partial_x F^{-1}$ is close to zero, which then leads to numerical instabilities in (4.4.3) and (4.4.5) outside of the region where μ_0 and μ_T have most of their mass. In the Black-Scholes and CEV examples of Section 4.4.1 and Section 4.4.2, we found this limitation in the local volatility surface to be less apparent when computing the option prices and implied volatilities by recovering both to a high accuracy in a larger range. This is perhaps due to the fact that the CDFs in the Black-Scholes case have analytic formulae, and in the CEV case the CDFs were constructed using 10^5 particles and antithetic variables, which gave us good accuracy, including at extremal values. This is in comparison with the Breeden-Litzenberger approach of Section 4.4.3, where interpolation and extrapolation are required to compute a continuous CDF from discrete options. When at the money, there are a large number of liquid options with strikes close together, so our pricing results are more accurate since the interpolation methods have less of an effect on our CDF construction. For very far in the money or out of the money options, pricing errors and sparsity of data mean that the method of interpolation has a larger effect, so the CDF we compute is further from the "true" CDF. Nevertheless, we recover the option prices and implied volatilities to a high degree of accuracy when we have a dense region of liquid options.

Bibliography

Abergel, Frédéric and Grégoire Loeper (2017). "Option pricing and hedging with liquidity costs and market impact". In: *Econophysics and Sociophysics: Recent Progress and Future Directions*. Springer, pp. 19–40.

Acciaio, Beatrice, Antonio Marini, and Gudmund Pammer (2023). "Calibration of the Bass Local Volatility model". arXiv: 2311.14567.

Ambrosio, Luigi and Alessio Figalli (2009). "On flows associated to Sobolev vector fields in Wiener spaces: an approach à la DiPerna–Lions". In: *Journal of Functional Analysis* 256.1, pp. 179–214. DOI: 10.1016/j.jfa.2008.05.007.

Ambrosio, Luigi, Nicola Gigli, and Giuseppe Savaré (2005). *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media.

Ambrosio, Luigi and Dario Trevisan (2014). "Well-posedness of Lagrangian flows and continuity equations in metric measure spaces". In: *Analysis & PDE* 7.5, pp. 1179–1234. DOI: 10.2140/apde.2014.7.1179.

Ametrano, Ferdinando and Marco Bianchetti (2013). "Everything you always wanted to know about multiple interest rate curve bootstrapping but were afraid to ask". DOI: 10.2139/ssrn.2219548.

Atlan, Marc (2006). "Localizing volatilities". arXiv: math/0604316.

Avellaneda, Marco, Craig Friedman, et al. (1997). "Calibrating volatility surfaces via relative-entropy minimization". In: *Applied Mathematical Finance* 4.1, pp. 37–64. DOI: 10.1080/135048697334827.

Avellaneda, Marco, Arnon Levy, and Antonio Parás (1995). "Pricing and hedging derivative securities in markets with uncertain volatilities". In: *Applied Mathematical Finance* 2.2, pp. 73–88. DOI: 10.1080/13504869500000005.

Bachelier, Louis (1900). "Théorie de la spéculation". In: *Annales scientifiques de l'École normale supérieure*. Vol. 17, pp. 21–86.

Backhoff-Veraguas, Julio, Mathias Beiglböck, Martin Huesmann, et al. (2020). "Martingale Benamou–Brenier: a probabilistic perspective". In: *The Annals of Probability* 48.5, pp. 2258–2289. DOI: 10.2307/26966038.

Backhoff-Veraguas, Julio, Mathias Beiglböck, Walter Schachermayer, et al. (2023). "The Structure of martingale Benamou-Brenier in $\mathbb{R}^d$". arXiv: 2306.11019.

Backhoff-Veraguas, Julio, Grégoire Loeper, and Jan Obłój (2024). "Geometric Martingale Benamou-Brenier transport and geometric Bass martingales". arXiv: 2406.04016.

Backhoff-Veraguas, Julio and Gudmund Pammer (2022). "Stability of martingale optimal transport and weak optimal transport". In: *The Annals of Applied Probability* 32.1, pp. 721–752. DOI: 10.1214/21-AAP1694.

Backhoff-Veraguas, Julio, Gudmund Pammer, and Walter Schachermayer (2024). "The Gradient Flow of the Bass Functional in Martingale Optimal Transport". arXiv: 2407.18781.

Backhoff-Veraguas, Julio, Walter Schachermayer, and Bertram Tschiderer (2023). "The Bass functional of martingale transport". arXiv: 2306.11019.

Bain, Alan, Matthieu Mariapragassam, and Christoph Reisinger (2021). "Calibration of local-stochastic and path-dependent volatility models to vanilla and no-touch options". In: *Journal of Computational Finance* 24.4. DOI: 10.21314/JCF.2020.400.

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen (1997). "Empirical performance of alternative option pricing models". In: *The Journal of finance* 52.5, pp. 2003–2049. DOI: 10.1111/j.1540-6261.1997.tb02749.x.

Barles, Guy and Espen Jakobsen (2002). "On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 36.1, pp. 33–54. DOI: 10.1051/m2an:2002002.

Barles, Guy and Panagiotis Souganidis (1991). "Convergence of approximation schemes for fully nonlinear second order equations". In: *Asymptotic Analysis* 4.3, pp. 271–283. DOI: 10.3233/ASY-1991-4305.

Bartl, Daniel et al. (2019). "Duality for pathwise superhedging in continuous time". In: *Finance and Stochastics* 23.3, pp. 697–728. DOI: 10.1007/s00780-019-00395-2.

Bass, Richard (1983). "Skorokhod imbedding via stochastic integrals". In: *Séminaire de probabilités de Strasbourg XVII 1981/82* 17, pp. 221–224. DOI: 10.1007/BFb0068318.

Beiglböck, Mathias, Alexander Cox, and Martin Huesmann (2017). "Optimal transport and Skorokhod embedding". In: *Inventiones mathematicae* 208, pp. 327–400. DOI: 10.1007/s00222-016-0692-2.

— (2020). "The geometry of multi-marginal Skorokhod embedding". In: *Probability Theory and Related Fields* 176, pp. 1045–1096. DOI: 10.1007/s00440-019-00935-z.

Beiglböck, Mathias, Pierre Henry-Labordère, and Friedrich Penkner (2013). "Model-independent bounds for option prices—a mass transport approach". In: *Finance and Stochastics* 17.3, pp. 477–501. DOI: 10.1007/s00780-013-0205-8.

Beiglböck, Mathias, Benjamin Jourdain, et al. (2023). "Stability of the weak martingale optimal transport problem". In: *The Annals of Applied Probability* 33.6B, pp. 5382–5412. DOI: 10.1214/23-AAP1950.

Beiglböck, Mathias and Nicolas Juillet (2016). "On a problem of optimal transport under marginal martingale constraints". In: *The Annals of Probability* 44.1, pp. 42–106. DOI: 10.1214/14-AOP966.

Beiglböck, Mathias, Tongseok Lim, and Jan Obłój (2019). "Dual attainment for the martingale transport problem". In: *Bernoulli* 25.3, pp. 1640–1658. DOI: 10.3150/17-BEJ1015. URL: https://doi.org/10.3150/17-BEJ1015.

Beiglböck, Mathias, Marcel Nutz, and Nizar Touzi (2016). "Complete Duality for Martingale Optimal Transport on the Line". In: *Annals of Probability* 45.5, pp. 3038–3074. DOI: 10.1214/16-AOP1131.

Beiglböck, Mathias, Gudmund Pammer, and Lorenz Riess (2024). "Change of numeraire for weak martingale transport". arXiv: 2406.07523.

Benamou, Jean-David and Yann Brenier (2000). "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In: *Numerische Mathematik* 84.3, pp. 375–393. DOI: 10.1007/s002119900117.

Benamou, Jean-David, Guillaume Carlier, et al. (2015). "Iterative Bregman projections for regularized transportation problems". In: *SIAM Journal on Scientific Computing* 37.2, A1111–A1138. DOI: 10.1137/141000439.

Benamou, Jean-David, Guillaume Chazareix, and Grégoire Loeper (2024). "From entropic transport to martingale transport, and applications to model calibration". arXiv: 2406.11537.

Benamou, Jean-David, Brittany Froese, and Adam Oberman (2010). "Two numerical methods for the elliptic Monge-Ampere equation". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 44.4, pp. 737–758. DOI: 10.1051/m2an/2010017.

— (2014). "Numerical solution of the optimal transportation problem using the Monge–Ampère equation". In: *Journal of Computational Physics* 260, pp. 107–126. DOI: 10.1016/j.jcp.2013.12.015.

Bentata, Amel and Rama Cont (2009). "Mimicking the marginal distributions of a semimartingale". arXiv: 0910.3992.

Berman, Robert (2020). "The Sinkhorn algorithm, parabolic optimal transport and geometric Monge–Ampère equations". In: *Numerische Mathematik* 145.4, pp. 771–836. DOI: 10.1007/s00211-020-01127-x.

Bianchetti, Marco and Mattia Carlicchi (2011). "Interest rates after the credit crunch: Multiple curve vanilla derivatives and sabr". DOI: 10.2139/ssrn.1783070.

Black, Fischer (1976). "The pricing of commodity contracts". In: *Journal of Financial Economics* 3.1-2, pp. 167–179. DOI: 10.1016/0304-405X(76)90024-6.

Black, Fischer and Piotr Karasinski (1991). "Bond and option pricing when short rates are lognormal". In: *Financial Analysts Journal* 47.4, pp. 52–59. DOI: 10.2469/faj.v47.n4.52.

Black, Fischer and Myron Scholes (1973). "The Pricing of Options and Corporate Liabilities". In: *The Journal of Political Economy* 81.3, pp. 637–654. DOI: 10.1086/260062.

Boenkost, Wolfram and Wolfgang M Schmidt (2005). "Cross currency swap valuation". DOI: 10.2139/ssrn.1375540.

Bokanowski, Olivier, Stefania Maroso, and Hasnaa Zidani (2009). "Some convergence results for Howard's algorithm". In: *SIAM Journal on Numerical Analysis* 47.4, pp. 3001–3026. DOI: 10.1137/08073041X.

Bonnans, J Frédéric and Housnaa Zidani (2003). "Consistency of generalized finite difference schemes for the stochastic HJB equation". In: *SIAM Journal on Numerical Analysis* 41.3, pp. 1008–1021. DOI: 10.1137/S0036142901387336.

Bouchard, Bruno, Grégoire Loeper, Halil Mete Soner, et al. (2019). "Second-order stochastic target problems with generalized market impact". In: *SIAM J. Control Optim.* 57.6, pp. 4125–4149. ISSN: 0363-0129. DOI: 10.1137/18M1196078.

Bouchard, Bruno, Grégoire Loeper, and Yiyi Zou (2016). "Almost-sure hedging with permanent price impact". In: *Finance Stoch.* 20.3, pp. 741–771. ISSN: 0949-2984. DOI: 10.1007/s00780-016-0295-1.

— (2017). "Hedging of covered options with linear market impact and gamma constraint". In: *SIAM Journal on Control and Optimization* 55.5, pp. 3319–3348. DOI: 10.1137/15M1054109.

Box, George and Norman Draper (1987). *Empirical model-building and response surfaces.* John Wiley & Sons.

Brace, Alan, Tim Dun, and Geoff Barton (2001). "Towards a central interest rate model". In: *Option Pricing, Interest Rates and Risk Management*, pp. 278–313.

Brace, Alan, Dariusz Gątarek, and Marek Musiela (1997). "The market model of interest rate dynamics". In: *Mathematical Finance* 7.2, pp. 127–155. DOI: 10.1111/1467-9965.00028.

Breeden, Douglas and Robert Litzenberger (1978). "Prices of state-contingent claims implicit in option prices". In: *Journal of business*, pp. 621–651. DOI: 10.2307/2352653.

Bregman, Lev M (1967). "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming". In: *USSR computational mathematics and mathematical physics* 7.3, pp. 200–217. DOI: 10.1016/0041-5553(67)90040-7.

Brenier, Yann (1987). "Décomposition polaire et réarrangement monotone des champs de vecteurs". In: *CR Acad. Sci. Paris Sér. I Math.* 305, pp. 805–808.

— (1991). "Polar factorization and monotone rearrangement of vector-valued functions". In: *Communications on pure and applied mathematics* 44.4, pp. 375–417. DOI: 10.1002/cpa.3160440402.

— (1997). "A homogenized model for vortex sheets". In: *Archive for Rational Mechanics and Analysis* 138.4, pp. 319–353. DOI: 10.1007/s002050050044.

— (1999). "Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 52.4, pp. 411–452. DOI: 10.1002/(SICI)1097-0312(199904)52:4%3C411::AID-CPA1%3E3.0.CO;2-3.

Brézis, Haim (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. eng. 1st ed. 2011. Universitext. New York, NY: Springer New York. ISBN: 1-283-47619-3.

Brigo, Damiano and Fabio Mercurio (2002). "Lognormal-mixture dynamics and calibration to market volatility smiles". In: *International Journal of Theoretical and Applied Finance* 5.04, pp. 427–446. DOI: 10.1142/S0219024902001511.

— (2007). *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Science & Business Media.

Brunick, Gerard and Steven Shreve (2013). "Mimicking an Itô process by a solution of a stochastic differential equation". In: *Annals of Applied Probability* 23.4, pp. 1584–1628. DOI: 10.1214/12-AAP881.

Caffarelli, Luis (1996). "Boundary regularity of maps with convex potentials–II". In: *Annals of mathematics* 144.3, pp. 453–496. DOI: 10.2307/2118564.

— (2017). "Allocation maps with general cost functions". In: *Partial differential equations and applications*. Routledge, pp. 29–35.

Cairns, Andrew (2018). *Interest rate models*. Princeton University Press.

Campi, Luciano, Ismail Laachir, and Claude Martini (2017). "Change of numeraire in the two-marginals martingale transport problem". In: *Finance and Stochastics* 21, pp. 471–486. DOI: 10.1007/s00780-016-0322-2.

Carlier, Guillaume (2003). "On a class of multidimensional optimal transportation problems". In: *Journal of convex analysis* 10.2, pp. 517–530.

Chizat, Lenaic et al. (2018). "Unbalanced optimal transport: Dynamic and Kantorovich formulations". In: *Journal of Functional Analysis* 274.11, pp. 3090–3123. DOI: 10.1016/j.jfa.2018.03.008.

Conze, Antoine and Pierre Henry-Labordere (2021). "Bass Construction with Multi-Marginals: Lightspeed Computation in a New Local Volatility Model". DOI: 10.2139/ssrn.3853085.

Corrias, Lucilla (1996). "Fast Legendre–Fenchel transform and applications to Hamilton–Jacobi equations and conservation laws". In: *SIAM journal on numerical analysis* 33.4, pp. 1534–1558. DOI: 10.1137/S0036142993260208.

Cox, John (1996). "The constant elasticity of variance option pricing model". In: *Journal of Portfolio Management*, pp. 15–17. DOI: 10.3905/jpm.1996.015.

Cox, John, Jonathan Ingersoll Jr, and Stephen Ross (1985). "A Theory of the Term Structure of Interest Rates". In: *Econometrica* 53.2, pp. 385–408. DOI: 10.2307/1911242.

Cox, John and Stephen Ross (1976). "The valuation of options for alternative stochastic processes". In: *Journal of financial economics* 3.1-2, pp. 145–166. DOI: 10.1016/0304-405X(76)90023-4.

Cozma, Andrei, Matthieu Mariapragassam, and Christoph Reisinger (2019). "Calibration of a hybrid local-stochastic volatility stochastic rates model with a control variate particle method". In: *SIAM Journal on Financial Mathematics* 10.1, pp. 181–213. DOI: 10.1137/17M1114570.

Crandall, Michael, Hitoshi Ishii, and Pierre-Louis Lions (1992). "User's guide to viscosity solutions of second order partial differential equations". In: *Bulletin of the American Mathematical Society* 27.1, pp. 1–67. DOI: 10.1090/S0273-0979-1992-00266-5.

Crépey, Stéphane, Zorana Grbac, Nathalie Ngor, et al. (2015). "A Lévy HJM multiple-curve model with application to CVA computation". In: *Quantitative Finance* 15.3, pp. 401–419. DOI: 10.1080/14697688.2014.942232.

Crépey, Stéphane, Zorana Grbac, and Hai-Nam Nguyen (2012). "A multiple-curve HJM model of interbank risk". In: *Mathematics and Financial Economics* 6.3, pp. 155–190. DOI: 10.1007/s11579-012-0083-4.

Cuturi, Marco (2013). "Sinkhorn distances: Lightspeed computation of optimal transport". In: *Advances in neural information processing systems* 26.

De March, Hadrien (2018). "Entropic approximation for multi-dimensional martingale optimal transport". In: arXiv: 1812.11104.

De March, Hadrien and Pierre Henry-Labordere (2019). "Building arbitrage-free implied volatility: Sinkhorn's algorithm and variants". arXiv: 1902.04456.

De Philippis, Guido and Alessio Figalli (2014). "The Monge–Ampère equation and its link to optimal transportation". In: *Bulletin of the American Mathematical Society* 51.4, pp. 527–580. DOI: 10.1090/S0273-0979-2014-01459-4.

Debrabant, Kristian and Espen Jakobsen (2013). "Semi-Lagrangian schemes for linear and fully non-linear diffusion equations". In: *Mathematics of Computation* 82.283, pp. 1433–1462. DOI: 10.2307/42002704.

Deelstra, Griselda and Grégory Rayée (2013). "Local volatility pricing models for long-dated FX derivatives". In: *Applied Mathematical Finance* 20.4, pp. 380–402. DOI: 10.1080/1350486X.2012.723516.

Delbaen, Freddy and Walter Schachermayer (1994). "A general version of the fundamental theorem of asset pricing". In: *Mathematische annalen* 300.1, pp. 463–520. DOI: 10.1007/BF01450498.

Derman, Emanuel and Iraj Kani (1994). "Riding on a smile". In: *Risk* 7.2, pp. 32–39.

— (1998). "Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility". In: *International journal of theoretical and applied finance* 1.01, pp. 61–110. DOI: 10.1142/S0219024998000059.

Doldi, Alessandro and Marco Frittelli (2023). "Entropy martingale optimal transport and nonlinear pricing–hedging duality". In: *Finance and Stochastics* 27.2, pp. 255–304. DOI: 10.1007/s00780-023-00498-x.

Dolinsky, Yan and Halil Mete Soner (2014). "Martingale optimal transport and robust hedging in continuous time". In: *Probability Theory and Related Fields* 160.1-2, pp. 391–427. DOI: 10.1007/s00440-013-0531-y.

Dupire, Bruno (1994). "Pricing with a smile". In: *Risk* 7.1, pp. 18–20.

— (1996). "A unified theory of volatility". In: *Derivatives Pricing: The Classic Collection*, pp. 185–196.

Dupire, Bruno (1997). "Pricing and hedging with smiles". In: *Mathematics of derivative securities* 1.1, pp. 103–111.

Eckstein, Stephan, Gaoyue Guo, et al. (2021). "Robust pricing and hedging of options on multiple assets and its numerics". In: *SIAM Journal on Financial Mathematics* 12.1, pp. 158–188. DOI: 10.1137/19M1286256.

Eckstein, Stephan and Michael Kupper (2021). "Computation of optimal transport and related hedging problems via penalization and neural networks". In: *Applied Mathematics & Optimization* 83.2, pp. 639–667. DOI: 10.1007/s00245-019-09558-1.

Engelmann, Bernd, Frank Koster, and Daniel Oeltz (2021). "Calibration of the Heston stochastic local volatility model: A finite volume scheme". In: *International Journal of Financial Engineering* 8.01, p. 2050048. DOI: 10.1142/S2424786320500486.

Engstrom, Linn, Sigrid Kallblad, and Johan Karlsson (2024). "Computation of Robust Option Prices via Structured Multi-Marginal Martingale Optimal Transport". arXiv: 2406.09959.

Ethier, Stewart and Thomas Kurtz (1986). *Markov processes: characterization and convergence*. John Wiley & Sons.

Figalli, Alessio (2008). "Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients". In: *Journal of Functional Analysis* 254.1, pp. 109–153. DOI: 10.1016/j.jfa.2007.09.020.

Filipovic, Damir (2009). *Term-Structure Models. A Graduate Course*. Springer.

Forsyth, Peter A and George Labahn (2007). "Numerical methods for controlled Hamilton-Jacobi-Bellman PDEs in finance". In: *Journal of Computational Finance* 11.2, p. 1. DOI: 10.21314/JCF.2007.163.

Forsyth, Peter A and Kenneth R Vetzal (2011). "Numerical methods for nonlinear PDEs in finance". In: *Handbook of computational finance*. Springer, pp. 503–528. DOI: 10.1007/978-3-642-17254-0_18.

Fortin, Michel and Roland Glowinski (1983). "Augmented Lagrangian methods, volume 15 of Studies in Mathematics and its Applications".

Franklin, Joel and Jens Lorenz (1989). "On the scaling of multidimensional matrices". In: *Linear Algebra and its applications* 114, pp. 717–735. DOI: 10.1016/0024-3795(89)90490-4.

Galichon, Alfred, Pierre Henry-Labordère, and Nizar Touzi (2014). "A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options". In: *The Annals of Applied Probability* 24.1, pp. 312–336. DOI: 10.1214/13-AAP925.

Gangbo, Wilfrid and Robert McCann (1996). "The geometry of optimal transportation". In: *Acta Mathematica* 177.2, pp. 113–161. DOI: 10.1007/BF02392620.

Gangbo, Wilfrid and Andrzej Święch (1998). "Optimal maps for the multidimensional Monge-Kantorovich problem". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 51.1, pp. 23–45. DOI: 10.1002/(SICI)1097-0312(199801)51:1<23::AID-CPA2>3.0.CO;2-H.

Ghosal, Promit and Marcel Nutz (2022). "On the Convergence Rate of Sinkhorn's Algorithm". arXiv: 2212.06000.

Ghosal, Promit, Marcel Nutz, and Espen Bernton (2022). "Stability of entropic optimal transport and Schrödinger bridges". In: *Journal of Functional Analysis* 283.9, p. 109622. DOI: 10.1016/j.jfa.2022.109622.

Guo, Gaoyue and Jan Obłój (2019). "Computational methods for martingale optimal transport problems". In: *The Annals of Applied Probability* 29.6, pp. 3311–3347. DOI: 10.2307/26891083.

Guo, Ivan, Nicolas Langrené, et al. (2022). "Portfolio optimization with a prescribed terminal wealth distribution". In: *Quantitative Finance* 22.2, pp. 333–347. DOI: 10.1080/14697688.2021.1967432.

Guo, Ivan and Grégoire Loeper (2021). "Path dependent optimal transport and model calibration on exotic derivatives". In: *The Annals of Applied Probability* 31.3, pp. 1232–1263. DOI: 10.1214/20-AAP1617.

Guo, Ivan, Grégoire Loeper, Jan Obłój, et al. (2022a). "Joint Modeling and Calibration of SPX and VIX by Optimal Transport". In: *SIAM Journal on Financial Mathematics* 13.1, pp. 1–31. DOI: 10.1137/20M1375905.

— (2022b). "Optimal transport for model calibration". In: *Risk Magazine*.

Guo, Ivan, Grégoire Loeper, and Shiyi Wang (2019). "Local volatility calibration by optimal transport". In: *2017 MATRIX Annals*. Springer, pp. 51–64. DOI: 10.1007/978-3-030-04161-8_5.

— (2022). "Calibration of local-stochastic volatility models by optimal transport". In: *Mathematical Finance* 32.1, pp. 46–77. DOI: 10.1111/mafi.12335.

Gutiérrez, Cristian and Haim Brezis (2001). *The Monge-Ampere equation*. Vol. 44. Springer.

Guyon, Julien (2020). "The joint S&P 500/VIX smile calibration puzzle solved". In: *Risk, April*.

— (2024). "Dispersion-constrained martingale Schrödinger problems and the exact joint S&P 500/VIX smile calibration puzzle". In: *Finance and Stochastics* 28.1, pp. 27–79. DOI: 10.1007/s00780-023-00524-y.

Guyon, Julien and Florian Bourgey (2022). "Fast Exact Joint S&P 500/VIX Smile Calibration in Discrete and Continuous Time". DOI: 10.2139/ssrn.4315084.

Guyon, Julien and Pierre Henry-Labordere (2011). "The smile calibration problem solved". DOI: 10.2139/ssrn.1885032.

Guyon, Julien and Pierre Henry-Labordère (2012). "Being particular about calibration". In: *Risk* 25.1, p. 88.

Gyöngy, István (1986). "Mimicking the one-dimensional marginal distributions of processes having an Itô differential". In: *Probability Theory and Related Fields* 71.4, pp. 501–516. DOI: 10.1007/BF00699039.

Gyöngy, István and Nikolai Krylov (1980). "On stochastic equations with respect to semimartingales I." In: *Stochastics: An International Journal of Probability and Stochastic Processes* 4.1, pp. 1–21. DOI: 10.1080/03610918008833154.

Hagan, Patrick et al. (2002). "Managing smile risk". In: *The Best of Wilmott* 1, pp. 249–296.

Harrison, J Michael and David M Kreps (1979). "Martingales and arbitrage in multiperiod securities markets". In: *Journal of Economic theory* 20.3, pp. 381–408. DOI: 10.1016/0022-0531(79)90043-7.

Heath, David, Robert Jarrow, and Andrew Morton (1992). "Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation". In: *Econometrica: Journal of the Econometric Society*, pp. 77–105. DOI: 10.2307/2951677.

Henrard, Marc (2007). "The Irony in the Derivatives Discounting". In: *Wilmott Magazine*, pp. 92–98.

— (2010). "The irony in derivatives discounting part II: The crisis". In: *Wilmott Journal* 2.6, pp. 301–316. DOI: 10.1002/wilj.39.

Henry-Labordère, Pierre (2009). "Calibration of local stochastic volatility models to market smiles: A Monte-Carlo approach". In: *Risk Magazine, September*.

Henry-Labordère, Pierre, Jan Obłój, et al. (2016). "The maximum maximum of a martingale with given n marginals". In: *Annals of Applied Probability* 26.1, pp. 1–44. DOI: 10.2307/43859592.

Henry-Labordère, Pierre and Nizar Touzi (2016). "An explicit martingale version of the one-dimensional Brenier theorem". In: *Finance and Stochastics* 20.3, pp. 635–668. DOI: 10.1007/s00780-016-0299-x.

Heston, Steven (1993). "A closed-form solution for options with stochastic volatility with applications to bond and currency options". In: *The Review of Financial Studies* 6.2, pp. 327–343. DOI: 10.1093/rfs/6.2.327.

Ho, Thomas SY and Sang-Bin Lee (1986). "Term structure movements and pricing interest rate contingent claims". In: *The Journal of Finance* 41.5, pp. 1011–1029. DOI: 10.1111/j.1540-6261.1986.tb02528.x.

Hobson, David (1998). "Robust hedging of the lookback option". In: *Finance and Stochastics* 2, pp. 329–347. DOI: 10.1007/s007800050044.

Hobson, David and Anthony Neuberger (2012). "Robust bounds for forward start options". In: *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 22.1, pp. 31–56. DOI: 10.1111/j.1467-9965.2010.00473.x.

Hok, Julien and Shih-Hau Tan (2019). "Calibration of local volatility model with stochastic interest rates by efficient numerical PDE methods". In: *Decisions in Economics and Finance* 42.2, pp. 609–637. DOI: 10.1007/s10203-019-00232-3.

Horn, Roger and Charles Johnson (2012). *Matrix analysis*. Cambridge university press.

Hou, Zhaoxu and Jan Obłój (2018). "Robust pricing–hedging dualities in continuous time". In: *Finance and Stochastics* 22.3, pp. 511–567. DOI: 10.1007/s00780-018-0363-9.

Huesmann, Martin and Dario Trevisan (2019). "A Benamou–Brenier formulation of martingale optimal transport". In: *Bernoulli* 25.4A, pp. 2729–2757. DOI: 10.3150/18-BEJ1069.

Hull, John (1994). "Numerical procedures for implementing term structure models II: Two-factor models". In: *The Journal of Derivatives* 2, pp. 37–48. DOI: 10.3905/jod.1994.407908.

Hull, John and Alan White (1987). "The pricing of options on assets with stochastic volatilities". In: *The journal of finance* 42.2, pp. 281–300. DOI: 10.1111/j.1540-6261.1987.tb02568.x.

— (1990). "Pricing interest-rate-derivative securities". In: *The Review of Financial Studies* 3.4, pp. 573–592. DOI: 10.1093/rfs/3.4.573.

— (1994). "Branching out". In: *Risk* 7.7, pp. 34–37.

— (1995). "A note on the models of Hull and White for pricing options on the term structure: Response". In: *The Journal of Fixed Income* 5.2, pp. 97–102. DOI: 10.3905/jfi.1995.408139.

Jacobs, Matt and Flavien Léger (2020). "A fast approach to optimal transport: The back-and-forth method". In: *Numerische Mathematik* 146.3, pp. 513–544. DOI: 10.1007/s00211-020-01154-8.

Jacod, Jean and Albert Shiryaev (2013). *Limit theorems for stochastic processes*. Vol. 288. Springer Science & Business Media.

Jamshidian, Farshid (1997). "LIBOR and swap market models and measures". In: *Finance and Stochastics* 1.4, pp. 293–330. DOI: 10.1007/s007800050026.

Jean-David, Songting Luo, and Hongkai Zhao (2010). "A compact upwind second order scheme for the eikonal equation". In: *Journal of Computational Mathematics*, pp. 489–516. DOI: 10.2307/43693919.

Jex, Mark, Robert Henderson, and David Wang (1999). "Pricing exotics under the smile". In: *Risk Magazine*, pp. 72–75.

Joseph, Benjamin, Grégoire Loeper, and Jan Obłój (2023a). "Calibration of Local Volatility Models with Stochastic Interest Rates using Optimal Transport". arXiv: 2305.00200.

— (2023b). "Joint Calibration of Local Volatility Models with Stochastic Interest Rates using Semimartingale Optimal Transport". arXiv: 2308.14473.

— (2023c). "The Measure Preserving Martingale Sinkhorn Algorithm". arXiv: 2310.13797.

Källblad, Sigrid, Xiaolu Tan, and Nizar Touzi (2017). "Optimal Skorokhod embedding given full marginals and Azéma-Yor peacocks". In: *Annals of Applied Probability* 27.2, pp. 686–719. DOI: 10.1214/16-AAP1191.

Kalotay, Andrew J, George O Williams, and Frank J Fabozzi (1993). "A model for valuing bonds and embedded options". In: *Financial Analysts Journal* 49.3, pp. 35–46. DOI: 10.2469/faj.v49.n3.35.

Kantorovich, Leonid (1942). "On the translocation of masses". In: *Dokl. Akad. Nauk. USSR (NS)*. Vol. 37, pp. 199–201.

Karatzas, Ioannis and Steven Shreve (2014). *Brownian motion and stochastic calculus*. Vol. 113. springer.

Karoui, Nicole El, Monique Jeanblanc-Picquè, and Steven Shreve (1998). "Robustness of the Black and Scholes formula". In: *Mathematical finance* 8.2, pp. 93–126. DOI: 10.1111/1467-9965.00047.

Kenyon, Chris (2010). "Short-rate pricing after the liquidity and credit shocks: including the basis". In: *Risk, November*.

Kijima, Masaaki, Keiichi Tanaka, and Tony Wong (2009). "A multi-quality model of interest rates". In: *Quantitative Finance* 9.2, pp. 133–145. DOI: 10.1080/14697680802624963.

Kim, Young-Heon and Brendan Pass (2014). "A general condition for Monge solutions in the multi-marginal optimal transport problem". In: *SIAM Journal on Mathematical Analysis* 46.2, pp. 1538–1550. DOI: 10.1137/130930443.

Klebaner, Fima (2012). *Introduction to stochastic calculus with applications*. World Scientific Publishing Company.

Knight, Philip (2008). "The Sinkhorn–Knopp algorithm: convergence and applications". In: *SIAM Journal on Matrix Analysis and Applications* 30.1, pp. 261–275. DOI: 10.1137/060659624.

Krylov, Nikolai (1984). "Once more about the connection between elliptic operators and Itô's stochastic equations". In: *Statistics and Control of Stochastic Processes, Steklov Seminar*, pp. 214–229.

Kullback, Solomon and Richard A Leibler (1951). "On information and sufficiency". In: *The annals of mathematical statistics* 22.1, pp. 79–86. DOI: 10.2307/2236703.

Lassalle, Rémi (2018). "Causal transport plans and their Monge–Kantorovich problems". In: *Stochastic Analysis and Applications* 36.3, pp. 452–484. DOI: 10.1080/07362994.2017.1422747.

Léonard, Christian (2014). "A survey of the Schrödinger problem and some of its connections with optimal transport". In: *Discrete Contin. Dyn. Syst.* 34.4, pp. 1533–1574. ISSN: 1078-0947,1553-5231. DOI: 10.3934/dcds.2014.34.1533.

Levin, Vladimir (1999). "Abstract cyclical monotonicity and Monge solutions for the general Monge–Kantorovich problem". In: *Set-Valued Analysis* 7.1, pp. 7–32. DOI: 10.1023/A:1008753021652.

Lieberman, Gary (1996). *Second order parabolic differential equations*. World scientific. DOI: 10.1142/3302.

Lim, Tongseok (2016). "Multi-martingale optimal transport". arXiv: 1611.01496.
Lions, Pierre-Louis (1983). "Optimal control of diffusion processes and Hamilton–Jacobi–Bellman equations part 2: viscosity solutions and uniqueness". In: *Communications in Partial Differential Equations* 8.11, pp. 1229–1276. DOI: 10.1080/03605308308820301.
Liu, Dong and Jorge Nocedal (1989). "On the limited memory BFGS method for large scale optimization". In: *Mathematical Programming* 45.1, pp. 503–528. DOI: 10.1007/BF01589116.
Loeper, Grégoire (2006). "The reconstruction problem for the Euler-Poisson system in cosmology". In: *Archive for Rational Mechanics and Analysis* 179.2, pp. 153–216. DOI: 10.1007/s00205-005-0384-3.
— (2009). "On the regularity of solutions of optimal transportation problems". In: *Acta mathematica* 202.2, pp. 241–283. DOI: 10.1007/s11511-009-0037-8.
— (2018). "Option pricing with linear market impact and nonlinear Black–Scholes equations". In: *The Annals of Applied Probability* 28.5, pp. 2664–2726. DOI: 10.2307/26542468.
Loeper, Grégoire and Fernando Quirós (2018). "Interior second derivative estimates for nonlinear diffusions". arXiv: 1812.11253.
Longstaff, Francis and Eduardo Schwartz (1992). "Interest rate volatility and the term structure: A two-factor general equilibrium model". In: *The Journal of Finance* 47.4, pp. 1259–1282. DOI: 10.1111/j.1540-6261.1992.tb04657.x.
Lucet, Yves (1997). "Faster than the fast Legendre transform, the linear-time Legendre transform". In: *Numerical Algorithms* 16, pp. 171–185. DOI: 10.1023/A:1019191114493.
Lyons, Terry (1995). "Uncertain volatility and the risk-free synthesis of derivatives". In: *Applied mathematical finance* 2.2, pp. 117–133. DOI: 10.1080/13504869500000007.
Ma, K and PA Forsyth (2017). "An unconditionally monotone numerical scheme for the two-factor uncertain volatility model". In: *IMA Journal of Numerical Analysis* 37.2, pp. 905–944. DOI: 10.1093/imanum/drw025.
Ma, Xi-Nan, Neil Trudinger, and Xu-Jia Wang (2005). "Regularity of potential functions of the optimal transportation problem". In: *Archive for rational mechanics and analysis* 177, pp. 151–183. DOI: 10.1007/s00205-005-0362-9.
McCann, Robert (1997). "A convexity principle for interacting gases". In: *Advances in mathematics* 128.1, pp. 153–179. DOI: 10.1006/aima.1997.1634.
Merton, Robert C (1969). "Lifetime portfolio selection under uncertainty: The continuous-time case". In: *The review of Economics and Statistics*, pp. 247–257. DOI: 10.2307/1926560.
— (1975). "Optimum consumption and portfolio rules in a continuous-time model". In: *Stochastic optimization models in finance*. Elsevier, pp. 621–661. DOI: 10.1016/B978-0-12-780850-5.50052-6.
Mikami, Toshio and Michèle Thieullen (2006). "Duality theorem for the stochastic optimal control problem". In: *Stochastic processes and their applications* 116.12, pp. 1815–1835. DOI: 10.1016/j.spa.2006.04.014.
Miltersen, Kristian, Klaus Sandmann, and Dieter Sondermann (1997). "Closed form solutions for term structure derivatives with log-normal interest rates". In: *The Journal of Finance* 52.1, pp. 409–430. DOI: 10.1111/j.1540-6261.1997.tb03823.x.
Monge, Gaspard (1781). "Mémoire sur la théorie des déblais et des remblais". In: *Histoire de l'Académie Royale des Sciences de Paris*.
Nesterov, Yurii (2018). *Lectures on convex optimization*. Vol. 137. Springer.

Nutz, Marcel (2022). "Introduction to Entropic Optimal Transport". In: URL: www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf.

Nutz, Marcel, Florian Stebegg, and Xiaowei Tan (2020). "Multiperiod martingale transport". In: *Stochastic Processes and their Applications* 130.3, pp. 1568–1615. DOI: 10.1016/j.spa.2019.05.010.

Nutz, Marcel and Johannes Wiesel (2022). "Entropic optimal transport: Convergence of potentials". In: *Probability Theory and Related Fields* 184.1, pp. 401–424. DOI: 10.1007/s00440-021-01096-8.

Oberman, Adam (2006). "Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton–Jacobi equations and free boundary problems". In: *SIAM Journal on Numerical Analysis* 44.2, pp. 879–895. DOI: 10.1137/S0036142903435235.

Obłój, Jan (2004). "The Skorokhod Embedding Problem and its offspring". In: *Probability Surveys* 1, pp. 321–392. DOI: 10.1214/154957804100000060.

Ögetbil, Orcan, Narayan Ganesan, and Bernhard Hientzsch (2022). "Calibrating local volatility models with stochastic drift and diffusion". In: *International Journal of Theoretical and Applied Finance* 25.02, p. 2250011. DOI: 10.1142/S021902492250011X.

Øksendal, Bernt (2003). *Stochastic differential equations: an introduction with applications*. Sixth edition. Springer Science & Business Media.

Pallavicini, Andrea and Marco Tarenghi (2010). "Interest-rate modeling with multiple yield curves". arXiv: 1006.4767.

Rebonato, Riccardo, Kenneth McKay, and Richard White (2009). *The SABR/LIBOR Market Model: Pricing, calibration and hedging for complex interest-rate derivatives*. John Wiley & Sons.

Reisinger, Christoph and Jan Hendrik Witte (2012). "On the use of policy iteration as an easy way of pricing American options". In: *SIAM Journal on Financial Mathematics* 3.1, pp. 459–478. DOI: 10.1137/110823328.

Ren, Yong, Dilip Madan, and Michael Qian (2007). "Calibrating and pricing with embedded local volatility models". In: *Risk Magazine* 20.9, p. 138.

Richardson, Steven and Song Wang (2006). "Numerical solution of Hamilton–Jacobi–Bellman equations by an exponentially fitted finite volume method". In: *Optimization* 55.1-2, pp. 121–140. DOI: 10.1080/02331930500530237.

Ritchken, Peter and L Sankarasubramanian (1995). "Volatility structures of forward rates and the dynamics of the term structure 1". In: *Mathematical finance* 5.1, pp. 55–72. DOI: 10.1111/j.1467-9965.1995.tb00101.x.

Rockafellar, R. Tyrrell (1970). *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J.

Rubinstein, Mark (1983). "Displaced diffusion option pricing". In: *The Journal of Finance* 38.1, pp. 213–217. DOI: 10.1111/j.1540-6261.1983.tb03636.x.

Rubner, Yossi, Carlo Tomasi, and Leonidas Guibas (1998). "A metric for distributions with applications to image databases". In: *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*. IEEE, pp. 59–66. DOI: 10.1109/ICCV.1998.710701.

— (2000). "The earth mover's distance as a metric for image retrieval". In: *International journal of computer vision* 40, pp. 99–121. DOI: 10.1023/A:1026543900054.

Samuelson, Paul A (1965). "Proof That Properly Anticipated Prices Fluctuate Randomly". In: *Management Review* 6.2. DOI: 10.1142/9789814566926_0002.

— (1973). "Proof that properly discounted present values of assets vibrate randomly". In: *The Bell Journal of Economics and Management Science*, pp. 369–374. DOI: 10.2307/3003046.

Santambrogio, Filippo (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Vol. 87. Birkhäuser.

Schachermayer, Walter and Bertram Tschiderer (2024). "The decomposition of stretched Brownian motion into Bass martingales". In: arXiv: 2406.10656.

Séjourné, Thibault, Gabriel Peyré, and François-Xavier Vialard (2023). "Unbalanced optimal transport, from theory to numerics". In: *Handbook of Numerical Analysis* 24, pp. 407–471. DOI: 10.1016/bs.hna.2022.11.003.

Sinkhorn, Richard (1964). "A relationship between arbitrary positive matrices and doubly stochastic matrices". In: *The annals of mathematical statistics* 35.2, pp. 876–879. DOI: 10.2307/2238545.

Soner, Halil Mete, Nizar Touzi, and Jianfeng Zhang (2013). "Dual formulation of second order target problems". In: *The Annals of Applied Probability* 23.1, pp. 308–347. DOI: 10.1214/12-AAP844. URL: https://doi.org/10.1214/12-AAP844.

Souganidis, Panagiotis E (1985). "Approximation schemes for viscosity solutions of Hamilton-Jacobi equations". In: *Journal of differential equations* 59.1, pp. 1–43. DOI: 10.1016/0022-0396(85)90136-6.

Strassen, V. (1965). "The Existence of Probability Measures with Given Marginals". In: *The Annals of Mathematical Statistics* 36.2, pp. 423–439. DOI: 10.1214/aoms/1177700153.

Stroock, Daniel and Srinivasa Varadhan (1979). *Multidimensional diffusion processes*. Vol. 233. Springer Science & Business Media.

Tan, Xiaolu and Nizar Touzi (2013). "Optimal transportation under controlled stochastic dynamics". In: *The Annals of Probability* 41.5, pp. 3201–3240. DOI: 10.2307/42919802.

Tolstoi, AN (1930). "Methods of finding the minimal total kilometrage in cargo transportation planning in space". In: *TransPress of the National Commissariat of Transportation* 1.1930, pp. 23–55.

Trevisan, Dario (2016). "Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients". In: *Electronic Journal of Probability* 21. DOI: 10.1214/16-EJP4453.

Tschiderer, Bertram (2024). "q-Bass martingales". arXiv: 2402.05669.

Vasicek, Oldrich (1977). "An equilibrium characterization of the term structure". In: *Journal of Financial Economics* 5.2, pp. 177–188. DOI: 10.1016/0304-405X(77)90016-2.

Villani, Cédric (2003). *Topics in optimal transportation*. 58. American Mathematical Soc.

— (2009). *Optimal transport: old and new*. Vol. 338. Springer.

Wang, Jian and Peter A Forsyth (2008). "Maximal use of central differencing for Hamilton–Jacobi–Bellman PDEs in finance". In: *SIAM Journal on Numerical Analysis* 46.3, pp. 1580–1601. DOI: 10.1137/060675186.

Wang, Shouyang, F Gao, and Kok Lay Teo (2000). "An upwind finite-difference method for the approximation of viscosity solutions to Hamilton-Jacobi-Bellman equations". In: *IMA Journal of Mathematical Control and Information* 17.2, pp. 167–178. DOI: 10.1093/imamci/17.2.167.

Wang, Song, Les Jennings, and Kok Lay Teo (2003). "Numerical solution of Hamilton-Jacobi-Bellman equations by an upwind finite volume method". In: *Journal of Global Optimization* 27, pp. 177–192. DOI: 10.1023/A:1024980623095.

Wiesel, Johannes (2023). "Continuity of the martingale optimal transport problem on the real line". In: *The Annals of Applied Probability* 33.6A, pp. 4645–4692. DOI: 10.1214/22-AAP1928.

Witte, Jan Hendrik and Christoph Reisinger (2011). "A penalty method for the numerical solution of Hamilton–Jacobi–Bellman (HJB) equations in finance". In: *SIAM Journal on Numerical Analysis* 49.1, pp. 213–231. DOI: 10.1137/100797606.

— (2012). "Penalty methods for the solution of discrete HJB equations—continuous control and obstacle problems". In: *SIAM Journal on Numerical Analysis* 50.2, pp. 595–625. DOI: 10.1137/110835840.

Wyns, Maarten and Jacques Du Toit (2017). "A finite volume–alternating direction implicit approach for the calibration of stochastic local volatility models". In: *International Journal of Computer Mathematics* 94.11, pp. 2239–2267. DOI: 10.1080/00207160.2017.1297805.