

---

# Calibration of Local Volatility Models with Stochastic Interest Rates using Optimal Transport

Benjamin Joseph · Grégoire Loeper ·  
Jan Obłój.

May 8, 2025

**Abstract** We develop a non-parametric, semimartingale optimal transport, calibration methodology for local volatility models with stochastic interest rate. The method finds a fully calibrated model which is the closest, in a way that can be defined by a general cost function, to a given reference model. We establish a general duality result which allows to solve the problem by optimising over solutions to a second order fully non-linear Hamilton-Jacobi-Bellman equation. Our methodology is analogous to Guo, Loeper, and Wang [29] and Guo et al. [31] but features a novel element of solving for discounted densities, or sub-probability measures. As an example, we apply the method to a sequential calibration problem, where a Vasicek model is already given for the interest rates and we seek to calibrate a stock price's local volatility model with volatility coefficient depending on time, the underlying and the short rate process, and the two processes driven by possibly correlated Brownian motions. The equity model is calibrated to any number of European options prices.

---

This research has been supported by BNP Paribas Global Markets and the EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation (EP/S023925/1).

For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

---

B. Joseph  
Mathematical Institute and Christ Church, University of Oxford  
E-mail: benjamin.joseph@maths.ox.ac.uk

G. Loeper  
BNP Paribas Global Markets  
E-mail: gregoire.loeper@bnpparibas.com

J. Obłój  
Mathematical Institute and St John's College, University of Oxford  
E-mail: jan.obloj@maths.ox.ac.uk

**Keywords** semimartingale optimal transport · model calibration · stochastic interest rates · computational methods · local volatility

**Mathematics Subject Classification** 91G20 · 91G80 · 60H30

**JEL Classification** C02 · C63 · G13

## 1 Introduction

Modelling involves inevitable trade-offs: “All models are wrong, some models are useful” as Box and Draper [13] put it. Models need to capture the important aspects of the system they represent but they also need to be tractable, and analytically and/or numerically solvable. In particular, calibration – picking model parameters which recover known outputs – is an essential part of any modelling process. It is a key challenge faced by financial industry practitioners on a daily basis, as their pricing models need to match market prices of liquid instruments before they can be used to price any bespoke or illiquid products.

In practice, models for a key underlying, such as the S&P500 index, will need to be calibrated to a large number of options with different maturities and strikes. This may be nigh impossible for a simple parametric model. Dupire [25] derived a formula to calibrate a local volatility model to an arbitrary number of options, establishing it as a benchmark for equities modelling (see also Atlan [4] for its stochastic interest rate extension). However, it came with its own shortcomings. Its calibration poses serious numerical challenges, see Bain, Mariapragassam, and Reisinger [7], and requires interpolation of the data as Dupire’s formula assumes a continuum of prices across strikes and maturities. Moreover, it has been criticised for wrong dynamic behaviour compared to suitable stochastic volatility models, see Hagan et al. [38]. This issue can lead to potentially costly mispricing of exotic products. Naturally, this motivated further research and led, in particular, to the introduction of local stochastic volatility models (LSV) or forward variance stochastic volatility models, see Bergomi [11] for details and historical references. While successful in many ways, these models also came with shortcomings, often related to an inability to calibrate to a new class of products. Prominent among these was the challenge to calibrate jointly to SPX and VIX options, see Guyon [34] for the literature review and discussion.

Monte Carlo methods to calibrate with Dupire’s formula have been used in Deelstra and Rayée [23], Hok and Tan [43], and Ögetbil, Ganesan, and Hientzsch [61]. Another approach is to discretise the non-linear Fokker-Planck equation and directly solve it by finite differences as in Ren, Madan, and Qian [62] or finite volume methods as in Wyns and Du Toit [71] and Engelmann, Koster, and Oeltz [26]. Dupire [24] derives a non-linear McKean SDE for the local volatility, which in fact recovered the mimicking result of Gyöngy [37] using a financial argument. This has been solved via Monte Carlo methods developed in Henry-Labordère [42] for an approximate calibration, and an exact calibration by the particle method of Guyon and Henry-Labordère [36,

35]. Cozma, Mariapragassam, and Reisinger [21] used variance reduction techniques to implement the particle method in a four factor model. All of these methods require a continuum of strikes and maturities to construct a surface of implied volatilities, which in practice results in interpolation and extrapolation of market data to build these surfaces. We adopt a different approach, without this requirement, by formulating the exact calibration of European options as discrete constraints within a convex optimisation problem.

Recently, a non-parametric exact calibration method based on optimal transport has been proposed, which aims to address the above challenges. This method uses semimartingale optimal transport as introduced by Tan and Touzi [67], which is the semimartingale version of the celebrated work of Benamou and Brenier [10] (see also Huesmann and Trevisan [44]). This approach has already been used in several contexts: for local volatility calibration in Guo, Loeper, and Wang [30]; for local stochastic volatility calibration in Guo, Loeper, and Wang [29], for joint VIX SPX calibration in Guo et al. [31] and for optimal investment in Guo et al. [33]. The most general formulation of the method, pointing to the breadth of its applications, is found in Guo and Loeper [28].

An overview of these results is given in Guo et al. [32]. It is worth noting that whilst the link with optimal transport was not recognised, a variational approach to calibrating local volatility models was first constructed already in Avellaneda et al. [5]. The core contribution of the method is to build a fully calibrated model while trying to preserve desirable features of a given model. In effect, we project a given reference model onto the set of calibrated models. As the calibration constraints depend on one-dimensional marginals, classical mimicking results, see Gyöngy [37] and Brunick and Shreve [18], allow us to restrict to Markovian models. This in turn allows us to use PDE methods to solve the dual problem.

The need for a calibrated interest rate model is common to all financial products involving future payments. While in a very low interest rate environment, which roughly held in the financial markets between 2009 and 2021, one may be tempted to ignore this need for short-dated products, it is no longer feasible for the current market conditions. It is therefore natural to extend the semimartingale optimal transport approach to a setup that includes stochastic rates. Our contribution here is to fill this gap and understand how to develop and calibrate joint models for rates and equities. This extension is not trivial, since the computation of the discount factor involves the whole path of the short rate, which renders the problem path dependent. We will show how to overcome this difficulty. We develop suitable duality results and seek to calibrate a stock price local volatility model with volatility coefficient depending on time, the underlying and the short rate process, and driven by a Brownian motion which can be correlated with the randomness driving the rates process. In a sequel paper Joseph, Loeper, and Obłój [51], we consider a simultaneous joint calibration problem. A particular difficulty here is in dealing with the path dependent discount terms while keeping the number of state variables, and thus the dimension of the problem, at  $d = 2$ . This is important for com-

putational reasons as the numerical methods rely on solving a non-linear HJB equation and pricing PDEs, which becomes computationally more difficult by standard techniques when the dimension  $d \geq 3$ . However, our duality result covers any interest rate model, at the expense of harder numerical complexity.

## 2 Preliminaries and Notation

We adopt the setup of Guo, Loeper, and Wang [29] and Guo et al. [31], who in turn used the formulation of Tan and Touzi [67]. Let  $E$  be a Polish space equipped with the Borel  $\sigma$ -algebra, let  $C(E)$  be the space of continuous functions on  $E$  and  $C_b(E)$  be the space of continuous bounded functions on  $E$ . Let  $\mathcal{M}(E)$  be the space of finite signed Borel measures endowed with the weak-\* topology, let  $\mathcal{M}_+(E) \subset \mathcal{M}(E)$  be the subset of non-negative finite Borel measures, and  $\mathcal{P}(E)$  be the set of Borel probability measures also under the weak-\* topology. Note that if  $E$  is compact, then the topological dual of  $C_b(E)^*$  is given by  $C_b(E)^* = \mathcal{M}(E)$ , but if  $E$  is non-compact, then  $C_b(E)^*$  is larger than  $\mathcal{M}(E)$ . Let  $\text{BV}(E)$  be the set of bounded variation functions on  $E$  and  $L^1(d\mu)$  be the space of  $\mu$ -integrable functions. For unambiguity we write  $C_b(E; \mathbb{R}^d)$ ,  $\mathcal{M}(E; \mathbb{R}^d)$ ,  $\text{BV}(E; \mathbb{R}^d)$ , and  $L^1(d\mu; \mathbb{R}^d)$  for the vector valued versions of those spaces (with an analogous definition for the matrix valued versions). Write  $\mathbb{S}^d$  for the set of  $d \times d$  symmetric matrices and  $\mathbb{S}_+^d \subset \mathbb{S}^d$  as the subset of positive semidefinite symmetric matrices. For  $a, b \in \mathbb{R}^d$  write  $a \cdot b$  for the inner product  $a^\top b$  and for  $A, B \in \mathbb{S}^d$  write  $A : B$  for their inner product  $\text{Tr}(A^\top B)$ . As a shorthand, we define  $\Lambda := [0, T] \times \mathbb{R}^d$  and  $\mathcal{X} := \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d$ , which will be used for the domain and range of the triple representing the law of the semimartingale, the drift and the volatility. Finally, denote the duality bracket between  $C_b(E)$  and  $C_b(E)^*$  by  $\langle \cdot, \cdot \rangle$ .

We fix a time horizon  $T > 0$  and dimension  $d > 1$ . We mainly work on the canonical space  $\Omega := C([0, T], \mathbb{R}^d)$  of continuous  $\mathbb{R}^d$ -valued paths on  $[0, T]$ , but sometimes we need to work on  $C([0, T], \mathbb{R}^{d+1})$ . The canonical process is denoted in two ways: either  $e$ ,  $e_t(\omega) = \omega_t$ , and the dimension is clear from the context, or as  $X$ ,  $X_t(\omega) = \omega_t$ , with the latter restricted to the  $d$ -dimensional setting. We take the canonical filtration  $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$  generated by  $e$ . We consider all probability measures  $\mathbb{P}$  on  $(\Omega, \mathcal{F}_T)$  such that  $X$  is an  $(\mathbb{F}, \mathbb{P})$ -semimartingale with decomposition

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

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

**Definition 2.1** We say that  $\mathbb{P}$  is characterised by  $(\alpha_t^\mathbb{P}, \beta_t^\mathbb{P})$  if

$$\alpha_t^\mathbb{P} = \frac{dA_t^\mathbb{P}}{dt}, \quad \beta_t^\mathbb{P} = \frac{d\langle M^\mathbb{P} \rangle_t}{dt} \quad dt \times \mathbb{P}(d\omega)\text{-a.e.},$$

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

The existence of progressively measurable  $(\alpha_t, \beta_t)$  is guaranteed since  $A_t^{\mathbb{P}}$  and  $M_t^{\mathbb{P}}$  are assumed to be Lebesgue absolutely continuous in  $t$ ,  $\mathbb{P}$ -a.s., see Jacod and Shiryaev [49, Proposition I.3.13, Proposition II.2.9]. Note that  $(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})$  is only determined up to  $d\mathbb{P} \times dt$ -almost everywhere. The set of probability measures  $\mathbb{P}$  satisfying the conditions above is denoted  $\mathcal{P}$ . We note that regular conditional probabilities exist on  $\Omega$  and we will use these implicitly, e.g.,  $\mathbb{E}_{t,x}^{\mathbb{P}}[\alpha_t^{\mathbb{P}}]$  or  $\mathbb{E}^{\mathbb{P}}[\alpha_t^{\mathbb{P}}|X_t = x]$  will denote the conditional expectation seen as a measurable function of  $(t, X_t)$  and evaluated at  $X_t = x$ . Finally, we note that, by Doob's martingale representation Theorem (see Karatzas and Shreve [53, Theorem 3.4.2]), possibly on an enlarged probability space, there exists a Brownian motion  $W^{\mathbb{P}}$  such that

$$X_t = X_0 + \int_0^t \alpha_s^{\mathbb{P}} ds + \int_0^t (\beta_s)^{1/2} dW_s^{\mathbb{P}}, \quad t \leq T.$$

The process  $X_t := (\mathbf{S}_t, X_t^r)$ , is composed of the short rate  $(X_t^r)_{t \in [0, T]}$  and a  $(d - 1)$ -dimensional process  $(\mathbf{S}_t)_{t \in [0, T]}$  corresponding to the underlying asset, such as the S&P 500, and extra state variables, e.g., extra assets, stochastic factors in the volatility functions, or multi-factors in the short rate. In specific examples, e.g., in section 4.1, the short rate is also denoted  $r_t$ . The state variable's  $x$ -coordinate corresponding to  $X^r$  is denoted  $x_r$  to stress that it represents the short rate. The stochastic discount factor is derived from the short rate,  $Y_t := \exp(-\int_0^t X_s^r ds)$ , and we refer to  $(X_t, Y_t)$  as the augmented process. Note that the augmented process can be expressed as an updating function in the sense of Brunick and Shreve [18, Definition 3.1]. We further consider the subset  $\mathcal{P}^1 \subset \mathcal{P}$  of measures  $\mathbb{P} \in \mathcal{P}$  which satisfy

$$\mathbb{E}^{\mathbb{P}} \left[ \int_0^T |\alpha_t^{\mathbb{P}}| + |\beta_t^{\mathbb{P}}| dt \right] < +\infty \quad \text{and} \quad X_t^r > -1 dt \otimes d\mathbb{P}\text{-a.e.}, \quad (2.1)$$

where  $|\cdot|$  is the  $L^1$  norm. We are not aware of any relevant model which would require the short rate to be unbounded from below so this part of the assumption poses no problem. It implies in particular that  $0 < Y_t < e^T$ ,  $t \in [0, T]$ , under all measures considered in this paper. The integrability assumption would equally be satisfied under any reasonable model and it is imposed so that we can apply Brunick and Shreve [18, Theorem 3.6]. These mimicking results, extending earlier works of Krylov [54] and Gyöngy [37], allow us to construct a Markov process with the same one-dimensional marginals as a given semi-martingale. As our constraints and cost will only depend on these marginals, it will allow us restrict our attention to Markov processes. Markovian projection relies on the classical result that the marginal law of a diffusion process is a distributional solution to the corresponding Fokker-Planck equation, with the converse result given in Figalli [27] where existence and uniqueness results are constructed for the corresponding SDE satisfied by a process with marginal law that is a weak solution to a Fokker-Planck equation. In addition, since we are considering a calibration problem with a stochastic interest rate, to avoid path dependent terms, we augment the process to include stochastic discount

factor terms. For simplicity, our argument will consider a one-factor short rate, but it is easy to adapt to a multi-factor or multi-curve setting.

**Theorem 2.2** *Let  $\mathbb{P} \in \mathcal{P}^1$  be a candidate model. There exist jointly measurable versions of the conditional expectations,  $\alpha_t(x, y) = \mathbb{E}_{t,x,y}^{\mathbb{P}}[\alpha_t^{\mathbb{P}}]$ ,  $\beta_t(x, y) = \mathbb{E}_{t,x,y}^{\mathbb{P}}[\beta_t^{\mathbb{P}}]$ ,  $dt \times d\mathbb{P}$ -a.e., and  $\tilde{\mathbb{P}} \in \mathcal{P}^1$  such that  $\alpha_t^{\tilde{\mathbb{P}}} = \alpha_t(X_t, Y_t)$  and  $\beta_t^{\tilde{\mathbb{P}}} = \beta_t(X_t, Y_t)$ ,  $dt \times d\tilde{\mathbb{P}}$ -a.e., and  $\tilde{\mathbb{P}} \circ (X_t, Y_t)^{-1} = \mathbb{P} \circ (X_t, Y_t)^{-1}$ .*

*Moreover, possibly on some other probability space, there exists a Markov process  $(X, Y)$  satisfying*

$$\begin{cases} dX_t = \alpha_t(X_t, Y_t) dt + [\beta_t(X_t, Y_t)]^{\frac{1}{2}} dW_t^{\mathbb{P}'}, & 0 \leq t \leq T, \\ dY_t = -Y_t X_t^r dt, & 0 \leq t \leq T, \\ X_0 = x_0, \\ Y_0 = 1. \end{cases} \quad (2.2)$$

such that  $(X_t, Y_t) \sim \mathbb{P} \circ (X_t, Y_t)^{-1}$  for all  $t \in [0, T]$ , and where  $W^{\mathbb{P}'}$  is a Brownian motion.

Finally, the marginal distributions,  $\rho_t = \rho(t, \cdot) = \mathbb{P} \circ (X_t, Y_t)^{-1}$ , are a weak solution to the Fokker-Planck equation

$$\begin{cases} \partial_t \rho_t + \nabla_x \cdot (\rho_t \alpha_t) - \frac{1}{2} \nabla_x^2 : (\rho_t \beta_t) - \partial_y(x_r y \rho_t) = 0, & \text{on } [0, T] \times \mathbb{R}^{d+1}, \\ \rho_0^{\mathbb{P}} = \delta_{X_0, Y_0}, & (x, y) \in \mathbb{R}^{d+1}. \end{cases} \quad (2.3)$$

**Definition 2.3** We let  $\mathcal{P}_{\text{loc}}^1$  denote the subset of measures in  $\mathcal{P}^1$  under which the semimartingale characteristics of  $X$  are given as measurable functions in  $(t, X_t, Y_t)$ .

*Remark 2.4* The first assertion in Theorem 2.2 follows from Brunick and Shreve [18, Theorem 7.1]. The second one then follows via the martingale representation theorem, and is Theorem 3.6 therein. The last assertion is obtained using Itô's formula. Note that  $\tilde{\mathbb{P}} \in \mathcal{P}_{\text{loc}}^1$  and is simply the distribution of  $X$  solving (2.2),  $\tilde{\mathbb{P}} = \mathbb{P}' \circ X^{-1}$ . With a slight abuse of language, we will refer to  $\mathcal{P}_{\text{loc}}^1$  as measures under which  $(X, Y)$  is a Markov process solving (2.2).

*Remark 2.5* A Borel curve  $(\nu_t)_{t \in [0, T]} \subset \mathcal{M}(\mathbb{R}^d)$  is narrowly continuous if for every  $f \in C_b(\mathbb{R}^d)$ , the map  $t \mapsto \int_{\mathbb{R}^d} f d\nu_t$  is continuous. Trevisan [68, Remark 2.3] (also Ambrosio, Gigli, and Savaré [2, Lemma 8.1.2]) show that any solution  $(\rho_t)_{t \in (0, T)}$  of (2.3) has a narrowly continuous representative  $(\tilde{\rho}_t)_{t \in [0, T]}$  such that  $\rho_t = \tilde{\rho}_t$  for almost all  $t \in (0, T)$ . Define  $\mathcal{L}$  as the diffusion operator associated with (2.3), then we also have

$$\int f_{t_2} d\tilde{\rho}_{t_2} - \int f_{t_1} d\tilde{\rho}_{t_1} = \int_{t_1}^{t_2} \int \partial_t f + (\mathcal{L}f)_t d\tilde{\rho}_t dt, \quad \text{for } 0 \leq t_1 < t_2 \leq T.$$

We therefore may assume without loss of generality that the solution to (2.3) is narrowly continuous.

Our calibration problem will be written as a minimization of a cost functional which represents a “distance” to a given favourite reference model and also ensures perfect calibration. We consider a strongly convex cost function  $F : \Lambda \times \mathbb{R}^d \times \mathbb{S}^d \rightarrow [0, +\infty]$ , proper and lower semicontinuous in  $(\alpha, \beta)$ . It will be set to take value  $+\infty$  outside of a set  $\Gamma$ , see for example (4.13). In particular we will always set  $F = +\infty$  if  $\beta \notin \mathbb{S}_+^d$  to ensure  $\beta$  is a legitimate covariance matrix. We use these properties implicitly, e.g., whenever we restrict to  $\beta \in \mathbb{S}_+^d$ . We also define the function  $\hat{F} : \Lambda \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d \rightarrow [0, +\infty]$  as  $\hat{F}(t, x, y, \alpha, \beta) = f(y)F(t, x, \alpha, \beta)$ , where the function  $f$  is specified by the discounting term in the payoffs of the options constraints.

The strong convexity assumption of  $F$  (see Nesterov [60, Definition 2.1.3]) means that for any subderivative  $\nabla$  performed over  $(\alpha, \beta) \in \mathbb{R}^d \times \mathbb{S}_+^d$  there exists  $C > 0$  such that for all  $(t, x, \alpha, \beta, \alpha', \beta') \in \Lambda \times \mathbb{R}^d \times \mathbb{S}_+^d \times \mathbb{R}^d \times \mathbb{S}_+^d$ , when  $F(t, x, \alpha, \beta) < \infty$  we have

$$\begin{aligned} F(t, x, \alpha', \beta') &\geq F(t, x, \alpha, \beta) + \langle \nabla F(t, x, \alpha, \beta), (\alpha' - \alpha, \beta' - \beta) \rangle \\ &\quad + C(\|\alpha - \alpha'\|_2^2 + \|\beta - \beta'\|_{\text{Fro}}^2). \end{aligned} \quad (2.4)$$

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

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

The Legendre-Fenchel transform of  $F$  (see for example Rockafellar [63, §12]) is given by

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

where the supremum is a priori over  $(\alpha, \beta) \in \mathbb{R}^d \times \mathbb{S}^d$  but can be restricted  $\beta \in \mathbb{S}_+^d$  by the comment above, or indeed can be later restricted to the set  $\Gamma$  as  $F = +\infty$  elsewhere. While  $F$  itself may not be differentiable with respect to  $(\alpha, \beta)$ , since  $F$  is strictly convex in  $(\alpha, \beta)$  we therefore have that  $F^*(t, x, a, b)$  is differentiable in  $(a, b)$  (see Rockafellar [63, Theorem 26.3]). For convenience, we will denote  $F(\alpha, \beta) := F(t, x, \alpha, \beta)$  and  $F^*(a, b) := F^*(t, x, a, b)$ .

### 3 Problem formulation

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

**Definition 3.1** Given an initial distribution  $\mu_0$ , expiry times  $\tau$ , market prices  $u$  corresponding to payoffs  $G$ , we introduce the set of calibrated measures

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

We denote  $\mathcal{P}_{\text{loc}}(\mu_0, \tau, u) = \mathcal{P}(\mu_0, \tau, u) \cap \mathcal{P}_{\text{loc}}^1$  the calibrated models under which the semimartingale characteristics of  $X$  are given as measurable functions of  $(t, X_t, Y_t)$ , see Remark 2.4.

We denote  $\tilde{\mathcal{P}}_{\text{loc}}(\mu_0, \tau, u) \subseteq \mathcal{P}_{\text{loc}}(\mu_0, \tau, u)$  the subset of measures under which the semimartingale characteristics of  $X$  are given as measurable functions of  $(t, X_t)$ .

We remark that we will usually take  $\mu_0 = \delta_{(X_0, 1)}$  where  $X_0$  are the observed initial values of our state variables.

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

**Proposition 3.2** *Given an initial distribution  $\mu_0$ , expiration times  $\tau$ , and market prices  $u$ , we have*

$$V := \inf_{\mathbb{P} \in \mathcal{P}(\mu_0, \tau, u)} \mathbb{E}^\mathbb{P} \left[ \int_0^T Y_t F(\alpha_t^\mathbb{P}, \beta_t^\mathbb{P}) dt \right] = \inf_{\mathbb{P} \in \tilde{\mathcal{P}}_{\text{loc}}(\mu_0, \tau, u)} \mathbb{E}^\mathbb{P} \left[ \int_0^T Y_t F(\alpha_t^\mathbb{P}, \beta_t^\mathbb{P}) dt \right]. \quad (3.1)$$

### 3.1 First Markovian reduction

Similarly to Guo, Loeper, and Wang [29, Proposition 3.4], since market constraints only depend on marginal distributions and the cost functional is convex, a combination of Markovian projection in Theorem 2.2 and Jensen's inequality readily shows that we can restrict our attention to Markov process in  $(t, X_t, Y_t)$ .

*Proof (of Proposition 3.2 (Part I))* Note that  $\tilde{\mathcal{P}}_{\text{loc}}(\mu_0, \tau, u) \subseteq \mathcal{P}(\mu_0, \tau, u)$  so the “ $\leq$ ” is trivial and also if  $\mathcal{P}(\mu_0, \tau, u)$  is empty then both sides are equal  $+\infty$ . We show now the reverse inequality but with  $\mathcal{P}_{\text{loc}}(\mu_0, \tau, u)$  replacing  $\tilde{\mathcal{P}}_{\text{loc}}(\mu_0, \tau, u)$ . Take  $\mathbb{P} \in \mathcal{P}(\mu_0, \tau, u)$  and use Theorem 2.2, and Remark 2.4, to find the corresponding  $\tilde{\mathbb{P}} \in \mathcal{P}_{\text{loc}}(\mu_0, \tau, u)$  such that  $(X, Y)$  has the same marginals under  $\mathbb{P}$  and  $\tilde{\mathbb{P}}$ . Using the tower property and Jensen's inequality as  $(\alpha, \beta) \mapsto F(\alpha, \beta)$  is convex, we have

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

This gives the desired inequality.

### 3.2 A ‘discounted density’ transformation and superposition principle

While we could easily reduce to Markovian models in the state process  $(X, Y)$ , this is not satisfactory. In particular, in Section 5.2 where we consider a local volatility model with a short rate, we would obtain a three dimensional fully nonlinear PDE instead of a two dimensional one, which involves a substantially increased computational effort. We thus want to further reduce to Markov process  $X$ , i.e., to obtain (3.1). This requires novel arguments to deal with the stochastic discount term. Instead of working with probability measures, we will work with discounted densities, i.e., sub-probability measures. When a  $\mathbb{P} \in \mathcal{P}_{loc}^1$  is fixed, we write  $\rho, \alpha, \beta$ , for  $\rho^\mathbb{P}, \alpha^\mathbb{P}, \beta^\mathbb{P}$ .

**Definition 3.3** Let  $\mathbb{P} \in \mathcal{P}_{loc}^1$  with the semimartingale characteristics of  $X$  given by  $\alpha_t^\mathbb{P} = \alpha_t(X_t, Y_t)$ ,  $\beta_t^\mathbb{P} = \beta_t(X_t, Y_t)$ . Let  $\rho_t(dx, dy)$  and  $\rho_t^X(dx)$  be, respectively, the marginal distribution of  $(X_t, Y_t)$  and of  $X_t$ ,  $t \in [0, T]$ . Let  $\rho_t(dx, dy) = \zeta_{t,x}(dy)\rho_t^X(dx)$ , i.e.,  $\zeta_{t,x}$  is the law of  $Y_t$  conditional on  $\{X_t = x\}$ . Define the ‘discounted’ density, drift and volatility for  $(t, x) \in [0, T] \times \mathbb{R}^d$

$$\rho_t^D(dx) = D_t(x)\rho_t^X(dx), \quad \text{where } D_t(x) = \int_{\mathbb{R}} y\zeta_{t,x}(dy), \quad (3.2)$$

$$\alpha_t^D(x) = \frac{\mathbb{E}[Y_t \alpha_t(X_t, Y_t)|X_t = x]}{\mathbb{E}[Y_t|X_t = x]} = \int_{\mathbb{R}} y\alpha_t(x, y)\frac{\zeta_{t,x}(dy)}{D_t(x)}, \quad (3.3)$$

$$\beta_t^D(x) = \frac{\mathbb{E}[Y_t \beta_t(X_t, Y_t)|X_t = x]}{\mathbb{E}[Y_t|X_t = x]} = \int_{\mathbb{R}} y\beta_t(x, y)\frac{\zeta_{t,x}(dy)}{D_t(x)}. \quad (3.4)$$

By construction,  $\rho_t^D$  inherits narrow continuity from  $\rho_t$ . In addition, using (2.1) and (3.2)–(3.4), we have

$$\begin{aligned} \int_0^T \int |\alpha^D(x)| + |\beta^D(x)|\rho_t^D(dx) dt &= \int_0^T \int |y\alpha(x, y)| + |y\beta(x, y)|\rho_t(dx, dy) dt \\ &\leq e^T \mathbb{E}^\mathbb{P} \left[ \int_0^T |\alpha_t^\mathbb{P}| + |\beta_t^\mathbb{P}| dt \right] < +\infty, \end{aligned} \quad (3.5)$$

so that  $\alpha^D$  and  $\beta^D$  are  $dt \times \rho^D$ -integrable.

A key tool in the dimension reduction is a “discounted” version of the superposition principle of Trevisan [68, Theorem 2.5]. We first define the augmented martingale problem.

**Definition 3.4** Let  $f \in C_b^{1,2}([0, T] \times \mathbb{R}^{d+1}; \mathbb{R})$ , and given measurable functions  $a : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{S}_+^d$ , and  $c : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ , define the operator

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

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

$$\int \int_0^T (|a_t| \circ e_t + |b_t| \circ e_t + |c_t| \circ e_t) dt d\boldsymbol{\eta} < \infty \quad (3.6)$$

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

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

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

We now state the “discounted superposition principle”.

**Theorem 3.5** *Let  $(\nu_t)_{t \in [0, T]} \in C([0, T]; \mathcal{M}(\mathbb{R}^d))$  be a narrowly continuous solution of the general discounted Fokker-Planck equation on  $[0, T] \times \mathbb{R}^d$*

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

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

The proof of Theorem 3.5 follows closely the steps in the proof of Trevisan [68, Theorem 2.5], which itself has a similar structure to Ambrosio, Gigli, and Savaré [2, Theorem 8.2.1], Ambrosio and Figalli [1, Theorem 4.5], Figalli [27, Theorem 2.6], Ambrosio and Trevisan [3, Theorem 7.1]. We defer the proof to the Appendix but summarise here the main steps. First, in section A.1, we argue the result holds for smooth coefficients and constructs smooth approximating sequences in an analogous way to Trevisan [68]. Then, Section A.2 asserts the compactness of solutions for the augmented martingale problem as a direct consequence of Trevisan [68, Section A.2], and establishes convergence via an easy modification of Trevisan [68, Section A.3]. For the generalisation to smooth and bounded coefficients, the adaptation of the proof is easy with the existence of a solution to aMP( $a, b, c$ ) with a given initial law following directly from the Trevisan [68, Theorem A.6], and uniqueness following from a simple adaptation of Trevisan [68, Theorem A.7], with the approach in obtaining estimates an easy modification. The generalisation to bounded coefficients relies on the de la Vallée Poussin criterion and Trevisan [68, Corollary A.5], with our modification again being the handling of the  $c$  term — here we rely on the already established tightness and superposition results for the augmented martingale problem, as it is just the usual martingale problem for  $(X, Y)$ , and the fact that  $Y$  is strictly positive. The estimates for the  $x$ -coordinates are obtained in the same way as Trevisan [68], and the estimates for the  $y$ -coordinate use the function  $\log(y\chi_R)$  where  $\chi_R$  is some cutoff function in order to eliminate the  $y$  term in the augmented diffusion operator. The locally bounded coefficients case then follows a similar modified approach to the bounded coefficients case, as in Trevisan [68].

### 3.3 Second Markovian reduction

We use now the tools introduced above to finish the proof of Proposition 3.2.

**Proposition 3.6** *Let  $\mathbb{P} \in \mathcal{P}_{loc}^1$  and  $\rho^D, \alpha^D, \beta^D$  be given as in Definition 3.3. Then  $\rho_t^D$  is a weak solution to the ‘discounted’ version of the Fokker-Planck equation for  $(t, x) \in [0, T] \times \mathbb{R}^d$*

$$\partial_t \rho_t^D(x) + \nabla_x \cdot [\alpha_t^D(x) \rho_t^D(x)] - \frac{1}{2} \nabla_x^2 : [\beta_t^D(x) \rho_t^D(x)] + x_r \rho_t^D(x) = 0. \quad (3.7)$$

Moreover, there exists a local volatility model  $\tilde{\mathbb{P}} \in \mathcal{P}^1$  under which the semi-martingale characteristics of  $X$  are given by  $\alpha_t^{\tilde{\mathbb{P}}} = \alpha_t^D(X_t)$  and  $\beta_t^{\tilde{\mathbb{P}}} = \beta_t^D(X_t)$   $dt \times d\tilde{\mathbb{P}}$ - a.e. and such that for any payoff function  $G \in C_b(\mathbb{R}^d; \mathbb{R})$  and any  $t \in [0, T]$

$$\mathbb{E}^{\mathbb{P}}[Y_t G(X_t)] = \mathbb{E}^{\tilde{\mathbb{P}}}[Y_t G(X_t)]. \quad (3.8)$$

Furthermore, the value of the objective function decreases:

$$\mathbb{E}^{\mathbb{P}} \left[ \int_0^T Y_t F(\alpha_t(X_t, Y_t), \beta_t(X_t, Y_t)) dt \right] \geq \mathbb{E}^{\tilde{\mathbb{P}}} \left[ \int_0^T Y_t F(\alpha_t^D(X_t), \beta_t^D(X_t)) dt \right].$$

The notion of solution to (3.7) is in the spirit of Trevisan [68, Definition 2.2] but with an extra  $-x_r f$  term in the diffusion operator of Trevisan [68, Definition 2.1].

*Proof (of Proposition 3.2 (Part II))* Proposition 3.6 instantly gives us that

$$\inf_{\mathbb{P} \in \mathcal{P}_{loc}(\mu_0, \tau, u)} \mathbb{E}^{\mathbb{P}} \left[ \int_0^T Y_t F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}}) dt \right] \geq \inf_{\mathbb{P} \in \tilde{\mathcal{P}}_{loc}(\mu_0, \tau, u)} \mathbb{E}^{\mathbb{P}} \left[ \int_0^T Y_t F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}}) dt \right],$$

and hence we have equality since  $\tilde{\mathcal{P}}_{loc}(\mu_0, \tau, u) \subseteq \mathcal{P}_{loc}(\mu_0, \tau, u)$ , which completes the proof of Proposition 3.2.

*Proof (of Proposition 3.6)* Recall the notation of Definition 3.3. Since  $\rho_t$  solves (2.3), we have for  $(t, x) \in (0, T) \times \mathbb{R}^d$

$$\begin{aligned} \int_{\mathbb{R}} y (\partial_t \rho_t(x, y) + \nabla_x \cdot [\rho_t(x, y) \alpha_t(x, y)] - \frac{1}{2} \nabla_x^2 : [\rho_t(x, y) \beta_t(x, y)] \\ - \partial_y [x_r y \rho_t(x, y)]) dy = 0. \end{aligned}$$

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

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

The same calculation gives us

$$\int_{\mathbb{R}} y \nabla_x^2 : [\rho_t(x, y) \beta_t(x, y)] dy = \nabla_x^2 : [\rho_t^D(x) \beta_t^D(x)].$$

Finally, integration by parts gives  $\int_{\mathbb{R}} y \partial_y [x_r y \rho_t(x, y)] dy = -x_r \rho_t^D(x)$ , so we have that  $\rho_t^D$  solves (3.7). Recall also the integrability condition (3.5) holds. We can thus apply Theorem 3.5 with  $\nu = \rho^D$  to obtain another probability measure,  $\tilde{\rho} \in \mathcal{P}(C([0, T]; \mathbb{R}^{d+1}))$  solving aMP( $\alpha^D, \beta^D, x_r$ ) such that  $\tilde{\rho}_t^D(\cdot) = \int_{\mathbb{R}} y \tilde{\rho}_t(\cdot, dy) = \rho_t^D(\cdot)$ . By Guo and Loeper [28, Lemma 3.4], for the canonical process  $(\tilde{X}, \tilde{Y})$  under  $\tilde{\rho}$ ,  $\tilde{X}$  is a local volatility process with the semimartingale characteristics  $\alpha_t^D(\tilde{X}_t)$  and  $\beta_t^{\tilde{\mathbb{P}}} = \beta_t^D(\tilde{X}_t)$ , and further  $\tilde{Y}_t = \tilde{Y}_0 \exp(-\int_0^t \tilde{X}_s^r ds)$ ,  $dt \times d\tilde{\rho}$ -a.e. We let  $\tilde{\mathbb{P}}$  be the projection of  $\tilde{\rho}$  on the first  $d$  coordinates. Note that  $\tilde{\mathbb{P}} \in \mathcal{P}^1$  by (3.6),  $X$  has the desired semimartingale characteristics under  $\tilde{\mathbb{P}}$  and also  $\tilde{\rho}_t$  is the distribution of  $(X_t, Y_t)$  under  $\tilde{\mathbb{P}}$ . Then, given a payoff function  $G \in C_b(\mathbb{R}^d; \mathbb{R})$ , we have for any fixed  $t \in [0, T]$

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

so (3.8) holds. Finally, we also have

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} \left[ \int_0^T Y_t F(\alpha_t(X_t, Y_t), \beta_t(X_t, Y_t)) dt \right] &= \int_0^T \int_{\mathbb{R}^{d+1}} y F(\alpha_t(x, y), \beta_t(x, y)) \zeta_{t,x}(dy) \rho_t^X(dx) dt, \\ &= \int_0^T \int_{\mathbb{R}^{d+1}} F(\alpha_t(x, y), \beta_t(x, y)) \frac{y}{D_t(x)} \zeta_{t,x}(dy) D_t(x) \rho_t^X(dx) dt, \\ &\geq \int_0^T \int_{\mathbb{R}^d} F \left( \frac{\int_{\mathbb{R}} y \alpha_t(x, y) \zeta_{t,x}(dy)}{D_t(x)}, \frac{\int_{\mathbb{R}} y \beta_t(x, y) \zeta_{t,x}(dy)}{D_t(x)} \right) D_t(x) \rho_t^X(dx) dt, \\ &= \int_0^T \int_{\mathbb{R}^d} F(\alpha_t^D(x), \beta_t^D(x)) \rho_t^D(dx) dt, \\ &= \int_0^T \int_{\mathbb{R}^d} F(\alpha_t^D(x), \beta_t^D(x)) \tilde{\rho}_t^D(dx) dt, \tag{3.9} \\ &= \mathbb{E}^{\tilde{\mathbb{P}}} \left[ \int_0^T Y_t F(\alpha_t^D(X_t), \beta_t^D(X_t)) dt \right]. \tag{3.10} \end{aligned}$$

#### 4 The dual problem

Proposition 3.2 asserted that without any loss of generality, we can restrict our attention to  $\mathbb{P} \in \tilde{\mathcal{P}}_{\text{loc}}(\mu_0, \tau, u)$ . By conditioning on  $X_t$ , we can express the value function as an integral against the discounted density, as shown above in (3.9)-(3.10). Accordingly, from now on, we always work with the discounted versions of the density, drift, and volatility. We drop the superscript  $D$  for notational ease. The problem we need to solve in the RHS of (3.1) is thus given equivalently as follows:

**Problem 4.1** The value function for the Primal Problem is given by

$$V := \inf_{\rho, \alpha, \beta} \int_0^T \int_{\mathbb{R}^d} F(\alpha_t(x), \beta_t(x)) \rho_t(dx) dt, \quad (4.1)$$

where the infimum is taken over  $(\rho, \alpha, \beta) \in C([0, T]; \mathcal{M}(\mathbb{R}^d)) \times L^1(d\rho_t dt; \mathbb{R}^d) \times L^1(d\rho_t dt; \mathbb{S}^d)$ , subject to the following constraints in the sense of distributions for  $(t, x) \in [0, T] \times \mathbb{R}^d$

$$\begin{aligned} \partial_t \rho_t(x) + \nabla_x [\rho_t(x) \alpha_t(x)] - \frac{1}{2} \nabla_x^2 : [\rho_t(x) \beta_t(x)] + x_r \rho_t(x) &= 0, \\ \text{for } i = 1, \dots, n, \quad \int_{\mathbb{R}^d} G_i(x) \rho_{\tau_i}(dx) &= u_i, \\ \rho_0(\cdot) &= \mu_0. \end{aligned}$$

We note that Problem 4.1 is in fact an example of unbalanced optimal transport, with the source term in the continuity equation being  $-x_r \rho$ , see Chizat et al. [19] and Séjourné, Peyré, and Vialard [65]. To solve this constrained optimisation problem, we will use a duality method inspired by Huesmann and Trevisan [44], Guo, Loeper, and Wang [29], and Guo et al. [31]. This proof is similar in spirit to the duality argument used in Brenier [16, Proposition 2.7], which considers dual of a relaxation of the minimal geodesic problem as an alternative approach to solving the Euler equations of an incompressible fluid. This technique had previously been applied in the proof of Brenier [15, Theorem 3.2] when seeking for the existence variational solutions of “homogenised vortex sheet equations”. The technique is also used to formulate the dual of a variational problem involving the Euler-Poisson system of equations in Loeper [57]. The proof relies mainly on the Fenchel-Rockafellar duality theorem (see Rockafellar [63, Chapter 31]), and an adjustment to make the problem convex. As our primal problem is quite similar to Guo, Loeper, and Wang [29], the approach used there can be adapted to our setting. The following result uses the notion of viscosity solution from Definition 4.8 below.

**Theorem 4.2** *The dual expression for the value function  $V$  is*

$$V = \sup_{\lambda} \left\{ \lambda \cdot u - \int_{\mathbb{R}^d} \varphi^\lambda(0, x) d\mu_0 \right\}, \quad (4.2)$$

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

$$\partial_t \varphi - x_r \varphi + \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i} + F^* \left( \nabla_x \varphi, \frac{1}{2} \nabla_x^2 \varphi \right) = 0, \quad (t, x) \in [0, T] \times \mathbb{R}^d, \quad (4.3)$$

with terminal condition  $\varphi(T, \cdot) = 0$ . If  $V$  is finite, then the infimum in Problem 4.1 is attained. If the supremum is attained for some  $\lambda^* \in \mathbb{R}^n$ , with  $\varphi^* \in \text{BV}([0, T]; C_b^2(\mathbb{R}^d))$  such that  $\varphi^*(T, \cdot) = 0$  solving the corresponding HJB equation, and  $(\rho^*, \alpha^*, \beta^*)$  being the optimal solution of Problem 4.1, then  $(\alpha^*, \beta^*)$  is given by

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

*Remark 4.3* Solving the dual problem in Theorem 4.2 yields the primal optimisers  $(\alpha^*, \beta^*)$ . These characterise the distribution of the Markov process  $(X, Y)$  solving (2.2) with these coefficients, i.e.,  $\rho^*$  solves (2.3), and the associated measure  $\mathbb{P}^*$  on  $\Omega$ ,  $\rho_t^* = \mathbb{P}^* \circ (X_t, Y_t)^{-1}$ . We can then compute the associated discount factor  $D_t^*(x)$  and apply the transformation (3.2) to obtain that  $D_t^*(x)\rho_t^*(x)$  is a solution of (3.7).

*Remark 4.4* It is not presently known if the supremum in (4.2) is attained for a generic cost function  $F$  — in general, this is an open problem in (semi)martingale optimal transport. In the static case of martingale OT, Beiglböck, Lim, and Obłój [9] show that dual attainment is delicate and depends on convexity and growth properties of the cost in the second marginal. In the Bass martingale setting of Backhoff-Veraguas et al. [6, Section 7], that is under the assumption of irreducibility of marginals, the authors establish the attainability of the static dual corresponding to a specific weak martingale OT problem. This static problem is equivalent to the dynamic formulation of Tan and Touzi [67] and Huesmann and Trevisan [44], see also Joseph, Loeper, and Obłój [52].

We prove the duality result above in the remainder of this section through a series of lemmas. Our first observation is that the objective function (4.1) is not jointly convex in  $(\rho, \alpha, \beta)$ . We define the measures  $\mathcal{A} := \rho\alpha$ ,  $\mathcal{B} := \rho\beta$ , so  $\mathcal{A}$  and  $\mathcal{B}$  are absolutely continuous with respect to  $\rho$ . Then, the objective function is convex in  $(\rho, \mathcal{A}, \mathcal{B})$  with constraints that are affine in  $(\rho, \mathcal{A}, \mathcal{B})$ . This arises from the classical notion that the function  $\bar{f}(z_1, z_2, z_3) := z_3 f(\frac{z_1}{z_3}, \frac{z_2}{z_3})$  is convex in  $(z_1, z_2, z_3)$  whenever  $f$  is convex in  $(z_1, z_2)$  on the set  $\{z_3 > 0\}$ . Note also that we write  $d\rho$  for  $\rho_t(dx)dt$ ,  $d\mathcal{A}$  for  $\alpha_t(x)\rho_t(dx)dt$  and  $d\mathcal{B}$  for  $\beta_t(x)\rho_t(dx)dt$ . Moreover, our constraints in Problem 4.1 can be formulated in the weak sense as:

$$\begin{aligned} \int_A \partial_t \varphi d\rho + \nabla \varphi \cdot d\mathcal{A} + \frac{1}{2} \nabla^2 \varphi : d\mathcal{B} - x_r \varphi d\rho + \int_{\mathbb{R}^d} \varphi d\mu_0 &= 0, \\ \int_A \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i} d\rho - \sum_{i=1}^n \lambda_i u_i &= 0. \end{aligned} \quad (4.4)$$

for any smooth compactly supported test function  $\varphi \in C_c^\infty(\Lambda)$  with  $\varphi(T, \cdot) = 0$  and  $\lambda \in \mathbb{R}^n$ . The terminal condition on  $\varphi$  arises when integrating by parts to derive (4.4), since we need the  $\rho_T(\cdot)$  boundary term to vanish as we do not have a priori knowledge of  $\rho_T(\cdot)$ . Therefore we can write Problem 4.1 as the following saddle point problem

**Problem 4.5**

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

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

We now want to find a functional with convex conjugate equal to (4.1), and another that is the remainder of the infimum in Problem 4.5. To do this, we use the following terminology from Huesmann and Trevisan [44] in the proof of the duality theorem. Denote  $\text{BV}_T([0, T]; C_b^2(\mathbb{R}^d))$  as the set of  $\varphi \in \text{BV}([0, T]; C_b^2(\mathbb{R}^d))$  such that  $\varphi(T, \cdot) = 0$ .

**Definition 4.6** We say that the triple  $(\gamma, a, b) \in C_b(\Lambda; \mathcal{X})$  is represented by  $(\varphi, \lambda) \in \text{BV}_T([0, T]; C_b^2(\mathbb{R}^d)) \times \mathbb{R}^n$  if

$$\begin{aligned} \gamma + \partial_t \varphi - x_r \varphi + \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i} &= 0, \\ a + \nabla \varphi &= 0, \\ b + \frac{1}{2} \nabla^2 \varphi &= 0. \end{aligned}$$

Since  $(\gamma, a, b) \in C_b(\Lambda; \mathcal{X})$ , the presence of the dirac delta functions give that  $t \mapsto \varphi(t, \cdot)$  is of bounded variation on  $[0, T]$  with jump discontinuities at  $t = \tau_i$ , which we denote as  $\varphi \in \text{BV}_T([0, T]; C_b^2(\mathbb{R}^d))$ , since we require  $\varphi$  to be at least  $C^2$  in space. Proceeding in an analogous way to Guo, Loeper, and Wang [29] to obtain the dual problem, first define functionals  $\Phi : C_b(\Lambda; \mathcal{X}) \rightarrow \mathbb{R} \cup \{+\infty\}$  and  $\Psi : C_b(\Lambda; \mathcal{X}) \rightarrow \mathbb{R} \cup \{+\infty\}$  by

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

**Lemma 4.7** *The objective function  $V$  can be expressed in terms of  $\Phi$  and  $\Psi$  as*

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

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

We remark that switching from  $\varphi \in C_c^\infty(\Lambda)$  (with  $\varphi(T, \cdot) = 0$ ) to  $\varphi \in BV_T([0, T]; C_b^2(\mathbb{R}^d))$  does not change the value of the supremum, which will be formalised by the notion of viscosity solutions later in Definition 4.8

*Proof* We first evaluate  $\Phi^*$  on  $(\rho, \mathcal{A}, \mathcal{B}) \in \mathcal{M}(\Lambda; \mathcal{X})$  such that  $\rho \in \mathcal{M}_+(\Lambda)$  and  $\mathcal{A}, \mathcal{B} \ll \rho$  as follows

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

where the exchange of the integral and the supremum is justified as in Guo, Loeper, and Wang [29, Lemma A.1]. Furthermore, the arguments in Guo, Loeper, and Wang [29, Lemma A.1] also show that  $\Phi^* = \infty$  for  $(\rho, \mathcal{A}, \mathcal{B}) \in C_b(\Lambda; \mathcal{X})^*$  which are not of the form above. In summary:

$$\Phi^*(\rho, \mathcal{A}, \mathcal{B}) = \begin{cases} \int_\Lambda F\left(\frac{d\mathcal{A}}{d\rho}, \frac{d\mathcal{B}}{d\rho}\right) d\rho & \text{if } (\rho, \mathcal{A}, \mathcal{B}) \in \mathcal{M}(\Lambda, \mathcal{X}), \\ +\infty & \rho \in \mathcal{M}_+(\Lambda), \mathcal{A}, \mathcal{B} \ll \rho \\ & \text{otherwise.} \end{cases} \quad (4.5)$$

We now compute  $\Psi^* : C_b(\Lambda; \mathcal{X})^* \rightarrow \mathbb{R} \cup \{+\infty\}$ :

$$\begin{aligned} \Psi^*(\rho, \mathcal{A}, \mathcal{B}) &= \sup_{(\gamma, a, b) \in C_b(\Lambda; \mathcal{X})} \left\{ \langle (\gamma, a, b), (\rho, \mathcal{A}, \mathcal{B}) \rangle - \int_{\mathbb{R}^d} \varphi(0, x) d\mu_0 + \sum_{i=1}^n \lambda_i u_i \right\} \\ &= \sup_{\varphi, \lambda} \left\{ \left\langle \left( x_r \varphi - \partial_t \varphi - \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i}, -\nabla \varphi, -\frac{1}{2} \nabla_x^2 \varphi \right), (\rho, \mathcal{A}, \mathcal{B}) \right\rangle \right. \\ &\quad \left. - \int_{\mathbb{R}^d} \varphi(0, x) d\mu_0 + \sum_{i=1}^n \lambda_i u_i \right\} \\ &= \sup_{\varphi, \lambda} \left\{ \left[ \int_\Lambda x_r \varphi d\rho - \partial_t \varphi d\rho - \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i} d\rho - \nabla_x \varphi d\mathcal{A} - \frac{1}{2} \nabla_x^2 \varphi : d\mathcal{B} \right] \right. \\ &\quad \left. - \int_{\mathbb{R}^d} \varphi(0, x) d\mu_0 + \sum_{i=1}^n \lambda_i u_i \right\}, \end{aligned} \quad (4.6)$$

where  $(\gamma, a, b)$  are represented by  $(\varphi, \lambda)$  in the sense of Definition 4.6.

Thus, since  $\Phi^*$  is independent of  $(\varphi, \lambda)$  we can simply add (4.5) and (4.6) together to get the argument of the infimum in Problem 4.5, so

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

where the restriction  $\rho \in \mathcal{M}_+(\Lambda), \mathcal{A}, \mathcal{B} \ll \rho$  are ensured automatically since otherwise  $\Phi^* = +\infty$ . To finish the proof, we invoke Guo, Loeper, and Wang [29, Lemma A.2] to conclude that

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

Now we apply the Fenchel-Rockefeller duality theorem as formulated in Villani [69, Theorem 1.9]. We first note that as the constraints in the functionals  $\Phi$  and  $\Psi$  are affine, the functionals are clearly convex in  $(\gamma, a, b)$ . We now check the conditions of the theorem at the point  $(0, O^{d \times 1}, O^{d \times d})$  which is represented by  $(\varphi, \lambda) = (e^{tx_r}, O^{n \times 1})$  where  $O$  refers to the zero matrix of appropriate dimension. Since  $F$  is non-negative

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

Therefore, we have  $\Phi(0, O^{d \times 1}, O^{d \times d}) = 0$  and  $(0, O^{d \times 1}, O^{d \times d})$  is a point of continuity of  $\Phi$  since  $F^*$  is continuous. Moreover, since  $\varphi(t, x) = e^{tx_r}$  and  $\mu_0$  is a probability measure

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

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

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

Thus rearranging we obtain

$$\begin{aligned}
V &= \inf_{(\rho, \mathcal{A}, \mathcal{B}) \in C_b(\Lambda; \mathcal{X})^*} \{\Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B})\} \\
&= \sup_{(\gamma, a, b) \in C_b(\Lambda; \mathcal{X})} \{-\Phi(-\gamma, -a, -b) - \Psi(\gamma, a, b)\} \\
&= \sup_{\substack{(\gamma, a, b) \in C_b(\Lambda; \mathcal{X}) \\ -\gamma + F^*(-a, -b) \leq 0 \\ (\gamma, a, b) \text{ is represented by } (\varphi, \lambda)}} \left\{ \sum_{i=1}^n \lambda_i u_i - \int_{\mathbb{R}^d} \varphi(0, x) d\mu_0 \right\} \\
&= \sup_{\substack{(\lambda, \varphi) \in \mathbb{R}^n \times \text{BV}_T([0, T]; C_b^2(\mathbb{R}^d)) \\ \partial_t \varphi - x_r \varphi + \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i} + F^*(\nabla_x \varphi, \frac{1}{2} \nabla_x^2 \varphi) \leq 0}} \left\{ \sum_{i=1}^n \lambda_i u_i - \int_{\mathbb{R}^d} \varphi(0, x) d\mu_0 \right\}. \tag{4.7}
\end{aligned}$$

Now, to obtain equality in the HJB equation constraint and thus the HJB equation (4.3) and dual formulation in Theorem 4.2, we adapt the classical notion of viscosity solutions from Lions [55] to include the required jump discontinuities, in analogy to Guo, Loeper, and Wang [29]. First define disjoint intervals  $I_k := [\tau_{k-1}, \tau_k]$  with  $\tau_0 = 0$ , with  $\bigcup_{k=1}^n I_k = [0, T]$ .

**Definition 4.8** For any  $\lambda \in \mathbb{R}^n$ , we say  $\varphi \in \text{BV}_T([0, T]; C_b(\mathbb{R}^d))$  is a viscosity subsolution (supersolution) of (4.3) if  $\varphi|_{I_k \times \mathbb{R}^d} \in C_b(I_k; C_b(\mathbb{R}^d))$  is a classical (continuous) viscosity subsolution (supersolution) of (4.3) in  $I_k \times \mathbb{R}^d$  and for all  $k = 1, \dots, n$  has jump discontinuities:

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

With terminal condition  $\varphi(T, \cdot) = 0$ . In addition,  $\varphi \in \text{BV}_T([0, T]; C_b(\mathbb{R}^d))$  is a viscosity solution of (4.3) if it is both a viscosity subsolution and viscosity supersolution of (4.3).

The expression for  $V$  in (4.7) involved supersolutions to (4.3) and the first step is to show that we can restrict to viscosity solutions, as stated in the first part of Theorem 4.2. For this we follow the proof of Guo, Loeper, and Wang [29, Proposition 3.5]. Guo, Loeper, and Wang [29, Remark 3.9] provides a comparison principle, which is the classical comparison principle of viscosity solutions applied to  $\varphi$  on  $I_k \times \mathbb{R}^d$  for each  $k$ . Using this, one can deduce existence and uniqueness of solutions to (4.3) via Crandall, Ishii, and Lions [22]. The comparison principle also implies that  $V$  in (4.2) is smaller than the supremum over viscosity solutions. We then use the smoothing argument from Bouchard, Loeper, and Zou [12], which shows that any viscosity solution of (4.3) can be approached by smooth supersolutions. This, together with (4.7) shows that  $V$  is larger than the supremum over viscosity solutions. This allows us to conclude the proof of the first point of Theorem 4.2. We now seek to obtain the form of the optimal  $(\alpha, \beta)$  for the second part of Theorem 4.2.

**Lemma 4.9** *If the supremum in Theorem 4.2 is attained for some  $\lambda^*$  with  $\varphi^* \in \text{BV}_T([0, T]; C_b^2(\mathbb{R}^d))$  solving the corresponding HJB equation, and  $(\rho^*, \alpha^*, \beta^*)$  being the optimal solution of Problem 4.1, then  $(\alpha^*, \beta^*)$  is given by:*

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

*Proof* Let  $(\alpha^*, \beta^*)$  be the optimal solution of Problem 4.1, then  $(\rho^*, \rho^* \alpha^*, \rho^* \beta^*)$  also achieves the infimum in Problem 4.5. Assume that  $\lambda^*$  is the optimal solution solving (4.2) with corresponding solution to (4.3)  $\varphi^*$ . Then,  $\lambda^*$  achieves the supremum in Problem 4.5, so with our optimal solution we may rewrite Problem 4.5 as

$$\begin{aligned} V = \int_A & \left( F(\alpha^*, \beta^*) - \partial_t \varphi^* - \nabla_x \varphi^* \cdot \alpha^* - \frac{1}{2} \nabla_x^2 \varphi^* : \beta^* + x_r \varphi^* \right. \\ & \left. - \sum_{i=1}^n \lambda_i^* G_i(x) \delta_{\tau_i} \right) d\rho^* - \int_{\mathbb{R}^d} \varphi^* d\mu_0 + \sum_{i=1}^n \lambda_i^* u_i. \end{aligned} \quad (4.8)$$

Since  $(\varphi^*, \lambda^*)$  are optimal, we have from Theorem 4.2 that

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

Therefore, (4.8) is equivalent to

$$\begin{aligned} 0 = \int_A & \left( F(\alpha^*, \beta^*) - \partial_t \varphi^* - \nabla_x \varphi^* \cdot \alpha^* - \frac{1}{2} \nabla_x^2 \varphi^* : \beta^* + x_r \varphi^* - \sum_{i=1}^n \lambda_i^* G_i(x) \delta_{\tau_i} \right) d\rho^* \\ = \int_A & \left( F(\alpha^*, \beta^*) + F^*(\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^*) - \nabla_x \varphi^* \cdot \alpha^* - \frac{1}{2} \nabla_x^2 \varphi^* : \beta^* \right) d\rho^*. \end{aligned} \quad (4.9)$$

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

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

Note that

$$(\bar{\alpha}, \bar{\beta}) = \nabla F^* \left( \nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^* \right) = \arg \max_{(a, b) \in \mathbb{R}^d \times \mathbb{S}_+^d} \left\{ a \cdot \nabla_x \varphi^* + b : \frac{1}{2} \nabla_x^2 \varphi^* - F(a, b) \right\}. \quad (4.10)$$

Thus,

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

Since  $F$  is convex, its Legendre transform is an involution (see Rockafellar [63, Corollary 12.2.1]), so taking the double Legendre transform, we have

$$(A, B) := \nabla F(\bar{\alpha}, \bar{\beta}) = \nabla F^{**}(\bar{\alpha}, \bar{\beta}) = \arg \max_{(a,b) \in \mathbb{R}^d \times \mathbb{S}_+^d} \{a \cdot \bar{\alpha} + b : \bar{\beta} - F^*(a, b)\}.$$

Therefore,

$$\begin{aligned} F^{**}(\bar{\alpha}, \bar{\beta}) &= A \cdot \bar{\alpha} + B : \bar{\beta} - F^*(A, B) \\ &= A \cdot \bar{\alpha} + B : \bar{\beta} - \max_{(x,y) \in \mathbb{R}^d \times \mathbb{S}_+^d} \{A \cdot x + B : y - F(x, y)\}. \end{aligned}$$

Since  $F(\bar{\alpha}, \bar{\beta}) = F^{**}(\bar{\alpha}, \bar{\beta})$ , from (4.10) we have that for the affine terms to cancel, we need  $(A, B) = (\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^*)$ . So, substituting into (4.9) we have

$$0 = \int_A (F(\alpha^*, \beta^*) - F(\bar{\alpha}, \bar{\beta}) - \nabla_x \varphi^* \cdot (\alpha^* - \bar{\alpha}) - \frac{1}{2} \nabla_x^2 \varphi^* : (\beta^* - \bar{\beta})) d\rho^*. \quad (4.11)$$

Since  $F(\alpha, \beta)$  is assumed to be strongly convex in  $(\alpha, \beta)$ , we have from (2.4) that for some constant  $C > 0$

$$F(\alpha^*, \beta^*) - F(\bar{\alpha}, \bar{\beta}) \geq \langle \nabla F(\bar{\alpha}, \bar{\beta}), (\alpha^* - \bar{\alpha}, \beta^* - \bar{\beta}) \rangle + C(\|\alpha^* - \bar{\alpha}\|^2 + \|\beta^* - \bar{\beta}\|^2).$$

Applying this inequality to (4.11) and noting that  $\nabla F(\bar{\alpha}, \bar{\beta}) = (\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^*)$  gives us

$$0 \geq \int_A C(\|\alpha^* - \bar{\alpha}\|^2 + \|\beta^* - \bar{\beta}\|^2) d\rho^* \geq 0.$$

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

#### 4.1 Sequential Calibration Setup

We now specify the setting in which we seek to apply Theorem 4.2. We first specify the state variables of our model: we consider  $d = 2$ ,  $X_t = (r_t, Z_t)$ , where  $r_t$  is the short rate and  $Z_t$  is the log-stock price. We start with a setting in which a model for the short rate is fixed and has already been calibrated. We refer to this as a “sequential calibration” problem. Our aim is then to calibrate a local volatility model for the stock price, where the local volatility function can depend on the short rate. For this we will employ the OT methodology developed above. The joint calibration problem in which both the short rate and the stock price are calibrated simultaneously using the OT methodology, and in particular the short rate model parameters can depend on the stock price, is considered in our sequel paper Joseph, Loeper, and Obłój [51].

Specifically, we take now a given pre-calibrated Hull-White model for the interest rate, see Hull and White [47, 46], and local volatility dynamics for the log-price

$$\begin{aligned} dZ_t &= r_t - \frac{1}{2} \sigma^2(t, Z_t, r_t) + \sigma(t, Z_t, r_t) dW_t^1, \\ dr_t &= (b(t) - ar_t) dt + \sigma_r dW_t^2, \\ d\langle W^1, W^2 \rangle_t &= \xi(t, Z_t, r_t) dt, \end{aligned}$$

where  $a, \sigma_r > 0$  are constants and  $b(\cdot)$  is a function of time, calibrated so that the dynamics of  $(r_t)_{0 \leq t \leq T}$  match the market data (e.g., suitable interest rates caps and floors). Both  $a$  and  $\sigma_r$  being positive constants is not a particularly restrictive constraint as remarked in Brigo and Mercurio [17] and Hull and White [45]. Note that  $b(t)$  will therefore need to be calibrated to fit the term structure of interest rates seen in the market. Our aim now is to calibrate  $\sigma(t, Z_t, r_t)$  and  $\xi(t, Z_t, r_t)$  using our OT-methodology. In order to calibrate the local volatility and correlation, we will want to find a cost function that forces  $\alpha_t^{\mathbb{P}}$  and  $\beta_t^{\mathbb{P}}$  to take the form above. As discussed before, we achieve this by using a functional form for  $F$  as long as  $(\alpha, \beta) \in \Gamma$  for some convex set  $\Gamma$ , with  $F = +\infty$  otherwise. The set  $\Gamma$  will enforce in particular that  $\beta$  is positive semidefinite and symmetric, and  $\beta_{22} = \sigma_r^2$ , which defines a set that is convex in  $\beta$ . With no further constraints on  $\beta_{12}$ , we obtain a non-explicit form of the Legendre transform of the cost function,  $F^*$ , which in general will require a numerical solver at each  $(t, x)$ . To reduce the computational difficulty of the numerical solution, as in Guo, Loeper, and Wang [29], we first restrict the correlation to the following form, which still keeps the set  $\Gamma$  convex

$$\xi(t, Z_t, r_t) = \frac{\sigma_r}{\sigma(t, Z_t, r_t)} \xi_{\text{ref}}(t, Z_t, r_t), \quad \text{for } t \in [0, T], \quad (4.12)$$

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

$$\begin{aligned} \Gamma(t, Z_t, r_t) = \left\{ (\alpha, \beta) \in \mathbb{R}^2 \times \mathbb{S}^2 : \alpha_1 &= r_t - \frac{1}{2}\beta_{11}, \alpha_2 = b(t) - ar_t, \right. \\ \left. \beta_{12} = \beta_{21} = \xi_{\text{ref}}\sigma_r^2, \beta_{22} = \sigma_r^2 \right\}, \end{aligned} \quad (4.13)$$

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

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

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

$$F(t, Z, r, \alpha, \beta) = \begin{cases} H(\beta_{11}, \bar{\sigma}^2, \xi_{\text{ref}}^2 \sigma_r^2), & \text{if } (\alpha, \beta) \in \Gamma(t, Z_t, r_t), \\ +\infty, & \text{otherwise.} \end{cases} \quad (4.14)$$

*Remark 4.10* It is easy to check that the function  $F$  defined in (4.14) satisfies the assumptions for the duality proof.

This cost function will ensure that we retain the Hull-White model in the interest rate, while also matching the market prices for the call options by calibrating the volatility of the stock. Additionally, we wish to enforce that the matrix  $\beta$  from our model characteristics is positive definite and that  $\xi_{\text{ref}}$  remains a correlation function, and we achieve this by setting  $s = \xi_{\text{ref}}^2 \sigma_r^2$  as an argument of  $H$  in the definition of  $F$ . Applying Theorem 4.2, we have the following dual formulation with the given cost function  $F(\alpha, \beta)$

**Problem 4.11**

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

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

$$\begin{aligned} \sum_{i=1}^n \lambda_i (\exp(z) - K_i)^+ \delta_{\tau_i} + \partial_t \varphi + \sup_{\beta_{11}} \left\{ (r - \frac{1}{2} \beta_{11}) \partial_z \varphi + (b(t) - ar) \partial_r \varphi + \frac{1}{2} \beta_{11} \partial_{zz}^2 \varphi \right. \\ \left. + \bar{\xi} \sigma_r^2 \partial_{zr}^2 \varphi + \frac{1}{2} \sigma_r^2 \partial_{rr}^2 \varphi - r \varphi - H(\beta_{11}, \bar{\sigma}^2, \bar{\xi}^2 \sigma_r^2) \right\} = 0. \end{aligned} \quad (4.15)$$

**Lemma 4.12** *The optimal characteristic in the HJB equation (4.15),  $\beta_{11}^*$  is given by*

$$\begin{aligned} \beta_{11}^*(t, z, r) = \bar{\xi}^2 \sigma_r^2 + & \left( \frac{(\bar{\sigma}^2 - \bar{\xi}^2 \sigma_r^2)^p (\varphi_{zz} - \varphi_z)}{4(p^2 - 1)} \right. \\ & \left. + \frac{1}{2} \sqrt{\left( \frac{(\bar{\sigma}^2 - \bar{\xi}^2 \sigma_r^2)^p (\varphi_{zz} - \varphi_z)}{2(p^2 - 1)} \right)^2 + 4(\bar{\sigma}^2 - \bar{\xi}^2 \sigma_r^2)^{2p}} \right)^{\frac{1}{p}}. \end{aligned} \quad (4.16)$$

*Proof* By differentiating the argument of the supremum in (4.15) with respect to  $\beta_{11}$ , we notice that solving the supremum over  $\beta_{11}$  in (4.15) is equivalent to solving the equation for  $x$ :

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

Computing the right hand term and rearranging, we have

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

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

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

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

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

$$\begin{cases} \partial_t \psi + (r - \frac{1}{2}\beta_{11}^*)\partial_z \psi + (b(t) - ar)\partial_r \psi + \frac{1}{2}\beta_{11}^*\partial_{zz}^2 \psi \\ \quad + \bar{\xi}\sigma_r^2\partial_{zr}^2 \psi + \sigma_r^2\partial_{rr}^2 \psi - r\psi = 0, \\ \psi(\tilde{\tau}, z, r) = \tilde{G}(z, r). \end{cases} \quad (4.17)$$

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

## 5 Numerical Method

In this section, we outline the numerical method used to solve the dual formulation in Theorem 4.2. We will use and adapt the methods presented in Guo, Loeper, and Wang [29] and Guo et al. [31] for ease of implementation. Analogously to those two works, in order to compute the supremum in 4.2, we must solve an HJB equation for a given  $\lambda$ , and then update the  $\lambda$  using an optimisation algorithm. To speed up the convergence, we compute the gradients with respect to  $\lambda$  of the dual objective function.

**Lemma 5.1** *Suppose Problem 4.1 is admissible and define the dual objective function as*

$$L(\lambda) = \lambda \cdot u - \varphi^\lambda(0, Z_0, r_0). \quad (5.1)$$

*Then the gradients of the dual objective function are given by*

$$\partial_{\lambda_i} L(\lambda) = u_i - \mathbb{E}^{\mathbb{P}} \left[ e^{-\int_0^{\tau_i} r_s ds} G_i(Z_{\tau_i}) \right]. \quad (5.2)$$

The gradients are obtained via the same method in Guo, Loeper, and Wang [29, Lemma 4.5], but with the discounting appearing in the expectation as a result of the  $-r\varphi$  term in the HJB equation from the Feynman-Kac formula. The interpretation is the same here, that the gradients represent the difference between the model and market prices. We briefly outline the numerical method from Guo, Loeper, and Wang [29] and Guo et al. [31] which can be directly applied here. We are first given an initial guess  $\lambda$ , which will usually be taken to be a zero vector, and then solve (4.15) to obtain  $\varphi(0, Z_0, r_0)$ . Since the HJB equation (4.15) has dirac deltas at the times  $(\tau_i)_{i=1,\dots,n}$ , the solution also has jump discontinuities in time. However, in-between these jump discontinuities, the solution is continuous so can be solved using standard techniques backwards in time, and then the jump discontinuities can be incorporated into a terminal condition at the expiration time. That is, we take the final jump discontinuity as a terminal condition, solve the HJB equation up to the next

calibrating option expiry, and then add the jump discontinuity to the solution as a new terminal condition at the next jump discontinuity in the same way as given in Definition 4.8 since we understand the solution in the viscosity sense. Then, given  $\varphi(0, Z_0, r_0)$ , we calculate the objective value (5.1), and apply an optimisation algorithm to update the  $\lambda$  to a new guess. As in Guo, Loeper, and Wang [29] and Guo et al. [31], the L-BFGS algorithm of Liu and Nocedal [56] displayed good convergence, and with the gradients calculated in Lemma 5.1 the convergence can be made faster. We remark that to compute (5.2), the  $\mathbb{E}^{\mathbb{P}} \left[ e^{-\int_0^{\tau_i} r_s \, ds} G_i(Z_{\tau_i}) \right]$  term is simply the model price of the  $i^{\text{th}}$  option. This expectation can either be computed via Monte Carlo methods, or using standard pricing PDE techniques by solving (4.17) with  $\beta_{11}^*$  obtained from the solution of the HJB equation. This process is then repeated until  $\|\nabla_{\lambda} L(\lambda)\|_{\infty} < \varepsilon$  for some specified tolerance  $\varepsilon > 0$ , which corresponds to the model prices matching all of the market prices.

We perform the constant rescaling in the short rate variable  $r_t \mapsto Rr_t$  where we choose  $R = 100$  for stability reasons in the finite difference approximations to make it of the same order as the other coordinate. In addition, we rescale the calibrating option prices and their payoffs by their vegas computed from their Black-Scholes implied volatility. This not only helps the stability of the numerical method by reducing the magnitude of the jump discontinuities, but it also converts pricing errors into implied volatility errors since the vega represents how much the option price will change as the volatility changes by 1%.

### 5.1 Numerically Solving the HJB Equation

We now outline how to solve the HJB equation (4.15). In Barles and Souganidis [8], it has been shown that monotonicity, stability and consistency of a numerical scheme guarantees convergence locally uniformly so long as there exists a comparison principle for the analytic solution. We use a policy iteration method (see Ma and Forsyth [58]) similar to that in Guo, Loeper, and Wang [29] and Guo et al. [31], where we solve the HJB equation using an implicit finite difference method, with central difference approximations for the spatial derivatives from In't Hout and Foulon [48]. We choose a boundary far away enough such that the boundary conditions do not have a significant effect on the HJB equation solution, and our boundary conditions are such that the second derivative of  $\varphi$  does not change with time between each maturity. That is, for a subsequence of the calibrating option maturity times  $(\tau_{i_k})_{k=1,\dots,m}$  such that for  $k = 1, \dots, m$  all  $\tau_{i_k}$  are distinct, with  $\tau_{i_0} = 0$ , we set for  $t \in (\tau_{i_{k-1}}, \tau_{i_k}]$ , for each  $k = 1, \dots, m$ ,

$$\partial_{zz}^2 \varphi(t, z, r) = \partial_{zz}^2 \varphi(\tau_i, z, r), \quad \text{for } z \in \{z_{\min}, z_{\max}\}, r \in [r_{\min}, r_{\max}], \quad (5.3)$$

$$\partial_{rr}^2 \varphi(t, z, r) = \partial_{rr}^2 \varphi(\tau_i, z, r), \quad \text{for } z \in [z_{\min}, z_{\max}], r \in \{r_{\min}, r_{\max}\}. \quad (5.4)$$

We started with this boundary condition and compared it with constant Dirichlet boundary conditions  $\varphi(t, x) = C_{i_k}$ , for some constant  $C_{i_k} \in \mathbb{R}$ , all  $t \in (\tau_{i_{k-1}}, \tau_{i_k}]$ , and  $x$  on the boundary of our rectangular domain, and observed no significant change. In addition, we use the boundary conditions (5.3)-(5.4) in the linearised pricing PDE (4.17). At each time step, we start with the value of  $\varphi$  at the previous time step, we then approximate the PDE coefficients  $\alpha$  and  $\beta$  using Lemma 4.9 and Lemma 4.12, and then solve the HJB equation at that time step using one step of a fully implicit spatially second order finite difference scheme. We then use Lemma 4.9 and Lemma 4.12 again at the same time step to approximate the PDE coefficients again with the new value of  $\varphi$ . This process is repeated until convergence within some specified tolerance, after which we proceed to the next time step. Once we have solved the HJB equation at all time steps, we then use the value of  $\beta_{11}$  computed from solving the HJB equation, compute  $\beta_{12}$  from the convexity adjustment, and use them to solve the linearised model pricing PDE via the ADI method to generate the model prices. In the simulated data example, we use a discretisation on a uniform  $100 \times 100$  spatial grid of  $[4, 5] \times [0, 5]$  for the log-stock and rescaled interest rates, and partition the time interval into year fractions at the resolution of one day, so that  $dt = \frac{1}{365}$ . We use the minimisation package, minfunc, of Schmidt [64] for the implementation of L-BFGS.

The above numerical method is summarised in Algorithm 1. We let  $0 = t_0 < t_1 < \dots < t_N = T$  be a discretisation of the time interval  $[0, T]$  such that the expiration times of the calibrating instruments  $(\tau_i)_{i=1,\dots,n}$  are a subset of the discretisation times. We let  $\varepsilon_1$  be the tolerance of the model prices to the calibrating prices, so that  $\|\nabla_\lambda L(\lambda)\|_\infty < \varepsilon_1$ , where the gradients are given in Lemma 5.1, and we let  $\varepsilon_2$  be the tolerance of the policy iteration for approximating the optimal characteristics.

## 5.2 Numerical Results

We present numerical results showcasing the performance of our proposed calibration method. We use simulated data to investigate the advantages and drawbacks of our method, and in particular its dependence on the reference model  $\bar{\sigma}$  in (4.14). We use the CEV (constant elasticity of variance) model of Cox [20], with different sets of parameters for the model used to generate the option prices and for our reference model. The underlying  $S_t = \exp(Z_t)$ ,  $0 \leq t \leq T$ , thus solves the following stochastic differential equation:

$$dS_t = r_t S_t dt + \sigma(t, S_t) S_t dW_t^1,$$

with  $\sigma(t, S_t) = \sigma S_t^{\gamma-1}$ , where  $\sigma \geq 0$  and  $\gamma \geq 0$  are both constants. We remark that this is a special case of the SABR (“stochastic  $\alpha, \beta, \rho$ ”) model derived in Hagan et al. [38] as a stochastic volatility extension of the CEV model.

We solve a pricing PDE to compute the generating model prices and consider the following instruments:

1. calls on the underlying with an expiration of 60 days;

**Algorithm 1:** Policy iteration algorithm.

---

**Data:** Input an initial  $\lambda$  and market prices  $u_i$ .  
**Result:** Calibrated model prices, optimal characteristics

```

1 while || $\nabla_\lambda L(\lambda)$ || $_\infty > \varepsilon_1$  do
2   /* Solve the HJB equation backwards in time */
3   for  $k = N - 1, \dots, 0$  do
4     /* Terminal Conditions - adding  $\lambda$  multiplied by the payoff */
5     if  $t_{k+1} = \tau_i$  for some  $i = 1, \dots, n$  then
6       |  $\varphi_{t_{k+1}} \leftarrow \varphi_{t_{k+1}} + \sum_{i=1}^n \lambda_i G_i \mathbb{1}_{\{t_{k+1}=\tau_i\}}$ 
7     end
8     /* Policy iteration to approximate the optimal characteristics */
9      $\varphi_{t_k}^{\text{new}} \leftarrow \varphi_{t_{k+1}}$  // Approximate using previous time step
10    while || $\varphi_{t_k}^{\text{new}} - \varphi_{t_k}^{\text{old}}$ ||  $> \varepsilon_2$  do
11      |  $\varphi_{t_k}^{\text{old}} \leftarrow \varphi_{t_k}^{\text{new}}$  // Store the old value of  $\varphi$ 
12      | Approximate  $\beta_{11}^*$  using Lemma 4.9 and Lemma 4.12 with  $\varphi_{t_k}^{\text{old}}$ .
13      | // Use old values to approximate optimal characteristics
14      | Use  $\beta_{11}^*$  to compute  $\alpha_1^*$  and  $\beta_{12}^*$ , then plug into (4.15) to remove the
15        supremum and solve using one step of an implicit finite difference
16        method, and set the solution to  $\varphi_{t_k}^{\text{new}}$ .
17    end
18    |  $\varphi_{t_k} \leftarrow \varphi_{t_k}^{\text{new}}$  // Save the solution once the  $\varphi$  has converged to the
19      optimal solution
20  end
21  /* Computing the model prices and gradients */
22  Compute the model prices by solving the pricing PDE (4.17) using the ADI
23  method.
24  Compute the gradients (5.2).
25  Use the L-BFGS algorithm to update  $\lambda$ .
26 end

```

---

2. calls on the underlying with an expiration of 120 days.

We note that the payoffs of these options are not smooth and cause instabilities when computing the derivatives in the terminal conditions of the HJB equation. Thus, we use the smoothed version of the call option payoffs, which for  $\varepsilon \ll K$  is given by:

$$(S_T - K)^+ \approx \frac{S_T - K}{2} \left[ \tanh\left(\frac{S_T - K}{\varepsilon}\right) + 1 \right].$$

We keep the interest rate model parameters the same throughout since that is assumed to be given.

We present two numerical examples, the “good” reference model where the initial reference model is parametrically close to the generating model, and the “bad” reference model where it is not parametrically close to the generating model and in particular has a correlation with a different sign. Note that in the extreme case when the generating model and the reference model are the same, the calibration procedure will stop instantly and recover the generating model. The parameters for all the models are summarised in Table 5.1. Our numerical methods converged for both the “good” and “bad” reference model

Parameter	Value	Interpretation
$Z_0$	$\log(92)$	Initial log-underlying price
$r_0$	$0.025 \times 100$	Initial short rate scaled by $R = 100$
$\varepsilon_1$	$1 \times 10^{-4}$	Tolerance for the difference in model and market implied volatility
$\varepsilon_2$	$1 \times 10^{-8}$	Tolerance for the policy iteration approximation of the optimal characteristics
$p$	4	Exponent in the cost function
$\sigma$	0.78	Volatility scaling of generating CEV model
$\gamma$	0.9	Power law in generating CEV model
$\theta(t)$	$ar_0 + \frac{\sigma_r^2}{2a}(1 - e^{-2at})$	Initial term structure of Hull-White generating model
$a$	0.4	Speed of mean reversion of Hull-White generating model
$\sigma_r$	0.05	Volatility of Hull-White generating model
$\xi$	-0.6	Instantaneous correlation between short rate and log-stock in generating model
$\bar{\sigma}_{\text{good}}$	0.9	Volatility scaling of the “good” reference CEV model
$\bar{\gamma}_{\text{good}}$	0.9	Power law in the “good” reference CEV model
$\bar{\xi}_{\text{good}}$	-0.4	Instantaneous correlation between short rate and log-stock in the “good” reference model
$\bar{\sigma}_{\text{bad}}$	1.2	Volatility scaling of the “bad” reference CEV model
$\bar{\gamma}_{\text{bad}}$	0.78	Power law in the “bad” reference CEV model
$\bar{\xi}_{\text{bad}}$	0.4	Instantaneous correlation between short rate and log-stock in the “bad” reference model

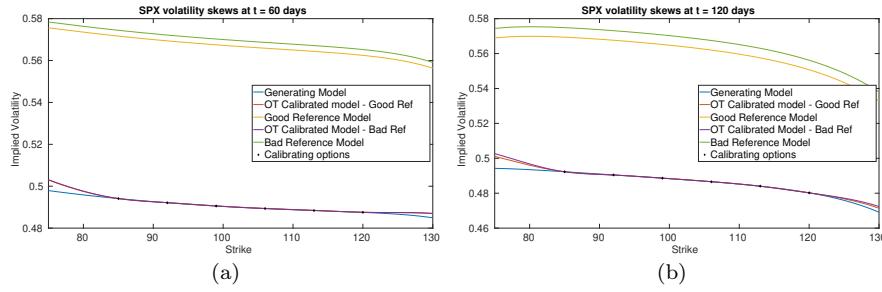
**Table 5.1** Parameter values for the simulated data example.

case, calibrating all the call options to a tolerance of  $10^{-4}$ , with the calibrated model implied volatility replicating the prices of the generating model. The generating and calibrated model implied volatility skews are indistinguishable for both maturities within the range  $K = [85, 120]$  in which our options’ strikes were taken, indicating that our optimal transport model replicates the observed simulated data both at and in-between the calibrating option strikes. The dependence on the reference model is only observed outside of this range, which is to be expected. The plots are given in Figure 5.1 with the exact results summarised in Table 5.2.

A feature of our method is that it is designed to find a calibrated model closest to the reference model, in the sense of minimising (4.1). This results in a spiky volatility surface – the method tries to stick to the reference model whilst making sharp deviations needed to calibrate to the given options’ prices. This is not necessarily a desired feature and we propose to smooth the surface to obtain a reasonable volatility surface while matching the market data. We used an iterative procedure: once we obtained convergence of the model prices to the generating model prices within the tolerance of  $\varepsilon_1$ , we then stored those characteristics as the new reference model, applied an interpolation method using cubic splines to smooth the surface and then restarted the process with the smoothed surface as our reference model. We remark that the final epoch of the calibration algorithm involved no interpolation, as this spoils calibration,

so the overall algorithm provided the OT calibrated surface after iterating through smoothed reference models. In addition to providing more reasonable model dynamics, the smoothed reference model iteration also improved the numerical stability of the algorithm, and thus allowed us to achieve a better calibration. This came at a significant computational cost due to the iterations of the calibration routine, however both ran on a standard notebook laptop in a matter of hours. Both reference models were smoothed around 10 times before changes in the volatility surface plots were no longer noticeable.

We remark that the aforementioned reference model iteration could, in principle, be circumvented via a regularisation technique by encoding a smoothing penalty into the cost function, such as via a second order Tikhonov regularisation method. Clearly this would change the optimal characteristics given in (4.16), but the optimal surface would then have this smoothing penalty encoded into it. However, this would require at all gridpoints to estimate the first and second derivatives of  $\beta_{11}$  via a central difference method, which would substantially increase the computational load beyond that of the reference model iteration method, and additionally potentially require extra state variables to account for the spatial derivatives of  $\beta$ .



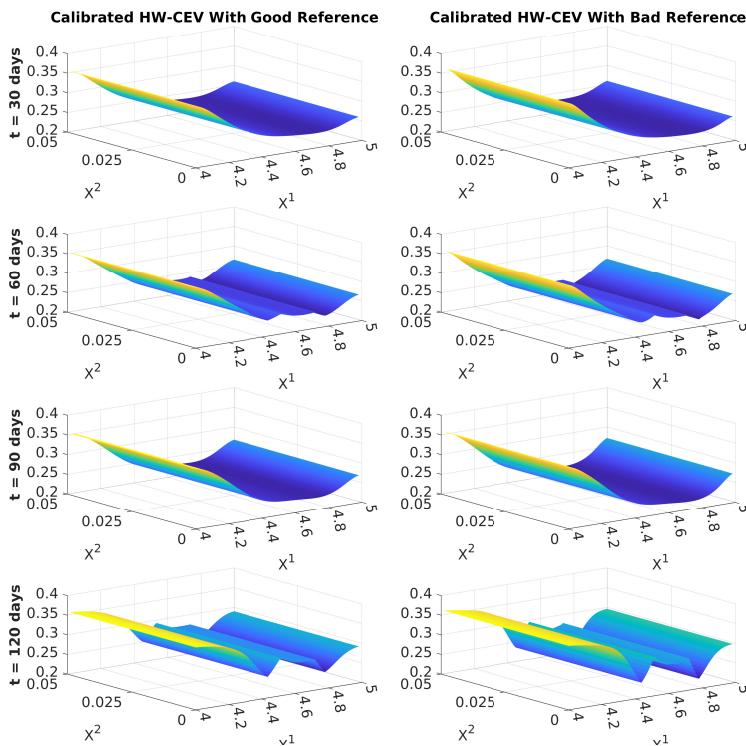
**Fig. 5.1** Implied volatility skews under the generating model, reference model and OT-calibrated models with both reference models and across two maturities.

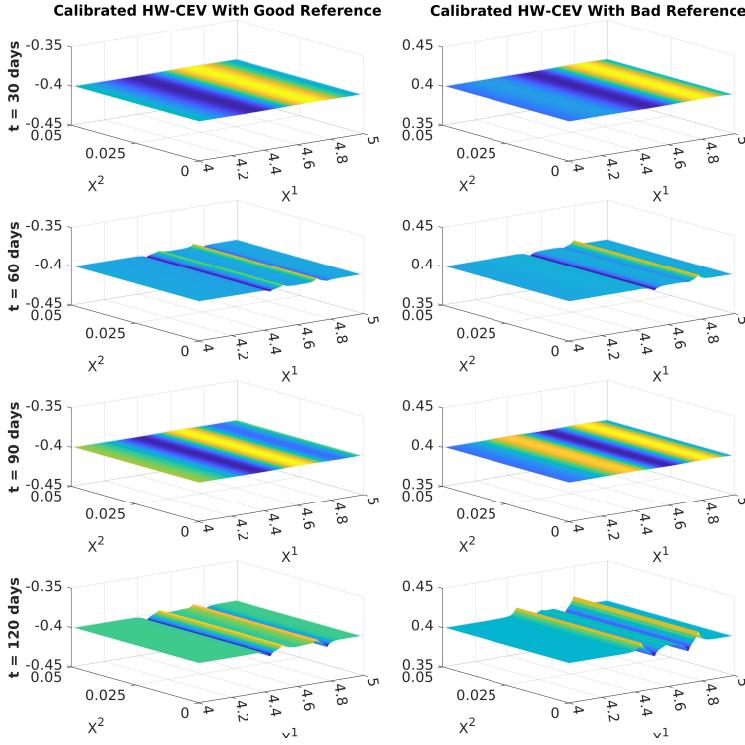
To better illustrate the features of the OT-calibrated model, we present plots of the surfaces of the model characteristics, both on their own – see Figures 5.2 and 5.3 – and superimposed with the characteristics of the generating and reference models for comparison, see Figures 5.4 and 5.5. Broadly speaking, as discussed above, the OT-calibrated model is close to the generating one in the region specified by the data and close to the reference one otherwise. In addition, the surfaces we obtain for the correlation  $\xi$  are highly dependent on the reference model, as one may expect from the convexity adjustment in (4.12). The choice of reference value will therefore mean that  $\xi$  is always close to the reference model in this case with some fluctuations obtained through the change in the value  $\beta_{11}$ . Consequently, the modeller’s (or trader’s) insight in specifying a correct correlation (not least its sign) is, relatively speaking, more

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

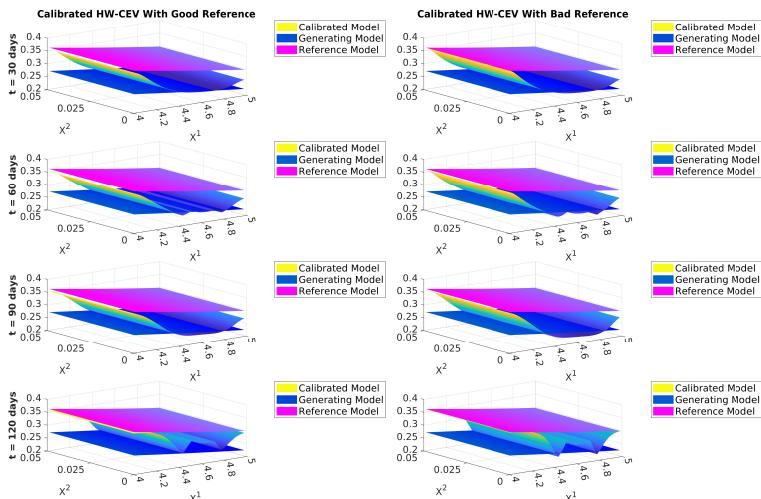
**Table 5.2** Table of the generating and calibrated model prices and implied volatilities.

important as the vanilla option prices' market data alone will not necessarily help to correct a widely wrong reference guess.

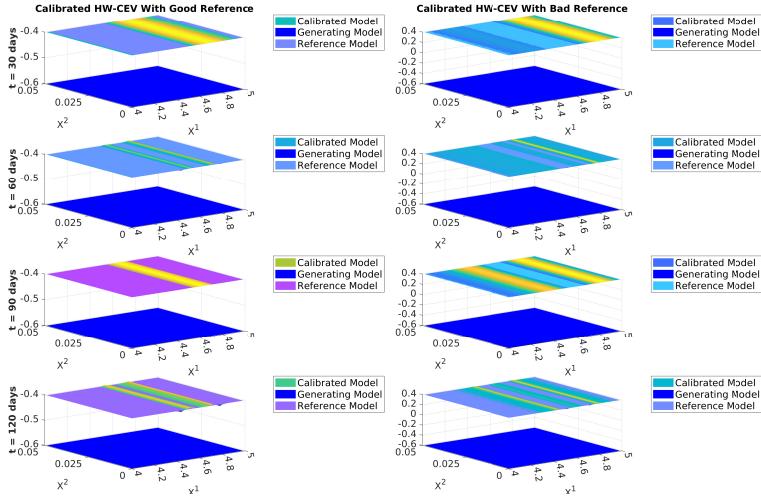
**Fig. 5.2** Plots of  $\beta_{11}$  at  $t = 30, 60, 90$ , and  $120$  days for the calibrated model.



**Fig. 5.3** Plots of  $\xi$  at  $t = 30, 60, 90$ , and  $120$  days for the calibrated model.



**Fig. 5.4** Plots of  $\beta_{11}$  at  $t = 30, 60, 90$ , and  $120$  days for the calibrated model compared with the generating model.



**Fig. 5.5** Plots of  $\xi$  at  $t = 30, 60, 90$ , and  $120$  days for the calibrated model compared with the generating model.

## 6 Conclusions

This paper developed a non-parametric optimal-transport driven method to calibrate stock price model in a stochastic interest rate environment. By switching to sub-probability measures representing discounted densities the method manages not to increase the dimension of the state variables. The method is flexible and can accommodate different calibrating instruments. We developed a generic duality result and applied it in the context of Hull-White short rate model and a local volatility stock price model with short rate dependence. We illustrated the numerical performance of the method considering a sequential calibration problem: the short rate model is calibrated first in a parametric setting and it then feeds into our non-parametric local volatility calibration. We highlighted how some model features are pinned down well via the market prices of options. However others, notably the correlation between the Brownian motions driving the rates and the stock price, are much more sensitive to the modeller's choice of their reference model. In a follow up paper Joseph, Loeper, and Obłój [51], we apply the duality result developed here to consider simultaneous OT-driven calibration for a joint model for the stochastic interest rates and the stock price process, and demonstrate its performance on real market data. We note that here and in Joseph, Loeper, and Obłój [51], we restrict ourselves to short-rate models to keep the dimension of the HJB equation to two state variables. It would be natural and interesting to extend this setup to include more stochastic factors, e.g., stochastic volatility for the stock prices. Each extra factor would raise the dimension of the HJB and imply significant numerical challenge. It would also be of interest to consider more realistic fixed income models, such as market models of Brace,

Gatarek, and Musiela [14], Miltersen, Sandmann, and Sondermann [59], and Jamshidian [50] or more recent multi-curve framework models, see Henrard [41, 40], where the post 2008 financial crisis effects of the LIBOR-OIS spreads are taken into account. This however would drastically increase the dimension of our state process. It would likely necessitate novel tools to solve the resulting PDEs, such as recent machine learning methods, see Han, Jentzen, and Weinan [39] and Weinan, Han, and Jentzen [70]. We believe these are very interesting avenues for future research.

### Competing Interests

The authors declare no competing interests.

### A Proof of the Discounted Superposition Principle

First, we define two operators associated with the discounted Fokker-Planck equation (3.7) and the augmented Fokker-Planck equation (2.3). Given  $a : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{S}_+^d$ ,  $c : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ , define

$$\begin{aligned} C([0, T]; \mathbb{R}^d) &\ni f \mapsto \mathcal{L}^D f = a \cdot \nabla_x f + \frac{1}{2} b : \nabla_x^2 f - c f, \\ C([0, T]; \mathbb{R}^{d+1}) &\ni f \mapsto \mathcal{L}^A f = a \cdot \nabla_x f + \frac{1}{2} b : \nabla_x^2 f + y c \partial_y f. \end{aligned}$$

Note that if  $(\nu_t)_{t \in [0, T]} \subset \mathcal{P}(\mathbb{R}^{d+1})$  solves the augmented Fokker-Planck equation associated with  $\mathcal{L}^A$ , denoted aFP( $a, b, c$ ), then  $(\nu_t^D)_{t \in [0, T]} \subset \mathcal{M}(\mathbb{R}^d)$  with  $\nu_t^D = \int_{\mathbb{R}} y \nu_t(\cdot, dy)$  solves the discounted Fokker-Planck equation associated with  $\mathcal{L}^D$ , denoted dFP( $a, b, c$ ). This is easily seen since the coefficients  $a, b, c$  have no  $y$ -dependence.

We construct a sequence of smooth solutions to (3.7),  $(\nu_t^n)_{t \in [0, T]} \subset \mathcal{M}(\mathbb{R}^d)$ , with coefficients  $a^n \in C^\infty([0, T] \times \mathbb{R}^d; \mathbb{R}^d)$ ,  $b^n \in C^\infty([0, T] \times \mathbb{R}^d; \mathbb{S}_+^d)$ , and  $c^n \in C^\infty([0, T] \times \mathbb{R}^d; \mathbb{R})$  for which we can easily show Theorem 3.5. We therefore can construct a sequence of diffusion operators  $\mathcal{L}^{A,n}$  corresponding to aMP( $a^n, b^n, c^n$ ) which has solution  $\boldsymbol{\eta}^n$  such that  $\int_{\mathbb{R}} y \eta_t^n(\cdot, dy) = \nu_t^n$ . The tightness and convergence of  $\boldsymbol{\eta}^n$  is then shown, and then we use the arguments to generalise the coefficients  $a, b, c$ .

#### A.1 Smooth Approximation of the Augmented Martingale Problem

**Lemma A.1** *Let  $(\nu_t)_{t \in [0, T]} \subset \mathcal{M}(\mathbb{R}^d)$  be a smooth solution of dFP( $a, b, c$ ) with coefficients  $a \in C^\infty([0, T] \times \mathbb{R}^d; \mathbb{R}^d)$ ,  $b \in C^\infty([0, T] \times \mathbb{R}^d; \mathbb{S}_+^d)$ ,  $c \in C^\infty([0, T] \times \mathbb{R}^d; \mathbb{R})$ . Then there exists a solution  $\boldsymbol{\eta} \in \mathcal{P}(C([0, T]; \mathbb{R}^{d+1}))$  to aMP( $a, b, c$ ) such that  $\int_{\mathbb{R}} y \eta_t(\cdot, dy) = \nu_t(\cdot)$ .*

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

$$\begin{cases} d\tilde{X}_t = a_t(\tilde{X}_t) dt + [b_t(\tilde{X}_t)]^{\frac{1}{2}} dW_t^{\tilde{\mathbb{P}}}, & 0 \leq t \leq T, \\ \tilde{X}_0 = x_0. \end{cases} \quad (\text{A.1})$$

Since  $a$  and  $b$  are taken to be smooth, we automatically have existence and uniqueness of (A.1) (see Stroock and Varadhan [66, Corollary 6.3.3]). Define  $\tilde{Y}_t = \exp(-\int_0^t c_s(\tilde{X}_s) ds)$ . Then, by the same reasoning as Theorem 2.2, we have that the curve of marginal laws of the augmented process  $(\tilde{X}, \tilde{Y})$  at  $t$ , given by  $\eta_t = (e_t)_\# \boldsymbol{\eta}$  solves the Fokker-Planck equation for  $(t, x, y) \in [0, T] \times \mathbb{R}^{d+1}$

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

We now define the discounted version of  $\eta_t$  by  $\eta_t^D(x) := \int_{\mathbb{R}} y \eta_t(x, y) dy$ , then by a similar calculation to that in the proof of Lemma 3.6, we have that  $(\eta_t^D)_{t \in [0, T]}$  solves the discounted Fokker-Planck equation (3.7). By a similar argument to Figalli [27, Proposition 4.1], we have uniqueness of (3.7), so  $\eta_t^D = \nu_t$ . By Itô's formula, it is immediate that  $\boldsymbol{\eta} \in \mathcal{P}(C([0, T]; \mathbb{R}^{d+1}))$  solves aMP( $a, b, c$ ), so the result follows.

We let  $(\nu_t)_{t \in [0, T]} \subset \mathcal{M}(\mathbb{R}^d)$  be a narrowly continuous solution of dFP( $a, b, c$ ) for  $a \in L^1(d\nu_t dt; \mathbb{R}^d)$ ,  $b \in L^1(d\nu_t dt; \mathbb{S}^d)$ , and  $c \in L^1(d\nu_t dt; \mathbb{R})$ . We first build a smooth sequence  $\nu_t^n$  approximating  $\nu_t$  for all  $t \in [0, T]$ , associated with diffusion operators  $\mathcal{L}^{D,n}$ . Then by Lemma A.1, there exists a sequence of solutions to the augmented martingale problem,  $\boldsymbol{\eta}^n$ , associated with  $\mathcal{L}^{A,n}$ , such that  $\int_{\mathbb{R}} y \eta_t^n(\cdot, dy) = \nu_t^n$ . As in Trevisan [68, Section A.1], we construct the approximations in two different ways for the proof of the general case.

#### Pushforward via Smooth Maps:

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

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

Note that  $\pi$  preserves uniform bounds on  $a, b, c$ . Additionally, by an application of the chain rule to compute  $\mathcal{L}^D(f \circ \pi)$ , it is easy to see that  $(\pi_\#\nu_t)_{t \in [0, T]} \subset \mathcal{M}(\mathbb{R}^d)$  solves dFP( $\pi(a), \pi(b), \pi(c)$ ), and therefore by Lemma A.1 there exists  $\boldsymbol{\eta}$  a solution to aMP( $\pi(a), \pi(b), \pi(c)$ ) such that  $\pi_\#\nu_t = \int_{\mathbb{R}} y \eta_t(\cdot, dy)$ .

#### Mollification by Convolutions:

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

$$a_t^\kappa := \frac{d[(a_t\nu_t) * \kappa]}{d[\nu_t * \kappa]}, \quad b_t^\kappa := \frac{d[(b_t\nu_t) * \kappa]}{d[\nu_t * \kappa]}, \quad c_t^\kappa := \frac{d[(c_t\nu_t) * \kappa]}{d[\nu_t * \kappa]}$$

By Trevisan [68, Lemma A.1] we have that  $a^\kappa, b^\kappa$ , and  $c^\kappa$  preserve integrability, and have uniformly bounded first and second spatial derivatives on compact

sets. Therefore,  $(\nu_t * \kappa)_{t \in [0, T]} \subset \mathcal{M}(\mathbb{R}^d)$  solves  $\text{dFP}(a^\kappa, b^\kappa, c^\kappa)$ . Additionally, by Lemma A.1 there exists  $\boldsymbol{\eta}$  a solution to  $\text{aMP}(a^\kappa, b^\kappa, c^\kappa)$  such that  $\nu_t * \kappa = \int_{\mathbb{R}} y \eta_t(\cdot, dy)$ .

## A.2 Compactness and Convergence of the Augmented Martingale Problem

We now assume that  $\boldsymbol{\eta}^n$  is a sequence of superposition solutions corresponding to an approximation  $\nu_t^n = \pi_\#^n \nu_t$  in the pushforward case or  $\nu_t^n = \nu_t * \kappa^n$  in the mollification case. In the mollification case, we assume that  $\kappa^n \rightarrow \delta_0$  narrowly as  $n \rightarrow \infty$ . In the pushforward case we assume that  $\pi_n \rightarrow \text{Id}$  locally uniformly, and  $\nabla \pi^n \rightarrow \text{Id}$ , and  $\nabla^2 \pi^n \rightarrow 0$  pointwise. Assume also that the sequences of derivatives are uniformly bounded so there exists  $C \geq 0$  such that  $|\nabla \pi^n|, |\nabla^2 \pi^n| \leq C$ . Compactness for solutions of the augmented martingale problem then follows directly from Trevisan [68, Section A.2] since we could recast the superposition principle in the setting of Trevisan [68, Theorem 2.5] by considering superposition solutions to  $\text{aFP}(a, b, c)$  and then the tightness arguments are directly applicable.

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

$$\int \int_s^t (\mathcal{L}_r^A \tilde{f}) \circ e_r dr d\boldsymbol{\eta} = \int_s^t \int_{\mathbb{R}^d} (\mathcal{L}_r^D f_r) \int_{\mathbb{R}} y \eta_r(dx, dy) dr.$$

From now, we use the shorthand  $d\nu = d\nu_t dt$  for notational ease. Now, let  $\overline{\mathcal{L}^D}$  be a discounted diffusion operator with coefficients  $\bar{a}, \bar{b}, \bar{c}$  continuous and compactly supported. In view of the arguments of Trevisan [68, Section A.3], due to narrow convergence of  $\boldsymbol{\eta}^n$ , we only need to provide bounds on

$$\limsup_{n \rightarrow \infty} \int |\mathcal{L}^{D,n} f - \overline{\mathcal{L}^D} f| d\nu^n + \int |\mathcal{L}^D f - \overline{\mathcal{L}^D} f| d\nu. \quad (\text{A.2})$$

### Pushforward via Smooth Maps:

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

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

Since  $\pi^n$  converges to the identity, we have that (A.2) is bounded above by

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

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

#### Mollification by Convolution:

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

$$\bar{a}^n := \frac{d[(\bar{a}\nu) * \kappa^n]}{d[\nu * \kappa^n]}, \quad \bar{b}^n := \frac{d[(\bar{b}\nu) * \kappa^n]}{d[\nu * \kappa^n]}, \quad \bar{c}^n := \frac{d[(\bar{c}\nu) * \kappa^n]}{d[\nu * \kappa^n]}.$$

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

$$\begin{aligned} \int |\bar{\mathcal{L}}^{D,n} f - \bar{\mathcal{L}}^D f| d\nu^n &\leq \int |\bar{a}^n - \bar{a}| d\nu^n + \int |\bar{b}^n - \bar{b}| d\nu^n + \int |\bar{c}^n - \bar{c}| d\nu^n, \\ &\leq 3 \int \bar{\omega}(z) \kappa^n(z) dz \rightarrow 0. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \int |\mathcal{L}^{D,n} f - \bar{\mathcal{L}}^D f| d\nu^n &= \limsup_{n \rightarrow \infty} \int |\mathcal{L}^{D,n} f - \bar{\mathcal{L}}^{D,n} f| d\nu^n \\ &\leq \int |a - \bar{a}| + |b - \bar{b}| + |c - \bar{c}| d\nu. \end{aligned}$$

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

### A.3 Generalisation of Drift and Diffusion Coefficients

We now use the approximation and convergence described in Section A.1 and Section A.2 to generalise the coefficients  $a, b$  and prove Theorem 3.5.

#### A.3.1 Smooth and Bounded Coefficients:

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

$$\int_0^T \|a_t\|_{C_b^2(\mathbb{R}^d)} + \|b_t\|_{C_b^2(\mathbb{R}^d)} + \|c_t\|_{C_b^2(\mathbb{R}^d)} dt < \infty. \quad (\text{A.3})$$

**Lemma A.2** Let  $a, b, c$  be Borel maps satisfying (A.3). Then for every  $\bar{\nu} \in \mathcal{P}(\mathbb{R}^d)$  there exists  $\eta \in \mathcal{P}(C([0, T]; \mathbb{R}^{d+1}))$  solving aMP( $a, b, c$ ) such that  $\int_{\mathbb{R}} y \eta_0(\cdot, dy) = \bar{\nu}(\cdot)$ .

The proof is the same as Trevisan [68, Theorem A.6], but by considering the augmented process  $(X, Y)$  where the regular martingale problem (see Trevisan [68, Definition 2.4]) is given by aMP( $a, b, c$ ).

**Lemma A.3** Let  $a : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{S}_+^d$ , and  $c : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  be Borel maps such that

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

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

The proof is similar to that of Trevisan [68, Theorem A.7], so we provide a sketch.

*Proof* Let  $g \in C_c^\infty([0, T] \times \mathbb{R}^d)$  be non-negative. Let  $\chi_R : \mathbb{R}^d \rightarrow [0, 1]$  be a cutoff function and let  $a_R^\varepsilon, b_R^\varepsilon, c_R^\varepsilon$  be a mollification with respect to time and space of  $a\chi_R, b\chi_R$ , and  $c\chi_R$ , so that  $a_R^\varepsilon, b_R^\varepsilon, c_R^\varepsilon$  satisfy (A.3). Let  $\mathcal{L}_R^{D, \varepsilon}$  be the diffusion operator associated with dFP( $a_R^\varepsilon, b_R^\varepsilon, c_R^\varepsilon$ ). Let  $f^\varepsilon$  solve

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

Recall the definition of  $\mathcal{L}^S$  in Section A.1 in the construction of the push-forward via smooth maps approximation. Since  $\nu$  solves dFP( $a, b, c$ ),  $f^\varepsilon \in C_b^{1,2}(\mathbb{R}^d)$  and  $f^\varepsilon, \nu_0 \leq 0$ , we have

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

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

### A.3.2 Bounded Coefficients:

This case follows in the same manner as the bounded case in Trevisan [68]. We now assume that  $a, b, c$  satisfy

$$\int_0^T \sup_{x \in \mathbb{R}^d} |a_t(x)| + \sup_{x \in \mathbb{R}^d} |b_t(x)| + \sup_{x \in \mathbb{R}^d} |c_t(x)| dt < \infty. \quad (\text{A.4})$$

Using the regularisation kernel  $\kappa = \exp(-\sqrt{1+|x|^2})$  and  $\kappa^\varepsilon = \varepsilon^{-n}\kappa(x/\varepsilon)$ , we have with  $\nu_t^\varepsilon = \nu_t * \kappa^\varepsilon$  that  $(\nu_t^\varepsilon)_{t \in [0,T]}$  solves  $\text{dFP}(a^\varepsilon, b^\varepsilon, c^\varepsilon)$  with coefficients satisfying (A.3). Thus, existence of  $\boldsymbol{\eta}^\varepsilon \in \mathcal{P}(C([0,T]; \mathbb{R}^{d+1}))$  solving  $\text{aMP}(a^\varepsilon, b^\varepsilon, c^\varepsilon)$  as a discounted superposition solution to  $\nu^\varepsilon$  follows. Since  $\nu_0^\varepsilon$  is a narrowly convergent sequence, there exists  $\theta : \mathbb{R} \rightarrow \mathbb{R}$  increasing with  $\lim_{x \rightarrow \infty} \theta(x) = \infty$  such that  $\sup_{\varepsilon > 0} \int \theta(|x|) d\nu_0^\varepsilon \leq 1$ . The proof of tightness of  $\boldsymbol{\eta}^\varepsilon$  is then an easy modification of that in Trevisan [68], where the de la Vallée Poussin criterion modifies (A.4) to

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

for  $\Theta : [0, \infty) \rightarrow [0, \infty)$  convex non-decreasing with  $\lim_{x \rightarrow \infty} x^{-1}\Theta(x) = \infty$ . Thus, in analogy to Trevisan [68, Corollary 2.5], with  $f^i = (x_i \chi_R) \circ e_t$  for  $i = 1, \dots, d$  and  $f^{d+1} = [\log(y \chi_R)] \circ e_t$ . Note that the function  $f^{d+1}$  is always well-defined and real valued since  $Y$  is assumed to be supported on  $(0, e^T)$ . Then, for the functional  $\Psi : C([0, T]; \mathbb{R}) \rightarrow [0, \infty]$ , we have for  $i = 1, \dots, d$ , with  $x_R^i = x_i \chi_R$ :

$$\begin{aligned} \int \Psi(x_R^i \circ \gamma) d\boldsymbol{\eta}^\varepsilon(\gamma) &\leq \int \theta(|x_R^i|) d\eta_0^\varepsilon + \int_0^T \int [\Theta(|\mathcal{L}_t^{S,\varepsilon} x_R^i|) \\ &\quad + \Theta(b_t^\varepsilon : \nabla x_R^i \otimes \nabla x_R^i)] d\eta_t^\varepsilon dt, \\ \int \Psi(f^{d+1} \circ \gamma) d\boldsymbol{\eta}^\varepsilon(\gamma) &\leq \int \theta(|f^{d+1}|) d\eta_0^\varepsilon + \int_0^T \int \Theta(|c_t^\varepsilon|) d\eta_t^\varepsilon dt \end{aligned}$$

We now note that  $\eta_t^\varepsilon = \nu_t^{A,\varepsilon}$  where  $(\nu_t^{A,\varepsilon})_{t \in [0,T]}$  solves  $\text{aFP}(a^\varepsilon, b^\varepsilon, c^\varepsilon)$ , and further that by construction  $\nu_0^{A,\varepsilon} = \nu_0^\varepsilon$ . Thus, letting  $R \rightarrow \infty$ , noting that the  $y$ -marginal of  $\eta_0^\varepsilon$  is  $\delta_1$ , by the regular superposition principle (Trevisan [68, Theorem 2.5]), we have the uniform bounds:

$$\begin{aligned} \int \Psi(x_R^i \circ \gamma) d\boldsymbol{\eta}^\varepsilon(\gamma) &\leq \int \theta(|x_R^i|) d\nu_0^\varepsilon + \int_0^T \int [\Theta(|\mathcal{L}_t^{S,\varepsilon} x_R^i|) \\ &\quad + \Theta(b_t^\varepsilon : \nabla x_R^i \otimes \nabla x_R^i)] d\nu_t^{A,\varepsilon} dt, \\ \int \Psi(f^{d+1} \circ \gamma) d\boldsymbol{\eta}^\varepsilon(\gamma) &\leq \theta(0) + \int_0^T \int \Theta(|c_t^\varepsilon|) d\nu_t^{A,\varepsilon} dt \end{aligned}$$

Thus, by Trevisan [68, Lemma A.1], we have the uniform bounds

$$\begin{aligned} \int \Psi(\gamma^i) d\eta^\varepsilon(\gamma) &\leq 1 + \int_0^T \Theta\left(\sup_{x \in \mathbb{R}^d} |a_t(x)|\right) + \Theta\left(\sup_{x \in \mathbb{R}^d} |b_t(x)|\right) dt, \\ \int \Psi(\log(\gamma^{d+1})) d\eta^\varepsilon(\gamma) &\leq \theta(0) + \int_0^T \Theta\left(\sup_{x \in \mathbb{R}^d} |c_t(x)|\right) dt. \end{aligned}$$

Since  $\gamma \mapsto \Psi(\log(\gamma^{d+1})) + \sum_{i=1}^d \Psi(\gamma^i)$  is coercive in  $C([0, T]; \mathbb{R}^{d+1})$ , we have tightness of  $\eta^\varepsilon$ . The limit of the approximations obtaining a discounted superposition solution follows directly by the results of Section A.2.

### A.3.3 Locally Bounded Coefficients:

We now assume that  $a, b, c$  satisfy

$$\int_0^T \sup_{x \in B} \{|a_t(x)| + |b_t(x)| + |c_t(x)|\} dt < \infty, \quad \text{for every bounded borel } B \subset \mathbb{R}^d. \quad (\text{A.5})$$

Let  $M \geq 1$  and  $\chi_M$  be a cutoff function. Define  $\pi_M : \mathbb{R}^d \rightarrow \mathbb{R}^d$  by  $\pi_M(x) = x\chi_M(x)$ . Thus, by (A.5), we have  $|\mathcal{L}^D(\pi_M^i)| \leq \|\pi_M^i\|_{C^2} \sup_{|x| \leq 2M} \{|a(x)| + |b(x)| + |x_r c(x)|\}$  and  $|b : \nabla \pi_M^i \otimes \nabla \pi_M^j| \leq \|\pi_M^i\|_{C^1} \sup_{|x| \leq M} |b(x)|$ . Thus with  $\nu_t^M := \pi_\#^M \nu_t$ , we have  $(\nu_t^M)_{t \in [0, T]}$  solves  $\text{DFP}(\pi^M(a), \pi^M(b), \pi^M(c))$  with coefficients satisfying (A.4) so we have  $\eta^M \in \mathcal{P}(C([0, T]; \mathbb{R}^{d+1}))$  solving aMP( $\pi^M(a), \pi^M(b), \pi^M(c)$ ) as a discounted superposition solution to  $\nu^M$ . The argument of tightness is similar to that of the previous section with the modified bound to be

$$\int_0^T \int \Theta_1(|a_t|) + \Theta_2(|b_t|) + \Theta_3(|c_t|) d\nu_t dt < \infty$$

As before, we obtain the following inequalities

$$\begin{aligned} \int \Psi(x_R^i \circ \gamma) d\eta^M(\gamma) &\leq \int \theta(|x_R^i|) d\eta_0^M + \int_0^T \int [\Theta(|\mathcal{L}_t^{S,M} x_R^i|) \\ &\quad + \Theta(b_t^M : \nabla x_R^i \otimes \nabla x_R^i)] d\eta_t^M dt, \\ \int \Psi(f^{d+1} \circ \gamma) d\eta^M(\gamma) &\leq \int \theta(|f^{d+1}|) d\eta_0^M + \int_0^T \int \Theta(|c_t^M|) d\eta_t^M dt. \end{aligned}$$

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

$$\begin{aligned} \int \Psi(\gamma^i) d\eta^M(\gamma) &\leq 1 + \int_0^T \int \Theta_1(|a_t^i|) + \Theta_2(|b_t^{i,i}|) d\nu_t^A dt \\ \int \Psi(\log(\gamma^{d+1})) d\eta^M(\gamma) &\leq \theta(0) + \int_0^T \int \Theta_3(|c_t|) d\nu_t^A dt. \end{aligned}$$

The limit of approximations obtaining a discounted superposition solution follows directly by Section A.2.

#### A.3.4 General Case:

The approximation is achieved via convolution, with the approximating  $\nu^\varepsilon$  having superposition solutions since the approximating coefficients satisfy (A.5). The tightness follows an identical argument, with similar bounds to the bounded coefficient case obtained. The limit of the approximations obtaining a discounted superposition solution follows directly by Section A.2. This concludes the proof of Theorem 3.5.

## References

- [1] Luigi Ambrosio and Alessio Figalli. “On flows associated to Sobolev vector fields in Wiener spaces: an approach à la DiPerna–Lions”. In: *Journal of Functional Analysis* 256.1 (2009), pp. 179–214. DOI: 10.1016/j.jfa.2008.05.007.
- [2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.
- [3] Luigi Ambrosio and Dario Trevisan. “Well-posedness of Lagrangian flows and continuity equations in metric measure spaces”. In: *Analysis & PDE* 7.5 (2014), pp. 1179–1234. DOI: 10.2140/apde.2014.7.1179.
- [4] Marc Atlan. *Localizing volatilities*. 2006. arXiv: math/0604316.
- [5] Marco Avellaneda et al. “Calibrating volatility surfaces via relative-entropy minimization”. In: *Applied Mathematical Finance* 4.1 (1997), pp. 37–64. DOI: 10.1080/135048697334827.
- [6] Julio Backhoff-Veraguas et al. *The Structure of martingale Benamou-Brenier in  $\mathbb{R}^d$* . 2023. arXiv: 2306.11019.
- [7] Alan Bain, Matthieu Mariapragassam, and Christoph Reisinger. “Calibration of local-stochastic and path-dependent volatility models to vanilla and no-touch options”. In: *Journal of Computational Finance* 24.4 (2021). DOI: 10.21314/JCF.2020.400.
- [8] Guy Barles and Panagiotis Souganidis. “Convergence of approximation schemes for fully nonlinear second order equations”. In: *Asymptotic Analysis* 4.3 (1991), pp. 271–283. DOI: 10.3233/ASY-1991-4305.
- [9] Mathias Beiglböck, Tongseok Lim, and Jan Obłój. “Dual attainment for the martingale transport problem”. In: *Bernoulli* 25.3 (2019), pp. 1640–1658. DOI: 10.3150/17-BEJ1015. URL: <https://doi.org/10.3150/17-BEJ1015>.
- [10] Jean-David Benamou and Yann Brenier. “A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem”. In: *Numerische Mathematik* 84.3 (2000), pp. 375–393. DOI: 10.1007/s002119900117.
- [11] Lorenzo Bergomi. *Stochastic volatility modeling*. CRC press, 2015.

- [12] Bruno Bouchard, Grégoire Loeper, and Yiyi Zou. “Hedging of covered options with linear market impact and gamma constraint”. In: *SIAM Journal on Control and Optimization* 55.5 (2017), pp. 3319–3348. DOI: 10.1137/15M1054109.
- [13] George Box and Norman Draper. *Empirical model-building and response surfaces*. John Wiley & Sons, 1987.
- [14] Alan Brace, Dariusz Gątarek, and Marek Musiela. “The market model of interest rate dynamics”. In: *Mathematical Finance* 7.2 (1997), pp. 127–155. DOI: 10.1111/1467-9965.00028.
- [15] Yann Brenier. “A homogenized model for vortex sheets”. In: *Archive for Rational Mechanics and Analysis* 138.4 (1997), pp. 319–353. DOI: 10.1007/s002050050044.
- [16] Yann Brenier. “Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations”. In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 52.4 (1999), pp. 411–452. DOI: 10.1002/(SICI)1097-0312(199904)52:4%3C411::AID-CPA1%3E3.0.CO;2-3.
- [17] Damiano Brigo and Fabio Mercurio. *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Science & Business Media, 2007.
- [18] Gerard Brunick and Steven Shreve. “Mimicking an Itô process by a solution of a stochastic differential equation”. In: *Annals of Applied Probability* 23.4 (2013), pp. 1584–1628. DOI: 10.1214/12-AAP881.
- [19] Lenaic Chizat et al. “Unbalanced optimal transport: Dynamic and Kantorovich formulations”. In: *Journal of Functional Analysis* 274.11 (2018), pp. 3090–3123. DOI: 10.1016/j.jfa.2018.03.008.
- [20] John Cox. “The constant elasticity of variance option pricing model”. In: *Journal of Portfolio Management* (1996), pp. 15–17. DOI: 10.3905/jpm.1996.015.
- [21] Andrei Cozma, Matthieu Mariapragassam, and Christoph Reisinger. “Calibration of a hybrid local-stochastic volatility stochastic rates model with a control variate particle method”. In: *SIAM Journal on Financial Mathematics* 10.1 (2019), pp. 181–213. DOI: 10.1137/17M1114570.
- [22] Michael Crandall, Hitoshi Ishii, and Pierre-Louis Lions. “User’s guide to viscosity solutions of second order partial differential equations”. In: *Bulletin of the American Mathematical Society* 27.1 (1992), pp. 1–67. DOI: 10.1090/S0273-0979-1992-00266-5.
- [23] Griselda Deelstra and Grégory Rayée. “Local volatility pricing models for long-dated FX derivatives”. In: *Applied Mathematical Finance* 20.4 (2013), pp. 380–402. DOI: 10.1080/1350486X.2012.723516.
- [24] Bruno Dupire. “A unified theory of volatility”. In: *Derivatives Pricing: The Classic Collection* (1996), pp. 185–196.
- [25] Bruno Dupire. “Pricing with a smile”. In: *Risk* 7.1 (1994), pp. 18–20.
- [26] Bernd Engelmann, Frank Koster, and Daniel Oeltz. “Calibration of the Heston stochastic local volatility model: A finite volume scheme”. In:

- International Journal of Financial Engineering* 8.01 (2021), p. 2050048. DOI: 10.1142/S2424786320500486.
- [27] Alessio Figalli. “Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients”. In: *Journal of Functional Analysis* 254.1 (2008), pp. 109–153. DOI: 10.1016/j.jfa.2007.09.020.
- [28] Ivan Guo and Grégoire Loeper. “Path dependent optimal transport and model calibration on exotic derivatives”. In: *The Annals of Applied Probability* 31.3 (2021), pp. 1232–1263. DOI: 10.1214/20-AAP1617.
- [29] Ivan Guo, Grégoire Loeper, and Shiyi Wang. “Calibration of local-stochastic volatility models by optimal transport”. In: *Mathematical Finance* 32.1 (2022), pp. 46–77. DOI: 10.1111/mafi.12335.
- [30] Ivan Guo, Grégoire Loeper, and Shiyi Wang. “Local volatility calibration by optimal transport”. In: *2017 MATRIX Annals*. Springer, 2019, pp. 51–64. DOI: 10.1007/978-3-030-04161-8\_5.
- [31] Ivan Guo et al. “Joint Modeling and Calibration of SPX and VIX by Optimal Transport”. In: *SIAM Journal on Financial Mathematics* 13.1 (2022), pp. 1–31. DOI: 10.1137/20M1375905.
- [32] Ivan Guo et al. “Optimal transport for model calibration”. In: *Risk Magazine* (2022).
- [33] Ivan Guo et al. “Portfolio optimization with a prescribed terminal wealth distribution”. In: *Quantitative Finance* 22.2 (2022), pp. 333–347. DOI: 10.1080/14697688.2021.1967432.
- [34] Julien Guyon. “The joint S&P 500/VIX smile calibration puzzle solved”. In: *Risk, April* (2020).
- [35] Julien Guyon and Pierre Henry-Labordère. “Being particular about calibration”. In: *Risk* 25.1 (2012), p. 88.
- [36] Julien Guyon and Pierre Henry-Labordère. *The smile calibration problem solved*. 2011. DOI: 10.2139/ssrn.1885032.
- [37] István Gyöngy. “Mimicking the one-dimensional marginal distributions of processes having an Itô differential”. In: *Probability Theory and Related Fields* 71.4 (1986), pp. 501–516. DOI: 10.1007/BF00699039.
- [38] Patrick Hagan et al. “Managing smile risk”. In: *The Best of Wilmott* 1 (2002), pp. 249–296.
- [39] Jiequn Han, Arnulf Jentzen, and E Weinan. “Solving high-dimensional partial differential equations using deep learning”. In: *Proceedings of the National Academy of Sciences* 115.34 (2018), pp. 8505–8510. DOI: 10.1073/pnas.1718942115.
- [40] Marc Henrard. “The irony in derivatives discounting part II: The crisis”. In: *Wilmott Journal* 2.6 (2010), pp. 301–316. DOI: 10.1002/wilj.39.
- [41] Marc Henrard. “The Irony in the Derivatives Discounting”. In: *Wilmott Magazine* (2007), pp. 92–98.
- [42] Pierre Henry-Labordère. “Calibration of local stochastic volatility models to market smiles: A Monte-Carlo approach”. In: *Risk Magazine, September* (2009).
- [43] Julien Hok and Shih-Hau Tan. “Calibration of local volatility model with stochastic interest rates by efficient numerical PDE methods”. In:

- Decisions in Economics and Finance* 42.2 (2019), pp. 609–637. DOI: 10.1007/s10203-019-00232-3.
- [44] Martin Huesmann and Dario Trevisan. “A Benamou–Brenier formulation of martingale optimal transport”. In: *Bernoulli* 25.4A (2019), pp. 2729–2757. DOI: 10.3150/18-BEJ1069.
- [45] John Hull and Alan White. “A note on the models of Hull and White for pricing options on the term structure: Response”. In: *The Journal of Fixed Income* 5.2 (1995), pp. 97–102. DOI: 10.3905/jfi.1995.408139.
- [46] John Hull and Alan White. “Branching out”. In: *Risk* 7.7 (1994), pp. 34–37.
- [47] John Hull and Alan White. “Pricing interest-rate-derivative securities”. In: *The Review of Financial Studies* 3.4 (1990), pp. 573–592. DOI: 10.1093/rfs/3.4.573.
- [48] KJ In’t Hout and S Foulon. “ADI finite difference schemes for option pricing in the Heston model with correlation”. In: *International Journal of Numerical Analysis & Modeling* 7.2 (2010), pp. 303–320.
- [49] Jean Jacod and Albert Shiryaev. *Limit theorems for stochastic processes*. Vol. 288. Springer Science & Business Media, 2013.
- [50] Farshid Jamshidian. “LIBOR and swap market models and measures”. In: *Finance and Stochastics* 1.4 (1997), pp. 293–330. DOI: 10.1007/s007800050026.
- [51] Benjamin Joseph, Grégoire Loeper, and Jan Obłój. *Joint Calibration of Local Volatility Models with Stochastic Interest Rates using Semimartingale Optimal Transport*. 2023. arXiv: 2308.14473.
- [52] Benjamin Joseph, Grégoire Loeper, and Jan Obłój. *The Measure Preserving Martingale Sinkhorn Algorithm*. 2023. arXiv: 2310.13797.
- [53] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*. Vol. 113. Springer, 2014.
- [54] Nikolai Krylov. “Once more about the connection between elliptic operators and Itô’s stochastic equations”. In: *Statistics and Control of Stochastic Processes, Steklov Seminar*. 1984, pp. 214–229.
- [55] Pierre-Louis Lions. “Optimal control of diffusion processes and Hamilton–Jacobi–Bellman equations part 2: viscosity solutions and uniqueness”. In: *Communications in Partial Differential Equations* 8.11 (1983), pp. 1229–1276. DOI: 10.1080/03605308308820301.
- [56] Dong Liu and Jorge Nocedal. “On the limited memory BFGS method for large scale optimization”. In: *Mathematical Programming* 45.1 (1989), pp. 503–528. DOI: 10.1007/BF01589116.
- [57] Grégoire Loeper. “The reconstruction problem for the Euler–Poisson system in cosmology”. In: *Archive for Rational Mechanics and Analysis* 179.2 (2006), pp. 153–216. DOI: 10.1007/s00205-005-0384-3.
- [58] K Ma and PA Forsyth. “An unconditionally monotone numerical scheme for the two-factor uncertain volatility model”. In: *IMA Journal of Numerical Analysis* 37.2 (2017), pp. 905–944. DOI: 10.1093/imanum/drw025.

- 
- [59] Kristian Miltersen, Klaus Sandmann, and Dieter Sondermann. “Closed form solutions for term structure derivatives with log-normal interest rates”. In: *The Journal of Finance* 52.1 (1997), pp. 409–430. DOI: 10.1111/j.1540-6261.1997.tb03823.x.
  - [60] Yurii Nesterov. *Lectures on convex optimization*. Vol. 137. Springer, 2018.
  - [61] Orcan Ögetbil, Narayan Ganesan, and Bernhard Hientzsch. “Calibrating local volatility models with stochastic drift and diffusion”. In: *International Journal of Theoretical and Applied Finance* 25.02 (2022), p. 2250011. DOI: 10.1142/S021902492250011X.
  - [62] Yong Ren, Dilip Madan, and Michael Qian. “Calibrating and pricing with embedded local volatility models”. In: *Risk Magazine* 20.9 (2007), p. 138.
  - [63] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.
  - [64] Mark Schmidt. “minFunc: unconstrained differentiable multivariate optimization in Matlab”. In: (2005). URL: <http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html>.
  - [65] Thibault Séjourné, Gabriel Peyré, and François-Xavier Vialard. “Unbalanced optimal transport, from theory to numerics”. In: *Handbook of Numerical Analysis* 24 (2023), pp. 407–471. DOI: 10.1016/bs.hna.2022.11.003.
  - [66] Daniel Stroock and Srinivasa Varadhan. *Multidimensional diffusion processes*. Vol. 233. Springer Science & Business Media, 1979.
  - [67] Xiaolu Tan and Nizar Touzi. “Optimal transportation under controlled stochastic dynamics”. In: *The Annals of Probability* 41.5 (2013), pp. 3201–3240. DOI: 10.2307/42919802.
  - [68] Dario Trevisan. “Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients”. In: *Electronic Journal of Probability* 21 (2016). DOI: 10.1214/16-EJP4453.
  - [69] Cédric Villani. *Topics in optimal transportation*. 58. American Mathematical Soc., 2003.
  - [70] E Weinan, Jiequn Han, and Arnulf Jentzen. “Algorithms for solving high dimensional PDEs: from nonlinear Monte Carlo to machine learning”. In: *Nonlinearity* 35.1 (2021), pp. 278–310. DOI: 10.1088/1361-6544/ac337f.
  - [71] Maarten Wyns and Jacques Du Toit. “A finite volume–alternating direction implicit approach for the calibration of stochastic local volatility models”. In: *International Journal of Computer Mathematics* 94.11 (2017), pp. 2239–2267. DOI: 10.1080/00207160.2017.1297805.