

Joint Modelling and Calibration of SPX and VIX by Optimal Transport

Ivan Guo^{1,2}, Grégoire Loeper^{1,2}, Jan Obłój³, and Shiyi Wang¹

¹School of Mathematics, Monash University, Clayton, VIC, Australia

²Centre for Quantitative Finance and Investment Strategies,
Monash University, Clayton, VIC, Australia

³University of Oxford, Oxford, United Kingdom

September 14, 2020

Abstract

This paper addresses the joint calibration problem of SPX options and VIX options or futures. We show that the problem can be formulated as a semimartingale optimal transport problem under a finite number of discrete constraints, in the spirit of [arXiv:1906.06478]. We introduce a PDE formulation along with its dual counterpart. The solution, a calibrated diffusion process, can be represented via the solutions of Hamilton–Jacobi–Bellman equations arising from the dual formulation. The method is tested on both simulated data and market data. Numerical examples show that the model can be accurately calibrated to SPX options, VIX options and VIX futures simultaneously.

1 Introduction

The CBOE Volatility Index (VIX), also known as the stock market's "fear gauge", reflects the expectations of investors on the volatility of the S&P500 index (SPX) over the next 30 days. Although the index in itself is not a tradable asset, its derivatives such as futures and options are highly liquid. Since the VIX options started trading in 2006, researchers and practitioners have been putting a lot of effort in jointly calibrating models to the SPX and VIX options prices. It has proven to be a challenging problem. As noted by many authors (e.g., [27, 36]), inconsistencies might appear between the volatility-of-volatility inferred from SPX and VIX.

In the literature, the first attempt at jointly calibrating with continuous models¹ was made by Gatheral [14], who considered a two-factor stochastic volatility model. Other attempts include a Heston model with stochastic volatility-of-volatility by Fouque and Saporito [13] and a regime-switching stochastic volatility model by Goutte et al. [16]. In addition, many authors have tried incorporating jumps into the SPX dynamics, see, e.g., [3, 9, 28, 34, 35]. However, even with jumps, these models have yet to achieve satisfactory accuracy, particularly for short maturities. This leads to a natural question of whether there exists a continuous model which can capture the SPX and VIX smiles simultaneously. In [1, 21], Acciaio and Guyon provide a necessary condition for the existence of such continuous models. Their work was followed by the contribution of Gatheral

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

et al. [15] who recently found an instance of such continuous models called the quadratic rough Heston model. Note that apart from continuous models, Guyon [22] accurately reproduced the SPX and VIX smiles by modelling the distributions of SPX in discrete time.

Recently, the theory of optimal transport was adapted to solve problems in robust hedging and pricing both in discrete and in continuous time models, see [5, 24]. It has proved a powerful tool since then and its applications were extended to non-parametric model calibration. In particular, the discrete-time martingale optimal transport has been applied to derive model-independent bounds on VIX derivatives by De Marco and Henry-Labordere [10]. The theory has been further used to calibrate the non-parametric discrete-time model proposed by Guyon [22]. Continuous-time optimal transport was applied by three of the authors of this paper to the calibration of local volatility [20] and local-stochastic volatility models [19] to European options. Furthermore, in [17], the first two authors have extended the semimartingale optimal transport problem [37] to a more general path-dependent setting. Their work expands the available calibration instruments from European options to path-dependent options, such as Asian options, barrier options and lookback options.

In this paper, we introduce a time continuous formulation of the joint calibration problem. Instead of directly modelling the instantaneous volatility or the VIX index, we consider a semimartingale X whose first element X^1 is the logarithm of the SPX price and whose second element X^2 is defined as the expectation of the forward quadratic variation of X^1 . By doing so, the calibration exercise only depends on the marginals of X at fixed times, and the joint calibration problem falls into the class of the semimartingale optimal transport problem studied in [19]. As a corollary of the superposition principle of Trevisan [38] (or earlier Figalli [12] for the bounded coefficients case), for any probability measure such that the drift and diffusion of X are adapted processes, there exists another measure under which the semimartingale X reduces to a time-inhomogeneous diffusion and has the same marginals at fixed times under both measures. It is worth noting that the idea of using diffusion processes to mimic an Itô process by matching their marginals at fixed times traces back to the classical mimicking theorem of Gyöngy [23], which was later extended by Brunick and Shreve [8] to remove the conditions of nondegeneracy and boundedness on the covariance of the Itô process. Based on this result, as shown in [19], it is sufficient to look for solutions among such diffusion processes. This allows us to deduce a PDE formulation of the problem along with its dual counterpart. The latter naturally gives rise to Hamilton–Jacobi–Bellman (HJB) equations which can be used to represent the solutions to the original problem.

In terms of numerical aspects, pricing of VIX derivatives involves evaluating the square root of a conditional expectation. This requires nested Monte Carlo or least square Monte Carlo methods. Nested Monte Carlo has good accuracy, but is computationally expensive. Least square Monte Carlo is efficient, but it is difficult to determine the sign of the error, which can be a useful piece of information in risk management. In the previous work of two authors of this paper [18], the least square Monte Carlo approach was adapted for computing the duality bounds of VIX derivatives. In this paper, by taking X^2 as the forward quadratic variation of X^1 , we can use conventional Monte Carlo methods or PDE methods to calculate the prices of VIX options and futures. Let us point out that, by defining suitable state variables, our results are applicable to any calibration problem in which the calibration instruments have payoffs in the form of a function of a conditional expectation.

The paper is organised as follows. Section 2 introduces some basic notations and the formulation of the problem. Section 3 presents the main results including the localisation result, the PDE formulation and the dual formulation. Section 4 describes the numerical method in detail. Finally, in Section 5, we provide numerical examples with both simulated data and market data.

2 Problem formulation

2.1 Preliminaries

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

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

$$X_t = X_0 + A_t + M_t, \quad \langle X \rangle_t = \langle M \rangle_t = B_t, \quad \mathbb{P}\text{-a.s.}, \quad (1)$$

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

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

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

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

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

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

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

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

2.2 The joint calibration problem

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

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

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

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

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

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

Note that the second term on the right-hand side of (2) is the T -futures price on X^1 at time t and hence is a martingale. It follows that the modelling setting we just described is captured by probability measures $\mathbb{P} \in \mathcal{P}^1$ characterised by (α, β) such that

$$\alpha_t = \begin{bmatrix} -\frac{1}{2}\sigma_t^2 \\ -\frac{1}{2}\sigma_t^2 \end{bmatrix} \quad \text{and} \quad \beta_t = \begin{bmatrix} \sigma_t^2 & (\beta_t)_{12} \\ (\beta_t)_{12} & (\beta_t)_{22} \end{bmatrix}, \quad 0 \leq t \leq T, \quad (3)$$

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

Remark 2.1. We note that this is a fully non-parametric description of all the models in \mathcal{P}^1 compatible with the market setting described above. In particular, we do not specify the dynamics of the volatility $(\sigma_t)_{t \leq T}$. Since we model the SPX price and the expected forward quadratic variation, the semimartingale X can reproduce the volatility smiles of a wide range of stochastic volatility models.

In order to restrict the probability measures to those characterised by (α, β) of the form (3), we can define a cost function that penalises characteristics that are not in the following convex set:

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

Define the convex cost function F as follows:

$$F(\alpha, \beta) = \begin{cases} \sum_{i,j=1}^2 (\beta_{ij} - \bar{\beta}_{ij})^2 & \text{if } (\alpha, \beta) \in \Gamma, \\ +\infty & \text{otherwise,} \end{cases} \quad (4)$$

where $\bar{\beta}$ is a matrix of some reference values for β . Note that $\bar{\beta}$ may depend on (t, X_t) . Then, F is finite if and only if (α, β) is in the form of (3). Furthermore, F allows for stability across calibration exercises through specification of a reference model $\bar{\beta}$. Employing F as the cost function, our aim will be to find a model which is the closest to $\bar{\beta}$ among the ones which calibrate fully to the given market data. We comment further on the significance of $\bar{\beta}$ below in Section 5.

The calibration instruments we consider are SPX European options, VIX options and VIX futures. The market prices of these derivatives can be imposed as constraints on X . Let G be a vector of m number of SPX option payoff functions². For example, if the i -th option is a put option with a strike K_i , then the payoff function $G_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by $G_i(x) = \max(K_i - \exp(x_1), 0)$. Let $u^{SPX} \in \mathbb{R}^m$ be the SPX option prices and $\tau \in [0, T]^m$ be the vector of their maturities. The prices u^{SPX} can be imposed on X by restricting \mathbb{P} to probability measures that satisfy

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

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

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

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

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

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

The VIX index at t_0 is defined to be the square root of the expected realised variance over the next 30 days (i.e., $T - t_0 = 30$ days), that is

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

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

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

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

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

$$X_0^2 = \mathbb{E}^{\mathbb{P}} \left(\frac{1}{2} \int_0^T \sigma_s^2 ds \right) = \int_0^{\hat{f}} \frac{\mathbb{E}^{\mathbb{P}}(k - S_T)^+}{k^2} dk + \int_{\hat{f}}^{\infty} \frac{\mathbb{E}^{\mathbb{P}}(S_T - k)^+}{k^2} dk,$$

where $\hat{f} = \mathbb{E}^{\mathbb{P}}(S_T)$ is the T -forward price of the SPX index (e.g., see [30]). If X_0^2 is not observable from the market, we can treat it as a parameter. Now, putting all the constraints together, we define a set of probability measures $\mathcal{P}(X_0, G, H, \tau, t_0, T, u^{SPX}, u^{VIX,f}, u^{VIX}) \subset \mathcal{P}^1$ as follows:

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

For simplicity, we write \mathcal{P}_{joint} as a shorthand for $\mathcal{P}(X_0, G, H, \tau, t_0, T, u^{SPX}, u^{VIX,f}, u^{VIX})$. Any $\mathbb{P} \in \mathcal{P}_{joint}$ is a feasible risk-neutral measure under which the semimartingale X reproduces

the market prices. If \mathcal{P}_{joint} is empty, it means that the market data is not compatible with a continuous-time semimartingale model. Adopting the convention $\inf \emptyset = +\infty$, we formulate the joint calibration problem as a *semimartingale optimal transport problem under a finite number of discrete constraints*, as studied in [19]:

Problem 1. Given $X_0, G, H, \tau, t_0, T, u^{SPX}, u^{VIX,f}$ and u^{VIX} , solve

$$V := \inf_{\mathbb{P} \in \mathcal{P}_{joint}} \mathbb{E}^{\mathbb{P}} \int_0^T F(\alpha_s^{\mathbb{P}}, \beta_s^{\mathbb{P}}) ds. \quad (5)$$

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

Remark 2.2. Let Y be an \mathcal{F}_T -measurable random variable. By identifying X_t^2 as a function of X_t^1 and $\mathbb{E}^{\mathbb{P}}(Y | \mathcal{F}_t)$, our results apply to any model calibration problem where the payoffs of the calibration instruments can be expressed as functions of X_t^1 and X_t^2 .

2.3 An example: the Heston model

The Heston model [25] is a one-factor stochastic volatility model which directly models the spot price S_t and the instantaneous variance ν_t under the risk-neutral measure. The model dynamics are given by

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

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

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

$$X_t^2 = \mathbb{E}^{\mathbb{P}} \left(\frac{1}{2} \int_t^T \nu_s ds \middle| \mathcal{F}_t \right) = \frac{1 - e^{-\kappa(T-t)}}{2\kappa} (\nu_t - \theta) + \frac{1}{2}\theta(T-t). \quad (6)$$

Define $A(t, \kappa) := (1 - e^{-\kappa(T-t)})/\kappa$, then a simple rearrangement of (6) gives that

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

The above equation establishes a one-to-one relation between ν_t and X_t^2 at time t . Applying Itô's formula to X_t^2 , we have

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

Therefore, the Heston model can be reformulated as

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

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

$$(\alpha_t, \beta_t) = \left(\begin{bmatrix} -\frac{1}{2}\nu(t, X_t^2, \kappa, \theta) \\ -\frac{1}{2}\nu(t, X_t^2, \kappa, \theta) \end{bmatrix}, \begin{bmatrix} \nu(t, X_t^2, \kappa, \theta) & \frac{1}{2}\eta\omega A(t, \kappa)\nu(t, X_t^2, \kappa, \theta) \\ \frac{1}{2}\eta\omega A(t, \kappa)\nu(t, X_t^2, \kappa, \theta) & \frac{1}{4}\omega^2 A(t, \kappa)^2 \nu(t, X_t^2, \kappa, \theta) \end{bmatrix} \right). \quad (7)$$

Further, it is easy to check that $\mathbb{E}^\mathbb{P}\nu(t, X_t^2, \kappa, \theta) < \infty$ and hence $\mathbb{P} \in \mathcal{P}^1$. The characteristics (7) will be used in the numerical example provided in Section 5 for generating simulated option prices and will also be used as a reference model.

3 Main results

This section is devoted to present our main results. By following [19], we first present a localisation result which shows that the optimal transportation cost can be achieved by a set of Markov processes. Focusing only on these Markov processes, we introduce a PDE formulation. Furthermore, we deduce a dual formulation and find the optimal characteristics as a by-product.

3.1 Localisation

In this section, we show that if Problem 1 is admissible then the optimal transportation cost V can be found by minimising (5) over a subset of probability measures under which X is a (time in-homogeneous) Markov processes. Before proceeding, we introduce some notations for brevity. Denote by $\mathbb{E}_{t,x}^\mathbb{P}$ the conditional expectation $\mathbb{E}^\mathbb{P}(\cdot | X_t = x)$. For any square matrix $\beta \in \mathbb{S}_+^2$, we write $\beta^{\frac{1}{2}}$ such that $\beta = \beta^{\frac{1}{2}}(\beta^{\frac{1}{2}})^\top$. Now, let us restate Lemma 3.2 of [19].

Lemma 3.1. *Let $\mathbb{P} \in \mathcal{P}^1$ and $\rho^\mathbb{P} = \rho^\mathbb{P}(t, \cdot) = \mathbb{P} \circ X_t^{-1}$ be the marginal distribution of X_t under \mathbb{P} , $t \leq T$. Then $\rho^\mathbb{P}$ is a weak solution to the Fokker–Planck equation:*

$$\begin{cases} \partial_t \rho^\mathbb{P} + \nabla_x \cdot (\rho^\mathbb{P} \mathbb{E}_{t,x}^\mathbb{P} \alpha_t^\mathbb{P}) - \frac{1}{2} \sum_{i,j} \partial_{ij} (\rho^\mathbb{P} (\mathbb{E}_{t,x}^\mathbb{P} \beta_t^\mathbb{P})_{ij}) = 0 & \text{in } [0, T] \times \mathbb{R}^2, \\ \rho_0^\mathbb{P} = \delta_{X_0} & \text{in } \mathbb{R}^2. \end{cases} \quad (8)$$

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

$$dX_t = \alpha^{\mathbb{P}'}(t, X_t) dt + (\beta^{\mathbb{P}'}(t, X_t))^{\frac{1}{2}} dW_t^{\mathbb{P}'}, \quad 0 \leq t \leq T, \quad (9)$$

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

Lemma 3.1 is a corollary of the superposition principle of Trevisan [38] and Figalli [12]. It is worth noting that the idea of using diffusion processes to mimic an Itô process by matching their

marginals at fixed times (also called Markovian projection in the literature) traces back to the classical mimicking theorem of Gyöngy [23], which was later extended by Brunick and Shreve [8] to remove the conditions of nondegeneracy and boundedness on the covariance of the Itô process.

Let $\mathcal{P}_{joint}^{loc} \subset \mathcal{P}_{joint}$ be a subset of probability measures under which X is Markov processes in the form of (9). In other words, any $\mathbb{P}' \in \mathcal{P}_{joint}^{loc}$ is characterised by $(\mathbb{E}_{t,x}^{\mathbb{P}'}, \alpha_t^{\mathbb{P}'}, \beta_t^{\mathbb{P}'})$ for some $\mathbb{P} \in \mathcal{P}^1$. Moreover, under \mathbb{P}' , X has an initial marginal δ_{X_0} and is fully calibrated to the market prices given in \mathcal{P}_{joint} . Applying Proposition 3.4 of [19], we have the following proposition for the joint calibration problem:

Proposition 3.2 (Localisation). *Given \mathcal{P}_{joint} and $\mathcal{P}_{joint}^{loc}$, if Problem 1 is admissible, then*

$$V = \inf_{\mathbb{P} \in \mathcal{P}_{joint}} \mathbb{E}^{\mathbb{P}} \int_0^T F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}}) dt = \inf_{\mathbb{P} \in \mathcal{P}_{joint}^{loc}} \mathbb{E}^{\mathbb{P}} \int_0^T F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}}) dt.$$

3.2 PDE formulation

For any $\mathbb{P} \in \mathcal{P}_{joint}^{loc}$, the characteristics are function of the state variable X_t and time t . As is classical in the theory of diffusions, this allows us to leverage PDE methods to describe Problem 1 and to use conventional numerical methods to find its solutions.

Proposition 3.3. *If Problem 1 is admissible, then*

$$V = \inf_{\rho, \alpha, \beta} \int_0^T \int_{\mathbb{R}^2} F(\alpha(t, x), \beta(t, x)) \rho(t, dx) dt, \quad (10)$$

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

$$\partial_t \rho(t, x) + \nabla_x \cdot (\rho(t, x) \alpha(t, x)) - \frac{1}{2} \sum_{i,j} \partial_{ij}(\rho(t, x) \beta_{ij}(t, x)) = 0 \quad \text{in } [0, T] \times \mathbb{R}^2, \quad (11)$$

$$\int_{\mathbb{R}^2} G_i(x) \rho(\tau_i, dx) = u_i^{SPX} \quad \forall i = 1, \dots, m, \quad (12)$$

$$\int_{\mathbb{R}^2} J(x) \rho(t_0, dx) = u^{VIX,f}, \quad (13)$$

$$\int_{\mathbb{R}^2} (H_i \circ J)(x) \rho(t_0, dx) = u_i^{VIX} \quad \forall i = 1, \dots, n, \quad (14)$$

$$\int_{\mathbb{R}^2} \xi(x) \rho(T, dx) = 0, \quad (15)$$

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

Proof. This proposition follows immediately from Lemma 3.1. The interchange of integrals in the objective is justified by Fubini's theorem. For the weak continuity of measure ρ in time we refer the reader to [32]. \square

The PDE formulation can be solved by the alternating direction method of multipliers (ADMM) which was originally used in [6] to solve the classical optimal transport. This method was extended to a one-dimensional martingale optimal transport problem in [20] and to instationary mean field games with diffusion in [2]. However, for problems with diffusions, the ADMM method requires to solve a fourth-order PDE with a bi-Laplacian operator. In this paper, we work on an alternative dual formulation derived by following the arguments in [19]. This will be presented in the next subsection.

3.3 Dual formulation

Although the PDE formulation is not a convex problem, it can be made convex by considering the triple of measures $(\rho, \mathcal{A}, \mathcal{B}) := (\rho, \rho\alpha, \rho\beta)$. By doing so, the objective function (10) is convex in $(\rho, \mathcal{A}, \mathcal{B})$. Moreover, the initial condition and the constraints (11) to (15) are linear in $(\rho, \mathcal{A}, \mathcal{B})$ and hence produce a convex feasible set. In consequence, the classical tools of convex analysis can be applied. Following Theorem 3.6 and Corollary 3.11 of [19], we introduce a dual formulation.

Let $\lambda^{SPX} \in \mathbb{R}^m$, $\lambda^{VIX,f} \in \mathbb{R}$, $\lambda^{VIX} \in \mathbb{R}^n$ and $\lambda^\xi \in \mathbb{R}$ be the Lagrange multipliers of constraints (12) to (15), respectively. To avoid confusion with the Dirac measure δ used previously, we denote by \mathcal{D} the Dirac delta function in the sense of distributions. The dual formulation is given as follows:

Theorem 3.4 (Duality). *If Problem 1 is admissible, then the infimum in (10) is attained and is equal to*

$$V = \sup_{(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^\xi) \in \mathbb{R}^{m+n+2}} \lambda^{SPX} \cdot u^{SPX} + \lambda^{VIX,f} u^{VIX,f} + \lambda^{VIX} \cdot u^{VIX} - \phi(0, X_0), \quad (16)$$

where ϕ is the viscosity solution to the HJB equation:

$$\begin{aligned} \partial_t \phi(t, x) + F^*(\nabla_x \phi(t, x), \frac{1}{2} \nabla_x^2 \phi(t, x)) &= - \sum_{i=1}^m \lambda_i^{SPX} G_i(x) \mathcal{D}(t - \tau_i) \\ &\quad - \lambda^{VIX,f} J(x) \mathcal{D}(t - t_0) - \sum_{i=1}^n \lambda_i^{VIX} (H_i \circ J)(x) \mathcal{D}(t - t_0) - \lambda^\xi \xi(x) \mathcal{D}(t - T) \quad \text{in } [0, T] \times \mathbb{R}^2, \end{aligned} \quad (17)$$

with the terminal condition $\phi(T, \cdot) = 0$. Moreover, if the supremum is attained by some λ^{SPX} , $\lambda^{VIX,f}$, λ^{VIX} and λ^ξ for which the associated solution to (17) is $\phi^* \in BV([0, T], C_b^2(\mathbb{R}^2))$, then the optimal (α, β) is

$$(\alpha^*, \beta^*) = \nabla F^*(\nabla_x \phi^*, \frac{1}{2} \nabla_x^2 \phi^*). \quad (18)$$

Theorem 3.4 is an application of the Fenchel–Rockafellar duality theorem [39, Theorem 1.9]. Due to the presence of \mathcal{D} in the source terms, the viscosity solution ϕ satisfies (17) in the sense of distributions³. Moreover, ϕ has possible discontinuities at t_0 , T and τ_i , $i = 1, \dots, m$. The numerical solution to (17) is described in detail in Section 4. For the cost function F defined in (4), the convex conjugate F^* is given in Lemma A.1.

In the dual formulation, the supremum can be solved by a standard optimisation algorithm. As pointed out in [19, Lemma 4.5], the convergence can be improved by providing the gradients of the objective.

Lemma 3.5. *Suppose Problem 1 is admissible and let*

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

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

³For the precise definition of viscosity solutions to (17) and the corresponding comparison principle, we refer the reader to [19, Section 3.3].

and the model prices:

$$\partial_{\lambda_i^{SPX}} L = u_i^{SPX} - \mathbb{E}^{\mathbb{P}} G_i(X_{\tau_i}), \quad i = 1, \dots, m, \quad (19)$$

$$\partial_{\lambda^{VIX,f}} L = u^{VIX,f} - \mathbb{E}^{\mathbb{P}} J(X_{t_0}), \quad (20)$$

$$\partial_{\lambda_i^{VIX}} L = u_i^{VIX} - \mathbb{E}^{\mathbb{P}} (H_i \circ J)(X_{t_0}), \quad i = 1, \dots, n, \quad (21)$$

$$\partial_{\lambda^\xi} L = -\mathbb{E}^{\mathbb{P}} \xi(X_T). \quad (22)$$

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

$$\begin{cases} \partial_t \phi' + \alpha \cdot \nabla_x \phi' + \frac{1}{2} \beta : \nabla_x^2 \phi' = 0, & \text{in } [0, \mathcal{T}) \times \mathbb{R}^2, \\ \phi'(\mathcal{T}, \cdot) = \mathcal{G}. \end{cases} \quad (23)$$

When applying Lemma 3.5, we shall be using (23) m times for $(\mathcal{G}, \mathcal{T}) = (G_i, \tau_i)$, $i = 1, \dots, m$, once for $(\mathcal{G}, \mathcal{T}) = (J, t_0)$, n times for $(\mathcal{G}, \mathcal{T}) = (H_i \circ J, t_0)$, $i = 1, \dots, n$, and once for $(\mathcal{G}, \mathcal{T}) = (\xi, T)$. We shall simply refer to this as solving the linear pricing PDEs (23). Naturally, once the optimal model (α^*, β^*) is found, the above can be used not only to verify that it is indeed calibrated but also to compute other option prices under the model.

4 Numerical methods

4.1 Solving the dual formulation

The numerical method proposed in [19] can be directly applied to solve the dual formulation, albeit with a number of caveats. Let us first recall the numerical method. Given an initial guess $(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^\xi)$, we solve the HJB equation (17) to get $\phi(0, X_0)$ and hence to calculate the objective value. Due to the presence of the Dirac delta functions \mathcal{D} , ϕ might be discontinuous in time. The HJB equation can be solved in several time intervals in which, in each interval, the solution ϕ is continuous in both time and space, and the source terms with \mathcal{D} can be incorporated into the terminal conditions. For example, if we consider SPX options with maturities t_0 and T ,

the HJB equation (17) can be reformulated as follows:

$$\left\{ \begin{array}{l} \partial_t \phi + \sup_{\beta \in \mathbb{S}_+^2} \left(-\frac{1}{2} \beta_{11} \partial_{x_1} \phi - \frac{1}{2} \beta_{11} \partial_{x_2} \phi + \frac{1}{2} \beta_{11} \partial_{x_1 x_1} \phi \right. \\ \quad \left. + \beta_{12} \partial_{x_1 x_2} \phi + \frac{1}{2} \beta_{22} \partial_{x_2 x_2} \phi - \sum_{i,j=1}^2 (\beta_{ij} - \bar{\beta}_{ij})^2 \right) = 0 \quad \text{in } [t_0, T), \\ \phi(T^-, \cdot) = \sum_{i=1}^m \lambda_i^{SPX} G_i \mathbf{1}(t = T) + \lambda^\xi \xi, \end{array} \right. \quad (24)$$

$$\left\{ \begin{array}{l} \partial_t \phi + \sup_{\beta \in \mathbb{S}_+^2} \left(-\frac{1}{2} \beta_{11} \partial_{x_1} \phi - \frac{1}{2} \beta_{11} \partial_{x_2} \phi + \frac{1}{2} \beta_{11} \partial_{x_1 x_1} \phi \right. \\ \quad \left. + \beta_{12} \partial_{x_1 x_2} \phi + \frac{1}{2} \beta_{22} \partial_{x_2 x_2} \phi - \sum_{i,j=1}^2 (\beta_{ij} - \bar{\beta}_{ij})^2 \right) = 0 \quad \text{in } [0, t_0), \\ \phi(t_0^-, \cdot) = \phi(t_0, \cdot) + \sum_{i=1}^m \lambda_i^{SPX} G_i \mathbf{1}(t = t_0) + \lambda^{VIX,f} J + \sum_{i=1}^n \lambda_i^{VIX} (H_i \circ J). \end{array} \right. \quad (25)$$

We then calculate the gradients of the objective by Lemma 3.5, in which the linear pricing PDEs (23) are solved by an alternating direction implicit (ADI) method (see e.g., [26]). Once we have the gradient values, we update $(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^\xi)$ by moving them against their gradients or by supplying gradients to an optimisation algorithm. Notably, the L-BFGS algorithm [31] was employed and showed good convergence. The above steps are repeated until some optimality condition is met. The numerical method is summarised in Appendix B.

4.2 Solving HJB equations

In terms of numerical schemes for HJB equations, in their seminal work, Barles and Souganidis [4] have established a convergence that requires schemes to be monotone. Since then, a wide literature on monotone schemes has developed. For multidimensional HJB equations, it is usually difficult to construct a monotone scheme because of the cross partial derivative terms. To ensure monotonicity, the explicit wide stencil schemes were studied by Bonnans and Zidani [7] and by Debrabant and Jakobsen [11]; however, the stability of explicit schemes are restricted by some CFL condition. In [33], Ma and Forsyth proposed an implicit wide stencil finite difference scheme with a local coordinate rotation which is unconditionally stable. They also maximised the use of the fixed point stencil and the central finite difference scheme to improve the order of accuracy while preserving the monotonicity of the scheme.

In this paper, we solve the HJB equations by a fully implicit finite difference method with central-difference schemes for approximating both first- and second-order derivatives. We discretise the time interval, and then, at each time step, we approximate β by Lemma A.1. Once the optimal β has been found, the fully nonlinear HJB equation reduces to a linear PDE which can be solved by the standard implicit finite difference method. When approximating β , we start with an arbitrary ϕ to approximate the derivatives of ϕ . Next we solve the linearised PDE and plug the solution back into the supremum to approximate β at the same time. The above procedure is repeated until ϕ converges, then we proceed to the next time step. This successive approximation is known as policy iteration in the literature. A good approximation to the initial ϕ is the one from the previous time step, which makes ϕ converge within a few iterations.

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

conditions to equations (24) and (25):

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

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

In addition, we set a sufficiently large computational domain to further reduce the impact of the boundary conditions. Since the linear pricing PDEs are related to the HJB equation, the boundary conditions of (23) are modified in a similar manner.

As noted in [29], the standard finite difference schemes are non-monotone unless the diffusion matrix is diagonally dominated. In spite of being non-monotone in general, this scheme has the advantage of second-order accuracy for smooth solutions and ease of implementation compared to sophisticated monotone schemes. In fact, the variance of X_t^2 is much smaller than the variance of X_t^1 , especially when t is close to T . Thus, we scale up X_t^2 by performing a simple change of variables: $(X^1, X^2) \mapsto (X^1, KX^2)$ with $K > 1$. In the numerical example of the next section we take $K = 40$. Although the diffusion matrix is not diagonally dominated and the scheme is still non-monotone in general, it shows good stability and convergence for this problem after the scaling.

4.3 Smoothing the volatility skews

It is clear from the formulation of Problem 1 that the reference $\bar{\beta}$ influences, potentially in a very significant way, the solution. This is also confirmed by our numerics, see Section 5.1 below. However, in practice, a good selection of the reference $\bar{\beta}$ might not be available. Assume that there exists a $\mathbb{P}_{mkt} \in \mathcal{P}_{joint}^{loc}$, characterised by $(\alpha_{mkt}, \beta_{mkt})$, which describes the real market dynamics. When $\bar{\beta}$ is far away from β_{mkt} , even though the optimised model matches all the calibrating option prices, the optimal β may still be very different from β_{mkt} . In the numerical experiment, we observed spiky volatility surfaces and hump-shaped model volatility skews. This is not surprising because the optimiser is trying to match the model prices to the calibrating option prices while keeping β close to $\bar{\beta}$.

Denote by $F^{\bar{\beta}}$ the cost function defined in (4) with reference $\bar{\beta}$. Let $V(\bar{\beta})$ be the optimal objective value of Problem 1 with cost function $F^{\bar{\beta}}$. If $V(\bar{\beta}) < \infty$, by Theorem 3.4, $V(\bar{\beta})$ is equal to the optimal objective value of the dual formulation with $(F^{\bar{\beta}})^*$ in the HJB equation (17). Let $R(\bar{\beta})$ be some regularisation term that measures the smoothness of $\bar{\beta}$. In order to smooth out the volatility surfaces and the model volatility skews, it is natural to consider the following problem:

$$\arg \inf_{\bar{\beta} \in L^1(d\rho_t dt, \mathbb{S}_+^2)} V(\bar{\beta}) + R(\bar{\beta}). \quad (26)$$

While we might not actually solve this problem, it motivates our *reference measure iteration* method. We start with an initial reference $\bar{\beta}^0$ and numerically solve the dual formulation with cost function $F^{\bar{\beta}^0}$. Then an optimal $(\beta^*)^0$ is obtained as a by-product of solving (17). Next, we smooth $(\beta^*)^0$ by a simple moving average over (t, X^1, X^2) with bandwidths of (l_t, l_{x_1}, l_{x_2}) . In the numerical examples, we set $(l_t, l_{x_1}, l_{x_2}) = (1, 3, 3)$. Next, we set the smoothed $(\beta^*)^0$ to $\bar{\beta}^1$ and solve the dual formulation with $\bar{\beta}^1$. The above steps are repeated until the model volatility skews are smooth enough.

Let us call the optimisation of solving (16) as the *inner iteration* and call the optimisation of solving (26) as the *outer iteration*. For the outer iteration, if the optimal $\bar{\beta}$ that achieves the

infimum in (26) is not very smooth, bandwidths (l_t, l_{x_1}, l_{x_2}) with large values might cause the optimiser searching around the optimal $\bar{\beta}$ forever. Thus, (l_t, l_{x_1}, l_{x_2}) can be intuitively interpreted as the "step size" for the outer iteration. Moreover, in practice, we can apply an early stop technique by only running for a few iterations for the inner iteration. By doing so, the optimiser is alternating between the inner iteration and the outer iteration. We include this procedure in our numerical routines presented in the next section.

5 Numerical experiments

5.1 Simulated data

In this section, we present a numerical example to demonstrate our method. We generate some calibrating options and futures prices from a Heston model with given parameters $(\kappa, \theta, \omega, \eta)$, and we call this model the *generating model*. Next, we calibrate the semimartingale X to these simulated prices by solving the dual formulation. In this case, we know that there exists such a probability measure $\mathbb{P} \in \mathcal{P}^1$ that X can be fully calibrated to the simulated prices under \mathbb{P} , i.e., $\mathcal{P}_{joint}^{loc} \neq \emptyset$. Recall that the interest rates and dividends are set to null. The characteristics of \mathbb{P} are given by (7) and the calibrating options and futures prices are computed by solving the linear pricing PDEs (23).

Recall that Problem 1, combined with Proposition 3.2, looks for a Markovian diffusion model which minimises a certain distance to a *reference model* $\bar{\beta}$ subject to being calibrated. In this section we not only show that our approach is feasible but also investigate the potential influence of the choice of the reference $\bar{\beta}$. Specifically, we consider two reference models:

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

$$\bar{\beta}(t, X_t^1, X_t^2) = \begin{bmatrix} \nu(t, X_t^2, \bar{\kappa}, \bar{\theta}) & \frac{1}{2}\bar{\eta}\bar{\omega}A(t, \bar{\kappa})\nu(t, X_t^2, \bar{\kappa}, \bar{\theta}) \\ \frac{1}{2}\bar{\eta}\bar{\omega}A(t, \bar{\kappa})\nu(t, X_t^2, \bar{\kappa}, \bar{\theta}) & \frac{1}{4}\bar{\omega}^2A(t, \bar{\kappa})^2\nu(t, X_t^2, \bar{\kappa}, \bar{\theta}) \end{bmatrix}; \quad (27)$$

(b) a model with constant reference values:

$$\bar{\beta}(t, X_t^1, X_t^2) = \begin{bmatrix} \bar{\beta}_{11} & \bar{\beta}_{12} \\ \bar{\beta}_{12} & \bar{\beta}_{22} \end{bmatrix}. \quad (28)$$

The optimal models (α^*, β^*) obtained using these two reference values will be referred to, respectively, as the *OT-calibrated model with a Heston reference* and the *OT-calibrated model with a constant reference*. These should not be confused with the generating (Heston) model. The idea behind the selection of candidates is to analyse the significance of $\bar{\beta}$ by comparing the results between two cases: (a) the dynamics of the reference model are close to the true dynamics, (b) the dynamics of the reference model are very different from the true dynamics. Note that in (a), if $(\bar{\kappa}, \bar{\theta}, \bar{\omega}, \bar{\eta}) = (\kappa, \theta, \omega, \eta)$, the supremum in (16) is achieved by a null vector $\mathbf{0} \in \mathbb{R}^{m+n+2}$ and hence $V = 0$. In this case, the OT-calibrated model quickly recovers the generating model.

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

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

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

Table 1: Parameter values and interpretations for the simulated data example.

Note that we also need to consider the singular contract (i.e., $\mathbb{E}^{\mathbb{P}}\xi(X_T) = 0$) to ensure that the dynamics of X are correct. All the parameter values and their interpretations are given in Table 1.

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

Ideally, we want the calibrated model to have at most 1 basis point error in implied volatility for both SPX options and VIX options. However, in our method, we can only calibrate the model to option prices instead of implied volatility. Therefore, we scale the payoff functions and option prices by dividing them by their Black–Scholes vegas, which roughly converts errors in option prices to errors in implied volatility. The optimisation algorithm will iterate until the maximum error between calibrating prices and model prices are below 0.0001, or until it cannot be further optimised. In addition, the volatility skews are smoothed by the reference measure iteration method introduced in Section 4.3.

The calibration results are shown in Table 2, and the volatility skews are given in Figure 1–2. We can see that the OT-calibrated models, both with the Heston reference and the constant reference, accurately capture the calibrating SPX options, VIX futures and VIX options prices. The errors, in implied volatility, of the SPX options are at most 1 basis point and of the VIX options are at most 10 basis points.

To verify if the model dynamics are correct, we perform a Monte Carlo simulation of X with the Euler scheme, and the results are shown in Figure 3–4. As demonstrated, $X_T^2 \approx 0$ in all three models, so we consider the constraints $X_T^2 = 0$ \mathbb{P} -a.s. are satisfied, and the model dynamics are correct. We note however that the dynamics of the three models are different. In fact, the OT-calibrated model with a constant reference is very different to the other two models. We display the volatility behaviour of the three models in Appendix C.

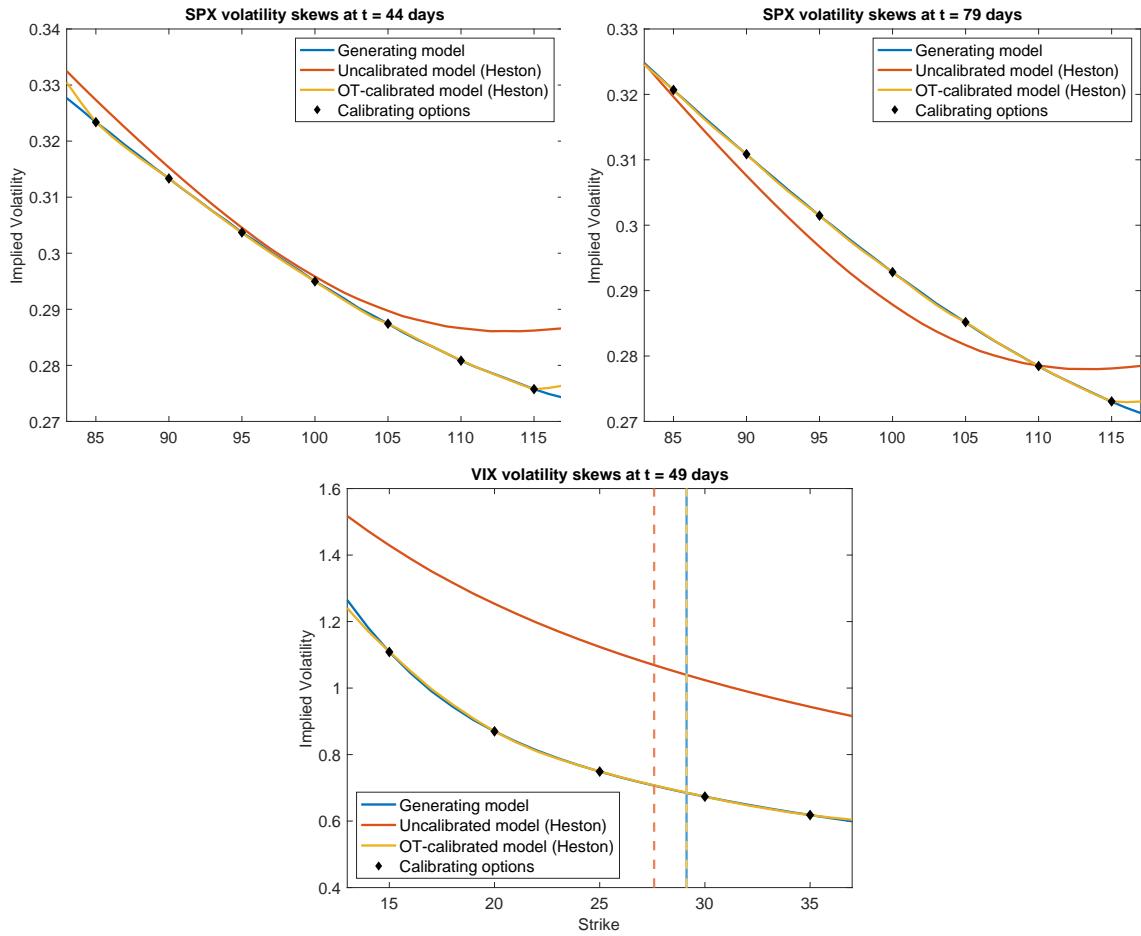


Figure 1: The volatility skews of SPX options at $t_0 - 5$ days = 44 days, SPX options at $T = 79$ days and VIX options at $t_0 = 49$ days for the simulated data example, including the implied volatility of the generating model, the uncalibrated Heston reference model and the OT-calibrated model with a Heston reference. The diamonds are the implied volatility of the calibrating options. The vertical lines are VIX futures prices.

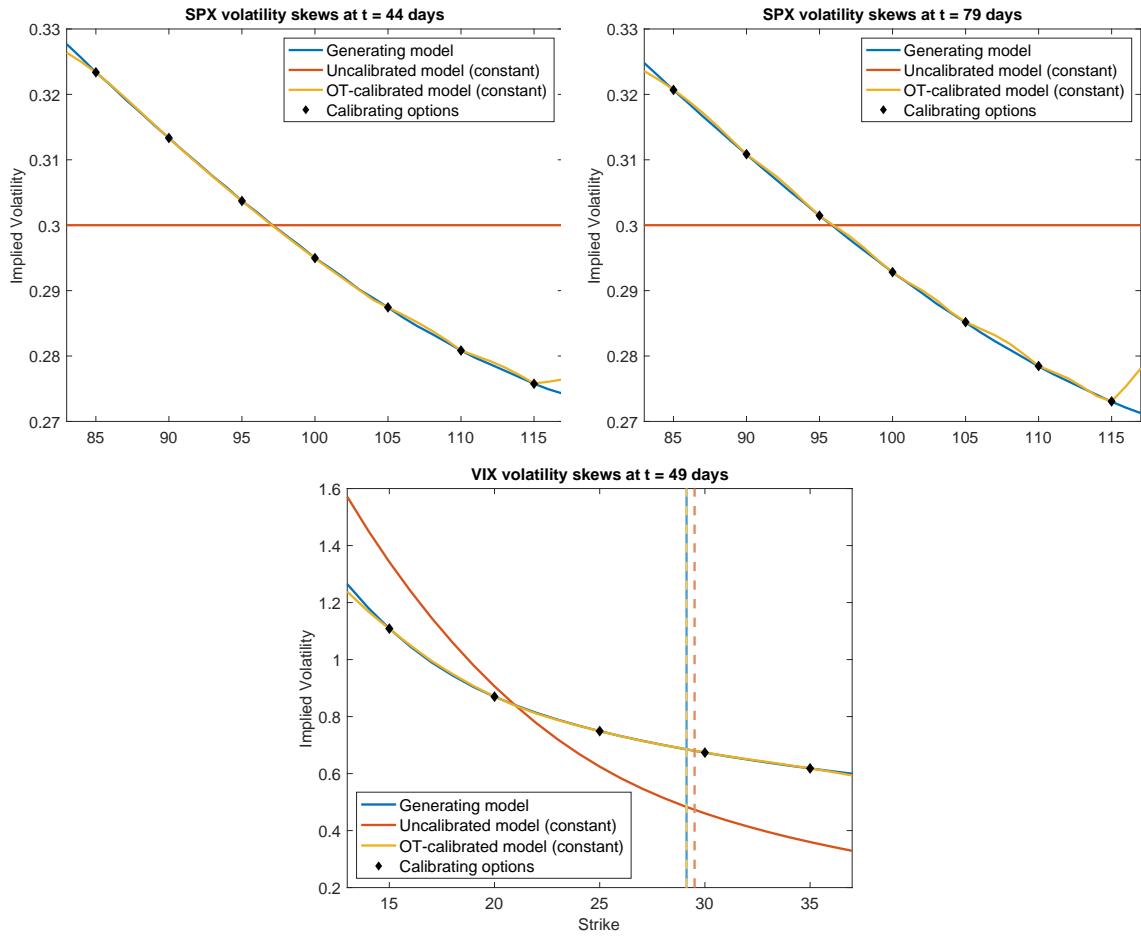


Figure 2: The volatility skews of SPX options at $t_0 - 5$ days = 44 days, SPX options at $T = 79$ days and VIX options at $t_0 = 49$ days for the simulated data example, including the implied volatility of the generating model, the uncalibrated constant reference model and the OT-calibrated model with a constant reference. The diamonds are the implied volatility of the calibrating options. The vertical lines are VIX futures prices.

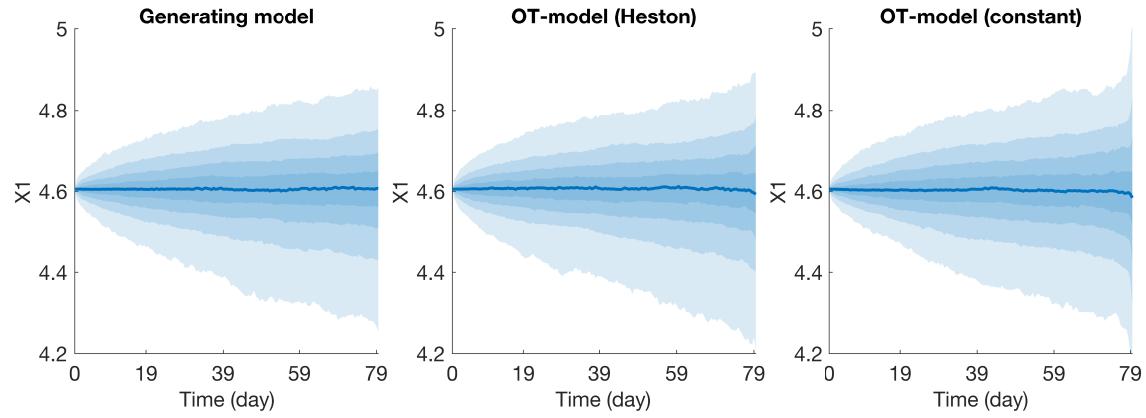


Figure 3: The simulations of X_t^1 for the simulated data example, including the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference.

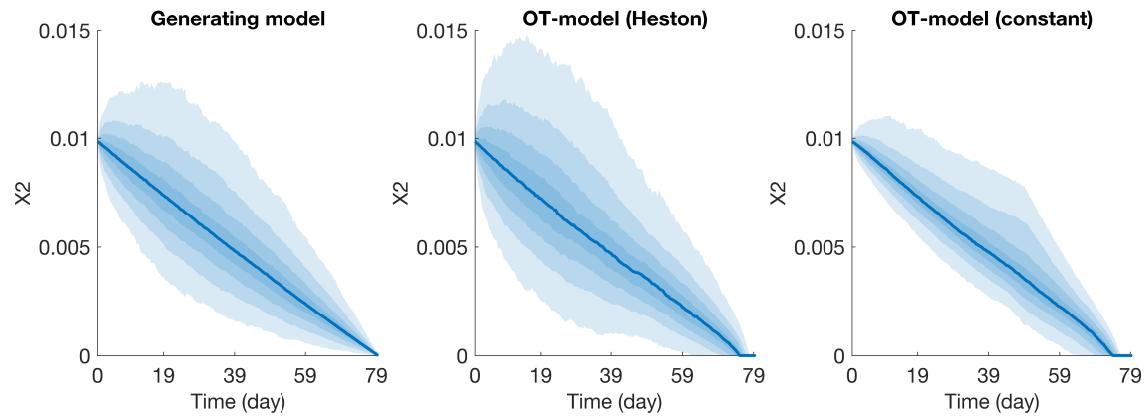


Figure 4: The simulations of X_t^2 for the simulated data example, including the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference.

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

Table 2: The calibration results of the simulated data example, including prices and implied volatility (IV) of the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference. The errors are shown in the parentheses.

5.2 Market data

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

Remark 5.1. For simplicity, we have assumed that the interest rates and dividends are null, and the spot price is a martingale under the risk-neutral measure. However, this assumption does not apply to the market data. To overcome this issue, we let X^1 be the logarithm of the T-forward price of the SPX index instead of the spot price. Then, we are interested in T-forward measures $\mathbb{P} \in \mathcal{P}^1$ under which $\exp(X^1)$ is a martingale.

The market data consists of monthly SPX options maturing at 17 days and 45 days and monthly VIX futures and options maturing at 15 days. The model is optimised with a Heston reference (27) with parameters given in Table 3. The parameters are obtained by (roughly) calibrating a standard Heston model to the SPX option prices. Our numerical procedure also includes smoothing of the reference value, as described in Section 4.3.

Parameter	X_0^1	X_0^2	$\bar{\kappa}$	$\bar{\theta}$	$\bar{\omega}$	$\bar{\eta}$
Value	3523.71	0.004	4.99	0.038	0.52	-0.99

Table 3: Parameter values for the market data example.

The OT-calibrated model volatility skews are plotted in Figure 5, and the simulation of X is given in Figure 6. From the plots, we can see that the OT-calibrated model accurately captures the market data while keeping $X_T^2 = 0$ \mathbb{P} -a.s. satisfied. The volatility behaviour is displayed in Appendix D.

Acknowledgements

The Centre for Quantitative Finance and Investment Strategies has been supported by BNP Paribas. The last author is supported by an Australian Government Research Training Program (RTP) Scholarship. Part of this research was carried out when J. Obłój was visiting Monash University and the Sydney Mathematical Research Institute and he gratefully acknowledges their hospitality.

A The convex conjugate F^*

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

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

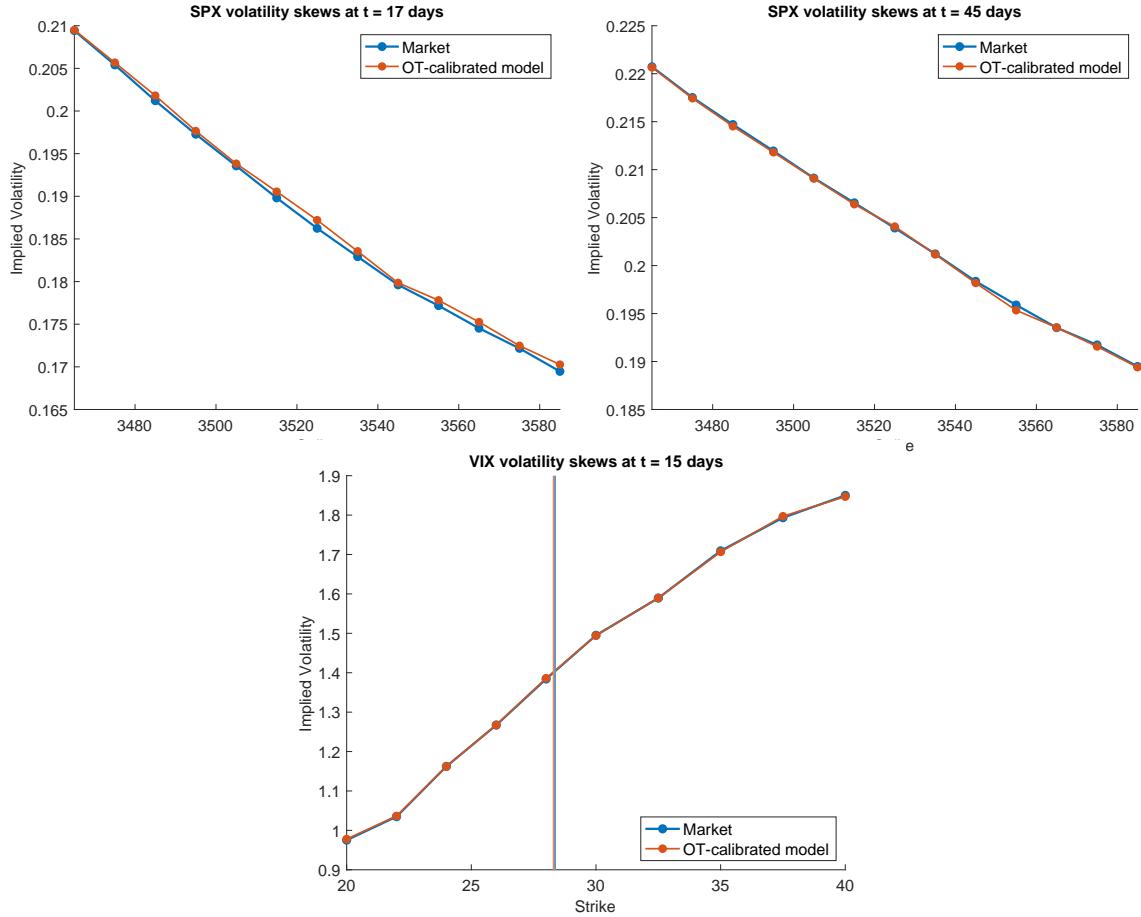


Figure 5: Approximated OT-calibrated model volatility skews of SPX options at $t_0 + 2$ days = 17 days, SPX options at $T = 45$ days and VIX options at $t_0 = 15$ days in the market data example. The vertical lines are VIX futures prices. Markers correspond to computed prices which are then interpolated with a piece-wise linear function.

We also define

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

and define

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

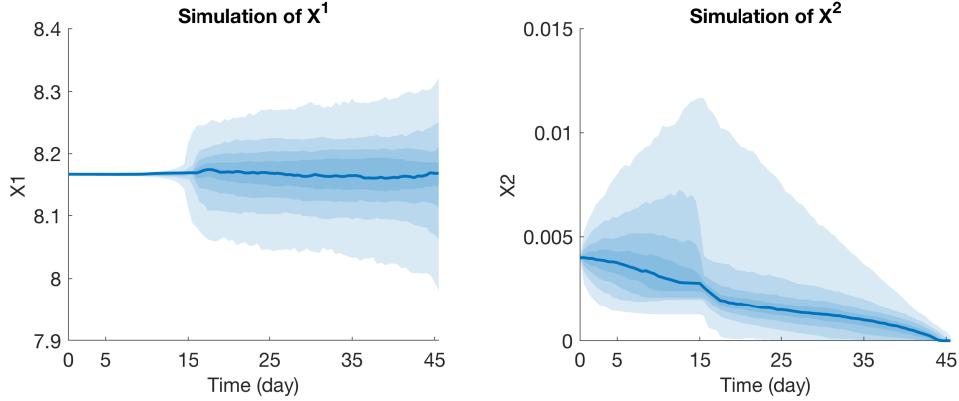


Figure 6: The simulations of the OT-calibrated model X in the market data example.

Lemma A.1. *The convex conjugate of F is*

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

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

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

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

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

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

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

$$(\beta_{11}, \beta_{12}, \beta_{22}) = \arg \inf_{(x, y, z) \in \mathbb{R}_{\geq 0} \times \mathbb{R} \times \mathbb{R}_{\geq 0}} \{(x - A)^2 + 2(y - B)^2 + (z - C)^2 \mid xz \geq y^2\}. \quad (29)$$

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

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

The inverse transformation is

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

In terms of (x', y', z') , the infimum in (29) can be reformulate as

$$\inf_{(x', y', z') \in W} 2(x' - \bar{x}')^2 + 2(y' - \bar{y}')^2 + 2(z' - \bar{z}')^2, \quad (30)$$

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

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

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

- (a) If $(\bar{x}', \bar{y}', \bar{z}') \in W$, the solution is $(x', y', z') = (\bar{x}', \bar{y}', \bar{z}')$.
- (b) If $\bar{x}'^2 + \bar{y}'^2 \leq \bar{z}'^2$, but $\bar{z}' < 0$. Then the solution should be on the boundary $z' = 0$, which also implies that $x' = y' = 0$.
- (c) Otherwise, the solution must be on the boundary of W :

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

By substituting $z' = \sqrt{x'^2 + y'^2}$ into (30) and solving the infimum, we find two stationary points:

$$\begin{aligned} (x'_+, y'_+, z'_+) &= \left(\frac{\bar{x}'}{2} + \frac{\bar{x}'\bar{z}'}{2\sqrt{\bar{x}'^2 + \bar{y}'^2}}, \frac{\bar{y}'}{2} + \frac{\bar{y}'\bar{z}'}{2\sqrt{\bar{x}'^2 + \bar{y}'^2}}, \sqrt{(x'_+)^2 + (y'_+)^2} \right), \\ (x'_-, y'_-, z'_-) &= \left(\frac{\bar{x}'}{2} - \frac{\bar{x}'\bar{z}'}{2\sqrt{\bar{x}'^2 + \bar{y}'^2}}, \frac{\bar{y}'}{2} - \frac{\bar{y}'\bar{z}'}{2\sqrt{\bar{x}'^2 + \bar{y}'^2}}, \sqrt{(x'_-)^2 + (y'_-)^2} \right). \end{aligned}$$

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

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

B Algorithm

Let $\pi^N := \{t_k : 0 \leq k \leq N\}$ be a discretisation of $[0, T]$ such that $0 = t^0 < t^1 < \dots < t^N = T$. We assume that each of t_0 and $\tau_i, i = 1, \dots, m$ coincides with some value in π^N . Denote by ϵ_1 the tolerance of the maximum of the gradients (19)–(22), and denote by ϵ_2 the tolerance for the policy iteration. Recall that ϵ_1 has an alternative interpretation as the tolerance of the maximum error between the calibrating prices and the model prices. In the numerical example presented in Section 5, $\epsilon_1 = 10^{-4}$ and $\epsilon_2 = 10^{-8}$. The numerical method described in Section 4 is summarised as the following algorithm.

Algorithm 1: The joint calibration algorithm

```

1 Set an initial  $(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^\xi)$ 
2 do
    /* Solving the HJB equation */
    for  $k = N - 1, \dots, 0$  do
        /* Terminal conditions */
        if  $\exists i = 1, \dots, m, t_{k+1} = \tau_i$  then
             $\phi_{t_{k+1}} \leftarrow \phi_{t_{k+1}} + \sum_{i=1}^m \lambda_i^{SPX} G_i \mathbf{1}(t_{k+1} = \tau_i)$  // SPX options
        end
        if  $t^{k+1} = t_0$  then
             $\phi_{t^{k+1}} \leftarrow \phi_{t^{k+1}} + \lambda^{VIX,f} J$  // VIX futures
             $\phi_{t^{k+1}} \leftarrow \phi_{t^{k+1}} + \sum_{i=1}^n \lambda_i^{VIX} (H_i \circ J)$  // VIX options
        end
        if  $t^{k+1} = T$  then
             $\phi_{t^{k+1}} \leftarrow \phi_{t^{k+1}} + \lambda^\xi \xi$  // Singular contract
        end
        /* Policy iteration */
         $\phi_{t_k}^{new} \leftarrow \phi_{t_{k+1}}$ 
        do
             $\phi_{t_k}^{old} \leftarrow \phi_{t_k}^{new}$ 
            Approximate  $\beta^*$  by Lemma A.1 with  $\phi_{t_k}^{old}$ 
            Solve the HJB equation (24) or (25) with  $\beta^*$  as a linearised PDE by the
            standard implicit finite difference method, and set the solution as  $\phi_{t_k}^{new}$ 
            while  $\|\phi_{t_k}^{new} - \phi_{t_k}^{old}\|_\infty > \epsilon_2$ 
             $\phi_{t_k} \leftarrow \phi_{t_k}^{new}$ 
        end
        /* Model prices and gradients */
        Calculate the model prices by solving equations (23) by the ADI method
        Calculate the gradients (19) to (22)
        Update  $(\lambda^{SPX}, \lambda^{VIX,f}, \lambda^{VIX}, \lambda^\xi)$  by the L-BFGS algorithm
25 while The maximum of the gradients (19) to (22) is greater than  $\epsilon_1$ 

```

C The diffusion process β for the simulated data example

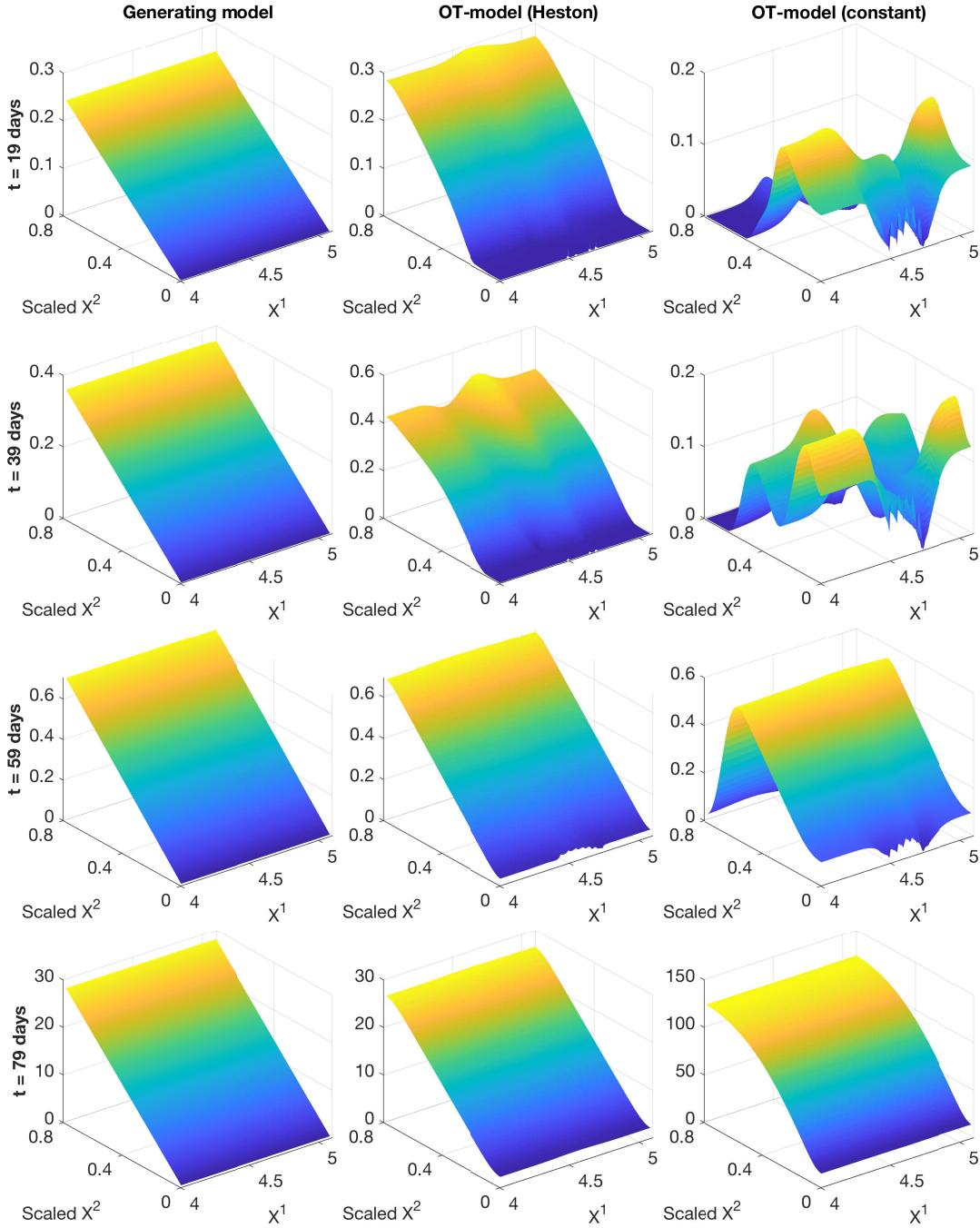


Figure 7: The functions $\beta_{11}(t, X^1, X^2)$ of the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference for the simulated data example.

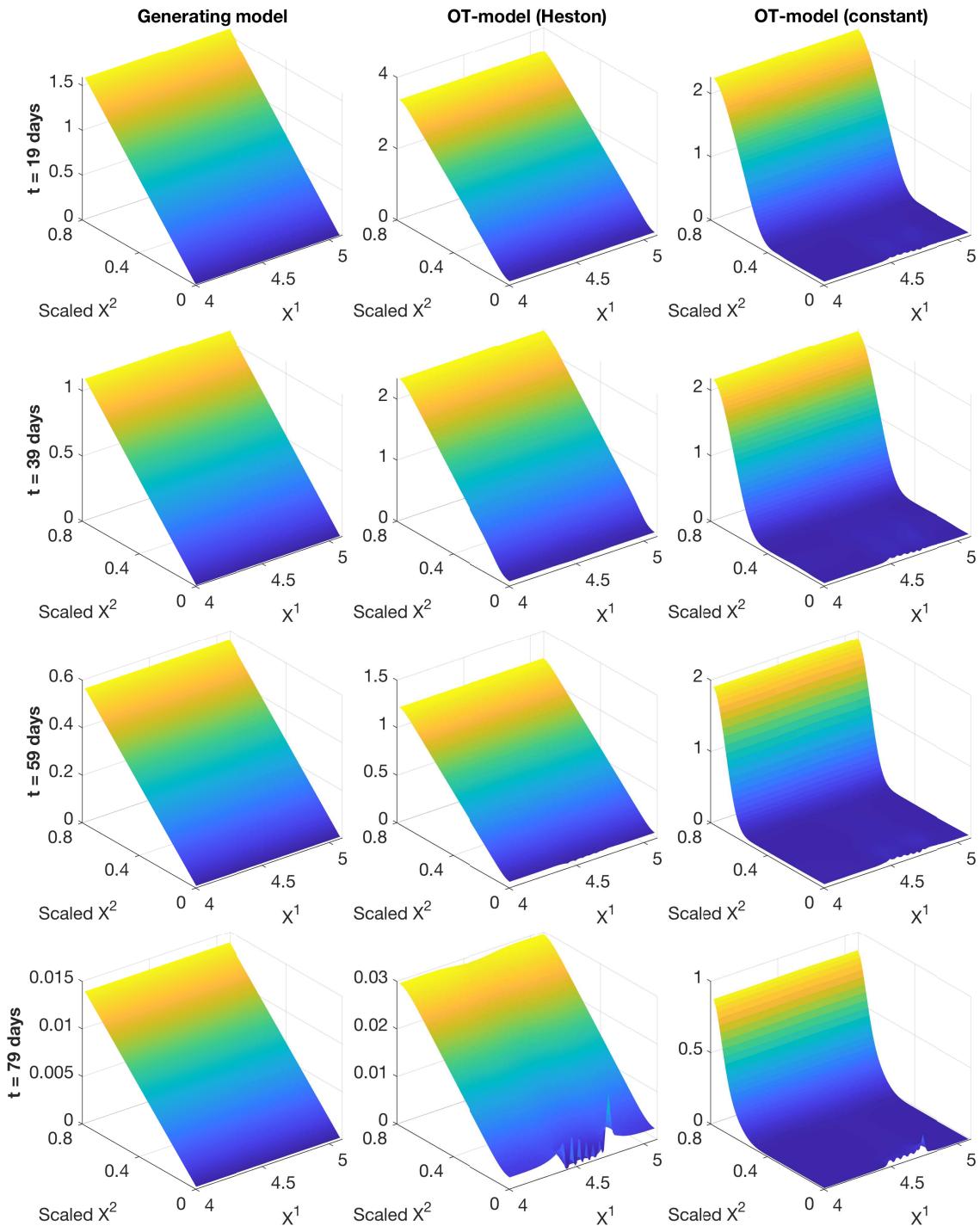


Figure 8: The functions $\beta_{22}(t, X^1, X^2)$ of the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference for the simulated data example.

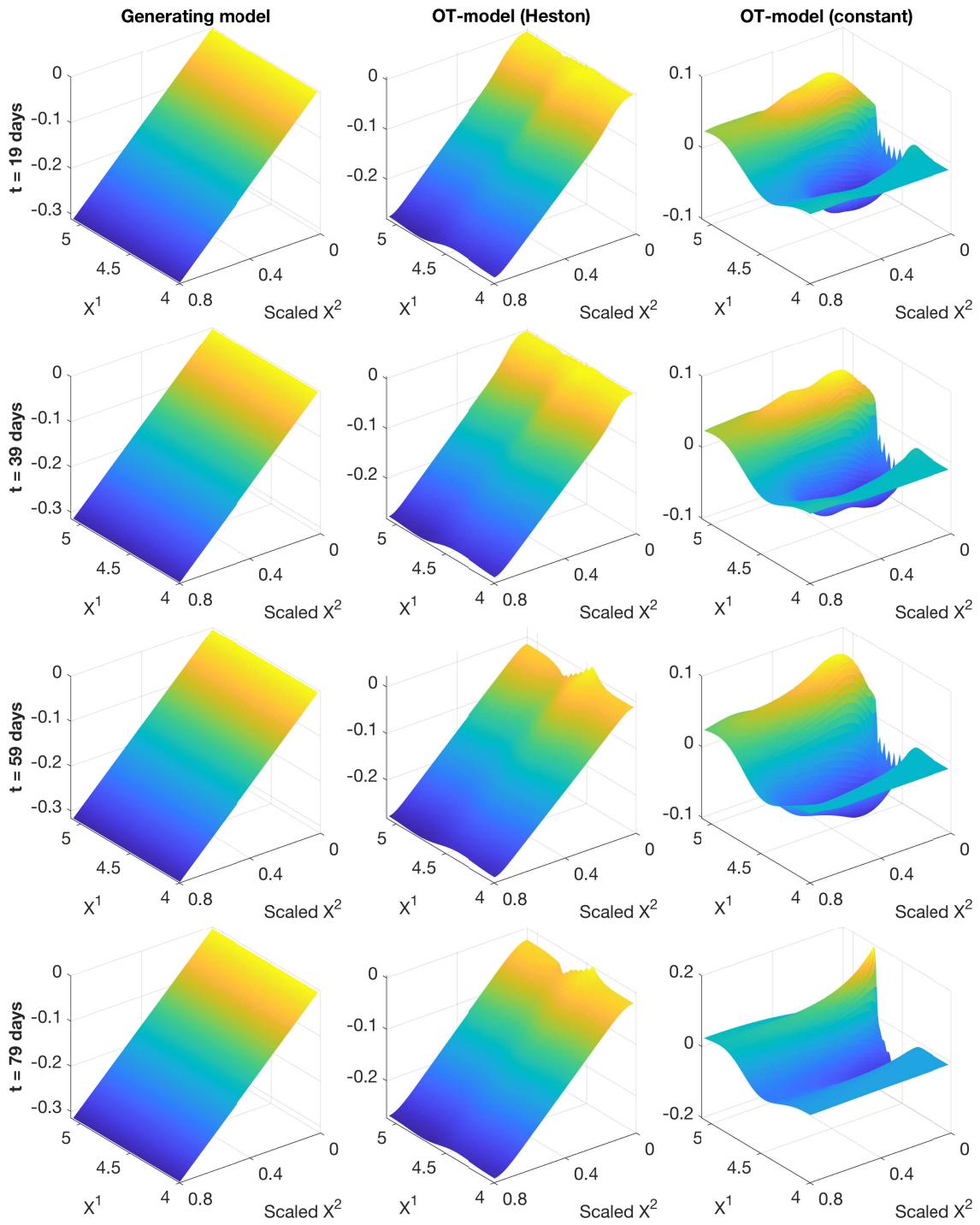


Figure 9: The functions $\beta_{12}(t, X^1, X^2)$ of the generating model, the OT-calibrated model with a Heston reference and the OT-calibrated model with a constant reference for the simulated data example.

D The diffusion process β for the market data example

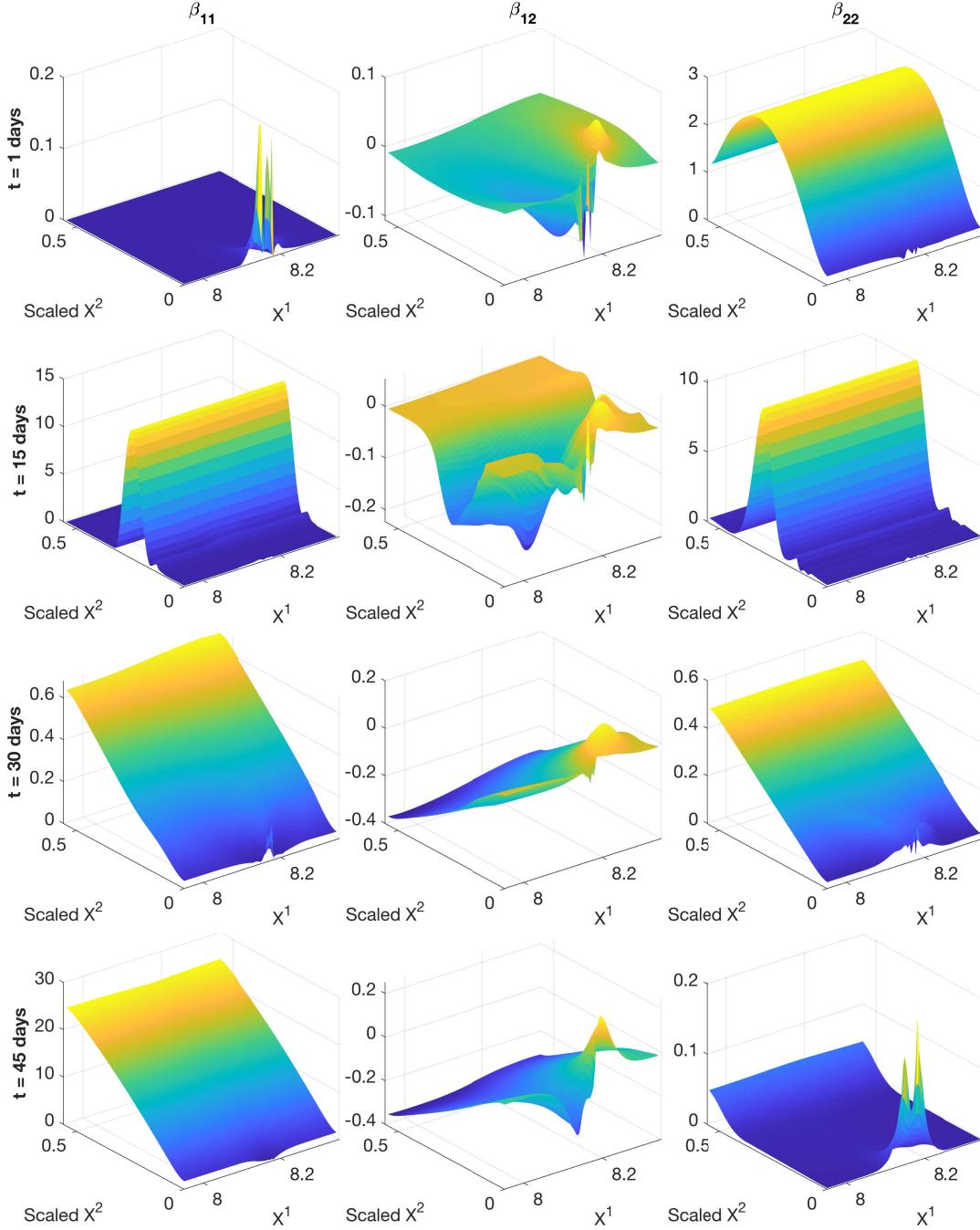


Figure 10: The functions $\beta_{11}(t, X^1, X^2)$, $\beta_{12}(t, X^1, X^2)$ and $\beta_{22}(t, X^1, X^2)$ of the OT-calibrated model for the market data example.

References

- [1] ACCIAIO, B., AND GUYON, J. Inversion of convex ordering: Local volatility does not maximize the price of VIX futures. *SIAM Journal on Financial Mathematics* 11, 1 (2020), SC–1.
- [2] ANDREEV, R. Preconditioning the augmented Lagrangian method for instationary mean field games with diffusion. *SIAM Journal on Scientific Computing* 39, 6 (2017), A2763–A2783.
- [3] BALDEAUX, J., AND BADRAN, A. Consistent modelling of VIX and equity derivatives using a 3/2 plus jumps model. *Applied Mathematical Finance* 21, 4 (2014), 299–312.
- [4] BARLES, G., AND SOUGANIDIS, P. E. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis* 4, 3 (1991), 271–283.
- [5] BEIGLBÖCK, M., HENRY-LABORDÈRE, P., AND PENKNER, F. Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics* 17, 3 (2013), 477–501.
- [6] BENAMOU, J.-D., AND BRENIER, Y. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik* 84, 3 (2000), 375–393.
- [7] BONNANS, J. F., AND ZIDANI, H. Consistency of generalized finite difference schemes for the stochastic HJB equation. *SIAM Journal on Numerical Analysis* 41, 3 (2003), 1008–1021.
- [8] BRUNICK, G., AND SHREVE, S. Mimicking an Itô process by a solution of a stochastic differential equation. *Annals of Applied Probability* 23, 4 (2013), 1584–1628.
- [9] CONT, R., AND KOKHOLM, T. A consistent pricing model for index options and volatility derivatives. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 23, 2 (2013), 248–274.
- [10] DE MARCO, S., AND HENRY-LABORDERE, P. Linking vanillas and VIX options: a constrained martingale optimal transport problem. *SIAM Journal on Financial Mathematics* 6, 1 (2015), 1171–1194.
- [11] DEBRABANT, K., AND JAKOBSEN, E. R. Semi-Lagrangian schemes for linear and fully nonlinear diffusion equations. *Mathematics of Computation* 82, 283 (2013), 1433–1462.
- [12] FIGALLI, A. Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *Journal of Functional Analysis* 254, 1 (2008), 109–153.
- [13] FOUCHE, J.-P., AND SAPORITO, Y. F. Heston stochastic vol-of-vol model for joint calibration of VIX and S&P 500 options. *Quantitative Finance* 18, 6 (2018), 1003–1016.
- [14] GATHERAL, J. Consistent modeling of SPX and VIX options. In *Bachelier congress* (2008), vol. 37, pp. 39–51.
- [15] GATHERAL, J., JUSSELIN, P., AND ROSENBAUM, M. The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. *arXiv preprint arXiv:2001.01789* (2020).
- [16] GOUTTE, S., ISMAIL, A., AND PHAM, H. Regime-switching stochastic volatility model: estimation and calibration to VIX options. *Applied Mathematical Finance* 24, 1 (2017), 38–75.
- [17] GUO, I., AND LOEPPER, G. Path dependent optimal transport and model calibration on exotic derivatives. *arXiv preprint arXiv:1812.03526* (2018).

- [18] GUO, I., AND LOEPER, G. Pricing bounds for volatility derivatives via duality and least squares Monte Carlo. *Journal of Optimization Theory and Applications* 179, 2 (2018), 598–617.
- [19] GUO, I., LOEPER, G., AND WANG, S. Calibration of local-stochastic volatility models by optimal transport. *arXiv preprint arXiv:1906.06478* (2019).
- [20] GUO, I., LOEPER, G., AND WANG, S. Local volatility calibration by optimal transport. In *2017 MATRIX annals*, vol. 2 of *MATRIX Book Ser*. Springer, Cham, 2019, pp. 51–64.
- [21] GUYON, J. Inversion of convex ordering in the VIX market. *Quantitative Finance* (2020), 1–27.
- [22] GUYON, J. The joint S&P 500/VIX smile calibration puzzle solved. *Risk, April* (2020).
- [23] GYÖNGY, I. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability Theory and Related Fields* 71, 4 (1986), 501–516.
- [24] HENRY-LABORDÈRE, P., AND TOUZI, N. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *The Annals of Applied Probability* 24, 1 (Feb. 2014), 312–336.
- [25] HESTON, S. L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* 6, 2 (1993), 327–343.
- [26] IN 'T HOUT, K. J., AND FOULON, S. ADI finite difference schemes for option pricing in the Heston model with correlation. *International Journal of Numerical Analysis and Modeling* 7, 2 (2010), 303–320.
- [27] JACQUIER, A., MARTINI, C., AND MUGURUZA, A. On VIX futures in the rough Bergomi model. *Quantitative Finance* 18, 1 (2018), 45–61.
- [28] KOKHOLM, T., AND STISEN, M. Joint pricing of VIX and SPX options with stochastic volatility and jump models. *The Journal of Risk Finance* (2015).
- [29] KUSHNER, H. J., AND DUPUIS, P. *Numerical methods for stochastic control problems in continuous time*, second ed., vol. 24 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 2001.
- [30] LIAN, G.-H., AND ZHU, S.-P. Pricing vix options with stochastic volatility and random jumps. *Decisions in Economics and Finance* 36, 1 (2013), 71–88.
- [31] LIU, D. C., AND NOCEDAL, J. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* 45, 3, (Ser. B) (1989), 503–528.
- [32] LOEPER, G. The reconstruction problem for the Euler-Poisson system in cosmology. *Archive for Rational Mechanics and Analysis* 179, 2 (2006), 153–216.
- [33] MA, K., AND FORSYTH, P. A. An unconditionally monotone numerical scheme for the two-factor uncertain volatility model. *IMA Journal of Numerical Analysis* 37, 2 (2017), 905–944.
- [34] PACATI, C., POMPA, G., AND RENÒ, R. Smiling twice: The Heston++ model. *Journal of Banking & Finance* 96 (2018), 185–206.

- [35] PAPANICOLAOU, A., AND SIRCAR, R. A regime-switching heston model for VIX and S&P 500 implied volatilities. *Quantitative Finance* 14, 10 (2014), 1811–1827.
- [36] SONG, Z., AND XIU, D. A tale of two option markets: State-price densities implied from S&P 500 and VIX option prices. *Unpublished working paper. Federal Reserve Board and University of Chicago* (2012).
- [37] TAN, X., AND TOUZI, N. Optimal transportation under controlled stochastic dynamics. *The Annals of Probability* 41, 5 (2013), 3201–3240.
- [38] TREVISAN, D. Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. *Electronic Journal of Probability* 21 (2016), Paper No. 22, 41.
- [39] VILLANI, C. *Topics in optimal transportation*, vol. 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.