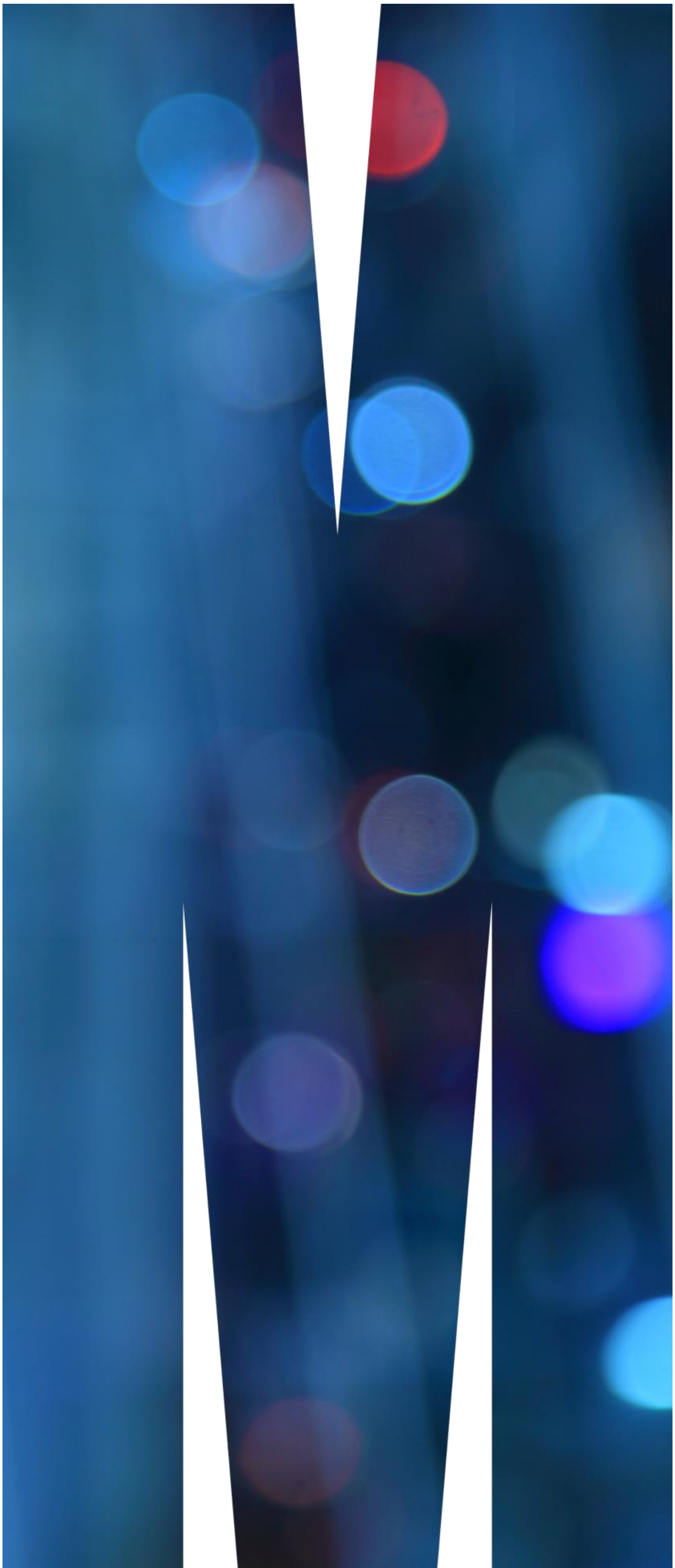
Calibration of Local-Stochastic Volatility Models by Optimal Transport

Ivan Guo, Grégoire Loeper, Shiyi Wang

Monash CQFIS Working Paper 2019

Abstract

In this paper, we study a semi-martingale optimal transport problem and its application to the calibration of Local-Stochastic Volatility (LSV) models. Rather than considering the classical constraints on marginal distributions at initial and final time, we optimise our cost function given the prices of a finite number of European options. We formulate the problem as a convex optimisation problem, for which we provide a dual formulation. Then we solve numerically the dual problem, which involves a fully non-linear Hamilton–Jacobi–Bellman equation. The method is tested by calibrating a Heston-like LSV model with simulated data and foreign exchange market data.



MONASH
CENTRE FOR
QUANTITATIVE
FINANCE AND
INVESTMENT
STRATEGIES

https://www.monash.edu/science/quantitative-finance/publication

Calibration of Local-Stochastic Volatility Models by Optimal Transport

Ivan Guo[1,2], Grégoire Loeper[1,2], and Shiyi Wang[1]

[1]School of Mathematics, Monash University, Australia
[2]Centre for Quantitative Finance and Investment Strategies*, Monash University, Australia

September 25, 2019

Abstract

In this paper, we study a semi-martingale optimal transport problem and its application to the calibration of Local-Stochastic Volatility (LSV) models. Rather than considering the classical constraints on marginal distributions at initial and final time, we optimise our cost function given the prices of a finite number of European options. We formulate the problem as a convex optimisation problem, for which we provide a dual formulation. Then we solve numerically the dual problem, which involves a fully non-linear Hamilton–Jacobi–Bellman equation. The method is tested by calibrating a Heston-like LSV model with simulated data and foreign exchange market data.

1 Introduction

Since the introduction of the Black–Scholes model, a lot of effort has been put on developing sophisticated volatility models that properly capture the market dynamics. In the space of equities and currencies, the most widely used models are the Local Volatility (LV) [10] and the Stochastic Volatility (SV) models [13, 20]. Introduced as an extension of the Black–Scholes model, the LV model can be exactly calibrated to any arbitrage-free implied volatility surface. Despite this feature, the LV model has often been criticised for its unrealistic volatility dynamics. The SV models tend to be more consistent with the market dynamics, but they struggle to fit short term market smiles and skews, and being parametric, they do not have enough degrees of freedom to match all vanilla market prices. A better fit can be obtained by increasing the number of stochastic factors in the SV models; however, this also increases the complexity of calibration and pricing.

Local-Stochastic Volatility (LSV) models, introduced in [23], naturally extend and take advantage of both approaches. The idea behind LSV models is to incorporate a local, non-parametric, factor into the SV models. Thus, while keeping consistent dynamics, the model can match all observed market prices (as long as one restricts to European claims). The determination of this local factor (also called *leverage*) is based on Gyöngy's mimicking theorem [17]. Research into the numerical calibration of LSV models has been developed in two different directions. One is based on a Monte Carlo approach, with Henry-Labordere [18], followed by Henry-Labordere and Guyon [16] using a so-called McKean's particle method.

* **Acknowledgements** The Centre for Quantitative Finance and Investment Strategies has been supported by BNP Paribas.

Another approach relies on solving the Fokker–Planck equation as in Ren et al. [29]. Engelmann and Wyns [11] used the finite volume method (FVM) to solve the partial differential equation (PDE), while Tian et al. [32] considered time-dependent parameters. A more recent study [35] considered a method that combines the FVM with alternating direction implicit (ADI) schemes.

All of the calibration methods mentioned above require the priori knowledge of the Local Volatility surface. This is usually obtained by using Dupire's formula [10] assuming the knowledge of vanilla options for all strikes and maturities. However, only a finite number of options are available in practice. Thus, an interpolation of the implied volatility surface or option prices is often needed, which can lead to inaccuracies and instabilities. Moreover, there is no a priori control on the regularity of the leverage function, and even its very existence remains an open problem, although some results have been obtained for small time [1] (see also Saporito et al. [30] who recently proposed to apply Tikhonov regularisation technique to the LSV calibration problem).

In the present work, inspired by the theory of optimal transport, we introduce a variational approach for calibrating LSV models that does not require any form of interpolation. In recent years, optimal transport theory has attracted the attention of many researchers. The problem was first addressed by Monge [27] in the context of civil engineering and was later given a modern mathematical treatment by Kantorovich [24]. In 2000, in a landmark paper [3], Benamou and Brenier introduced a time-continuous formulation of the problem, which they solved numerically by an augmented Lagrangian method. In [5] and [26], the dual formulation of the time-continuous optimal transport has been formally expressed and generalised as an application of the Fenchel–Rockafellar theorem [34]. The problem has then been extended to transport by semi-martingales. Tan and Touzi [31] studied the optimal transport problem for semi-martingales with constraints on the marginals at initial and final times. More recently, in [14], the first two authors further extend the semi-martingale optimal transport problem to a more general path-dependent setting.

In the area of mathematical finance, optimal transport theory has recently been applied to many different problems, including model-free bounds of exotics derivatives [19], robust hedging [8] and stochastic portfolio theory [28]. The authors in [15] explored its application to LV model calibration by adapting the augmented Lagrangian method of [3] to the semi-martingale transport. We also mention that a variational calibration method was proposed in [2] much earlier, although the connection with optimal transport was not established at that time.

In this paper, we further extend the approach of [15] and [14] to the calibration of LSV models. The calibration problem is formulated as a semi-martingale optimal transport problem. Unlike [31], we consider a finite number of discrete constraints given by the prices of European claims. As a consequence of Jensen's inequality, we show that the optimal diffusion process is Markovian in the state variables given by the initial SV model. This result leads to an equivalent PDE formulation. By following the duality theory of optimal transport introduced in [5] and a smoothing argument used in [4], we establish a dual formulation. We also provide a numerical method to solve a fully non-linear Hamilton–Jacobi–Bellman (HJB) equation.

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

2 Preliminaries

Given a Polish space E equipped with its Borel σ-algebra, let $C(E)$ be the space of continuous functions on E and $C_b(E)$ be the space of bounded continuous functions. Denote by $\mathcal{M}(E)$ the space of finite signed Borel measures endowed with the weak-$*$ topology. Let $\mathcal{M}_+(E) \subset \mathcal{M}(E)$ denote the subset of nonnegative measures. If E is compact, the topological dual of $C_b(E)$ is given by $C_b(E)^* = \mathcal{M}(E)$. More generally, if E is non-compact, $C_b(E)^*$ is larger than $\mathcal{M}(E)$. Let $\mathcal{P}(E)$ be the space of Borel probability measures, $BV(E)$ be the space of functions of bounded variation and $L^1(d\mu)$ be the space of μ-integrable functions. We also write $C_b(E,\mathbb{R}^d)$, $\mathcal{M}(E,\mathbb{R}^d)$, $BV(E,\mathbb{R}^d)$ and $L^1(d\mu,\mathbb{R}^d)$ as the vector-valued versions of their corresponding spaces. For convenience, let $\Lambda = [0,T]\times\mathbb{R}^d$ and $\mathcal{X} = \mathbb{R}\times\mathbb{R}^d\times\mathbb{S}^d$. We use the notation $\langle\cdot,\cdot\rangle$ to denote the duality bracket between $C_b(\Lambda,\mathcal{X})$ and $C_b(\Lambda,\mathcal{X})^*$.

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

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

where M is an $(\mathbb{F},\mathbb{P})$-martingale with quadratic variation $\langle X_t\rangle = \langle M_t\rangle = B_t$, and the processes A and B are $\mathbb{P}$-a.s. absolutely continuous with respect to t. Denote by $\mathbb{S}^d$ the set of $d\times d$ symmetric matrices and $\mathbb{S}^d_+ \subset \mathbb{S}^d$ the set of positive semidefinite matrices. For any matrices $A, B \in \mathbb{S}^d$, we write $A : B := \mathrm{tr}(A^\intercal B)$ for their scalar product. We say $(\alpha^{\mathbb{P}},\beta^{\mathbb{P}})$ are the characteristics of $\mathbb{P}$ if

$$\alpha^{\mathbb{P}}_t = \frac{dA^{\mathbb{P}}_t}{dt}, \quad \beta^{\mathbb{P}}_t = \frac{dB^{\mathbb{P}}_t}{dt},$$

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

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

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

Given a vector $\tau := (t_1,\dots,t_m) \in (0,T]^m$, denote by G a vector of m functions such that each function $G_i \in C_b(\mathbb{R})$ for all $i = 1,\dots,m$. Given a Dirac measure $\mu_0 = \delta_{x_0}$ and a vector $c \in \mathbb{R}^m$, we define $\mathcal{P}(\mu_0,\tau,c,G) \subset \mathcal{P}^1$ as follows:

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

For brevity, we also write $\mathbb{E}^{\mathbb{P}}_{t,x} := \mathbb{E}^{\mathbb{P}}[\,\cdot \mid X_t = x]$.

For technical reasons, we restrict ourselves to functions G_i in $C_b(\mathbb{R})$. In the context of volatility models calibration, G_i are discounted European payoffs. Although the call option payoff functions are not technically in $C_b(\mathbb{R})$, we only work with them in a truncated (compact) space in practice. Alternatively, one may consider only put options using put-call parity.

3 Formulations

3.1 From SDE to PDE

Now let us introduce the formulation of the semi-martingale optimal transport problem under discrete constraints.

Define the cost function $F:\Lambda\times\mathbb{R}^d\times\mathbb{S}^d\to\mathbb{R}\cup\{+\infty\}$ where $F(t,x,\alpha,\beta)=+\infty$ if $\beta\notin\mathbb{S}^d_+$, and $F(t,x,\alpha,\beta)$ is nonnegative, proper, lower semi-continuous and convex in (α,β). We also let F be coercive in the sense that there exist constant $p>1$ and $C>0$ such that

$$|\alpha|^p+|\beta|^p\le C(1+F(t,x,\alpha,\beta)),\qquad \forall(t,x)\in\Lambda.$$

Its convex conjugate $F^*:\Lambda\times\mathbb{R}^d\times\mathbb{S}^d\to\mathbb{R}\cup\{+\infty\}$ for (α,β) is then defined as

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

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

We are interested in the following minimisation problem:

Definition 3.1 (Problem 1)**.** *Given μ_0,τ,c and G, we want to find*

$$\inf_{\mathbb{P}\in\mathcal{P}(\mu_0,\tau,c,G)}\mathbb{E}^{\mathbb{P}}\int_0^T F(\alpha^{\mathbb{P}}_t,\beta^{\mathbb{P}}_t)\,dt.$$

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

The following lemma is an immediate consequence of Itô's formula and [33, Theorem 2.5] (see [12, Theorem 2.6] for a shorter proof in the case of bounded coefficients).

Lemma 3.2. *Let $\mathbb{P}\in\mathcal{P}^1$ and $\rho^{\mathbb{P}}$ be the law of $\mathbb{P}$. Defining $\alpha(t,x)=\mathbb{E}^{\mathbb{P}}_{t,x}\alpha^{\mathbb{P}}_t$ and $\beta(t,x)=\mathbb{E}^{\mathbb{P}}_{t,x}\beta^{\mathbb{P}}_t$, then $\rho^{\mathbb{P}}$ is a weak solution to the Fokker–Planck equation:*

$$\begin{cases}\partial_t\rho_t(x)+\nabla_x\cdot(\rho_t(x)\alpha(t,x))-\dfrac{1}{2}\displaystyle\sum_{i,j}\partial_{ij}(\rho_t(x)\beta_{ij}(t,x))=0 & \text{in }\Lambda,\\ \rho_0=\mu_0 & \text{in }\mathbb{R}^d.\end{cases}\tag{1}$$

Moreover, there exists a Markov process X such that, under another probability measure $\mathbb{P}'\in\mathcal{P}^1$, X has law $\rho^{\mathbb{P}'}=\rho^{\mathbb{P}}$ and solves

$$\begin{cases}dX_t=\alpha(t,X_t)\,dt+\sigma(t,X_t)\,dW^{\mathbb{P}'}_t,\\ X_0=x_0,\end{cases}\tag{2}$$

where $\sigma:\Lambda\to\mathbb{R}^{d\times d}$ is defined such that $\beta=\sigma\sigma^\intercal$ and $W^{\mathbb{P}'}$ is a $\mathbb{P}'$-Brownian motion.

Definition 3.3. *Define $\mathcal{P}_{loc}$ to be a subset of $\mathcal{P}^1$ such that, under $\mathbb{P}'\in\mathcal{P}_{loc}$, X is a Markov process that has law $\rho^{\mathbb{P}'}=\rho^{\mathbb{P}}$ and takes the form of (2) with coefficients $(\alpha(t,x),\beta(t,x))=(\mathbb{E}^{\mathbb{P}}_{t,x}\alpha^{\mathbb{P}}_t,\mathbb{E}^{\mathbb{P}}_{t,x}\beta^{\mathbb{P}}_t)$ for some $\mathbb{P}\in\mathcal{P}^1$.*

In the next proposition we write $(\mathbb{E}^{\mathbb{P}}_{t,x}\alpha^{\mathbb{P}}_t,\mathbb{E}^{\mathbb{P}}_{t,x}\beta^{\mathbb{P}}_t)$ as $(\alpha^{\mathbb{P}'}_{loc},\beta^{\mathbb{P}'}_{loc})$ for $\mathbb{P}\in\mathcal{P}^1$ with $\rho^{\mathbb{P}'}=\rho^{\mathbb{P}}$. The superscript $\mathbb{P}'$ indicates the dependence of $\mathbb{P}'$ via $\rho^{\mathbb{P}'}=\rho^{\mathbb{P}}$. Then, we establish the following result:

Proposition 3.4 (Localisation)**.** *Let $\alpha^{\mathbb{P}'}_{loc}=\mathbb{E}^{\mathbb{P}}_{t,x}\alpha^{\mathbb{P}}_t$ and $\beta^{\mathbb{P}'}_{loc}=\mathbb{E}^{\mathbb{P}}_{t,x}\beta^{\mathbb{P}}_t$. Then,*

$$\inf_{\mathbb{P}\in\mathcal{P}^1}\mathbb{E}^{\mathbb{P}}\int_0^T F(\alpha^{\mathbb{P}}_t,\beta^{\mathbb{P}}_t)\,dt=\inf_{\mathbb{P}'\in\mathcal{P}_{loc}}\mathbb{E}^{\mathbb{P}'}\int_0^T F(\alpha^{\mathbb{P}'}_{loc},\beta^{\mathbb{P}'}_{loc})\,dt.\tag{3}$$

Proof. By applying Jensen's inequality together with the tower property of conditional expectation, we have

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\int_0^T F(\alpha_t^{\mathbb{P}},\beta_t^{\mathbb{P}})\,dt &= \mathbb{E}^{\mathbb{P}}\int_0^T \left(\mathbb{E}^{\mathbb{P}}_{t,x}F(\alpha_t^{\mathbb{P}},\beta_t^{\mathbb{P}})\right)\,dt \\ &\geq \mathbb{E}^{\mathbb{P}}\int_0^T F(\mathbb{E}^{\mathbb{P}}_{t,x}\alpha_t^{\mathbb{P}},\mathbb{E}^{\mathbb{P}}_{t,x}\beta_t^{\mathbb{P}})\,dt.\end{aligned} \tag{4}$$

This shows that by localising $(\alpha^{\mathbb{P}},\beta^{\mathbb{P}})$, one lowers the transportation cost. Denote by $\rho^{\mathbb{P}}$ the law of $\mathbb{P}$. Applying Lemma 3.2, there exists $\mathbb{P}'\in\mathcal{P}^1$ and a Markov process that has law $\rho^{\mathbb{P}'}=\rho^{\mathbb{P}}$ and takes the form of (2) with coefficients $(\alpha_{loc}^{\mathbb{P}'},\beta_{loc}^{\mathbb{P}'})=(\mathbb{E}^{\mathbb{P}}_{t,x}\alpha_t^{\mathbb{P}},\mathbb{E}^{\mathbb{P}}_{t,x}\beta_t^{\mathbb{P}})$. Thus,

$$\mathbb{E}^{\mathbb{P}}\int_0^T F(\alpha_t^{\mathbb{P}},\beta_t^{\mathbb{P}})\,dt \geq \mathbb{E}^{\mathbb{P}'}\int_0^T F(\alpha_{loc}^{\mathbb{P}'},\beta_{loc}^{\mathbb{P}'})\,dt, \tag{5}$$

and $\mathbb{E}^{\mathbb{P}}$ is replaced by $\mathbb{E}^{\mathbb{P}'}$ on the right-hand side because $\rho^{\mathbb{P}}=\rho^{\mathbb{P}'}$. Since $\mathcal{P}_{loc}\subset\mathcal{P}^1$, taking the infimum over $\mathbb{P}\in\mathcal{P}^1$ on the left-hand side of (5) and over $\mathbb{P}'\in\mathcal{P}_{loc}$ on the right-hand side of (5), we obtain the required result. □

Proposition 3.4 shows that it suffices to consider only $(\alpha_{loc}^{\mathbb{P}'},\beta_{loc}^{\mathbb{P}'})$. Thus Problem 1 can be studied via PDE methods. Following the Benamou–Brenier formulation of the classical optimal transport [3], we introduce the following problem:

Proposition 3.5 (Problem 2)**.** *Given μ_0,τ,c and G, we want to minimise*

$$\mathcal{V}=\inf_{\rho,\alpha,\beta}\int_0^T\int_{\mathbb{R}^d}F(\alpha(t,x),\beta(t,x))\,d\rho(t,x)dt, \tag{6}$$

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

$$\partial_t\rho(t,x)+\nabla_x\cdot(\rho(t,x)\alpha(t,x))-\frac{1}{2}\sum_{i,j}\partial_{ij}(\rho(t,x)\beta_{ij}(t,x))=0, \tag{7}$$

$$\int_{\mathbb{R}^d}G_i(x)\,d\rho(t_i,x)=c_i,\ \forall i=1,\dots,m,\qquad \textit{and}\qquad \rho(0,\cdot)=\mu_0. \tag{8}$$

The interchange of integrals in (6) is justified by Fubini's theorem. For the weak continuity of measure ρ in time, the reader can refer to [26, Theorem 3].

3.2 Duality

This section examines the proof of the duality result by closely following [26, Section 3.2] (see also [5, 21]).

Theorem 3.6. *Assumed that $\mathcal{V}$ is finite. Then,*

$$\mathcal{V}=\sup_{\phi,\lambda}\sum_{i=1}^m\lambda_i c_i-\int_{\mathbb{R}^m}\phi(0,x)\,d\mu_0, \tag{9}$$

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

$$\partial_t\phi+\sum_{i=1}^m\lambda_i G_i\delta_{t_i}+F^*(\nabla_x\phi,\frac{1}{2}\nabla_x^2\phi)\leq 0\qquad \textit{in } [0,T)\times\mathbb{R}^d, \tag{10}$$

and $\phi(T,\cdot)=0$.

Proof. Define measures $\mathcal{A} := \rho\alpha$ and $\mathcal{B} := \rho\beta$, then $\mathcal{A}$ and $\mathcal{B}$ are absolutely continuous with respect to ρ. Assuming that there exists an admissible solution $(\bar{\rho}, \bar{\mathcal{A}} = \bar{\rho}\bar{\alpha}, \bar{\mathcal{B}} = \bar{\rho}\bar{\beta})$ of Problem 2 satisfying (7, 8), the constraints can be formulated in the following weak form:

$$\forall \phi \in C_c^\infty(\Lambda), \qquad \int_\Lambda \partial_t \phi \, (d\rho - d\bar{\rho}) + \nabla_x \phi \cdot (d\mathcal{A} - d\bar{\mathcal{A}}) + \frac{1}{2} \nabla_x^2 \phi : (d\mathcal{B} - d\bar{\mathcal{B}}) = 0, \tag{11}$$

$$\forall \lambda \in \mathbb{R}^m, \qquad \int_\Lambda \sum_{i=1}^m \lambda_i G_i \delta_{t_i} (d\rho - d\bar{\rho}) = 0. \tag{12}$$

Thus Problem 2 can be reformulated as the following saddle point problem:

$$\begin{aligned}\mathcal{V} = \inf_{\rho, \mathcal{A}, \mathcal{B}} \sup_{\phi, \lambda} \Bigg\{ \int_\Lambda F\left(\frac{d\mathcal{A}}{d\rho}, \frac{d\mathcal{B}}{d\rho}\right) d\rho - \sum_{i=1}^m \lambda_i G_i \delta_{t_i} (d\rho - d\bar{\rho}) \\ - \partial_t \phi (d\rho - d\bar{\rho}) - \nabla_x \phi \cdot (d\mathcal{A} - d\bar{\mathcal{A}}) - \frac{1}{2} \nabla_x^2 \phi : (d\mathcal{B} - \bar{\mathcal{B}}) \Bigg\}.\end{aligned}$$

Adopting the terminology of [21], we say the triple (r, a, b) is *represented* by (ϕ, λ) if it satisfies

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

If we choose (r, a, b) from $C_b(\Lambda, \mathcal{X})$, then the multiplier ϕ has possible jump discontinuities at $t = t_i$ due to the presence of the Dirac delta functions. Now, define functionals $\Phi : C_b(\Lambda, \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ and $\Psi : C_b(\Lambda, \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ as follows:

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

$$\Psi(r, a, b) = \begin{cases} \displaystyle\int_\Lambda r \, d\bar{\rho} + a \cdot d\bar{\mathcal{A}} + b : d\bar{\mathcal{B}} & \text{if } (r, a, b) \text{ is represented by } (\phi, \lambda) \in BV([0, T], C_b^2(\mathbb{R}^d)) \times \mathbb{R}^m, \\ +\infty & \text{otherwise.} \end{cases}$$

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

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

As shown in Lemma A.1, if we restrict Φ^* to $\mathcal{M}(\Lambda, \mathcal{X})$, then

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

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

$$\Psi^*(\rho, \mathcal{A}, \mathcal{B}) = \sup_{r, a, b} \langle (r, a, b), (\rho - \bar{\rho}, \mathcal{A} - \bar{\mathcal{A}}, \mathcal{B} - \bar{\mathcal{B}}) \rangle,$$

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

$$\Psi^*(\rho, \mathcal{A}, \mathcal{B}) = \sup_{\phi, \lambda} \langle (-\partial_t \phi - \sum_{i=1}^m \lambda_i G_i \delta_{t_i}, -\nabla_x \phi, -\frac{1}{2} \nabla_x^2 \phi), (\rho - \bar{\rho}, \mathcal{A} - \bar{\mathcal{A}}, \mathcal{B} - \bar{\mathcal{B}}) \rangle.$$

Therefore, the objective V can be expressed as

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

where the second equality is proved in Lemma A.2.

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

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

This shows that

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

Thus, at $(-1, O^{d\times 1}, O^{d\times d})$, Φ is continuous with respect to the uniform norm (since F^* is continuous in $\mathrm{dom}(F^*)$), and Ψ is finite. Furthermore, as the convex functionals Φ and Ψ take values in $(-\infty, +\infty]$, all of the required conditions are fulfilled to apply the Fenchel–Rockafellar duality theorem (see [6, Chapter 1]). We then obtain

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

and the infimum is in fact attained. Consequently,

$$\mathcal{V} = \sup_{r,a,b} \left\{ \int_\Lambda -r\, d\bar{\rho} - a\cdot d\bar{\mathcal{A}} - b : \bar{\mathcal{B}}\ ;\ -r + F^*(-a,-b) \le 0 \right\},$$

where the supremum is restricted to all (r,a,b) represented by $(\phi,\lambda) \in BV([0,T], C_b^2(\mathbb{R}^d)) \times \mathbb{R}^m$. Again, in terms of (ϕ,λ), this is equivalent to

$$\mathcal{V} = \sup_{\phi,\lambda} \int_\Lambda (\partial_t\phi + \sum_{i=1}^m \lambda_i G_i \delta_{t_i})\, d\bar{\rho} + \nabla_x\phi\cdot d\bar{\mathcal{A}} + \frac{1}{2}\nabla_x^2\phi : d\bar{\mathcal{B}},$$

where $(\phi,\lambda) \in BV([0,T], C_b^2(\mathbb{R}^d)) \times \mathbb{R}^m$ subject to (10). Recalling that the admissible solution $(\bar{\rho}, \bar{\mathcal{A}}, \bar{\mathcal{B}})$ satisfies (7, 8) and fixing $\phi(t=T) = 0$, we obtain the required result by integrating by parts. □

3.3 Viscosity solutions

Adopting the concept of viscosity solutions, it can be shown that the supremum of the objective with respect to ϕ is achieved by the viscosity solution of a HJB equation. In order to deal with the jump discontinuities in ϕ, assuming that T is the longest maturity ($T = \max t_k$), we define the viscosity solution in the following way.

Definition 3.7. *Denote by* $\mathrm{set}(\tau)$ *the set of entries of vector* τ *and by* K *the cardinality of* $\mathrm{set}(\tau)$. *Let* $t_0 = 0$, *we define disjoint intervals* $I_k := [t_{k-1}, t_k)$ *such that*

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

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

Definition 3.8 (Viscosity solution)**.** *For any $\lambda \in \mathbb{R}^m$, we say ϕ is a viscosity subsolution (resp,. supersolution) of*

$$\begin{cases} \partial_t \phi + \sum_{i=1}^m \lambda_i G_i \delta_{t_i} + F^*(\nabla_x \phi, \frac{1}{2}\nabla_x^2 \phi) = 0 & \text{in } [0,T) \times \mathbb{R}^d, \\ \phi_T = 0 & \text{in } \mathbb{R}^d, \end{cases} \tag{13}$$

if ϕ is a classical (continuous) viscosity subsolution (resp,. supersolution) of (13) in $I_k \times \mathbb{R}^d$ for all $k = 1, \dots, K$, and has jump discontinuities:

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

Then, ϕ is called a viscosity solution of (13) if ϕ is both a viscosity subsolution and a viscosity supersolution of (13).

Remark 3.9 (Comparison principle)**.** *The comparison principle still holds for viscosity solutions of (13). Let u and v be a viscosity subsolution and a viscosity supersolution of the equation (13), respectively. At the terminal time T, $u(T,\cdot) \le v(T,\cdot)$. Since $t_K = T$ is in* $\mathrm{set}(\tau)$ *and u, v have the same jump size at $\{T\} \times \mathbb{R}^d$, we get $u(T^-,\cdot) \le v(T^-,\cdot)$. Next, in the interval $I_K = [t_{K-1}, t_K)$, by the classical comparison principle, we get $u \le v$ on I_K. Applying this argument for all intervals I_k for $k = 1, \dots, K$, we conclude that*

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

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

Remark 3.10 (Existence and uniqueness)**.** *As a consequence of the comparison principle, there exists a unique viscosity solution of (13). The uniqueness is a direct consequence of the comparison principle. The existence can be obtained by the Perron's method (see [7]) under which the comparison principle is a key argument.*

Now we show that, with any $\lambda \in \mathbb{R}^m$, the optimal ϕ is the viscosity solution of the HJB equation (13). We prove the result by the smoothing technique used in [4]. The proof is similar to [4, Theorem 2.4], which we sketch here for completeness.

Corollary 3.11. *The dual formulation is equivalent to*

$$\mathcal{V} = \sup_{\lambda \in \mathbb{R}^m} \sum_{i=1}^m \lambda_i c_i - \int_{\mathbb{R}^d} \phi(0,x)\, d\mu_0, \tag{14}$$

where ϕ is the viscosity solution to the HJB equation (13). Moreover, if the supremum in ϕ is attained by some $\bar{\phi}$, the optimal (α, β) is given by

$$(\alpha, \beta) = \nabla F^*(\nabla_x \bar{\phi}, \frac{1}{2}\nabla_x^2 \bar{\phi}). \tag{15}$$

Proof. Denote by φ a viscosity solution of the equation (13) with any $\lambda \in \mathbb{R}^m$. From Remark 3.10, we know that such φ exists and is unique. The proof is divided into two steps.

Step 1. Assuming that there exists a sequence of supersolutions of (13) in $BV([0,T], C_b^2(\mathbb{R}^d))$ converging to φ pointwise, we can show that φ achieves the supremum with respect to ϕ in the objective of the dual (9). Let $\phi \in BV([0,T], C_b^2(\mathbb{R}^d))$ be any solution satisfies (10), and ϕ is also a (viscosity) supersolution of (13). By Remark 3.9, we have $\varphi(0,x) \le \phi(0,x)$ for all $x \in \mathbb{R}^d$, hence

$$\sum_{i=1}^m \lambda_i c_i - \int_{\mathbb{R}^d} \phi(0,x)\, d\mu_0 \le \sum_{i=1}^m \lambda_i c_i - \int_{\mathbb{R}^d} \varphi(0,x)\, d\mu_0. \tag{16}$$

The equality can be achieved in (16) by taking the supremum with respect to ϕ on the left-hand side of (16).

Step 2. Now, we shall construct the sequence of supersolutions required in Step 1. We first smooth the viscosity solution φ in space by a standard regularising kernel. By applying the result of [4] which relies critically on the fact that $F^*(a,b)$ is convex in (a,b), it can be shown that φ_ε are supersolutions of equation (13). If we send ε to 0, the supersolutions φ_ε converges to the viscosity solution φ pointwise. The desired sequence is then constructed.

The second part of the theorem is easy to check by the definition of convex conjugate. Therefore, the claim is proven. □

4 Calibration

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

$$\begin{cases} dZ_t = (r(t) - q(t) - \frac{1}{2}\sigma^2(t, Z_t, V_t))\,dt + \sigma(t, Z_t, V_t)\,dW_t^Z, \\ dV_t = \kappa(\theta - V_t)\,dt + \xi\sqrt{V_t}\,dW_t^V, \\ dW_t^Z dW_t^V = \eta(t, Z_t, V_t)\,dt, \end{cases} \tag{17}$$

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

Remark 4.1. *If the volatility $\sigma(t, Z, V) \equiv \sqrt{V}$ and the correlation $\eta(t, Z, V)$ is a constant, the LSV model reduces to a pure Heston model. Furthermore, if $\sigma(t, Z, V)$ is independent of the variable V, the model is equivalent to a local volatility model.*

Let the canonical process X_t be the vector of processes (Z_t, V_t) with an initial distribution $\mu_0 = (\delta_{Z_0}, \delta_{V_0})$. We want to find a probability measure $\mathbb{P} \in \mathcal{P}(\mu_0, \tau, c, G)$ characterised by (α, β) where

$$\alpha = \begin{bmatrix} (r - q - \sigma^2/2) \\ \kappa(\theta - V) \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} \sigma^2 & \eta\xi\sqrt{V}\sigma \\ \eta\xi\sqrt{V}\sigma & \xi^2 V \end{bmatrix}, \tag{18}$$

for the $\mathbb{F}$-adapted processes σ and η. On other words, we want the dynamics of X to be consistent with (17). Given m European options with prices $c \in \mathbb{R}_+^m$, maturities $\tau = (t_1, \ldots, t_m) \in (0, T]^m$ and discounted payoffs $G = (G_1, \ldots, G_m)$ where $G_i : \mathbb{R}_+ \to \mathbb{R}$ (e.g., $G_i(Z) = e^{-\int_0^{t_i} r(s)ds}(e^Z - K)^+$ for a European call of strike K and maturity t_i), the calibration problem can be formulated as

$$\inf_{\mathbb{P} \in \mathcal{P}(\mu_0, \tau, c, G)} \mathbb{E}^{\mathbb{P}} \int_0^T F(t, X_t, \alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})\,dt, \tag{19}$$

where F is a suitable convex cost function that forces (α, β) to take the form of (18).

One possible way to choose the cost function F is based on the idea of minimising the difference between each element of the covariance matrix β and a reference value while keeping β to be in $\mathbb{S}_+^2$. However, it is often impossible to find an explicit formula to approximate F^*.

Thus numerical optimisation is needed, which makes the method computationally expensive. To overcome this issue, we choose the correlation

$$\eta(t,Z,V) = \frac{\sqrt{V}}{\sigma}\bar{\eta},$$

where $\bar{\eta}$ is a constant correlation obtained (along with κ, θ, ξ) by calibrating a pure Heston model. In this case, β is positive semidefinite if and only if $\sigma^2 \geq \bar{\eta}^2 V$.

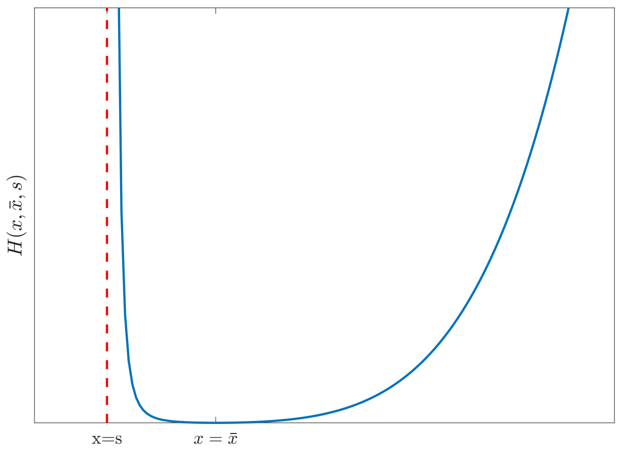


Figure 1: The plot of $H(x,\bar{x},s)$ for given $\bar{x}$ and s.

Definition 4.2. *Define function* $H : \mathbb{R}\times\mathbb{R}_+\times\mathbb{R}\to\mathbb{R}$ *such that*

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

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

Given $\bar{x}$ and s satisfying $\bar{x} > s$, the function H is convex in x and minimised at $\bar{x}$. It is finite only when $x > s$. A plot of H is given in Figure 1. Then, we define the cost function as follows.

Definition 4.3. *The cost function* $F : \mathbb{R}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}^2\times\mathbb{S}^2\to\mathbb{R}$ *is defined as*

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

where the convex set Γ *is defined as*

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

Remark 4.4. *The function* H *penalises deviations of the LSV model from a pure Heston model by choosing* $\bar{x} = V$ *(see Remark 4.1). This approach seeks to retain the attractive features of the Heston model while still matching all the market prices. We also set* $s = \bar{\eta}^2 V$ *to ensure that* β *remains positive semidefinite, so the correlation* η *is in* $[-1,1]$. *The set* Γ *forces the dynamics of the semi-martingale* X_t *to be consistent with the LSV model (17). In particular, it remains risk neutral.*

Applying Theorem 3.6 and Corollary 3.11, the equivalent PDE dual formulation of (19) can be derived:

$$\mathcal{V} = \sup_{\lambda\in\mathbb{R}^m} J(\lambda) := \sup_{\lambda\in\mathbb{R}^m} \sum_{i=1}^{m} \lambda_i c_i - \phi(0, Z_0, V_0), \tag{20}$$

where ϕ is the viscosity solution to the HJB equation

$$\begin{aligned}\partial_t\phi + \sum_{i=1}^{m} \lambda_i G_i \delta_{t_i} + \sup_{\beta_{11}} \Big\{ (r - q - \frac{1}{2}\beta_{11})\partial_Z\phi + \kappa(\theta - V)\partial_V\phi + \bar\eta\xi V\partial_{ZV}\phi \\ + \frac{1}{2}\beta_{11}\partial_{ZZ}\phi + \frac{1}{2}\xi^2 V\partial_{VV}\phi - H(\beta_{11}, V, \bar\eta^2 V)\Big\} = 0,\end{aligned} \tag{21}$$

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

The dual formulation is easier to solve than its primal. Given any $\lambda \in \mathbb{R}^m$, we can calculate $\phi(0, Z_0, V_0)$ by numerically solving the HJB equation (21). The optimal λ can be found by applying a standard optimisation algorithm. The convergence of the algorithm can be improved by providing the gradient of the objective.

Lemma 4.5. *The gradient of the objective with respect to λ_i can be formulated as:*

$$\partial_{\lambda_i} J(\lambda) = c_i - \mathbb{E}^{\mathbb{P}}_{0,(Z_0,V_0)} G_i(Z_{t_i}) \qquad \forall i = 1, \dots, m. \tag{22}$$

Proof. Since λ appears in (21), $\phi(0, Z_0, V_0)$ is also a function of λ. Differentiating (21) with respect to λ_i for any $i = 1, \dots, m$ and writing $\phi' = \partial_{\lambda_i}\phi$, we obtain the following PDE

$$\begin{aligned}\partial_t\phi' + (r - q - \frac{1}{2}\sigma^2)\partial_Z\phi' + \kappa(\theta - V)\partial_V\phi' + \bar\eta\xi V\partial_{ZV}\phi' \\ + \frac{1}{2}\sigma^2\partial_{ZZ}\phi' + \frac{1}{2}\xi^2 V\partial_{VV}\phi' = -G_i\delta_{t_i},\end{aligned} \tag{23}$$

where σ^2 is the optimal β_{11} that achieves the supremum in (21). With the terminal condition $\phi'(T,\cdot) = 0$, we solve (23) by the Feynman-Kac formula (see [25, Theorem 7.6]). Thus,

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

Differentiating $J(\lambda)$ with respect to λ_i completes the proof. □

Remark 4.6. *Note that $\mathbb{E}^{\mathbb{P}}_{0,(Z_0,V_0)} G_i(Z_{t_i})$ is the model price of the i-th European option. The gradient can be thus interpreted as the difference between market price and model price. Also, solving (23) with terminal condition $\phi'(T,\cdot) = 0$ is equivalent to solving*

$$\partial_t\phi' + (r - q - \frac{1}{2}\sigma^2)\partial_Z\phi' + \kappa(\theta - V)\partial_V\phi' + \bar\eta\xi V\partial_{ZV}\phi' + \frac{1}{2}\sigma^2\partial_{ZZ}\phi' + \frac{1}{2}\xi^2 V\partial_{VV}\phi' = 0, \tag{24}$$

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

Once the optimal (ϕ, λ) has been found, the optimal volatility σ^2 can be recovered by (15), which is equivalent to solving

$$(\partial_{ZZ}\phi - \partial_Z\phi)/2 = \partial_{\sigma^2} H(\sigma^2, V, \bar\eta^2 V), \tag{25}$$

for which a closed-form solution is available.

5 Numerical examples

In this section, we present a numerical method for solving the dual formulation. Let us first restrict the problem to a computational domain $Q \subseteq \mathbb{R}^2$, which should be large enough so that the density vanishes at the boundary. We consider a uniform mesh of the domain Q and a uniform mesh of the time interval $[0, T]$ with step size $\Delta t = T/N_T$ with $N_T \in \mathbb{N}$. Thus

$$t_k = k\Delta t, \quad \forall k = 0, \dots, N_T.$$

If we have the optimal σ_{t_k} and any $\lambda \in \mathbb{R}^m$, the HJB equation (21) becomes a linear PDE which can be numerically solved by an ADI finite difference method (see [22] for more details). To approximate the optimal $\sigma^2_{t_k}$, at each time step $t = t_k$, we solve (25) with $\phi = \phi_{t_{k+1}}$ via a closed-form formula. If t_{k+1} equals to the maturity of any input option, we incorporate the jump discontinuity by adding the jump size $\sum_{i=1}^m \lambda_i G_i \mathbb{1}(t_i = t_{k+1})$ into $\phi_{t_{k+1}}$.

Setting an initial λ, the objective $\mathcal{V}$ is maximised over $\lambda \in \mathbb{R}^m$ by an optimisation algorithm. At each optimisation iteration, we calculate ϕ_0 by the above arguments. The process is aided by numerically computing the gradient (22). We measure the optimality by the uniform norm of the gradient. Setting a threshold ϵ_1, the algorithm terminates when the following stopping criterion is reached:

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

The method is summarised in Algorithm 1.

5.1 Simulated data

With simulated data, we provide two numerical examples to demonstrate the calibration results. In both examples, the risk-free rate is set to a constant $r = 0.05$ and the dividend yield is set to $q = 0$. Let $Z_0 = \ln 100$ and $V_0 = 0.04$ for both models, we set the problem in the computational domain $Q = [Z_0 - 4\sqrt{V_0}, Z_0 + 4\sqrt{V_0}] \times [0, 0.5]$ with the time interval $[0, 1]$. For the spatial discretisation on Q, we use 51 points in Z-direction and 51 points in V-direction. For the time interval, we choose $N_T = 100$. The LSV model is calibrated to a set of European call options generated by the Heston model with given parameters. The

Algorithm 1: LSV calibration

Data: Market prices of European option
Result: A calibrated LSV model that matches all market prices
Set an initial λ;
do
  for $k = N_T - 1, \dots, 0$ **do**
    if t_{k+1} *is equal to the maturity of any input option.* **then**
      $\phi_{t_{k+1}} \leftarrow \phi_{t_{k+1}} + \sum_{i=1}^m \lambda_i G_i \mathbb{1}(t_i = t_{k+1})$;
    end
    Approximate the optimal $\sigma^2_{t_k}$ by solving (25) with $\phi_{t_{k+1}}$;
    Solve the HJB equation (21) by the ADI method at $t = t_k$;
  end
  Solve (24) to calculate the model prices by the ADI method;
  Calculate the gradient $\partial_\lambda J(\lambda)$ by (22);
  Update λ by an optimisation algorithm;
while $\|\partial_\lambda J(\lambda)\|_\infty > \epsilon_1$;

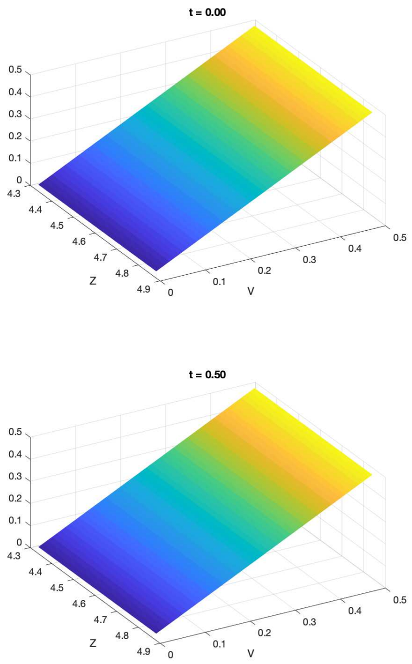


Figure 2: The volatility function $\sigma^2(t, Z, V)$ for Example 1.

option prices are calculated at maturities $t_i \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$ for strikes of all points in $[Z_0 - 1.4\sqrt{V_0}, Z_0 + 1.4\sqrt{V_0}]$. The parameter $p = 4$ is chosen for the cost function. For solving the HJB equation (21) and the backward pricing PDE (24), we use the ADI method of Douglas and Rachford [9] (also see [22]).

5.1.1 Example 1

We use the parameters $\kappa = 0.5, \theta = 0.04, \xi = 0.16$ and $\bar{\eta} = -0.4$ for both the LSV model and the Heston model. This example represents a trivial case, since we know that the optimal analytical solution is $\sigma^2(t, Z, V) = V$ and $\eta(t, Z, V) = \bar{\eta}$ if we use the same set of parameters for both models. With a threshold $\epsilon_1 = 10^{-6}$, we obtain the expected results (see Figure 2).

5.1.2 Example 2

In this example, we give different parameters to the LSV model and the Heston model (see Table 2). As in Remark 4.1, the LSV model can be converted to a LV model, so it has the ability to be exactly calibrated to all option prices generated by Heston model with any parameters. We obtain the result with threshold $\epsilon_1 = 0.0005$. Figures 4 and 3 visualise the volatility function $\sigma^2(t, Z, V)$ and the correlation function $\eta(t, Z, V)$. Figure 5 shows the implied volatility (IV) of the input and the model generated options. We can see that all the model IVs match their corresponding input IVs. The values of a subset of implied IVs are given in Table 1.

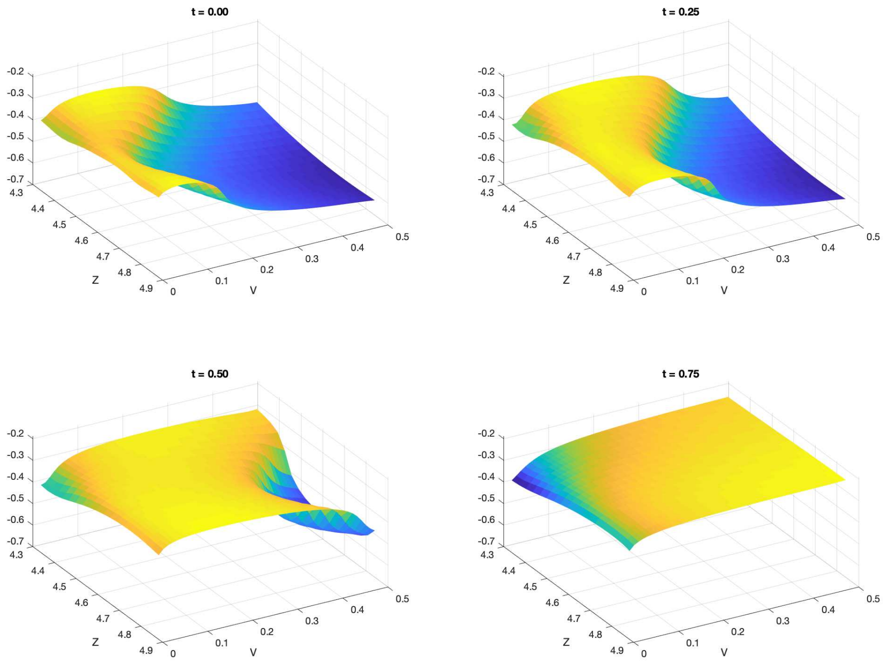


Figure 3: The correlation function $\eta(t, Z, V)$ in Example 2.

5.2 Market data

In this section, we calibrate the LSV model to the FX options data provided in [32]. The data are listed in Appendix A.4. The parameters we use are shown in Table 3. In this case, $2\kappa\theta/\xi^2 = 0.169 \ll 1$ and the Feller condition is strongly violated.

For the numerical settings, the computational domain is set to $Q = [-0.6, 1.0] \times [0, 2]$. For improving the accuracy while still keeping a reasonable computation time, we employ non-uniform meshes in both Z- and V-directions and put more points around (Z_0, V_0) (see [22, Section 2.2] for more details). For the time interval $[0, 5]$, we give 30 time steps with an equal step size between any two consecutive maturities.

In the simulated data example, F^* and the optimal $\sigma^2_{t_k}$ are simply approximated by using $\phi_{t_{k+1}}$ for a clear illustration. We therefore introduce an iterative algorithm to improve the convergence. It is described in Appendix A.3.

Setting the threshold $\epsilon_1 = 6 \times 10^{-6}$, we obtain an exact calibration. The maximum difference between the model IVs and market IVs is less than 1 basis point. Figure 6 shows the IVs of short-maturity options (1 month and 3 months) for the market data, the Heston model and the LSV model. Figure 7 shows the IVs of long-maturity options (2 years and 5 years).

Maturity	Log-strike	Input IV	Model IV	Error
T = 0.2	4.3492	0.2396	0.2396	5.05E-05
	4.4452	0.2291	0.2291	2.81E-06
	4.5732	0.2199	0.2199	4.08E-06
	4.7012	0.2138	0.2138	5.02E-06
	4.8292	0.2123	0.2124	8.48E-06
T = 0.4	4.3492	0.2488	0.2488	3.94E-06
	4.4452	0.2422	0.2422	1.16E-06
	4.5732	0.2359	0.2359	1.65E-07
	4.7012	0.2303	0.2303	9.82E-07
	4.8292	0.2257	0.2257	4.55E-06
T = 0.6	4.3492	0.2576	0.2576	2.84E-06
	4.4452	0.2523	0.2523	1.60E-06
	4.5732	0.2471	0.2471	7.56E-07
	4.7012	0.2423	0.2423	3.35E-07
	4.8292	0.2378	0.2378	3.24E-09
T = 0.8	4.3492	0.2646	0.2646	3.45E-06
	4.4452	0.2600	0.2600	8.69E-07
	4.5732	0.2555	0.2555	7.62E-07
	4.7012	0.2512	0.2512	1.42E-07
	4.8292	0.2472	0.2472	4.91E-09
T = 1.0	4.3492	0.2699	0.2699	8.28E-07
	4.4452	0.2659	0.2659	1.83E-06
	4.5732	0.2620	0.2620	1.17E-06
	4.7012	0.2581	0.2581	1.54E-06
	4.8292	0.2544	0.2544	9.24E-07

Table 1: A subset of the input IVs and model IVs in Example 2.

	κ	θ	ξ	$\bar{\eta}$
Heston	2.0	0.09	0.10	-0.6
LSV	0.5	0.04	0.16	-0.4

Table 2: The parameters used in the Heston model and the LSV model

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

Table 3: The parameters used in the market data example

A Appendix

A.1 Lemma A.1

Lemma A.1. *Define* $\Phi : C_b(\Lambda, \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ *by*

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

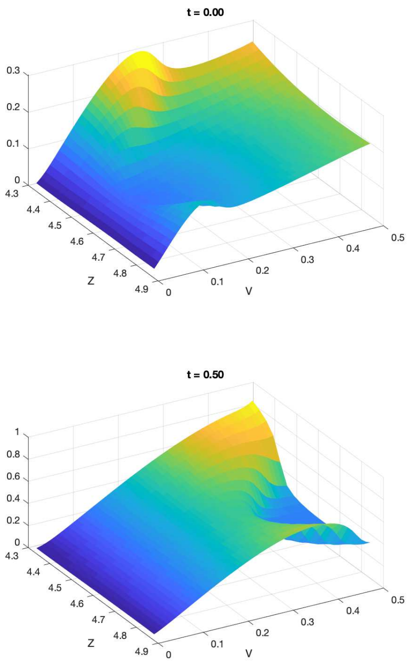


Figure 4: The volatility function $\sigma^2(t, Z, V)$ in Example 2.

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

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

Proof. For any $(\rho, \mathcal{A}, \mathcal{B}) \in C_b(\Lambda, \mathcal{X})^*$, the convex conjugate Φ^* is given by

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

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

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

To show that ρ is a nonnegative measure, we assume that $\rho(E) < 0$ for some measurable set $E \subset \Lambda$. By the density of C_b in L^1, there exists a sequence of functions in $C_b(\mathbb{R})$ that converges to $\mathbb{1}_E$. Let $r = -\lambda \mathbb{1}_E$ for some positive λ, $a = O^{d\times 1}$ and $b = O^{d\times d}$ where $O^{m\times n}$ denotes the null matrix of size $m \times n$. It is clear that the constraint $r + F^*(a, b) \le 0$ is satisfied at $(-\lambda \mathbb{1}_E, O^{d\times 1}, O^{d\times d})$ as F is nonnegative. If we send λ to infinity, the function Φ^* becomes unbounded.

To show $(\mathcal{A}, \mathcal{B}) \ll \rho$, we assume that there exists a measurable set E such that $(\mathcal{A}, \mathcal{B})(E) \neq 0$ but $\rho(E) = 0$. Again, we construct a sequence of continuous functions in $C_b(\Lambda)$ converging to $\mathbb{1}_E$ in $L^1(d\rho)$. Then let $r = -F^*(a, b)$, $a = \lambda \mathbb{1}_E I^{d\times 1}$ and $b = \lambda \mathbb{1}_E I^{d\times d}$ where $I^{m\times n}$ denotes the identity matrix of size $m \times n$. The function Φ^* goes to infinity if we send λ to $+\infty$ or $-\infty$, depending on the sign of $\mathcal{A}$ and $\mathcal{B}$ on E.

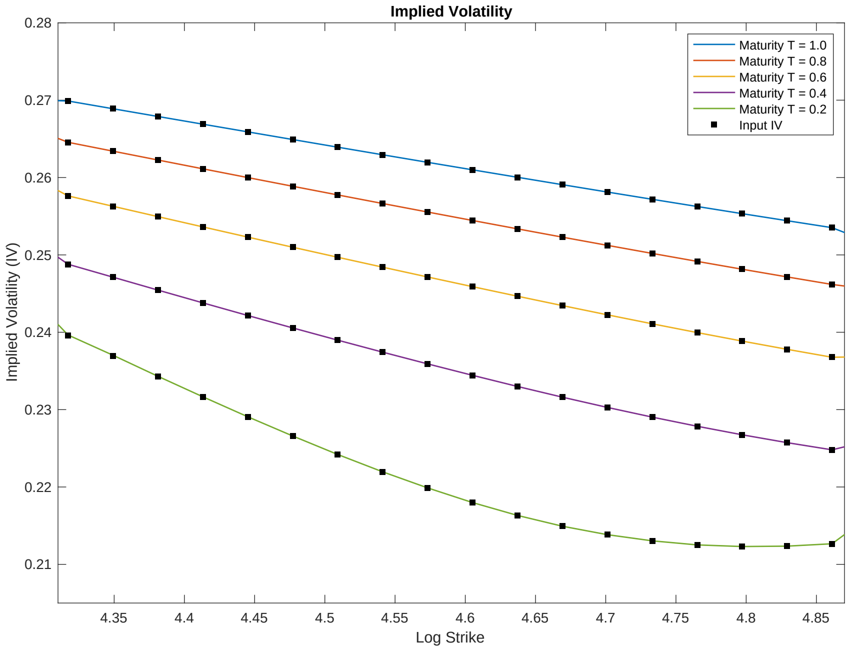


Figure 5: The input and the model generated IV for Example 2.

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

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

The interchange of the supremum and integral is made by choosing a sequence of continuous functions $(a, b)_n$ converging to $\nabla F(\frac{d\mathcal{A}}{d\rho}, \frac{d\mathcal{B}}{d\rho})$, then applying the dominated convergence theorem and the fact that F^* is continuous in its effective domain. The last equality holds since the cost function F equals to its biconjugate F^{**} according to the Fenchel–Moreau theorem (see [6, Theorem 1.11]). The proof is completed. $\square$

A.2 Lemma A.2

In this section, we prove that, the duality between spaces C_b and $\mathcal{M}$ can be extended to the non-compact space $[0, T] \times \mathbb{R}^d$ in this particular case. A similar argument for the Kantorovich duality of the classical optimal transport was made in [34, Appendix 1.3].

Lemma A.2. *Denote by K^o the set of (r, a, b) in $C_b(\Lambda, \mathcal{X})$ that can be represented by some (ϕ, λ) in $BV([0, T], C_b^2(\mathbb{R}^d)) \times \mathbb{R}^m$ (see the proof of Theorem 3.5 for the definition of*

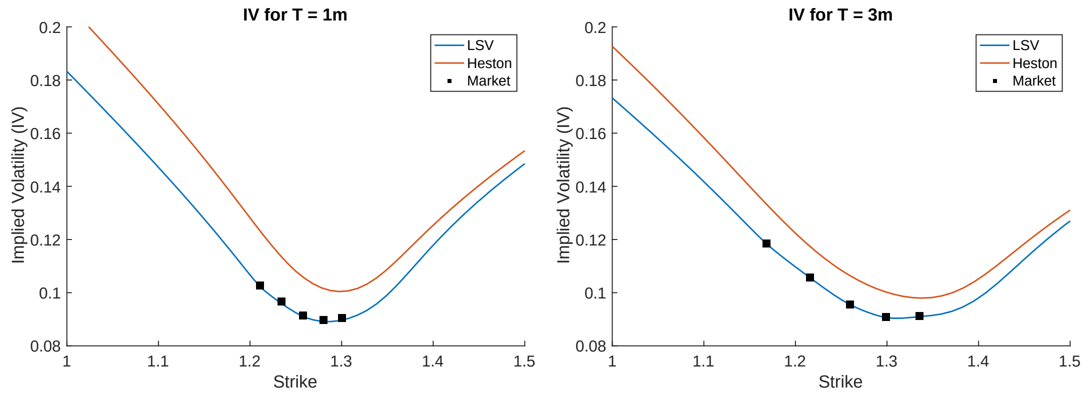


Figure 6: IV for options with maturities 1 month and 3 months

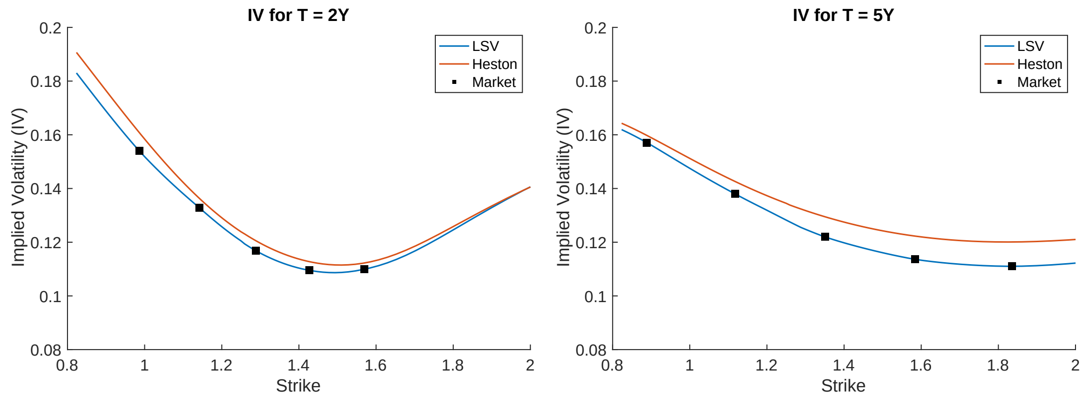


Figure 7: IV for options with maturities 2 years and 5 years

'represented'). Let $\Phi^* : C_b(\Lambda, \mathcal{X})^* \to \mathbb{R} \cup \{+\infty\}$ *and* $\Psi^* : C_b(\Lambda, \mathcal{X})^* \to \mathbb{R} \cup \{+\infty\}$ *be defined by*

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

$$\Psi^*(\rho, \mathcal{A}, \mathcal{B}) = \sup_{(r,a,b) \in K^o} \langle (r, a, b), (\rho - \bar{\rho}, \mathcal{A} - \bar{\mathcal{A}}, \mathcal{B} - \bar{\mathcal{B}}) \rangle.$$

Then,

$$\inf_{(\rho, \mathcal{A}, \mathcal{B}) \in C_b(\Lambda, \mathcal{X})^*} (\Phi^* + \Psi^*)(\rho, \mathcal{A}, \mathcal{B}) = \inf_{(\rho, \mathcal{A}, \mathcal{B}) \in \mathcal{M}(\Lambda, \mathcal{X})} (\Phi^* + \Psi^*)(\rho, \mathcal{A}, \mathcal{B}). \tag{26}$$

Proof. Let $C_0(\Lambda, \mathcal{X})$ be the space of continuous functions on Λ valued in $\mathcal{X}$ that vanish at infinity. We decompose $(\rho, \mathcal{A}, \mathcal{B}) = (\tilde{\rho}, \tilde{\mathcal{A}}, \tilde{\mathcal{B}}) + (\delta\rho, \delta\mathcal{A}, \delta\mathcal{B})$ such that $(\tilde{\rho}, \tilde{\mathcal{A}}, \tilde{\mathcal{B}}) \in \mathcal{M}(\Lambda, \mathcal{X})$ and $\langle (\phi_\rho, \phi_\mathcal{A}, \phi_\mathcal{B}), (\delta\rho, \delta\mathcal{A}, \delta\mathcal{B}) \rangle = 0$ for any $(\phi_\rho, \phi_\mathcal{A}, \phi_\mathcal{B}) \in C_0(\Lambda, \mathcal{X})$ [1]. Since, $\mathcal{M}(\Lambda, \mathcal{X})$ is a subset of $C_b(\Lambda, \mathcal{X})^*$, it follows that

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

Next, we show that the converse of the above inequality is also valid.

[1] The reader can refer to [34, Appendix 1.3] for the existence of such a decomposition.

For Φ^*, it follows by a standard approximation argument that

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

For Ψ^*, if we restrict its domain to $(\tilde{\rho},\tilde{\mathcal{A}},\tilde{\mathcal{B}})\in\mathcal{M}(\Lambda,\mathcal{X})$, then $\Psi^*=0$ if $(\tilde{\rho},\tilde{\mathcal{A}},\tilde{\mathcal{B}})$ satisfies the constraints (7) and (8) or $\Psi^*=+\infty$ otherwise. Whenever Ψ^* is finite,

$$\int_\Lambda r\,d(\tilde{\rho}-\bar{\rho}) + a\cdot d(\tilde{\mathcal{A}}-\bar{\mathcal{A}}) + b : d(\tilde{\mathcal{B}}-\bar{\mathcal{B}}) = 0 \qquad \forall (r,a,b)\in K^o. \tag{27}$$

The equation (27) holds in particular for (r,a,b) in the subset $K^o\cap C_0(\Lambda,\mathcal{X})$. Thus,

$$\begin{aligned}\Psi^*(\tilde{\rho},\tilde{\mathcal{A}},\tilde{\mathcal{B}}) &= \sup_{(r,a,b)\in K^o}\int_\Lambda r\,d(\tilde{\rho}-\bar{\rho}) + a\cdot d(\tilde{\mathcal{A}}-\bar{\mathcal{A}}) + b : d(\tilde{\mathcal{B}}-\bar{\mathcal{B}}) \\ &= \sup_{(r,a,b)\in K^o\cap C_0(\Lambda,\mathcal{X})}\int_\Lambda r\,d(\tilde{\rho}-\bar{\rho}) + a\cdot d(\tilde{\mathcal{A}}-\bar{\mathcal{A}}) + b : d(\tilde{\mathcal{B}}-\bar{\mathcal{B}}) \\ &= \sup_{(r,a,b)\in K^o\cap C_0(\Lambda,\mathcal{X})} \langle (r,a,b),(\rho-\bar{\rho},\mathcal{A}-\bar{\mathcal{A}},\mathcal{B}-\bar{\mathcal{B}})\rangle \\ &\leq \Psi^*(\rho,\mathcal{A},\mathcal{B})\end{aligned}$$

where the last inequality holds since $K^o\cap C_0(\Lambda,\mathcal{X})$ is a subset of K^o.

Therefore,

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

This completes the proof. □

A.3 Algorithm 2

To improve the convergence, we add an iteration step to handle the nonlinearity in the HJB equation. Let $\phi_{t_k}^{old}=\phi_{t_{k+1}}$, we approximate $\sigma_{t_k}^2$ using $\phi_{t_k}^{old}$. Then, we solve the HJB equation (21) with σ_{t_k} at $t=t_k$ to get $\phi_{t_k}^{new}$. If $\|\phi_{t_k}^{new}-\phi_{t_k}^{old}\|_2$ is greater than a threshold ϵ_2, we let $\phi_{t_k}^{old}=\phi_{t_k}^{new}$ and repeat above steps. Otherwise, we let $\phi_{t_k}=\phi_{t_k}^{new}$ and continue. The modified algorithm is described in Algorithm 2.

Algorithm 2: LSV calibration with nonlinear iterations

Data: Market prices of European option
Result: A calibrated LSV model that matches all market prices
Set an initial λ;
do
  for $k = N_T - 1, \ldots, 0$ **do**
    if t_{k+1} *is equal to the maturity of any input option.* **then**
      $\phi_{t_{k+1}} \leftarrow \phi_{t_{k+1}} + \sum_{i=1}^{m} \lambda_i G_i \mathbb{1}(t_i = t_{k+1})$;
    end
    Let $\phi_{t_k}^{old} = \phi_{t_{k+1}}$;
    do
      Approximate the optimal $\sigma_{t_k}^2$ by solving (25) with $\phi_{t_k}^{old}$;
      Solve the HJB equation (21) by the ADI method at $t = t_k$, and set the solution to $\phi_{t_k}^{new}$;
    while $\|\phi_{t_k}^{new} - \phi_{t_k}^{old}\|_2 > \epsilon_2$;
    Let $\phi_{t_k} = \phi_{t_k}^{new}$;
  end
  Solve (24) to calculate the model prices by the ADI method;
  Calculate the gradient $\partial_\lambda J(\lambda)$ by (22);
  Update λ by an optimisation algorithm;
while $\|\partial_\lambda J(\lambda)\|_\infty > \epsilon_1$;

A.4 FX option data

The market data used for calibration is EUR/USD option data quoted on 23 August 2012. The spot price $S_0 = 1.257$ USD per EUR. Maturities, strikes and implied volatilities are shown in Table 4. At each maturity, the options correspond to 10-delta calls, 25-delta calls, 50-delta calls, 25-delta puts and 10-delta puts. The interest yields are given in Table 5. Note that the yields data are yield to maturity. Thus we need to convert them to spot rates for our method.

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

Table 4: The EUR/USD option data quoted on 23 August 2012.

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

Table 5: The domestic and foreign yields (in %) quoted on 23 August 2012.

References

[1] Abergel, F., and Tachet, R. A nonlinear partial integro-differential equation from mathematical finance. *Discrete Contin. Dyn. Syst. 27*, 3 (2010), 907–917.

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

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

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

[5] Brenier, Y. Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. *Comm. Pure Appl. Math. 52*, 4 (1999), 411–452.

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

[7] Crandall, M. G., Ishii, H., and Lions, P.-L. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.) 27*, 1 (1992), 1–67.

[8] Dolinsky, Y., and Soner, H. M. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields 160*, 1-2 (2014), 391–427.

[9] Douglas, Jr., J., and Rachford, Jr., H. H. On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc. 82* (1956), 421–439.

[10] Dupire, B. Pricing with a smile. *Risk Magazine* (1994), 18–20.

[11] Engelmann, B., Koster, F., and Oeltz, D. Calibration of the Heston stochastic local volatility model: A finite volume scheme. *https://ssrn.com/abstract=1823769* (2011).

[12] Figalli, A. Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *J. Funct. Anal. 254*, 1 (2008), 109–153.

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

[14] Guo, I., and Loeper, G. Path dependent optimal transport and model calibration on exotic derivatives. *arXiv preprint arXiv:1812.03526* (2018).

[15] Guo, I., Loeper, G., and Wang, S. Local volatility calibration by optimal transport. *2017 MATRIX Annals 2* (2019), 51–64.

[16] Guyon, J., and Henry-Labordere, P. The smile calibration problem solved. *https://ssrn.com/abstract=1885032* (2011).

[17] Gyöngy, I. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probab. Theory Relat. Fields 71*, 4 (1986), 501–516.

[18] Henry-Labordere, P. Calibration of local stochastic volatility models to market smiles: A monte-carlo approach. *Risk Magazine* (2009).

[19] Henry-Labordère, P., and Touzi, N. An explicit martingale version of the one-dimensional brenier theorem. *Finance Stoch. 20*, 3 (2016), 635–668.

[20] Heston, S. L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud. 6*, 2 (1993), 327–343.

[21] Huesmann, M., and Trevisan, D. A Bnamou–Brenier formulation of martingale optimal transport. *arXiv preprint arXiv:1707.01493* (2017).

[22] In 't Hout, K. J., and Foulon, S. ADI finite difference schemes for option pricing in the Heston model with correlation. *Int. J. Numer. Anal. Model. 7*, 2 (2010), 303–320.

[23] Jex, M., Henderson, R., and Wang, D. Pricing exotics under the smile. *Risk Magazine* (1999), 72–75.

[24] Kantorovich, L. V. On a problem of Monge (in Russian). *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 3* (1948), 255–226.

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

[26] Loeper, G. The reconstruction problem for the Euler-Poisson system in cosmology. *Arch. Ration. Mech. Anal. 179*, 2 (2006), 153–216.

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

[28] Pal, S., Wong, T.-K. L., et al. Exponentially concave functions and a new information geometry. *Ann. Probab. 46*, 2 (2018), 1070–1113.

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

[30] Saporito, Y. F., Yang, X., and Zubelli, J. P. The calibration of stochastic local-volatility models: an inverse problem perspective. *Comput. Math. Appl. 77*, 12 (2019), 3054–3067.

[31] Tan, X., and Touzi, N. Optimal transportation under controlled stochastic dynamics. *Ann. Probab. 41*, 5 (2013), 3201–3240.

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

[33] Trevisan, D. Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. *Electron. J. Probab. 21* (2016), Paper No. 22, 41.

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

[35] Wyns, M., and Du Toit, J. A finite volume–alternating direction implicit approach for the calibration of stochastic local volatility models. *Int. J. Comput. Math. 94*, 11 (2017), 2239–2267.