THE JOINT S&P 500/VIX SMILE CALIBRATION PUZZLE SOLVED

JULIEN GUYON
QUANTITATIVE RESEARCH, BLOOMBERG L.P.

Abstract. Since VIX options started trading in 2006, many researchers have tried to build a model that jointly and exactly calibrates to the prices of S&P 500 (SPX) options, VIX futures, and VIX options. So far the best attempts, which used parametric continuous-time jump-diffusion models on the SPX, only produced an approximate fit. In this article we solve this longstanding puzzle using a completely different approach: a nonparametric discrete-time model. Given a VIX future maturity T_1, we build a joint probability measure on the SPX at T_1, the VIX at T_1, and the SPX at $T_2 = T_1 + 30$ days which is perfectly calibrated to the SPX smiles at T_1 and T_2, and the VIX future and VIX smile at T_1. Our model satisfies the martingality constraint on the SPX as well as the requirement that the VIX at T_1 is the implied volatility of the 30-day log-contract on the SPX. We prove by duality that the existence of such a model means that the SPX and VIX markets are jointly arbitrage-free.

The joint calibration puzzle is cast as a dispersion-constrained martingale transport problem which is solved using (an extension of) the Sinkhorn algorithm, in the spirit of De March and Henry-Labordère (2019). The algorithm identifies joint SPX/VIX arbitrages should they arise. Our numerical experiments show that the algorithm performs very well in both low and high volatility regimes. Finally we explain how to handle the fact that the VIX future and SPX option monthly maturities do not perfectly coincide, and how to extend the two-maturity model to include all available monthly maturities.

1. Introduction

Volatility indices, such as the VIX index [10], do not only serve as market-implied indicators of volatility. Futures and options on these indices are also widely used as risk-management tools to hedge the volatility exposure of options portfolios. The existence of a liquid market for these futures and options has led to the need for models that jointly calibrate to the prices of options on the underlying asset and the prices of volatility derivatives. Without such models, financial institutions could possibly arbitrage each other, and even market making desks within the same institution could do so, e.g., the VIX desk could arbitrage the S&P 500 (SPX) desk; and by using models that fail to correctly incorporate the prices of the liquid hedging instruments, such as SPX options, VIX futures and VIX options, exotic desks may misprice options, especially (but not only) those with payoffs that involve both the underlying and its volatility index.

In particular, since VIX options started trading in 2006, many researchers and practitioners have tried to build a model that jointly and exactly calibrates to the prices of SPX futures, SPX options, VIX futures, and VIX options. This is known to be a very challenging problem, especially for short maturities. In particular, the very large negative skew of short-term SPX options, which in continuous models implies a very large volatility of volatility, seems inconsistent with the comparatively low levels of VIX implied volatilities. For example the double mean-reverting model of Gatheral [21], though it is very flexible, cannot perfectly fit both the negative at-the-money (ATM) SPX skew—not large enough in absolute value—and the ATM VIX implied volatility—too large—for short maturities. One should decrease the volatility of volatility to decrease the latter, but this would also decrease the former, which is already too small.

Guyon's experiments [27] using very flexible models such as the skewed two-factor Bergomi model [5], the skewed rough Bergomi model, independently introduced by Guyon [26, 28] and De Marco [15], and their stochastic local volatility versions, also suggest that joint calibration could be out of the reach of classical continuous-time models with continuous SPX paths ("continuous models" for short): for short maturities, either the SPX smile is well fitted, but then the model ATM VIX implied volatility is too large; or the VIX smile is well calibrated, but then the model ATM SPX skew is too small (in absolute value). Song and Xiu [39] argued that "the state-of-the-art stochastic volatility models in the literature cannot capture the S&P 500 and VIX option prices simultaneously." Jacquier *et al* [32], who investigated the (unskewed) rough Bergomi model, also noted: "Interestingly, we observe a 20% difference between the [volatility-of-volatility] parameter obtained through VIX calibration and the one obtained through SPX. This suggests that the volatility of volatility in the SPX market is 20% higher when compared to VIX, revealing potential data inconsistencies (arbitrage?)." Note that this difference in implied volatilities of volatility does not indicate the existence of an arbitrage; it simply means that the model under consideration is inconsistent with market data. Even if it could be proved that no continuous model can jointly calibrate to the SPX and VIX smiles, it would not reveal any arbitrage; it would simply mean that although it is very large, this class of models is inconsistent with market data.

In the case of instantaneous VIX, Guyon [25, 28] derived a necessary and sufficient condition for continuous models to jointly calibrate to the SPX and VIX smiles: that the distribution of the Dupire market local variance [18] be smaller than the distribution of the (instantaneous) VIX squared in the convex order, at all times. Guyon [25, 28] also reported that for short maturities the distribution of the (true, i.e., 30-day) VIX squared in the Dupire market local volatility model is actually *larger* than the market-implied distribution of the (true) VIX squared in the convex order. He showed numerically [27, 28] that when the (typically negative) spot-vol correlation is large enough in absolute value, (a) traditional stochastic volatility models with large mean-reversion, and (b) rough volatility models with small Hurst exponent, can reproduce this inversion of convex ordering. The fact that large mean-reversion can generate this inversion of convex ordering is also supported by [29] where an expansion of the volatility of the VIX squared implied by VIX futures at order 5 in small volatility-of-volatility is derived in Bergomi models and compared with the expansion of the SPX ATM skew of [6]. Acciaio and Guyon [1] provide a mathematical proof that the inversion of convex ordering can be produced by continuous models. However, the inversion of convex ordering is only a necessary condition for the joint SPX/VIX calibration of continuous models; it is not sufficient. Other attempts at jointly calibrating with continuous models include Fouque and Saporito [19], but their approach does not apply to short maturities (below 4 months), for which VIX derivatives are most liquid and the joint calibration is most difficult; and Goutte *et al* [22], but the SPX smile used in their calibration tests is erroneous.[1]

Since it looks to be very difficult to jointly calibrate the SPX and VIX smiles with continuous models, many authors have incorporated jumps in the dynamics of the SPX, see, e.g., [37, 11, 36, 2, 34, 35, 3]. Jumps offer extra degrees of freedom to partly decouple the ATM SPX skew and the ATM VIX implied volatility. However so far all the attempts at solving the joint SPX/VIX smile calibration problem only produced imperfect, approximate fits.

In this article we solve this longstanding puzzle using a completely different approach: instead of postulating a parametric continuous-time (jump-)diffusion model on the SPX, we build a nonparametric discrete-time model. Discrete time allows us to easily decouple the ATM SPX skew and the ATM VIX implied volatility; going nonparametric gives us full flexibility for perfect calibration. Given a VIX future maturity T_1, we build a joint probability measure on the SPX at T_1, denoted by S_1, the VIX at T_1, denoted by V, and the SPX at $T_2 = T_1 + 30$ days, denoted by S_2, which is perfectly calibrated to the SPX smiles at T_1 and T_2, and the VIX future and VIX smile at T_1. Our model satisfies the martingality constraint on the SPX as well as the requirement that the VIX at T_1 is the implied volatility of the 30-day log-contract on the SPX. The discrete-time model is cast as a solution to a dispersion-constrained martingale transport problem which is solved using (an extension of) the Sinkhorn algorithm, in the spirit of the recent work by De March and Henry-Labordère [14].

In the first part of the paper, we formulate the superreplication problem for a general payoff $f(S_1, V, S_2)$. We include all vanilla options on S_1, V, and S_2 as (static) hedging instruments, as well as trading (dynamically, i.e., at T_1) in the SPX itself and the forward-starting log-contract. Moreover, we allow the deltas at T_1 to depend on the information available, that is, the SPX and the VIX index at T_1. Using similar arguments as in [4], we show that the price of the cheapest superreplicating portfolio is equal to the largest expectation of the payoff over the set $\mathcal{P}(\mu_1, \mu_V, \mu_2)$ of all the distributions of (S_1, V, S_2) that have the marginals μ_1, μ_V, and μ_2 implied by option prices, and that satisfy the martingality constraint on the SPX as well as the consistency condition that V is the implied volatility at T_1 of the 30-day log-contract on the SPX (see Definition 2). Hence the superreplication problem is dual to a dispersion-constrained martingale optimal transport problem. We use this duality result to prove that in this setting the SPX and VIX markets are jointly arbitrage-free if

[1]Only for the shortest maturity is the reported SPX smile correct; for later maturities the reported SPX smiles have much smaller skews (in absolute value) than the market smiles. The jointly calibrated model is well calibrated to the VIX smiles but actually fails to produce enough ATM SPX skew to calibrate to market smiles, which is consistent with the other numerical experiments reported above.

and only if the set $\mathcal{P}(\mu_1, \mu_V, \mu_2)$ of risk-neutral measures is not empty. We also characterize the absence of joint SPX/VIX arbitrages in terms of the existence of a coupling of μ_1 and μ_V such that two distributions in dimension two are in convex order. A sufficient condition for the existence of a joint SPX/VIX arbitrage is also given in the appendix, that uses a family of functionally generated superreplicating portfolios similar to the one introduced in [24].

In the second part of the paper, we build a jointly calibrating model. This is equivalent to building an element of the set $\mathcal{P}(\mu_1, \mu_V, \mu_2)$. To build such a model, and thus prove that the market is free of joint SPX/VIX arbitrage, we choose a reference model $\bar{\mu}$ and look for the model in $\mathcal{P}(\mu_1, \mu_V, \mu_2)$ that is closest to $\bar{\mu}$ in the entropic sense, see (5.1). By duality, this model is described in a nonparametric way by its Radon-Nikodym derivative w.r.t. $\bar{\mu}$; the Radon-Nikodym derivative is of Gibbs type: it is the exponential of the value of a portfolio made of the same hedging instruments as in the superreplication problem. This portfolio is the solution to a concave maximization problem, see (5.5). To numerically compute this portfolio, we use an extension of the Sinkhorn algorithm [38], which was recently revived by Cuturi [12] in the context of classical (discrete) optimal transport and very recently used by De March and Henry-Labordère [14] to quickly build arbitrage-free smile interpolations. The algorithm iterates the one-dimensional Newton method in different directions to converge to the optimum, which is where the gradient of the functional vanishes.

In the third and last part of the paper, we provide useful implementation details and report the results of our numerical experiments. Instead of working with abstract portfolios, we use combinations of calls and puts that are listed on the market. In this discrete portfolio context, it is apparent that the joint calibration condition is equivalent to the portfolio maximizing a certain concave function, see (6.5). Our numerical experiments show that the algorithm performs very well in both low and high volatility regimes. The smile calibrations are extremely accurate, while the martingality and consistency conditions are perfectly satisfied. We describe the calibrating model in terms of the joint distribution of (S_1, V), the corresponding local VIX function, the distribution of S_2 given S_1 and V, the distribution of the log-return $\ln \frac{S_2}{S_1}$ normalized by the VIX, and the implied volatility of forward-starting options. Finally we explain how to handle the fact that the VIX future and SPX option monthly maturities do not perfectly coincide, and how to extend the two-maturity model to include all available monthly maturities.

The remainder of this article is structured as follows. After introducing our discrete-time setting in Section 2, we describe in Section 3 the primal and dual superreplication problems and show that they yield the same value (absence of a duality gap). We apply this result in Section 4 to characterize the absence of joint SPX/VIX arbitrage. Section 5 describes our strategy to build an arbitrage-free, consistent model that jointly calibrates to the market smiles of the SPX and VIX indices. Section 6 provides implementation details, while we gather our numerical results in Section 7. Finally Section 8 deals with the n-maturity case, and Section 9 concludes. The appendix describes a family of functionally generated portfolios that superreplicate zero and gives sufficient conditions for the existence of a joint SPX/VIX arbitrage.

2. Setting and notation

For simplicity, we assume zero interest rates, repos, and dividends. We consider a VIX future maturity T_1. We take as given the full market smiles of the SPX index S at the two maturities T_1 and $T_2 := T_1 + 30$ days,[2] as well as the full market smile of the VIX index V at T_1. By full market smile we mean the continuum of all call prices $C(K)$ for strikes $K \geq 0$. For $i \in \{1, 2\}$ we denote $S_i := S_{T_i}$. We call *forward-starting log-contract* (FSLC for short) the financial derivative that pays $-\frac{2}{\tau} \ln \frac{S_2}{S_1}$ at T_2, where $\tau := T_2 - T_1 = 30$ days. We recall that, by definition of the VIX (substituting the strip of out-the-money options by the log-contract for simplicity), the price at T_1 of the FSLC is V^2. For convenience, we introduce the strictly convex function

$$L(x) := -\frac{2}{\tau} \ln x.$$

For each maturity T_i, $i \in \{1, 2\}$, absence of static SPX arbitrage (or butterfly arbitrage) is equivalent to the existence of a risk-neutral measure $\mu_i := \partial^2 C_i / \partial K^2$ such that the price of any vanilla option u_i written on S_i is the expectation $\mathbb{E}^i[u_i(S_i)] := \mathbb{E}^{\mu_i}[u_i(S_i)]$ of the payoff under μ_i, and we shall refer to μ_i as a smile as well. (The notation $\partial^2 C_i / \partial K^2$ refers to the second derivative measure of the convex function C_i.) Similarly,

[2]Remark 10 explains how to deal with the fact that listed SPX options do not exactly mature at T_1.

by absence of static VIX arbitrage, there exists a risk-neutral measure $\mu_V = \partial^2 C_V/\partial K^2$ such that the price of any vanilla option u_V written on V is the expectation $\mathbb{E}^V[u_V(V)] := \mathbb{E}^{\mu_V}[u_V(V)]$ of the payoff under μ_V.

Absence of dynamic SPX arbitrage (or calendar arbitrage) is equivalent to μ_1 and μ_2 being in convex order, i.e., $\mathbb{E}^1[f(S_1)] \le \mathbb{E}^2[f(S_2)]$ for any convex function $f : \mathbb{R}_{>0} \to \mathbb{R}$, even if we allow trading in the FSLC at T_1 (see [24, Theorem 3.4]). By absence of arbitrage, the price of S_i at time 0 is the initial value $S_0 > 0$ of the SPX. Moreover, we denote by $F_V \ge 0$ the value at time 0 of the VIX future maturing at T_1. Therefore, throughout the article, (S_1, V, S_2) denotes the identity on $\mathbb{R}_{>0} \times \mathbb{R}_{\ge 0} \times \mathbb{R}_{>0}$; μ_1, μ_2 are two probability measures on $\mathbb{R}_{>0}$ in convex order, with mean S_0; and μ_V is a probability measure on $\mathbb{R}_{\ge 0}$ with mean F_V. For the log-contracts and the VIX squared to have finite prices, we assume the following throughout the paper.

Assumption 1. *The given marginals μ_1, μ_2, μ_V satisfy*

$$\mathbb{E}^i[S_i] = S_0, \quad \mathbb{E}^i[|\ln S_i|] < \infty, \quad i \in \{1,2\}; \qquad \mathbb{E}^V[V] = F_V, \quad \mathbb{E}^V[V^2] < \infty.$$

3. Duality

3.1. **The primal problem.** It is well known that the knowledge of μ_1 and μ_2 gives little information on the prices of options $\mathbb{E}^\mu[g(S_1, S_2)]$, e.g., forward-starting options $\mathbb{E}^\mu[f(S_2/S_1)]$. The many couplings of μ_1 and μ_2 usually lead to a wide variety of prices. Computing the range of these prices is precisely the subject of classical optimal transport. Adding the no-arbitrage constraint that (S_1, S_2) is a martingale leads to tighter bounds, as this provides information on the conditional average of S_2/S_1 given S_1. This problem is then called martingale optimal transport, see [31] and the references therein. In this paper, S being the SPX index, we add VIX market data information to produce even tighter bounds. Indeed this now provides information on the conditional dispersion of S_2/S_1, which is controlled by the VIX V.

We consider a market with two trading dates ($T_0 = 0$ and T_1) where the financial instruments are the SPX (tradable at T_0 and T_1), the vanilla options on it with maturities T_1 and T_2 (tradable at T_0), the VIX future maturing at T_1 and the vanilla options on it also maturing at T_1 (tradable at T_0), as well as the FSLC (tradable at T_1). Note that we consider only static positions in vanilla options, but we allow dynamic trading, that is, trading at T_1, in the SPX and the FSLC. We are interested in deriving the optimal upper bound on the price of a generic payoff $f(S_1, V, S_2)$, given these instruments. The optimal lower bound is obtained by simply taking the negative of the payoff.

For any measurable function $\Delta : \mathbb{R}_{>0} \times \mathbb{R}_{\ge 0} \to \mathbb{R}$ let us introduce the shorthand notation

$$\Delta^{(S)}(s_1, v, s_2) := \Delta(s_1, v)(s_2 - s_1), \qquad \Delta^{(L)}(s_1, v, s_2) := \Delta(s_1, v)\left(L\left(\frac{s_2}{s_1}\right) - v^2\right)$$

for $s_1, s_2 > 0$ and $v \ge 0$. Let $\mathcal{U}$ denote the set of all integrable portfolios, i.e., the set of all measurable functions $u_1, u_2 : \mathbb{R}_{>0} \to \mathbb{R}$, $u_V : \mathbb{R}_{\ge 0} \to \mathbb{R}$, $\Delta_S, \Delta_L : \mathbb{R}_{>0} \times \mathbb{R}_{\ge 0} \to \mathbb{R}$ with $u_1 \in L^1(\mu_1)$, $u_V \in L^1(\mu_V)$, $u_2 \in L^1(\mu_2)$, and Δ_S, Δ_L bounded. Similarly as in De Marco and Henry-Labordère [16] and Guyon *et al* [24], the model-independent no-arbitrage upper bound for the derivative with payoff $f(S_1, V, S_2)$ is the smallest price at time 0 of a superreplicating portfolio,

$$P_f := \inf_{\mathcal{U}_f} \left\{\mathbb{E}^1[u_1(S_1)] + \mathbb{E}^V[u_V(V)] + \mathbb{E}^2[u_2(S_2)]\right\} \tag{3.1}$$

where $\mathcal{U}_f$ is the set of integrable superreplicating portfolios, i.e., that satisfy the superreplication constraint

$$\forall (s_1, s_2, v) \in \mathbb{R}^2_{>0} \times \mathbb{R}_{\ge 0}, \quad u_1(s_1) + u_V(v) + u_2(s_2) + \Delta_S^{(S)}(s_1, v, s_2) + \Delta_L^{(L)}(s_1, v, s_2) \ge f(s_1, v, s_2). \tag{3.2}$$

This linear program is known as the primal problem; P stands for "primal" and v stands for the value of the VIX at the future date T_1. At time T_1, delta-hedging in the SPX and in the FSLC is allowed. The respective deltas, $\Delta_S(s_1, v)$ and $\Delta_L(s_1, v)$, may depend on the values s_1 and v of the SPX and the VIX at T_1. Since the price at T_1 of the FSLC is v^2, the delta strategies are costless, and the price of the portfolio is $\mathbb{E}^1[u_1(S_1)] + \mathbb{E}^V[u_V(V)] + \mathbb{E}^2[u_2(S_2)]$.

3.2. **Duality.** In this section, we introduce the dual problem to superreplicating the payoff $f(S_1, V, S_2)$ and prove the absence of a duality gap as well as the existence of an extremal model.

Definition 2. *Let $\mathcal{P}(\mu_1, \mu_V, \mu_2)$ be the set of all the probability measures μ on $\mathbb{R}_{>0} \times \mathbb{R}_{\ge 0} \times \mathbb{R}_{>0}$ such that*

$$S_1 \sim \mu_1, \quad V \sim \mu_V, \quad S_2 \sim \mu_2, \quad \mathbb{E}^\mu[S_2|S_1, V] = S_1, \quad \mathbb{E}^\mu\left[L\left(\frac{S_2}{S_1}\right)\middle|S_1, V\right] = V^2. \tag{3.3}$$

The dual problem to superreplicating the payoff $f(S_1, V, S_2)$ is one of maximizing the expected payoff $\mathbb{E}^\mu[f(S_1, V, S_2)]$ over all probability measures $\mu \in \mathcal{P}(\mu_1, \mu_V, \mu_2)$:

$$D_f := \sup_{\mu \in \mathcal{P}(\mu_1, \mu_V, \mu_2)} \mathbb{E}^\mu[f(S_1, V, S_2)]. \tag{3.4}$$

We call (3.4) a *dispersion-constrained martingale optimal transport problem.* Indeed, since V is known at T_1, $\mathbb{E}^\mu[S_2|S_1, V] = S_1$ is the martingality condition of the SPX index, while $\mathbb{E}^\mu[L(S_2/S_1)|S_1, V] = V^2$ is the consistency condition, which expresses that the VIX at T_1 is the implied volatility of the 30-day log-contract on the SPX. The first equation imposes a condition on the average of the distribution of S_2 given S_1 and V, while the second equation imposes a condition on its dispersion around the average.

Martingale optimal transport, which was introduced in [4, 30] in discrete time and in [20] in continuous time, corresponds to the case where the VIX market is ignored. Compared to the classical two-period martingale optimal transport, our two-period dispersion-constrained martingale optimal transport problem has one extra variable (V, known at T_1) and one extra constraint (the dispersion constraint, which is controlled by V), and in both constraints the conditional expectations are taken w.r.t. S_1 and V instead of S_1 only. In [4] the authors prove absence of a duality gap for the martingale optimal transport in discrete time after reducing it to classical transport theory by dualizing the martingale constraint, and then using a minimax argument. The exact same technique of proof applies in our case. In particular, the set $\mathcal{P}(\mu_1, \mu_V, \mu_2)$ is compact in the weak topology, and we have

Theorem 3. *Let $f : \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0} \to \mathbb{R}$ be upper semicontinuous and satisfy*

$$|f(s_1, v, s_2)| \leq C\big(1 + s_1 + s_2 + |L(s_1)| + |L(s_2)| + v^2\big) \tag{3.5}$$

for some constant $C > 0$. Then

$$P_f := \inf_{\mathcal{U}_f} \left\{\mathbb{E}^1[u_1(S_1)] + \mathbb{E}^V[u_V(V)] + \mathbb{E}^2[u_2(S_2)]\right\} = \sup_{\mu \in \mathcal{P}(\mu_1, \mu_V, \mu_2)} \mathbb{E}^\mu[f(S_1, V, S_2)] =: D_f.$$

Moreover, $D_f \neq -\infty$ if and only if $\mathcal{P}(\mu_1, \mu_V, \mu_2) \neq \emptyset$, and in that case the supremum is attained.

We omit the proof, as it is a straightforward adaptation of the proof of Theorem 1 in [4].

Remark 4. As in [4], the proof actually shows that the portfolios can also be required to be continuous, without changing the value of P_f. Moreover, the vanilla payoffs u_1, u_V, and u_2 can be chosen to be linear combinations of finitely many call options, together with one position in the bond, one position in S_1, and one position in the VIX future, without changing the value of P_f.

Compared to unconstrained martingale optimal transport, adding VIX market data may possibly reveal a joint SPX/VIX arbitrage. As we explain in the next section, this corresponds to the situation where $\mathcal{P}(\mu_1, \mu_V, \mu_2) = \emptyset$. In the limiting case where $\mathcal{P}(\mu_1, \mu_V, \mu_2) = \{\mu_0\}$ is a singleton, the joint SPX/VIX market data information completely specifies the joint distribution of (S_1, S_2), in particular the price of forward-starting options.

The absence of a duality gap proves useful to characterize the absence of joint SPX/VIX arbitrage, which we introduce in the next section.

4. Joint SPX/VIX arbitrage

In this section we define what we mean by a joint SPX/VIX arbitrage and we characterize absence of joint arbitrage in two different ways. Recall that $\mathcal{U}_0$ denotes the set of integrable portfolios that superreplicate zero.

Definition 5. *An (S_1, S_2, V)-arbitrage is an element of $\mathcal{U}_0$ such that $\mathbb{E}^1[u_1(S_1)] + \mathbb{E}^V[u_V(V)] + \mathbb{E}^2[u_2(S_2)] < 0$.*

Since any such element can be scaled, we observe that there is an (S_1, S_2, V)-arbitrage in the market if and only if

$$P_0 := \inf_{\mathcal{U}_0} \left\{\mathbb{E}^1[u_1(S_1)] + \mathbb{E}^V[u_V(V)] + \mathbb{E}^2[u_2(S_2)]\right\} = -\infty. \tag{4.1}$$

4.1. **Consistent extrapolation of SPX and VIX smiles.** Recall Assumption 1. If $\mathbb{E}^V[V^2] \neq \mathbb{E}^2[L(S_2)] - \mathbb{E}^1[L(S_1)]$, there is a trivial (S_1, S_2, V)-arbitrage. For instance, if $\mathbb{E}^V[V^2] < \mathbb{E}^2[L(S_2)] - \mathbb{E}^1[L(S_1)]$, pick

$$u_1(s_1) = L(s_1), \quad u_2(s_2) = -L(s_2), \quad u_V(v) = v^2, \quad \Delta_S(s_1, v) = 0, \quad \Delta_L(s_1, v) = 1.$$

Therefore, throughout the rest of this article, we assume that

$$\mathbb{E}^V[V^2] = \mathbb{E}^2[L(S_2)] - \mathbb{E}^1[L(S_1)]. \tag{4.2}$$

This means that the price $\mathbb{E}^2[L(S_2)] - \mathbb{E}^1[L(S_1)]$ of the VIX squared inferred from SPX futures and options via the replication formula

$$\mathbb{E}^i[L(S_i)] = L(S_0) + \frac{2}{\tau}\int_0^{S_0} \frac{\mathbb{E}^i[(K - S_i)_+]}{K^2} dK + \frac{2}{\tau}\int_{S_0}^{\infty} \frac{\mathbb{E}^i[(S_i - K)_+]}{K^2} dK \tag{4.3}$$

coincides with the price $\mathbb{E}^V[V^2]$ of the VIX squared inferred from VIX futures and options using the replication formula [5]

$$\mathbb{E}^V[V^2] = F_V^2 + 2\int_0^{F_V} \mathbb{E}^V[(K - V)_+] dK + 2\int_{F_V}^{\infty} \mathbb{E}^V[(V - K)_+] dK. \tag{4.4}$$

Violations of (4.2) in the market have been reported, suggesting arbitrage opportunities, see, e.g., [7, Section 7.7.4]. However, both sides of (4.2) do not purely depend on market data. By (4.4) the l.h.s. depends on an (arbitrage-free) extrapolation of the smile of V beyond the last quoted strikes, while by (4.3) the r.h.s. depends on (arbitrage-free) extrapolations of the SPX smile at maturities T_1 and T_2.[3] The reported violations of (4.2) actually rely on some arbitrary smile extrapolations. Guyon [27] has shown that both sides of (4.2) can even be made arbitrarily large.

The joint SPX/VIX arbitrage opportunities identified in [16, Section 5.1] can also be traced back to a misalignment of $\mathbb{E}^2[L(S_2)] - \mathbb{E}^1[L(S_1)]$, which De Marco and Henry-Labordère denote σ_{12}^2, and $\mathbb{E}^V[V^2]$. In their numerical tests the authors use a VIX future value $F_V = 18.05\%$ that is very close to $\sigma_{12} = 18.15\%$. By the replication formula (4.4), since the authors assume that $\mathbb{E}^V[V^2] = \sigma_{12}^2$, this implies that VIX calls and puts must be cheap—they would need to be worth zero if F_V were equal to σ_{12}. This explains why the upper bound on the VIX implied volatility reported by the authors is so small, below market values. In fact, it is likely that $\mathbb{E}^V[V^2]$ computed from (4.4) be larger than $(18.15\%)^2$, whatever the extrapolation of VIX market implied volatilities used in (4.4). Once a larger value of σ_{12}, consistent with $\mathbb{E}^V[V^2]$, is used, the arbitrage disappears. Stated otherwise, as reported by the authors, their upper bound on VIX implied volatility is very sensitive to σ_{12}, a quantity that is not well defined, is highly dependent on the SPX smile extrapolations, and can actually be made arbitrarily large.

For these reasons it is crucial to extrapolate SPX and VIX smiles in a *consistent* way, so that (4.2) holds. Guyon [27] explained how to do just that, by first inter- and extrapolating the VIX smile, then building an SPX variance swap curve consistent with (4.4), computed for all available VIX options maturities, and finally devising an extrapolation of the SPX smile consistent with those VIX-inferred variance swap prices.[4]

4.2. **Characterization of absence of joint SPX/VIX arbitrage.** The following theorem, which is the main result of this section, states that absence of (S_1, S_2, V)-arbitrage is equivalent to the existence of risk-neutral measures $\mu \in \mathcal{P}(\mu_1, \mu_V, \mu_2)$, and that this is also equivalent to the existence of a coupling of μ_1 and μ_V under which $(S_1, L(S_1) + V^2)$ and $(S_2, L(S_2))$ are in convex order.

Theorem 6. *The following assertions are equivalent:*

(i) *The market is free of (S_1, S_2, V)-arbitrage,*

(ii) $\mathcal{P}(\mu_1, \mu_V, \mu_2) \neq \emptyset$,

(iii) *There exists a coupling ν of μ_1 and μ_V such that $\mathrm{Law}_\nu(S_1, L(S_1) + V^2)$ and $\mathrm{Law}_{\mu_2}(S_2, L(S_2))$ are in convex order, i.e., $\mathbb{E}^\nu[f(S_1, L(S_1)+V^2)] \le \mathbb{E}^2[f(S_2, L(S_2))]$ for any convex function $f : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}$.*

Proof. (i) $\iff$ (ii): By duality (Theorem 3), we have $P_0 = D_0$. Now, by definition, the market is free of (S_1, S_2, V)-arbitrage if and only if $P_0 = 0$, and from Theorem 3, $\mathcal{P}(\mu_1, \mu_V, \mu_2) \neq \emptyset$ if and only if $D_0 \neq -\infty$, in which case $D_0 = 0$.

[3] Both quantities also depend on smile interpolations between quoted strikes. However, the impact of (arbitrage-free) interpolations is much smaller than that of extrapolations.

[4] Those VIX-inferred variance swap prices usually lie within the bid-ask spread of real OTC variance swap quotes.

(ii) $\iff$ (iii): Let us define $M_1 = (S_1, L(S_1) + V^2)$ and $M_2 = (S_2, L(S_2))$ as well as

$$\mu_{M_2}(dx, dy) = \mu_2(dx)\delta_{L(x)}(dy).$$

Let $\Pi(\mu_1, \mu_V)$ denote the set of transport plans from μ_1 to μ_V, i.e., the set of all couplings of μ_1 and μ_V, or the set of all joint distributions of (S_1, V) that have the prescribed marginals μ_1 and μ_V. For $\nu \in \Pi(\mu_1, \mu_V)$, let us denote by $\mu^\nu_{M_1}$ the distribution of M_1 under ν and by $\mathcal{M}(\nu, \mu_2)$ the set of all probability measures μ on $\mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$ such that

$$M_1 \sim \mu^\nu_{M_1}, \qquad M_2 \sim \mu_{M_2}, \qquad \mathbb{E}^\mu [M_2|M_1] = M_1.$$

Then

$$\mathcal{P}(\mu_1, \mu_V, \mu_2) = \bigcup_{\nu \in \Pi(\mu_1, \mu_V)} \mathcal{M}(\nu, \mu_2).$$

By Strassen's theorem [40], each $\mathcal{M}(\nu, \mu_2)$ is nonempty if and only if $\mu^\nu_{M_1}$ and μ_{M_2} are in convex order, which yields the equivalence of (ii) and (iii). □

In the appendix we derive a sufficient condition for the existence of an (S_1, S_2, V)-arbitrage based on a family of functionally generated portfolios that superreplicate zero which is similar to the one introduced in [24].

In order to check the absence of joint SPX/VIX arbitrage, directly solving the linear problem (4.1) associated to Assertion (i) of Theorem 6 is numerically doable but not easy as one needs to try all possible $(u_1, u_V, u_2, \Delta_S, \Delta_V)$ and check the superreplication constraints for all $s_1, s_2 > 0$ and $v \geq 0$. Condition (iii) looks more handy, but numerically checking it is actually difficult as, in dimension two, the extreme rays of the convex cone of convex functions are dense in the cone [33], unlike in dimension one where the extreme rays are the call and put payoffs [8]. Instead, we will verify absence of (S_1, S_2, V)-arbitrage by building—numerically, but with high accuracy—an element of $\mathcal{P}(\mu_1, \mu_V, \mu_2)$, thus checking (ii).

5. Building a model in $\mathcal{P}(\mu_1, \mu_V, \mu_2)$

In this section we explain how to numerically build a model $\mu \in \mathcal{P}(\mu_1, \mu_V, \mu_2)$. We thus solve a longstanding puzzle in derivatives modeling: build an arbitrage-free model that jointly calibrates to the prices of SPX futures, SPX options, VIX futures, and VIX options.

Our strategy, inspired by the recent work of De March and Henry-Labordère [14], is the following. We assume that $\mathcal{P}(\mu_1, \mu_V, \mu_2) \neq \emptyset$ and try to build an element μ in this set. To this end, we fix a reference probability measure $\bar{\mu}$ on $\mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$ and look for the measure $\mu \in \mathcal{P}(\mu_1, \mu_V, \mu_2)$ that minimizes the relative entropy $H(\mu, \bar{\mu})$ of μ w.r.t. $\bar{\mu}$, also known as the Kullback-Leibler divergence:[5]

$$D_{\bar{\mu}} := \inf_{\mu \in \mathcal{P}(\mu_1, \mu_V, \mu_2)} H(\mu, \bar{\mu}), \qquad H(\mu, \bar{\mu}) := \begin{cases} \mathbb{E}^\mu \left[\ln \frac{d\mu}{d\bar{\mu}}\right] = \mathbb{E}^{\bar{\mu}} \left[\frac{d\mu}{d\bar{\mu}} \ln \frac{d\mu}{d\bar{\mu}}\right] & \text{if } \mu \ll \bar{\mu}, \\ +\infty & \text{otherwise.} \end{cases} \tag{5.1}$$

This is a strictly convex problem that can be solved after dualization using Sinkhorn's fixed point iteration [38]. Note that $D_{\bar{\mu}} \neq +\infty$ if and only if there exists $\mu \ll \bar{\mu}$ in $\mathcal{P}(\mu_1, \mu_V, \mu_2)$, and in that case the infimum defining $D_{\bar{\mu}}$ is attained.[6] Indeed, from the proof of Theorem D.13 in [17], $\mu \mapsto H(\mu, \bar{\mu})$ is lower semicontinuous in the weak topology. Since $\mathcal{P}(\mu_1, \mu_V, \mu_2)$ is compact in this topology, the infimum is attained.

Let $\mathcal{M}_1$ denote the set of probability measures on $\mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$. Introducing the Lagrange multipliers $u = (u_1, u_V, u_2, \Delta_S, \Delta_L) \in \mathcal{U}$ associated to the five constraints (3.3), and assuming that the inf and sup

[5] We recall that $\mu \ll \bar{\mu}$ if and only if for every event A, $\bar{\mu}(A) = 0 \implies \mu(A) = 0$.

[6] Note that the choice of the support of $\bar{\mu}$ matters. In particular, $\mathcal{P}(\mu_1, \mu_V, \mu_2) \cap \{\mu \in \mathcal{M}_1 | \mu \ll \bar{\mu}\}$ may be empty even when $\mathcal{P}(\mu_1, \mu_V, \mu_2) \neq \emptyset$. In practice we choose $\bar{\mu}$ having full support $\mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$.

operators can be swapped (absence of a duality gap), we have

$$D_{\bar\mu} = \inf_{\mu\in\mathcal{M}_1}\sup_{u\in\mathcal{U}}\Big\{H(\mu,\bar\mu)+\mathbb{E}^1[u_1(S_1)]+\mathbb{E}^V[u_V(V)]+\mathbb{E}^2[u_2(S_2)] -\mathbb{E}^\mu\Big[u_1(S_1)+u_V(V)+u_2(S_2)+\Delta_S^{(S)}(s_1,v,s_2)+\Delta_L^{(L)}(s_1,v,s_2)\Big]\Big\} \tag{5.2}$$

$$= \sup_{u\in\mathcal{U}}\inf_{\mu\in\mathcal{M}_1}\Big\{H(\mu,\bar\mu)+\mathbb{E}^1[u_1(S_1)]+\mathbb{E}^V[u_V(V)]+\mathbb{E}^2[u_2(S_2)] -\mathbb{E}^\mu\Big[u_1(S_1)+u_V(V)+u_2(S_2)+\Delta_S^{(S)}(s_1,v,s_2)+\Delta_L^{(L)}(s_1,v,s_2)\Big]\Big\}. \tag{5.3}$$

Now, the inner infimum can be exactly computed. For any random variable X such that $\mathbb{E}^{\bar\mu}[e^X]<+\infty$, let us denote by $\bar\mu_X$ the probability distribution defined by $\frac{d\bar\mu_X}{d\bar\mu}=\frac{e^X}{\mathbb{E}^{\bar\mu}[e^X]}$. We have

$$\inf_{\mu\in\mathcal{M}_1}\{H(\mu,\bar\mu)-\mathbb{E}^\mu[X]\} = \inf_{\mu\in\mathcal{M}_1}\mathbb{E}^\mu\left[\ln\frac{d\mu}{d\bar\mu}-X\right] = \inf_{\mu\in\mathcal{M}_1}\mathbb{E}^\mu\left[\ln\frac{d\mu}{d\bar\mu_X}+\ln\frac{d\bar\mu_X}{d\bar\mu}-X\right]$$
$$= \inf_{\mu\in\mathcal{M}_1}\mathbb{E}^\mu\left[\ln\frac{d\mu}{d\bar\mu_X}-\ln\mathbb{E}^{\bar\mu}[e^X]\right] = \inf_{\mu\in\mathcal{M}_1}H(\mu,\bar\mu_X)-\ln\mathbb{E}^{\bar\mu}[e^X] = -\ln\mathbb{E}^{\bar\mu}[e^X]$$

and the infimum is attained at $\mu=\bar\mu_X$ since for all $\mu\in\mathcal{M}_1$, $H(\mu,\bar\mu_X)\ge 0$ and $H(\mu,\bar\mu_X)=0$ if and only if $\mu=\bar\mu_X$. As a consequence,

$$D_{\bar\mu}=\sup_{u\in\mathcal{U}}\Psi_{\bar\mu}(u)=:P_{\bar\mu} \tag{5.4}$$

where for $u=(u_1,u_V,u_2,\Delta_S,\Delta_L)\in\mathcal{U}$, we have defined

$$\Psi_{\bar\mu}(u):=\mathbb{E}^1[u_1(S_1)]+\mathbb{E}^V[u_V(V)]+\mathbb{E}^2[u_2(S_2)]-\ln\mathbb{E}^{\bar\mu}\left[e^{u_1(S_1)+u_V(V)+u_2(S_2)+\Delta_S^{(S)}(S_1,V,S_2)+\Delta_L^{(L)}(S_1,V,S_2)}\right]. \tag{5.5}$$

If the supremum defining $P_{\bar\mu}$ is attained at $u^*=(u_1^*,u_V^*,u_2^*,\Delta_S^*,\Delta_L^*)$, then the infimum defining $D_{\bar\mu}$ is reached at

$$\mu^*(ds_1,dv,ds_2)=\bar\mu(ds_1,dv,ds_2)\frac{e^{u_1^*(s_1)+u_V^*(v)+u_2^*(s_2)+\Delta_S^{*(S)}(s_1,v,s_2)+\Delta_L^{*(L)}(s_1,v,s_2)}}{\mathbb{E}^{\bar\mu}\left[e^{u_1^*(S_1)+u_V^*(V)+u_2^*(S_2)+\Delta_S^{*(S)}(S_1,V,S_2)+\Delta_L^{*(L)}(S_1,V,S_2)}\right]}. \tag{5.6}$$

Our jointly calibrating model μ^* is built from $\bar\mu$ using a change of measure of Gibbs type: the Radon-Nikodym derivative is the exponential of the value of a portfolio made of the available hedging instruments, normalized. Note that the normalization factor can always be taken equal to 1 by adjusting the cash component of the portfolio.[7] As a consequence we will always work with a normalized version of $u^*\in\mathcal{U}$ such that

$$\mathbb{E}^{\bar\mu}\left[e^{u_1^*(S_1)+u_V^*(V)+u_2^*(S_2)+\Delta_S^{*(S)}(S_1,V,S_2)+\Delta_L^{*(L)}(S_1,V,S_2)}\right]=1. \tag{5.7}$$

The initial, difficult problem (5.1) of minimizing $H(\mu,\bar\mu)$ over μ in the constrained set $\mathcal{P}(\mu_1,\mu_V,\mu_2)$ has been reduced to the simpler problem (5.4) of maximizing the strictly concave function $\Psi_{\bar\mu}$ over u in the *unconstrained* set $\mathcal{U}$. If it exists, the optimum u^* simply cancels the gradient of $\Psi_{\bar\mu}$; the equations $\frac{\partial\Psi_{\bar\mu}}{\partial u_1(s_1)}=\frac{\partial\Psi_{\bar\mu}}{\partial u_V(v)}=\frac{\partial\Psi_{\bar\mu}}{\partial u_2(s_2)}=\frac{\partial\Psi_{\bar\mu}}{\partial\Delta_S(s_1,v)}=\frac{\partial\Psi_{\bar\mu}}{\partial\Delta_L(s_1,v)}=0$ read[8]

$$\begin{array}{rrcl}
\forall s_1>0, & u_1(s_1) & = & \Phi_1(s_1;u_V,u_2,\Delta_S,\Delta_L)\\
\forall v\ge 0, & u_V(v) & = & \Phi_V(v;u_1,u_2,\Delta_S,\Delta_L)\\
\forall s_2>0, & u_2(s_2) & = & \Phi_2(s_2;u_1,u_V,\Delta_S,\Delta_L)\\
\forall s_1>0,\ \forall v\ge 0, & 0 & = & \Phi_{\Delta_S}(s_1,v;\Delta_S(s_1,v),\Delta_L(s_1,v))\\
\forall s_1>0,\ \forall v\ge 0, & 0 & = & \Phi_{\Delta_L}(s_1,v;\Delta_S(s_1,v),\Delta_L(s_1,v))
\end{array} \tag{5.8}$$

[7]Note that $\Psi_{\bar\mu}$ is invariant by translation of u_1, u_V, and u_2: for any constant $c\in\mathbb{R}$, $\Psi_{\bar\mu}(u_1+c,u_V,u_2,\Delta_S,\Delta_L)=\Psi_{\bar\mu}(u_1,u_V,u_2,\Delta_S,\Delta_L)$ (and similarly with u_V and u_2); c corresponds to a cash position.

[8]Here for simplicity we assume that the distributions $\mu_1,\mu_V,\mu_2,\bar\mu$ are either discrete or continuous w.r.t. the Lebesgue measure. For instance, $\mu_1(s_1)$ denotes $\mu_1(\{s_1\})$ if μ_1 is discrete, or the density at point s_1 if μ_1 is continuous.

where, imposing the normalization (5.7),

$$\begin{aligned}
\Phi_1(s_1; u_V, \Delta_S, \Delta_L) &:= \ln\mu_1(s_1) - \ln\left(\int \bar{\mu}(s_1, dv, ds_2) e^{u_V(v)+u_2(s_2)+\Delta_S^{(S)}(s_1,v,s_2)+\Delta_L^{(L)}(s_1,v,s_2)}\right) \\
\Phi_V(v; u_1, \Delta_S, \Delta_L) &:= \ln\mu_V(v) - \ln\left(\int \bar{\mu}(ds_1, v, ds_2) e^{u_1(s_1)+u_2(s_2)+\Delta_S^{(S)}(s_1,v,s_2)+\Delta_L^{(L)}(s_1,v,s_2)}\right) \\
\Phi_2(s_2; u_1, u_V, \Delta_S, \Delta_L) &:= \ln\mu_2(s_2) - \ln\left(\int \bar{\mu}(ds_1, dv, s_2) e^{u_1(s_1)+u_V(v)+\Delta_S^{(S)}(s_1,v,s_2)+\Delta_L^{(L)}(s_1,v,s_2)}\right) \\
\Phi_{\Delta_S}(s_1, v; u_2, \delta_S, \delta_L) &:= \int \bar{\mu}(s_1, v, ds_2)(s_2 - s_1) e^{u_2(s_2)+\delta_S(s_2-s_1)+\delta_L\left(L\left(\frac{s_2}{s_1}\right)-v^2\right)} \\
\Phi_{\Delta_L}(s_1, v; u_2, \delta_S, \delta_L) &:= \int \bar{\mu}(s_1, v, ds_2)\left(L\left(\frac{s_2}{s_1}\right) - v^2\right) e^{u_2(s_2)+\delta_S(s_2-s_1)+\delta_L\left(L\left(\frac{s_2}{s_1}\right)-v^2\right)}.
\end{aligned}$$

Remark 7. Note that these are also the equations satisfied by the maximum of

$$\bar{\Psi}_{\bar{\mu}}(u) := \mathbb{E}^1[u_1(S_1)] + \mathbb{E}^V[u_V(V)] + \mathbb{E}^2[u_2(S_2)] - \mathbb{E}^{\bar{\mu}}\left[e^{u_1(S_1)+u_V(V)+u_2(S_2)+\Delta_S^{(S)}(S_1,V,S_2)+\Delta_L^{(L)}(S_1,V,S_2)}\right]$$

where, compared to $\Psi_{\bar{\mu}}(u)$, the logarithm has been removed. One could directly get that $D_{\bar{\mu}} = \sup_{u\in\mathcal{U}} \bar{\Psi}_{\bar{\mu}}(u)$ by using the set $\mathcal{M}_+$ of nonnegative measures instead of $\mathcal{M}_1$ in (5.2), and by computing the inner $\inf_{\mu\in\mathcal{M}_+}$ in (5.3) by differentiating w.r.t. $\frac{d\mu}{d\bar{\mu}}$. This is for instance the route followed by [13, 14]. In any case, the jointly calibrating model reads

$$\mu^*(ds_1, dv, ds_2) = \bar{\mu}(ds_1, dv, ds_2) e^{u_1^*(s_1)+u_V^*(v)+u_2^*(s_2)+\Delta_S^{*(S)}(s_1,v,s_2)+\Delta_L^{*(L)}(s_1,v,s_2)} \tag{5.9}$$

where $u^* = (u_1^*, u_V^*, u_2^*, \Delta_S^*, \Delta_L^*)$ is solution to (5.8).

Remark 8. Note that we could have simply postulated a model of the form (5.9); then the five conditions (3.3) translate into the five equations (5.8). The relative entropy argument explains how to naturally derive this particular form for μ. The fact that the five equations (5.8) correspond to the five constraints (3.3) is not surprising, as those equations correspond to canceling the gradient of the Lagrangian function with respect to the Lagrange multipliers $u_1, u_V, u_2, \Delta_S, \Delta_L$.

Sinkhorn's algorithm [38] was first used in the context of computational optimal transport by Cuturi [12]. It has recently been extended to computational martingale optimal transport by De March [13] and applied to the problem of quickly building arbitrage-free smiles by De March and Henry-Labordère [14]. (Alternative computational methods for martingale optimal transport problems are investigated in Guo and Obłój [23].) In our context, Sinkhorn's algorithm is an exponentially fast fixed point method that iterates computations of one-dimensional gradients to approximate the optimizer u^*. Starting from an initial guess $u^{(0)} = (u_1^{(0)}, u_V^{(0)}, u_2^{(0)}, \Delta_S^{(0)}, \Delta_L^{(0)})$, we recursively define $u^{(n+1)}$ knowing $u^{(n)}$ by

$$\begin{aligned}
\forall s_1 > 0, \quad & u_1^{(n+1)}(s_1) = \Phi_1(s_1; u_V^{(n)}, u_2^{(n)}, \Delta_S^{(n)}, \Delta_L^{(n)}) \\
\forall v \geq 0, \quad & u_V^{(n+1)}(v) = \Phi_V(v; u_1^{(n+1)}, u_2^{(n)}, \Delta_S^{(n)}, \Delta_L^{(n)}) \\
\forall s_2 > 0, \quad & u_2^{(n+1)}(s_2) = \Phi_2(s_2; u_1^{(n+1)}, u_V^{(n+1)}, \Delta_S^{(n)}, \Delta_L^{(n)}) \\
\forall s_1 > 0,\ \forall v \geq 0, \quad & 0 = \Phi_{\Delta_S}(s_1, v; u_2^{(n+1)}, \Delta_S^{(n+1)}(s_1, v), \Delta_L^{(n)}(s_1, v)) \\
\forall s_1 > 0,\ \forall v \geq 0, \quad & 0 = \Phi_{\Delta_L}(s_1, v; u_2^{(n+1)}, \Delta_S^{(n+1)}(s_1, v), \Delta_L^{(n+1)}(s_1, v))
\end{aligned}$$

until convergence. Note that each of the above five lines corresponds to a Bregman projection [9] in the space of measures. That is, with obvious notations, $\mu[\bar{\mu}; u_1^{(n+1)}, u_V^{(n)}, u_2^{(n)}, \Delta_S^{(n)}, \Delta_L^{(n)}]$ exactly satisfies the first constraint $S_1 \sim \mu_1$ (but not the other constraints), $\mu[\bar{\mu}; u_1^{(n+1)}, u_V^{(n+1)}, u_2^{(n)}, \Delta_S^{(n)}, \Delta_L^{(n)}]$ exactly satisfies the second constraint $V \sim \mu_V$ (but not the other constraints), and so on. When the algorithm converges, the limit $\mu^* = \mu[\bar{\mu}; u_1^*, u_V^*, u_2^*, \Delta_S^*, \Delta_L^*]$ satisfies all the constraints (3.3) at once.

If Sinkhorn's algorithm diverges, then $P_{\bar{\mu}} = +\infty$, so $D_{\bar{\mu}} = +\infty$, which means $\mathcal{P}(\mu_1, \mu_V, \mu_2) \cap \{\mu \in \mathcal{M}_1 | \mu \ll \bar{\mu}\} = \emptyset$. In practice, when $\bar{\mu}$ has full support, this is a sign that there likely exists a joint SPX/VIX arbitrage, which can then be identified by numerically solving the linear problem (4.1). We have never experienced this situation in our numerical tests, which covered both low and high volatility regimes.

6. Implementation details

A natural choice of a reference measure $\bar{\mu}$ is one that satisfies all the constraints (3.3) except $S_2 \sim \mu_2$, i.e., pick $\bar{\mu}$ in the set $\mathcal{P}(\mu_1, \mu_V)$ of all the probability distributions

$$\mu(ds_1, dv, ds_2) = \nu(ds_1, dv)\, T(s_1, v, ds_2) \tag{6.1}$$

where ν is a coupling of μ_1 and μ_V and the transition kernel $T(s_1, v, ds_2)$ satisfies

$$\int s_2\, T(s_1, v, ds_2) = s_1, \qquad \int L(s_2)\, T(s_1, v, ds_2) = L(s_1) + v^2 \tag{6.2}$$

for μ_1-a.e. $s_1 > 0$ and μ_V-a.e. $v \geq 0$. For instance, we may choose

$$\nu = \mu_1 \otimes \mu_V, \qquad T(s_1, v, ds_2) \text{ is the distribution of } s_1 \exp\left(v\sqrt{\tau}G - \frac{1}{2}v^2\tau\right), \tag{6.3}$$

where G denotes a standard Gaussian random variable, i.e., under $\bar{\mu}$, S_1 and V are independent, and given S_1 and V, S_2 is lognormal with mean S_1 and annualized volatility V.[9] Note that if we choose $\bar{\mu}$ as in (6.1) and (6.3) then Φ_1, Φ_V, Φ_{Δ_S}, and Φ_{Δ_L} under (5.8) can be replaced by

$$\begin{aligned}
\Phi_1(s_1; u_V, u_2, \Delta_S, \Delta_L) &:= -\ln\left(\int \mu_V(dv) T(s_1, v, ds_2) e^{u_V(v) + u_2(s_2) + \Delta_S^{(S)}(s_1,v,s_2) + \Delta_L^{(L)}(s_1,v,s_2)}\right) \\
\Phi_V(v; u_1, u_2, \Delta_S, \Delta_L) &:= -\ln\left(\int \mu_1(ds_1) T(s_1, v, ds_2) e^{u_1(s_1) + u_2(s_2) + \Delta_S^{(S)}(s_1,v,s_2) + \Delta_L^{(L)}(s_1,v,s_2)}\right) \\
\Phi_{\Delta_S}(s_1, v; u_2, \delta_S, \delta_L) &:= \int T(s_1, v, ds_2)(s_2 - s_1) e^{u_2(s_2) + \delta_S(s_2 - s_1) + \delta_L\left(L\left(\frac{s_2}{s_1}\right) - v^2\right)} \\
\Phi_{\Delta_L}(s_1, v; u_2, \delta_S, \delta_L) &:= \int T(s_1, v, ds_2)\left(L\left(\frac{s_2}{s_1}\right) - v^2\right) e^{u_2(s_2) + \delta_S(s_2 - s_1) + \delta_L\left(L\left(\frac{s_2}{s_1}\right) - v^2\right)}.
\end{aligned}$$

Practically, we consider vanilla payoffs u_1, u_V, and u_2 that are linear combinations of finitely many call options, together with one position in the bond, one position in S_1, and one position in the VIX future. That is, we consider market strikes $\mathcal{K} := (\mathcal{K}_1, \mathcal{K}_V, \mathcal{K}_2)$ and market prices (C_K^1, C_K^V, C_K^2) of vanilla options on S_1, V, and S_2 respectively, and we build the model

$$\begin{aligned}
&\mu^*_{\mathcal{K}}(ds_1, dv, ds_2) = \bar{\mu}(ds_1, dv, ds_2) \\
&\quad e^{c^* + \Delta_S^{0*} s_1 + \Delta_V^{0*} v + \sum_{K \in \mathcal{K}_1} a_K^{1*}(s_1 - K)_+ + \sum_{K \in \mathcal{K}_V} a_K^{V*}(v - K)_+ + \sum_{K \in \mathcal{K}_2} a_K^{2*}(s_2 - K)_+ + \Delta_S^{*(S)}(s_1,v,s_2) + \Delta_L^{*(L)}(s_1,v,s_2)}
\end{aligned}$$

where $\theta^* := (c^*, \Delta_S^{0*}, \Delta_V^{0*}, a^{1*}, a^{V*}, a^{2*}, \Delta_S^*, \Delta_L^*)$ maximizes

$$\begin{aligned}
&\bar{\Psi}_{\bar{\mu},\mathcal{K}}(\theta) := c + \Delta_S^0 S_0 + \Delta_V^0 F_V + \sum_{K \in \mathcal{K}_1} a_K^1 C_K^1 + \sum_{K \in \mathcal{K}_V} a_K^V C_K^V + \sum_{K \in \mathcal{K}_2} a_K^2 C_K^2 \\
&- \mathbb{E}^{\bar{\mu}}\left[e^{c + \Delta_S^0 S_1 + \Delta_V^0 V + \sum_{K \in \mathcal{K}_1} a_K^1 (S_1 - K)_+ + \sum_{K \in \mathcal{K}_V} a_K^V (V - K)_+ + \sum_{K \in \mathcal{K}_2} a_K^2 (S_2 - K)_+ + \Delta_S^{(S)}(S_1,V,S_2) + \Delta_L^{(L)}(S_1,V,S_2)}\right]
\end{aligned}$$

over the set Θ of portfolios $\theta := (c, \Delta_S^0, \Delta_V^0, a^1, a^V, a^2, \Delta_S, \Delta_L)$ such that $c, \Delta_S^0, \Delta_V^0 \in \mathbb{R}$, $a^1 \in \mathbb{R}^{\mathcal{K}_1}$, $a^V \in \mathbb{R}^{\mathcal{K}_V}$, $a^2 \in \mathbb{R}^{\mathcal{K}_2}$, and $\Delta_S, \Delta_L : \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ are bounded measurable functions of (s_1, v). This corresponds to solving the entropy minimization problem

$$D_{\bar{\mu},\mathcal{K}} := \inf_{\mu \in \mathcal{P}(\mathcal{K})} H(\mu, \bar{\mu}) = \sup_{\theta \in \Theta} \bar{\Psi}_{\bar{\mu},\mathcal{K}}(\theta) =: P_{\bar{\mu},\mathcal{K}}$$

where $\mathcal{P}(\mathcal{K})$ denotes the set of probability measures μ on $\mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$ such that

$$\begin{aligned}
&\mathbb{E}^\mu[S_1] = S_0, \quad \mathbb{E}^\mu[V] = F_V, \quad \forall K \in \mathcal{K}_1,\ \mathbb{E}^\mu\left[(S_1 - K)_+\right] = C_K^1, \quad \forall K \in \mathcal{K}_V,\ \mathbb{E}^\mu\left[(V - K)_+\right] = C_K^V, \\
&\forall K \in \mathcal{K}_2,\ \mathbb{E}^\mu\left[(S_2 - K)_+\right] = C_K^2, \quad \mathbb{E}^\mu\left[S_2 | S_1, V\right] = S_1, \quad \mathbb{E}^\mu\left[L\left(\frac{S_2}{S_1}\right)\middle| S_1, V\right] = V^2.
\end{aligned} \tag{6.4}$$

[9]Note that with this choice of $\bar{\mu}$, the denominator in (5.6) is infinite since, conditionally on (S_1, V), S_2 is lognormal. Despite this lack of integrability, our numerical experiments ran well. Choices of $\bar{\mu}$ guaranteeing integrability include for instance Gaussian distributions conditioned to be positive for S_1 and V, as well as for S_2 given (S_1, V).

One can directly check that the model $\mu^*_{\mathcal{K}}$ is a consistent arbitrage-free model that jointly calibrates the prices of SPX futures, SPX options, VIX futures, and VIX options. Indeed, if $\bar{\Psi}_{\bar{\mu},\mathcal{K}}$ reaches its maximum at θ^*, then θ^* is solution to $\frac{\partial \bar{\Psi}_{\bar{\mu},\mathcal{K}}}{\partial \theta_i}(\theta) = 0$, i.e.,

$$
\begin{aligned}
\frac{\partial \bar{\Psi}_{\bar{\mu},\mathcal{K}}}{\partial c} = 0: &\qquad \mathbb{E}^{\bar{\mu}}\left[\frac{d\mu^*_{\mathcal{K}}}{d\bar{\mu}}\right] = 1 \\
\frac{\partial \bar{\Psi}_{\bar{\mu},\mathcal{K}}}{\partial \Delta^0_S} = 0: &\qquad \mathbb{E}^{\bar{\mu}}\left[S_1\frac{d\mu^*_{\mathcal{K}}}{d\bar{\mu}}\right] = S_0 \\
\frac{\partial \bar{\Psi}_{\bar{\mu},\mathcal{K}}}{\partial \Delta^0_V} = 0: &\qquad \mathbb{E}^{\bar{\mu}}\left[V\frac{d\mu^*_{\mathcal{K}}}{d\bar{\mu}}\right] = F_V \\
\frac{\partial \bar{\Psi}_{\bar{\mu},\mathcal{K}}}{\partial a^1_K} = 0: &\qquad \mathbb{E}^{\bar{\mu}}\left[(S_1-K)_+\frac{d\mu^*_{\mathcal{K}}}{d\bar{\mu}}\right] = C^1_K \\
\frac{\partial \bar{\Psi}_{\bar{\mu},\mathcal{K}}}{\partial a^V_K} = 0: &\qquad \mathbb{E}^{\bar{\mu}}\left[(V-K)_+\frac{d\mu^*_{\mathcal{K}}}{d\bar{\mu}}\right] = C^V_K \\
\frac{\partial \bar{\Psi}_{\bar{\mu},\mathcal{K}}}{\partial a^2_K} = 0: &\qquad \mathbb{E}^{\bar{\mu}}\left[(S_2-K)_+\frac{d\mu^*_{\mathcal{K}}}{d\bar{\mu}}\right] = C^2_K \\
\frac{\partial \bar{\Psi}_{\bar{\mu},\mathcal{K}}}{\partial \Delta_S(s_1,v)} = 0: &\qquad \mathbb{E}^{\bar{\mu}}\left[(S_2-S_1)\frac{d\mu^*_{\mathcal{K}}}{d\bar{\mu}}\middle| S_1=s_1, V=v\right] = 0, \quad \forall s_1 \geq 0, v>0 \\
\frac{\partial \bar{\Psi}_{\bar{\mu},\mathcal{K}}}{\partial \Delta_L(s_1,v)} = 0: &\qquad \mathbb{E}^{\bar{\mu}}\left[\left(L\left(\frac{S_2}{S_1}\right)-V^2\right)\frac{d\mu^*_{\mathcal{K}}}{d\bar{\mu}}\middle| S_1=s_1, V=v\right] = 0, \quad \forall s_1 \geq 0, v>0.
\end{aligned} \tag{6.5}
$$

The first equation states that $\mu^*_{\mathcal{K}}$ is a probability measure, while the next seven equations state that it satisfies the seven constraints (6.4), so $\mu^*_{\mathcal{K}} \in \mathcal{P}(\mathcal{K})$. This is not surprising: in Lagrangian relaxation, canceling the gradient of the Lagrangian function with respect to the Lagrange multipliers is equivalent to enforcing the constraints.

The corresponding Sinkhorn iterations read

$$
\begin{aligned}
& c^{(n+1)} &=& \ \Phi_c(\Delta_S^{0,(n)}, \Delta_V^{0,(n)}, a^{1,(n)}, a^{V,(n)}, a^{2,(n)}, \Delta_S^{(n)}, \Delta_L^{(n)}) \\
& S_0 &=& \ \Phi_{\Delta^0_S}(c^{(n+1)}, \Delta_S^{0,(n+1)}, \Delta_V^{0,(n)}, a^{1,(n)}, a^{V,(n)}, a^{2,(n)}, \Delta_S^{(n)}, \Delta_L^{(n)}) \\
& F_V &=& \ \Phi_{\Delta^0_V}(c^{(n+1)}, \Delta_S^{0,(n+1)}, \Delta_V^{0,(n+1)}, a^{1,(n)}, a^{V,(n)}, a^{2,(n)}, \Delta_S^{(n)}, \Delta_L^{(n)}) \\
\forall K_i \in \mathcal{K}_1, \quad & C^1_{K_i} &=& \ \Phi_1(K_i; c^{(n+1)}, \Delta_S^{0,(n+1)}, \Delta_V^{0,(n+1)}, a^{1,(n,i)}, a^{V,(n)}, a^{2,(n)}, \Delta_S^{(n)}, \Delta_L^{(n)}) \\
\forall K_i \in \mathcal{K}_V, \quad & C^V_{K_i} &=& \ \Phi_V(K_i; c^{(n+1)}, \Delta_S^{0,(n+1)}, \Delta_V^{0,(n+1)}, a^{1,(n+1)}, a^{V,(n,i)}, a^{2,(n)}, \Delta_S^{(n)}, \Delta_L^{(n)}) \\
\forall K_i \in \mathcal{K}_2, \quad & C^2_{K_i} &=& \ \Phi_2(K_i; c^{(n+1)}, \Delta_S^{0,(n+1)}, \Delta_V^{0,(n+1)}, a^{1,(n+1)}, a^{V,(n+1)}, a^{2,(n,i)}, \Delta_S^{(n)}, \Delta_L^{(n)}) \\
\forall s_1 > 0, \ \forall v \geq 0, \quad & 0 &=& \ \Phi_{\Delta_S}(s_1, v; a^{2,(n+1)}, \Delta_S^{(n+1)}(s_1,v), \Delta_L^{(n)}(s_1,v)) \\
\forall s_1 > 0, \ \forall v \geq 0, \quad & 0 &=& \ \Phi_{\Delta_L}(s_1, v; a^{2,(n+1)}, \Delta_S^{(n+1)}(s_1,v), \Delta_L^{(n+1)}(s_1,v))
\end{aligned}
$$

where, for $\varepsilon \in \{1, V, 2\}$,

$$
a^{\varepsilon,(n,i)} := \left(a_{K_1}^{\varepsilon,(n+1)}, \ldots, a_{K_i}^{\varepsilon,(n+1)}, a_{K_{i+1}}^{\varepsilon,(n)}, \ldots, a_{K_{|\mathcal{K}_\varepsilon|}}^{\varepsilon,(n)}\right)
$$

with

$$\begin{aligned}
\Phi_c(\Delta_S^0, \Delta_V^0, a^1, a^V, a^2, \Delta_S, \Delta_L) &:= -\ln \int \bar{\mu}(ds_1, dv, ds_2)\, e_\theta^{-c} \\
\Phi_{\Delta_S^0}(c, \Delta_S^0, \Delta_V^0, a^1, a^V, a^2, \Delta_S, \Delta_L) &:= \int \bar{\mu}(ds_1, dv, ds_2)\, s_1 e_\theta \\
\Phi_{\Delta_V^0}(c, \Delta_S^0, \Delta_V^0, a^1, a^V, a^2, \Delta_S, \Delta_L) &:= \int \bar{\mu}(ds_1, dv, ds_2)\, v e_\theta \\
\Phi_1(K; c, \Delta_S^0, \Delta_V^0, a^1, a^V, a^2, \Delta_S, \Delta_L) &:= \int \bar{\mu}(ds_1, dv, ds_2)\, (s_1 - K)_+ e_\theta \\
\Phi_V(K; c, \Delta_S^0, \Delta_V^0, a^1, a^V, a^2, \Delta_S, \Delta_L) &:= \int \bar{\mu}(ds_1, dv, ds_2)\, (v - K)_+ e_\theta \\
\Phi_2(K; c, \Delta_S^0, \Delta_V^0, a^1, a^V, a^2, \Delta_S, \Delta_L) &:= \int \bar{\mu}(ds_1, dv, ds_2)\, (s_2 - K)_+ e_\theta \\
\Phi_{\Delta_S}(s_1, v; a^2, \delta_S, \delta_L) &:= \int \bar{\mu}(s_1, v, ds_2)\, (s_2 - s_1) e^{\sum_{K\in\mathcal{K}_2} a_K^2 (s_2-K)_+ + \delta_S(s_2-s_1) + \delta_L \left(L\left(\frac{s_2}{s_1}\right) - v^2\right)} \\
\Phi_{\Delta_L}(s_1, v; a^2, \delta_S, \delta_L) &:= \int \bar{\mu}(s_1, v, ds_2) \left(L\left(\frac{s_2}{s_1}\right) - v^2 \right) e^{\sum_{K\in\mathcal{K}_2} a_K^2 (s_2-K)_+ + \delta_S(s_2-s_1) + \delta_L \left(L\left(\frac{s_2}{s_1}\right) - v^2\right)}
\end{aligned}$$

where e_θ is a shorthand notation for

$$e_\theta := e^{c + \Delta_S^0 s_1 + \Delta_V^0 v + \sum_{K\in\mathcal{K}_1} a_K^1 (s_1-K)_+ + \sum_{K\in\mathcal{K}_V} a_K^V (v-K)_+ + \sum_{K\in\mathcal{K}_2} a_K^2 (s_2-K)_+ + \Delta_S^{(S)}(s_1,v,s_2) + \Delta_L^{(L)}(s_1,v,s_2)}$$

and e_θ^{-c} is the same expression without the c term.

We use $\theta^{(0)} = 0$ as the starting point of the Sinkhorn algorithm, and we numerically checked that the result does not depend on the initial guess, which is in line with [14, Theorem 4.5]. The above integrals are estimated using Gaussian quadrature; we use Gauss-Legendre quadrature when we integrate over s_1 and v, and Gauss-Hermite quadrature when we integrate over s_2. While the expression for $c^{(n+1)}$ is explicit, computing the other parameters requires using a one-dimensional root solver; we use Newton's algorithm. As an exception, for each point s_1 and v in the quadrature grid, $(\Delta_S^{(n+1)}(s_1, v), \Delta_L^{(n+1)}(s_1, v))$ are actually jointly computed using the Levenberg-Marquardt algorithm. Numerical convergence is typically reached after about a hundred iterations (a few minutes using Python code on a standard personal computer—Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz processor and 16GB of RAM), which provides a very accurate estimate of θ^*, hence $\mu_{\mathcal{K}}^*$.

If the Sinkhorn algorithm diverges, then $P_{\bar{\mu},\mathcal{K}} = +\infty$, so $D_{\bar{\mu},\mathcal{K}} = +\infty$, which means that $\mathcal{P}(\mathcal{K}) \cap \{\mu \in \mathcal{M}_1 | \mu \ll \bar{\mu}\} = \emptyset$. In practice, when $\bar{\mu}$ has full support, this is a sign that there likely exists a joint SPX/VIX arbitrage based only on $\mathcal{K}$. In such a case, a joint SPX/VIX arbitrage portfolio can be identified by numerically solving the linear problem (whose value is then $-\infty$)

$$\inf_{\Theta_0} \left\{ c + \Delta_S^0 S_0 + \Delta_V^0 F_V + \sum_{K\in\mathcal{K}_1} a_K^1 C_K^1 + \sum_{K\in\mathcal{K}_V} a_K^V C_K^V + \sum_{K\in\mathcal{K}_2} a_K^2 C_K^2 \right\}$$

where Θ_0 is the set of portfolios $\theta = (c, \Delta_S^0, \Delta_V^0, a^1, a^V, a^2, \Delta_S, \Delta_L)$ superreplicating zero, i.e., such that

$$c + \Delta_S^0 s_1 + \Delta_V^0 v + \sum_{K\in\mathcal{K}_1} a_K^1 (s_1 - K)_+ + \sum_{K\in\mathcal{K}_V} a_K^V (v-K)_+ + \sum_{K\in\mathcal{K}_2} a_K^2 (s_2 - K)_+ + \Delta_S^{(S)}(s_1, v, s_2) + \Delta_L^{(L)}(s_1, v, s_2) \geq 0.$$

Remark 9. For the sake of clarity we have so far assumed zero rates, repo, and dividends. Of course in practice, we use the market values F_1 and F_2 of the SPX futures maturing at T_1 and T_2. The first and last two conditions in (6.4) then read

$$\mathbb{E}^\mu[S_1] = F_1, \qquad \mathbb{E}^\mu[S_2 | S_1, V] = F_{12} S_1, \qquad \mathbb{E}^\mu \left[L\left(\frac{S_2}{F_{12} S_1}\right) \middle| S_1, V \right] = V^2,$$

where $F_{12} := F_2/F_1$, and the C_K are understood as undiscounted call prices.

Remark 10. One practical issue is that there are no listed SPX options maturing at T_1. Indeed, T_1 is the Wednesday that is exactly 30 days before the third Friday of the following month $m+1$, therefore the monthly SPX options maturity T_1' of month m is either two days after or five days before T_1. The rigorous treatment of this annoying fact requires that we introduce a fourth random variable S_1' representing the value of the SPX index at time T_1'. If T_1' is two days after T_1, we consider the primal portfolios

$$u_1(s_1') + u_V(v) + u_2(s_2) + \Delta_S(s_1,v)(s_1'-s_1) + \Delta_S'(s_1,v,s_1')(s_2-s_1') + \Delta_L(s_1,v)\left(L\left(\frac{s_2}{s_1}\right) - v^2\right)$$

and the corresponding dual risk-neutral probability measures $\mathcal{P}'(\mu_1,\mu_V,\mu_2)$ on $\mathbb{R}_{>0}\times\mathbb{R}_{\geq 0}\times\mathbb{R}_{>0}^2$ such that

$$V\sim\mu_V,\quad S_1'\sim\mu_1,\quad S_2\sim\mu_2,\quad \mathbb{E}^\mu\left[S_1'|S_1,V\right]=S_1,\quad \mathbb{E}^\mu\left[S_2|S_1,V,S_1'\right]=S_1',\quad \mathbb{E}^\mu\left[L\left(\frac{S_2}{S_1}\right)\middle|S_1,V\right]=V^2,$$

where (S_1,V,S_1',S_2) denotes the identity on $\mathbb{R}_{>0}\times\mathbb{R}_{\geq 0}\times\mathbb{R}_{>0}^2$. If T_1' is five days before T_1, the primal portfolios are

$$u_1(s_1') + u_V(v) + u_2(s_2) + \Delta_S'(s_1')(s_1-s_1') + \Delta_S(s_1',s_1,v)(s_2-s_1) + \Delta_L(s_1',s_1,v)\left(L\left(\frac{s_2}{s_1}\right) - v^2\right)$$

and the corresponding dual risk-neutral probability measures satisfy

$$S_1'\sim\mu_1,\quad V\sim\mu_V,\quad S_2\sim\mu_2,\quad \mathbb{E}^\mu\left[S_1|S_1'\right]=S_1',\quad \mathbb{E}^\mu\left[S_2|S_1',S_1,V\right]=S_1,\quad \mathbb{E}^\mu\left[L\left(\frac{S_2}{S_1}\right)\middle|S_1',S_1,V\right]=V^2.$$

This adds one dimension to the Gaussian quadratures and slows the Sinkhorn algorithm. Instead, we assume that SPX options mature exactly at T_1 and we define the SPX smile at T_1 by interpolating the total implied variance in maturity.

7. Numerical experiments

Let us first test our algorithm on SPX and VIX market data as of August 1, 2018. We first build the model $\mu^*_{\mathcal{K}}$ when $T_1 = 21$ days, the closest monthly VIX future maturity—usually the maturity for which joint calibration is most difficult. Figure 9.1 compares the model and market smiles of S_1, V, and S_2. The fit is remarkably accurate. Figure 9.1 also shows that the model is arbitrage-free: the martingale condition $\mathbb{E}^\mu[S_2|S_1,V]=F_{12}S_1$ and the consistency condition $\mathbb{E}^\mu[L(S_2/(F_{12}S_1))|S_1,V]=V^2$ are perfectly satisfied.

Figure 9.2 explores some properties of the model $\mu^*_{\mathcal{K}}$. We plot the joint distribution of (S_1,V) under $\mu^*_{\mathcal{K}}$, as well as the local VIX function $\mathrm{VIX}_{\mathrm{loc}}(s_1)$, which we define as

$$\mathrm{VIX}^2_{\mathrm{loc}}(S_1) := \mathbb{E}^{\mu^*_{\mathcal{K}}}\left[V^2\middle|S_1\right].$$

The local VIX has the usual pattern of local volatility, decreasing except for large s_1. We also plot the conditional distribution of S_2 given (s_1,v) under $\mu^*_{\mathcal{K}}$ for different values of (s_1,v).[10] This shows that for s_1 close to the money the conditional distribution has negative skewness, i.e., a fat tail to the left, whereas for large values of s_1 it has positive skewness, i.e., a fat tail to the right. For small values of s_1, it is close to the lognormal distribution $T(s_1,v,ds_2)$ in (6.1) for small v, but diverges from it for large v. We also graph the distribution of the normalized return

$$R := \frac{\ln\frac{S_2}{F_{12}S_1}}{V\sqrt{\tau}} + \frac{1}{2}V\sqrt{\tau}$$

under $\mu^*_{\mathcal{K}}$. This return is exactly Gaussian under $\bar\mu$, but has negative skewness under $\mu^*_{\mathcal{K}}$. Finally we also plot the smile of implied volatilities of forward starting call options $\left(\frac{S_2}{F_{12}S_1}-K\right)_+$ in the model $\mu^*_{\mathcal{K}}$. The forward starting smile is mostly V-shaped, with a minimum lying slightly out of the money. This is consistent with how the skewness of S_2 given s_1 depends on s_1. The model could also be used to price payoffs mixing SPX and VIX values, such as a "VIX Sharped Forward Starting SPX Call" $(S_2/S_1-K)_+/V$, a "VIX Sharped Forward Starting SPX Forward" $(S_2/S_1)/V$, the normalized return R, or options on them. Most importantly, our jointly calibrating model $\mu^*_{\mathcal{K}}$ looks very "standard"; it shows no extreme or pathological feature. The SPX and VIX smiles can be jointly calibrated with a model that looks familiar and reasonable.

In Figure 9.3 we report the optimal payoffs $u_1^*(s_1)=\sum_{K\in\mathcal{K}_1}a_K^{1*}(s_1-K)_+$, $u_V^*(v)=\sum_{K\in\mathcal{K}_V}a_K^{V*}(v-K)_+$, and $u_2^*(s_2)=\sum_{K\in\mathcal{K}_2}a_K^{2*}(s_2-K)_+$, as well as the optimal deltas $\Delta_S^*(s_1,v)$ and $\Delta_L^*(s_1,v)$. Note that u_1^* and u_2^* look like (smoothed) call options, and the surfaces Δ_S^* and Δ_L^* have a similar shape.

[10] The piecewise continuous shapes of the distributions are the result of working with portfolios made of finitely many call options.

Figures 9.4, 9.5 and 9.6 are the analogous of Figures 9.1, 9.2 and 9.3 for $T_1 = 49$ days, the second closest monthly VIX future maturity. Figures 9.7, 9.8 and 9.9 correspond to the calibration results (for $T_1 = 23$ days) as of December 24, 2018, when the VIX reached a high value of 36.07% after a 10-day period where the SPX index fell 11.3%. These figures show that our algorithm works equally well in a high volatility environment. Note that the payoffs u_1^*, u_V^*, u_2^* as well as the surfaces Δ_S^* and Δ_L^* have a similar shape for the two calibration dates reported here (low and high volatility regimes).

8. Extension to the multi-maturity case

So far we have considered only two maturities T_1 and $T_2 = T_1 + 30$ days. In this section we explain how to extend the model to include more market maturities.

To that end, for simplicity, we ignore the SPX/VIX maturity issue raised in Remark 10: we assume that monthly SPX options and VIX futures maturities T_i perfectly coincide and, for two consecutive months, are separated by exactly 30 days, $T_{i+1} - T_i = \tau$ for all $i \geq 1$. Assume that for each month i we are able to build a jointly calibrating model ν_i following the algorithm described in Section 6; here ν_i denotes the joint distribution of (S_i, V_i, S_{i+1}) where S_i and V_i denote the SPX and VIX values at T_i. Then we can build a calibrated model on $(S_i, V_i)_{i\geq 1}$ as follows: $(S_1, V_1, S_2) \sim \nu_1$; recursively we define the distribution of (V_{i+1}, S_{i+2}) given $(S_1, V_1, S_2, V_2, \ldots, S_i, V_i, S_{i+1})$ as the conditional distribution of (V_{i+1}, S_{i+2}) given S_{i+1} under ν_{i+1}. It is easy to check that the resulting model ν is arbitrage-free, consistent, and calibrated to all the SPX and VIX monthly market smiles μ_{S_i} and μ_{V_i}: for all $i \geq 1$,

$$S_i \sim \mu_{S_i}, \quad V_i \sim \mu_{V_i}, \quad \mathbb{E}^{\nu}\left[S_{i+1} | (S_j, V_j)_{1\leq j\leq i}\right] = S_i, \quad \mathbb{E}^{\nu}\left[L\left(\frac{S_{i+1}}{S_i}\right) \middle| (S_j, V_j)_{1\leq j\leq i}\right] = V_i^2.$$

Unfortunately one cannot rigorously deal with the exact monthly maturities of SPX options and VIX options by solving a sequence of independent monthly joint calibration problems and then gluing the solutions together as above, because of the possible overlap of $[T_i, T'_{i+1}]$ and $[T_{i+1}, T'_{i+2}]$ (using the notations of Remark 10). For instance, if $T_2 < T'_2$, then the influence of V_2 on S'_2 cannot be captured by the first-month model, whose variables do not include V_2. The rigorous treatment requires that our methodology be directly applied to the full vector $(V_i, S_i, S'_i)_{i\geq 1}$, which is doable in principle but impractical.

9. Conclusion

In this article we have for the first time built an arbitrage-free model that is perfectly consistent with market data on SPX futures, SPX options, VIX futures, and VIX options. As a consequence, we have proved that the market is free of joint SPX/VIX arbitrage. If joint arbitrages were to appear, our algorithm would detect and identify them.

Our model is nonparametric, providing full flexibility to perfectly calibrate the market smiles, and is formulated in discrete time, so as to easily decouple the ATM SPX skew and the ATM VIX implied volatility. It is specified by a joint probability distribution of the SPX at T_1, the VIX at T_1, and the SPX at $T_2 = T_1$ + 30 days, which is described in a nonparametric way by its Radon-Nikodym derivative w.r.t. a reference measure. The Radon-Nikodym derivative, which has a Gibbs structure, is itself expressed in terms of a portfolio made of the available hedging instruments: SPX futures and options, VIX futures and options, and dynamic trading in the SPX and the forward-starting log-contract. This portfolio, which maximizes a concave functional, is numerically computed using an extension of the Sinkhorn algorithm. The model satisfies the martingality constraint on the SPX as well as the consistency condition that the VIX at T_1 is the implied volatility of the 30-day log-contract on the SPX. It can in principle be extended to include all available market maturities.

Our numerical tests prove the efficiency of the algorithm in various volatility regimes. Our jointly calibrating model has familiar features and does not look weird or pathological in any way. Whether there exists a continuous-time model consistent with our three-dimensional probability measure, and how our model depends on the reference measure, are natural questions which are important for practical risk management purposes, and that we leave for further research.

Acknowledgements. We would like to thank Pierre Henry-Labordère for interesting discussions.

References

[1] Acciaio, B., Guyon, J.: *Inversion of convex ordering: Local volatility does not maximize the price of VIX futures*, SIAM J. Finan. Math., 11(1):SC1–SC13, 2020.

[2] Baldeaux, J, Badran, A.: *Consistent Modelling of VIX and Equity Derivatives Using a 3/2 plus Jumps Model*, Appl. Math. Finance, 21(4):299–312, 2014.

[3] Bardgett, C., Gourier, E., Leippold, M.: *Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets*, Journ. Financial Economics, 131(3):593–618, 2019.

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

[5] Bergomi, L.: *Smile Dynamics III*, Risk, March 2008.

[6] Bergomi, L., Guyon, J.: *Stochastic volatility's orderly smiles*, Risk, May, 2012.

[7] Bergomi, L.: *Stochastic Volatility Modeling*, Chapman & Hall/CRC Financial Mathematics Series, 2016.

[8] Blaschke, W, Pick, G.: *Distanzschätzungen im Funktionenraum II*, Math. Ann., 77:277–302, 1916.

[9] Bregman, L.M.: *A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming*, Ž. Vyčisl. Mat. i Mat. Fiz., 7:620–631, 1967.

[10] The CBOE volatility index-VIX, www.cboe.com/micro/vix/vixwhite.pdf (accessed on August 18, 2017).

[11] Cont, R., Kokholm, T.: *A consistent pricing model for index options and volatility derivatives*, Math. Finance, 23(2):248–274, 2013.

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

[13] De March, A.: *Entropic approximation for multi-dimensional martingale optimal transport*, preprint arXiv:1812.11104, 2018.

[14] De March, A., Henry-Labordère, P.: *Building arbitrage-free implied volatility: Sinkhorn's algorithm and variants*, SSRN preprint 3326486, 2019.

[15] De Marco, S.: *Volatility derivatives in (rough) forward variance models*, presentation at the Bachelier congress, Dublin, July 2018.

[16] De Marco, S., Henry-Labordère, P.: *Linking vanillas and VIX options: A constrained martingale optimal transport problem*, SIAM Journal on Financial Mathematics, 6:1171–1194, 2015.

[17] Dembo, A., Zeitouni, O.: *Large Deviations Techniques and Applications*, 2nd ed., Springer, New York, 1998.

[18] Dupire, B.: *Pricing with a smile*, Risk, January, 1994.

[19] Fouque, J.-P., Saporito, Y.: *Heston Stochastic Vol-of-Vol Model for Joint Calibration of VIX and S&P 500 Options*, preprint arXiv:1706.00873, 2017.

[20] Galichon, A., Henry-Labordère, P., Touzi, N.: *A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options*, Ann. Appl. Probab., 24(1):312–336, 2014.

[21] Gatheral, J.: *Consistent modeling of SPX and VIX options*, presentation at the Bachelier congress, London, July 18, 2008.

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

[23] Guo, G., Obłój, J.: *Computational methods for martingale optimal transport problems*, preprint arXiv:1710.07911, 2017.

[24] Guyon, J., Menegaux, R., Nutz, M.: *Bounds for VIX futures given S&P 500 smiles*, Finan. & Stoch., 21(3):593–630, 2017.

[25] Guyon, J.: *On the joint calibration of SPX and VIX options*, presentation at Jim Gatheral's 60th birthday conference, New York, October 14, 2017.

[26] Guyon, J.: *On the joint calibration of SPX and VIX options*, presentation at Imperial College London, March 28, 2018.

[27] Guyon, J.: *On the joint calibration of SPX and VIX options*, presentation at QuantMinds 2018, Lisbon, May 16, 2018.

[28] Guyon J.: *Inversion of Convex Ordering in the VIX Market*, to appear in Quantitative Finance, preprint available at `ssrn.com/abstract=3504022`, 2019.

[29] Guyon, J.: *The VIX future in Bergomi models*, preprint, 2020.

[30] Henry-Labordère, P.: *Automated Option Pricing: Numerical Methods*, Intern. Journ. of Theor. and Appl. Finance, 16(8):1–27, 2013.

[31] Henry-Labordère, P.: *Model-free Hedging: A Martingale Optimal Transport Viewpoint*, Chapman & Hall/CRC Financial Mathematics Series, 2017.

[32] Jacquier, A., Martini, C., Muguruza, A.: *On the VIX futures in the rough Bergomi model*, SSRN preprint 2900248, 2017.

[33] Johansen, S.: The extremal convex functions. Math. Scand., 34:61–68, 1974.

[34] Kokholm, T., Stisen, M.: *Joint pricing of VIX and SPX options with stochastic volatility and jump models*, The Journal of Risk Finance, 16(1):27–48, 2015.

[35] Pacati, C., Pompa, P., Renò, R.: *Smiling Twice: The Heston++ Model*, Journal of Banking & Finance, 96:185–206, 2018.

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

[37] Sepp, A.: *Achieving consistent modeling of VIX and equity derivatives*, Imperial College Mathematical Finance Seminar, November 2, 2011.

[38] Sinkhorn, R.: *Diagonal equivalence to matrices with prescribed row and column sums*, The American Mathematical Monthly, 74(4):402–405, 1967.

[39] Song, Z., Xiu, D.: *A tale of two option markets: State-price densities implied from S&P 500 and VIX option prices*, SSRN preprint 2976966, 2012.

[40] Strassen, V.: *The existence of probability measures with given marginals*, Ann. Math. Statist., 36:423–439, 1965.

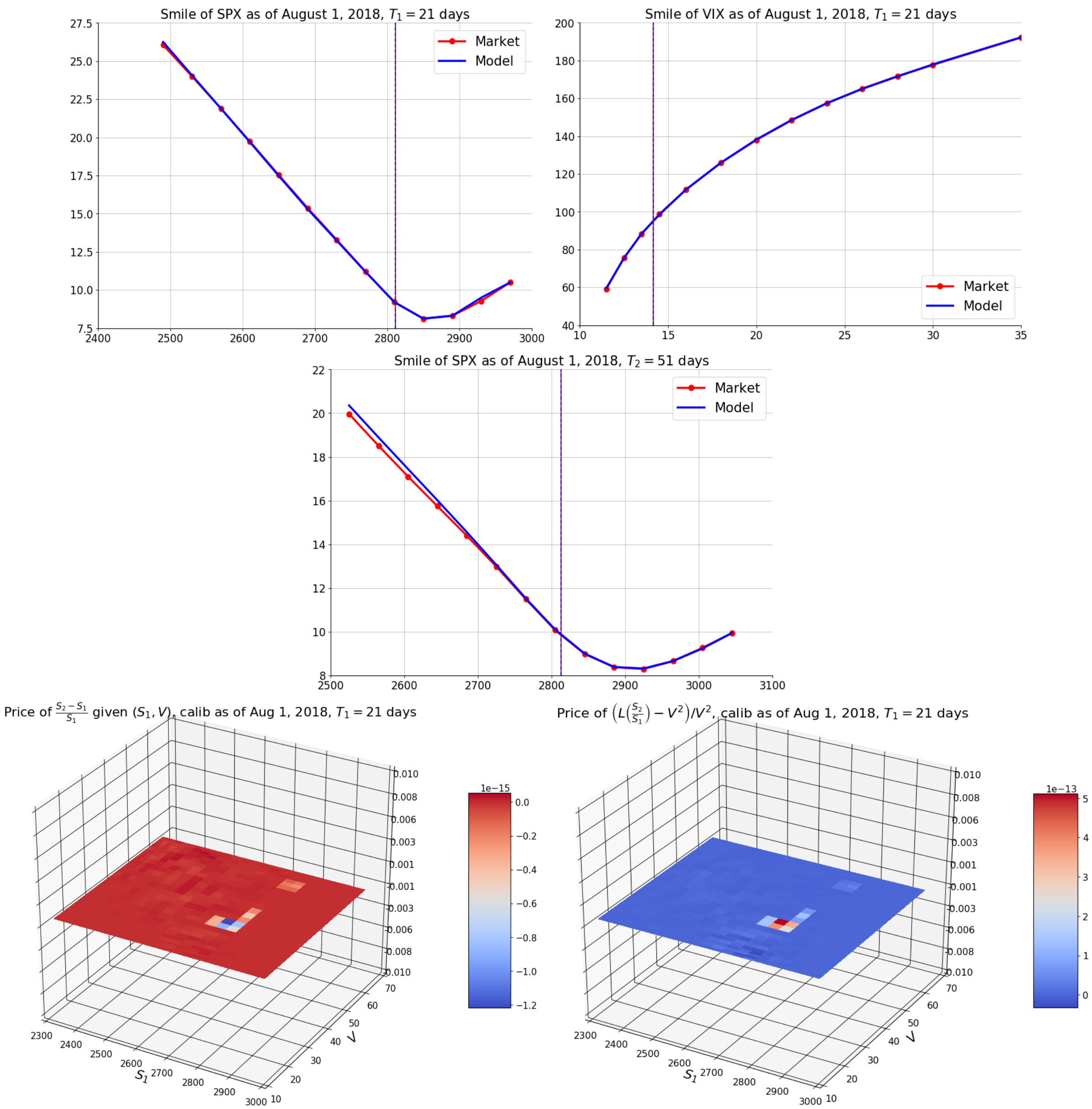

FIGURE 9.1. Top: Futures and smiles of S_1, V, and S_2 in the calibrated model $\mu^*_{\mathcal{K}}$ vs market futures and market smiles. Bottom: Model price of $(S_2 - S_1)/S_1$ (left) and of $(L(S_2/S_1) - V^2)/V^2$ (right) given $(S_1, V) = (s_1, v)$ as a function of (s_1, v) in the quadrature grid. Calibration as of August 1, 2018; $T_1 = 21$ days

[41] Villani, C.: *Topics in optimal transportation*, American Mathematical Society, Graduate Studies in Mathematics 58, 2003.

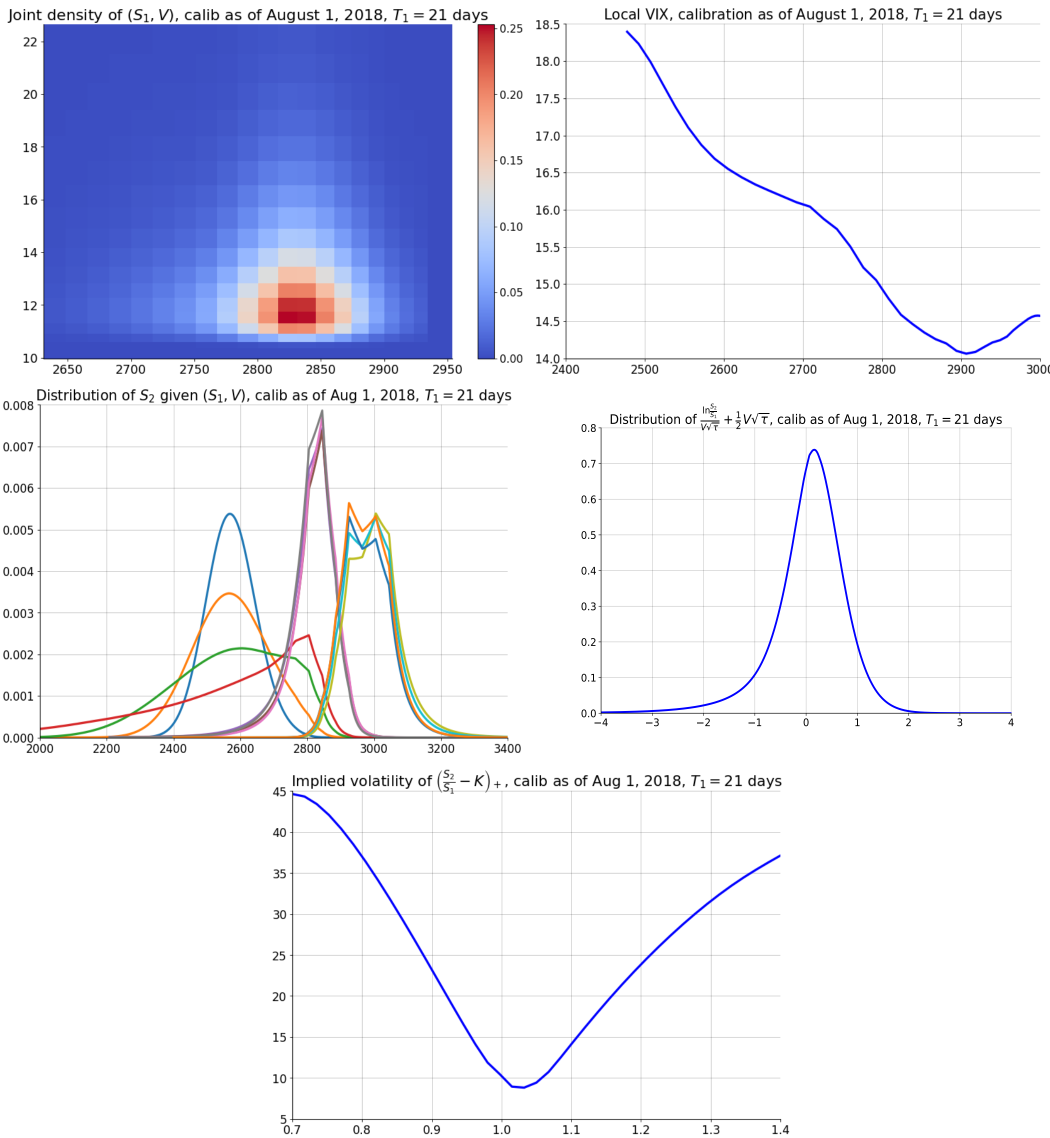

FIGURE 9.2. Some properties of the calibrated model model $\mu^*_{\mathcal{K}}$. Top: joint distribution of (S_1, V) and local VIX function $\mathrm{VIX}_{\mathrm{loc}}(s_1)$. Middle: conditional distribution of S_2 given (s_1, v) for different vales of (s_1, v): $s_1 \in \{2571, 2808, 3000\}$, $v \in \{10.10, 15.30, 23.20, 35.72\}\%$, and distribution of the normalized return R. Bottom: smile of forward starting call options $(S_2/S_1 - K)_+$. Calibration as of August 1, 2018; $T_1 = 21$ days

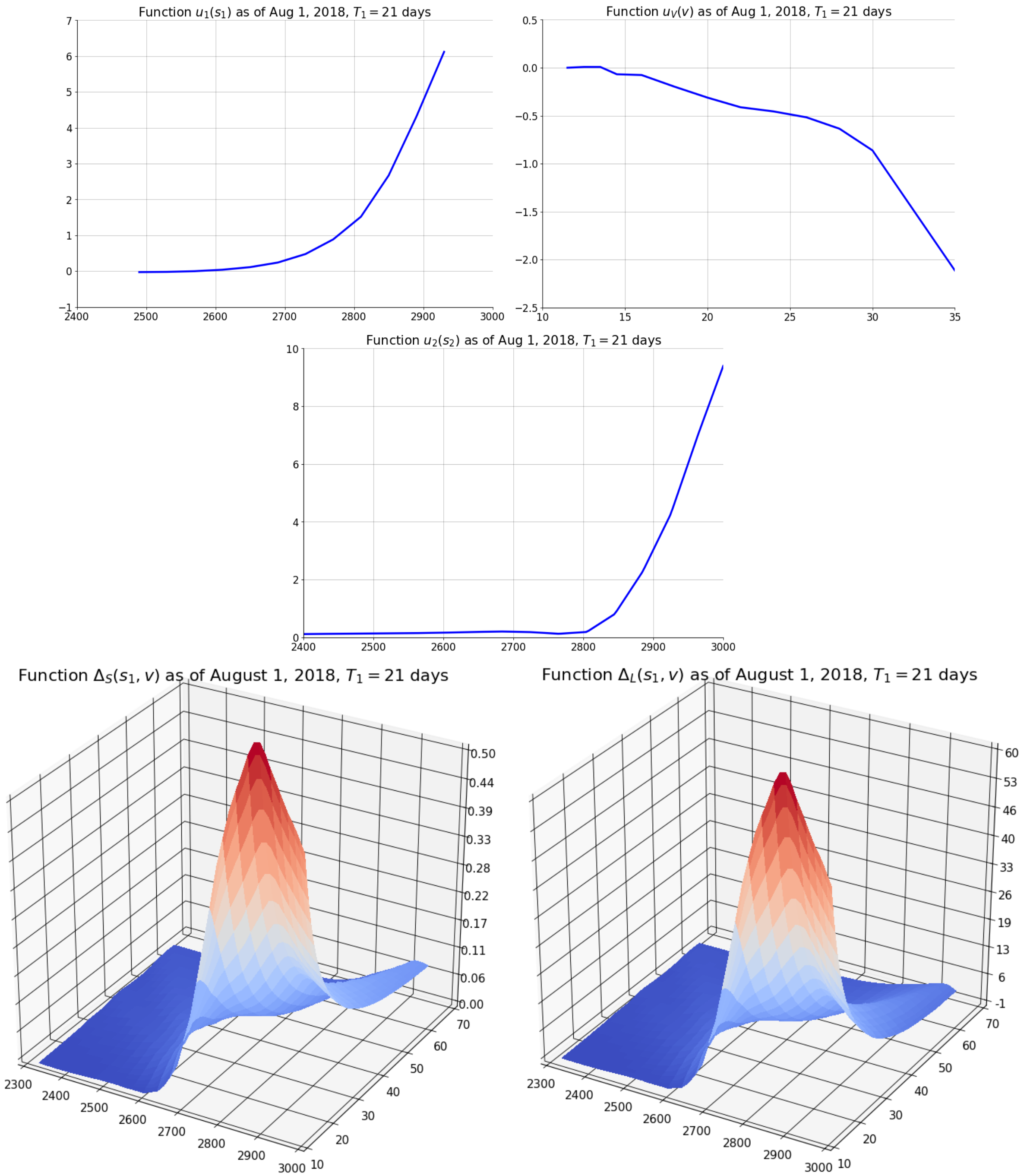

FIGURE 9.3. Description of the calibrated model model $\mu^*_{\mathcal{K}}$. Top: optimal functions u^*_1, u^*_V and u^*_2. Bottom: optimal functions $\Delta^*_S(s_1, v)$ and $\Delta^*_L(s_1, v)$ for (s_1, v) in the quadrature grid. Calibration as of August 1, 2018; $T_1 = 21$ days

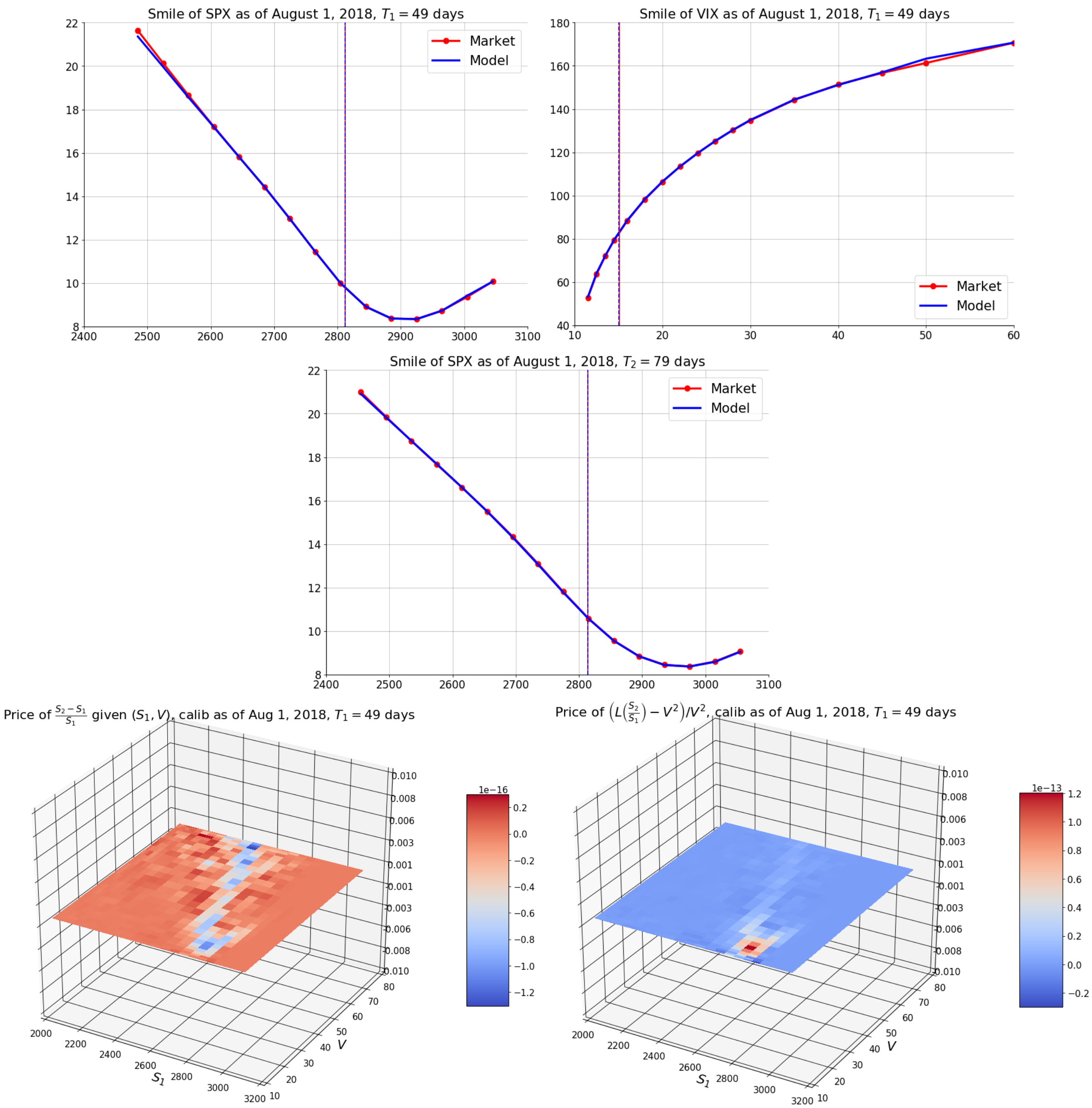

FIGURE 9.4. Top: Futures and smiles of S_1, V, and S_2 in the calibrated model $\mu^*_{\mathcal{K}}$ vs market futures and market smiles. Bottom: Model price of $(S_2 - S_1)/S_1$ (left) and of $(L(S_2/S_1) - V^2)/V^2$ (right) given $(S_1, V) = (s_1, v)$ as a function of (s_1, v) in the quadrature grid. Calibration as of August 1, 2018; $T_1 = 49$ days

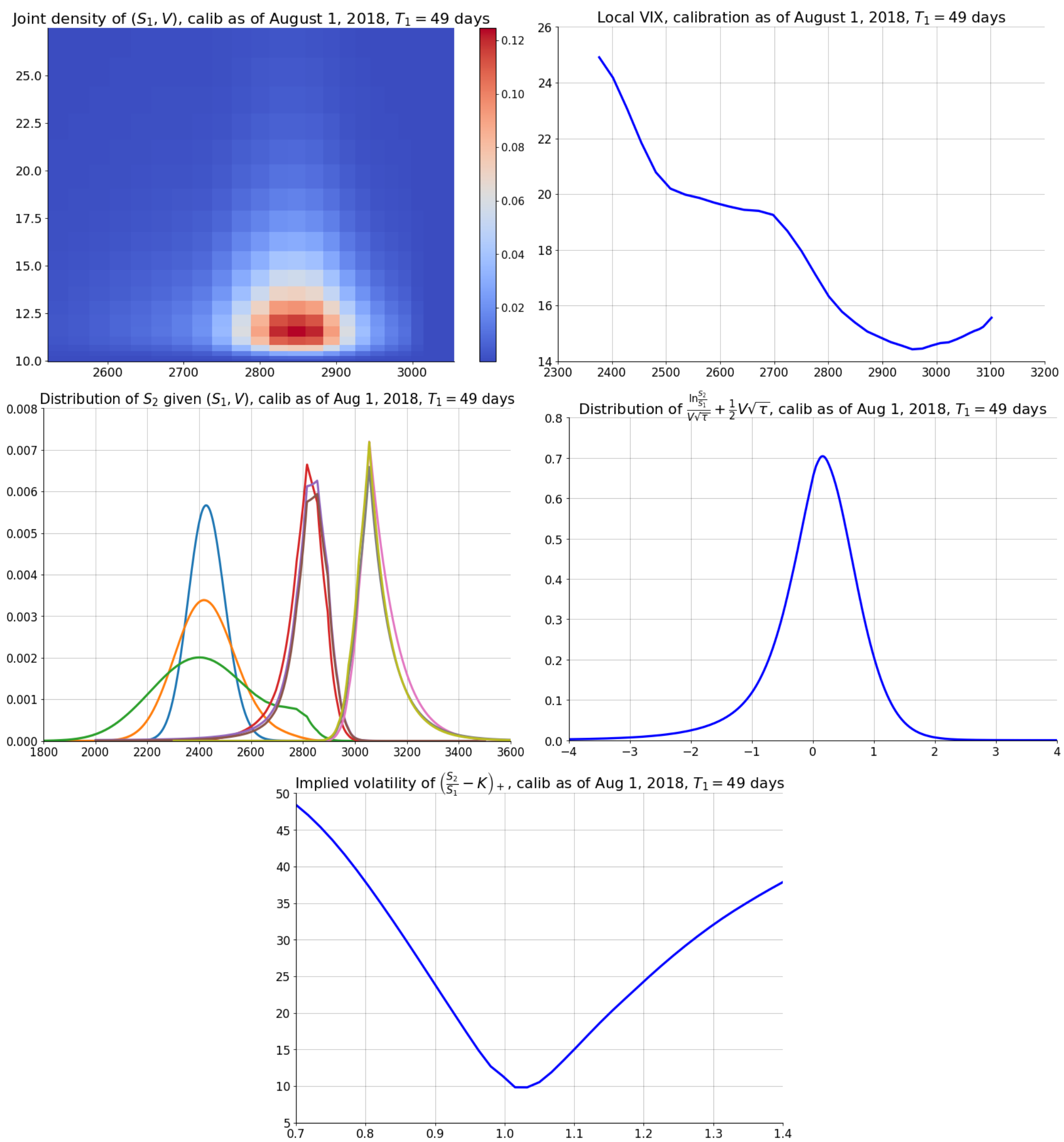

FIGURE 9.5. Some properties of the calibrated model model $\mu^*_{\mathcal{K}}$. Top: joint distribution of (S_1, V) and local VIX function $\mathrm{VIX}_{\mathrm{loc}}(s_1)$. Middle: conditional distribution of S_2 given (s_1, v) for different vales of (s_1, v): $s_1 \in \{2428, 2801, 3102\}$, $v \in \{10.14, 17.32, 28.25\}\%$, and distribution of the normalized return R. Bottom: smile of forward starting call options $(S_2/S_1 - K)_+$. Calibration as of August 1, 2018; $T_1 = 49$ days

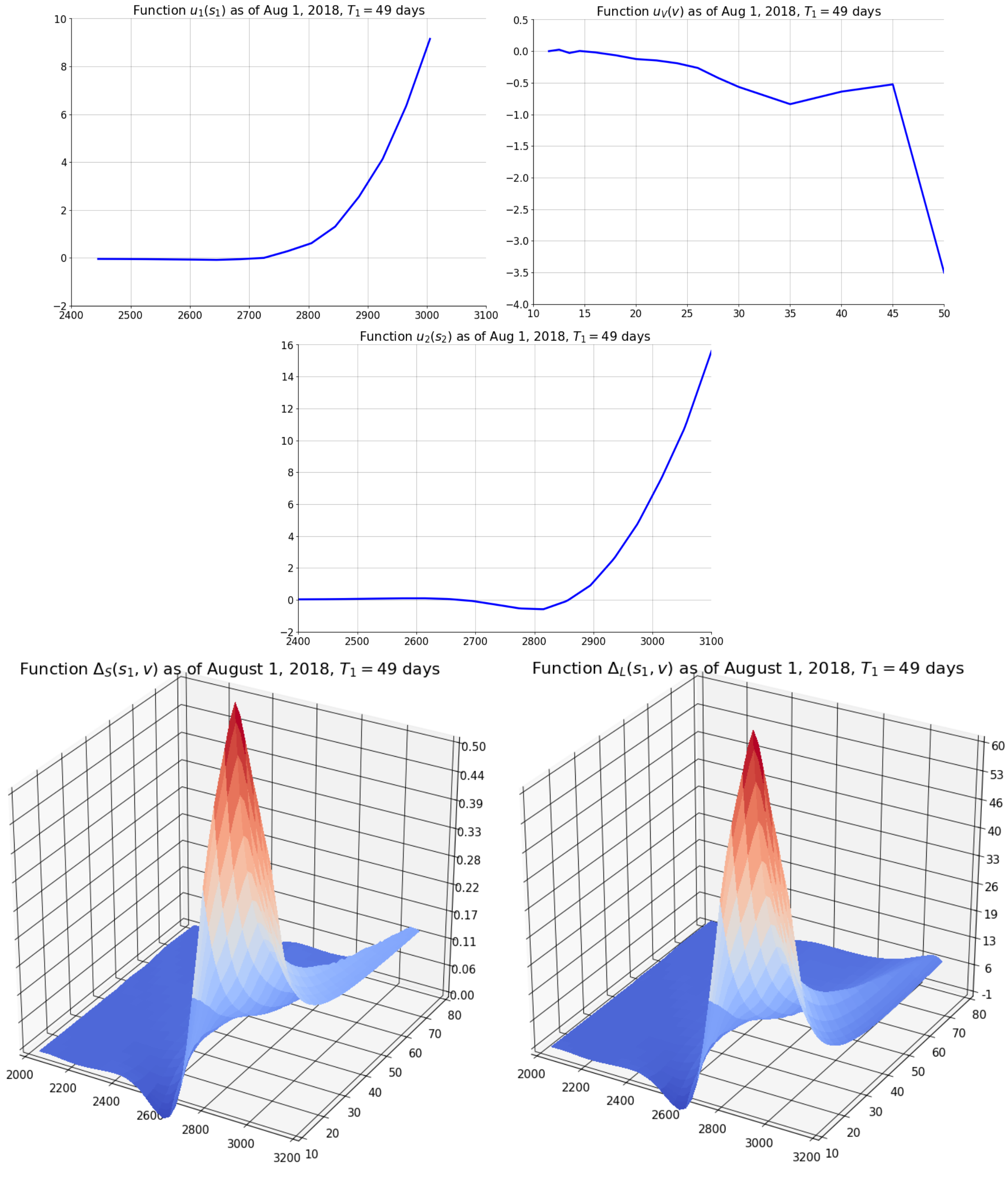

FIGURE 9.6. Description of the calibrated model model $\mu^*_{\mathcal{K}}$. Top: optimal functions u_1^*, u_V^* and u_2^*. Bottom: optimal functions $\Delta^*_S(s_1, v)$ and $\Delta^*_L(s_1, v)$ for (s_1, v) in the quadrature grid. Calibration as of August 1, 2018; $T_1 = 49$ days

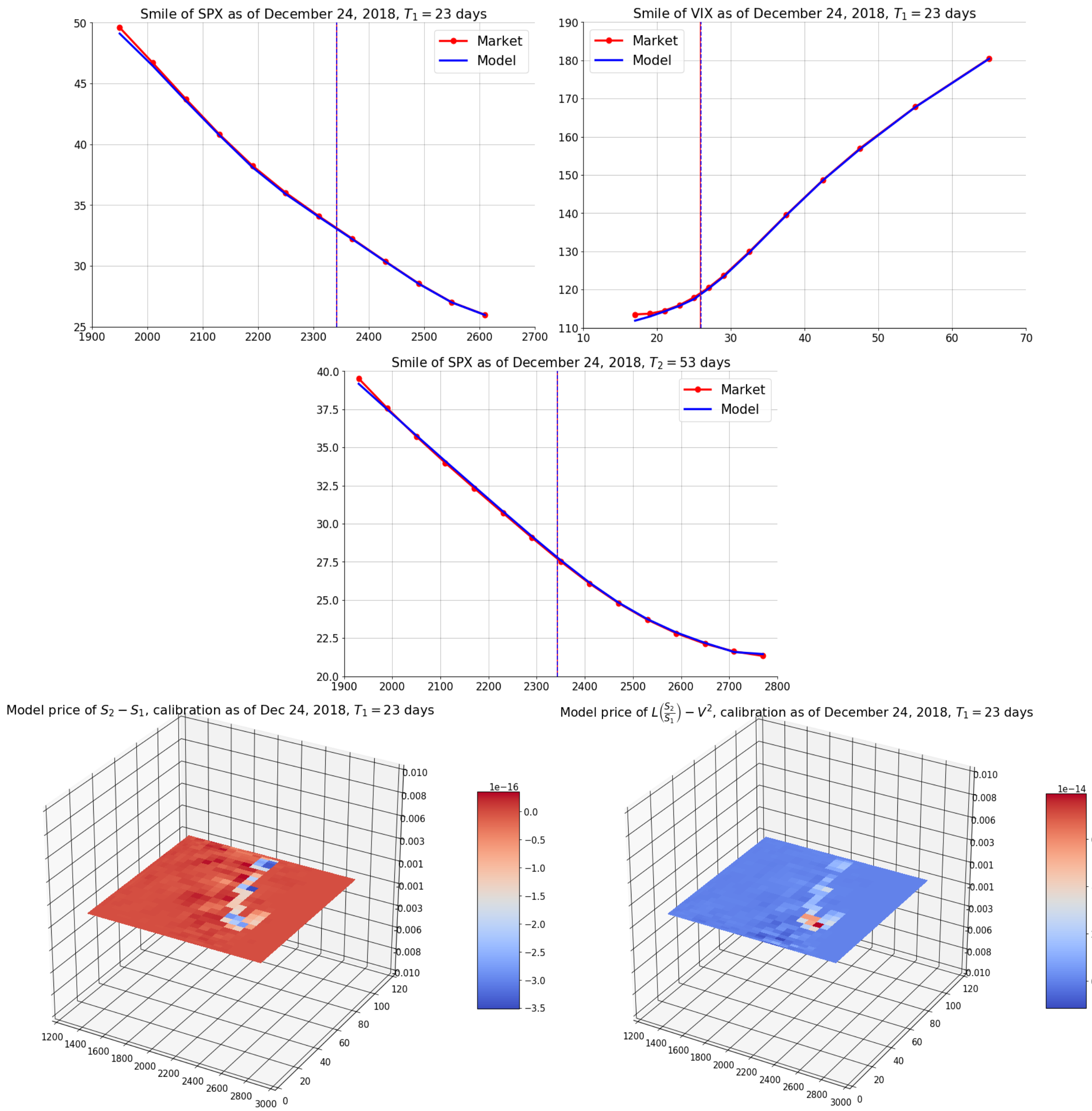

FIGURE 9.7. Top: Futures and smiles of S_1, V, and S_2 in the calibrated model $\mu^*_{\mathcal{K}}$ vs market futures and market smiles. Bottom: Model price of $(S_2 - S_1)/S_1$ (left) and of $(L(S_2/S_1) - V^2)/V^2$ (right) given $(S_1, V) = (s_1, v)$ as a function of (s_1, v) in the quadrature grid. Calibration as of December 24, 2018; $T_1 = 23$ days

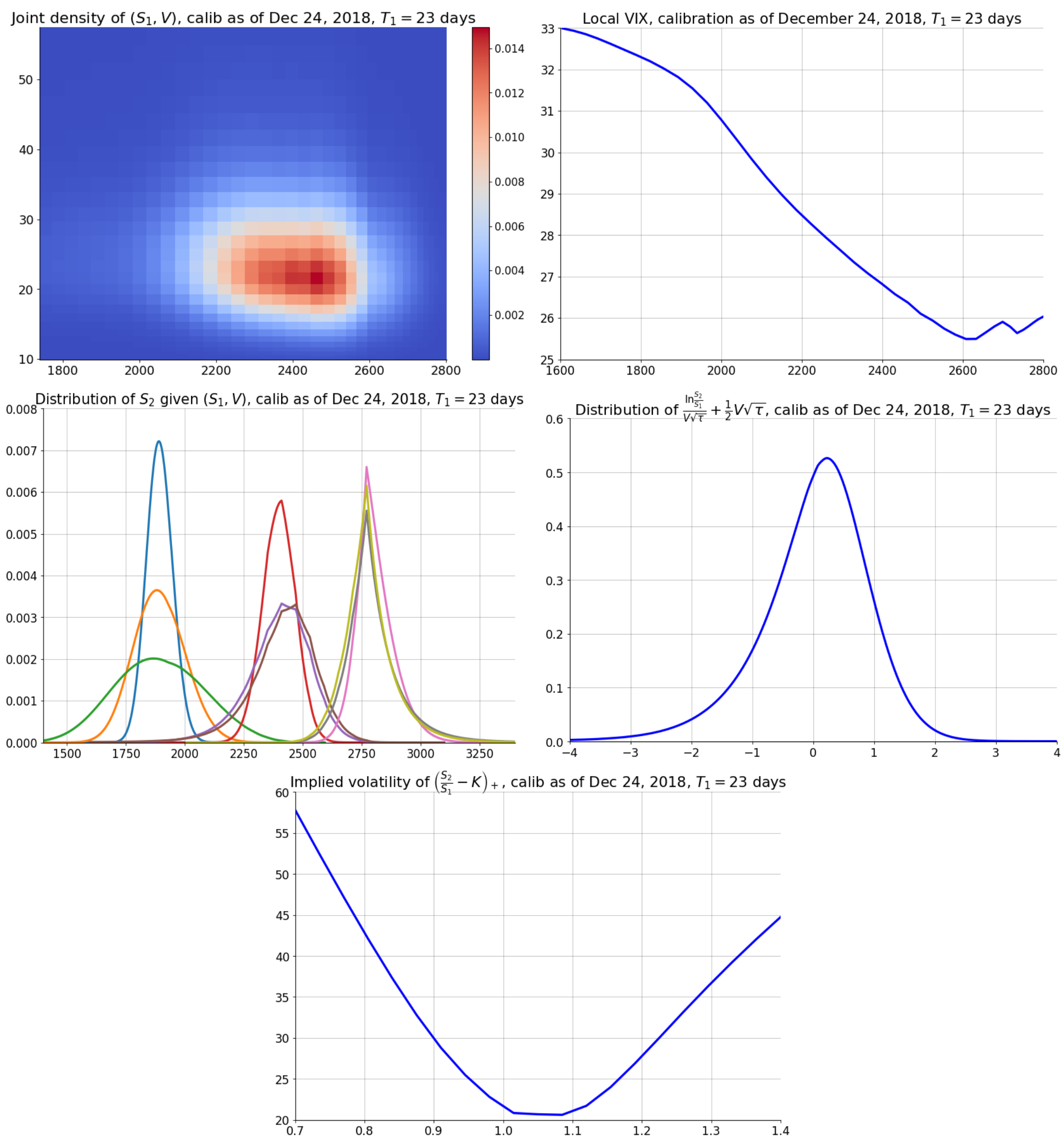

FIGURE 9.8. Some properties of the calibrated model model $\mu^*_{\mathcal{K}}$. Top: joint distribution of (S_1, V) and local VIX function $\mathrm{VIX}_{\mathrm{loc}}(s_1)$. Middle: conditional distribution of S_2 given (s_1, v) for different vales of (s_1, v): $s_1 \in \{1892, 2398, 2807\}$, $v \in \{10.20, 20.02, 34.98\}\%$, and distribution of the normalized return R. Bottom: smile of forward starting call options $(S_2/S_1 - K)_+$. Calibration as of December 24, 2018; $T_1 = 23$ days

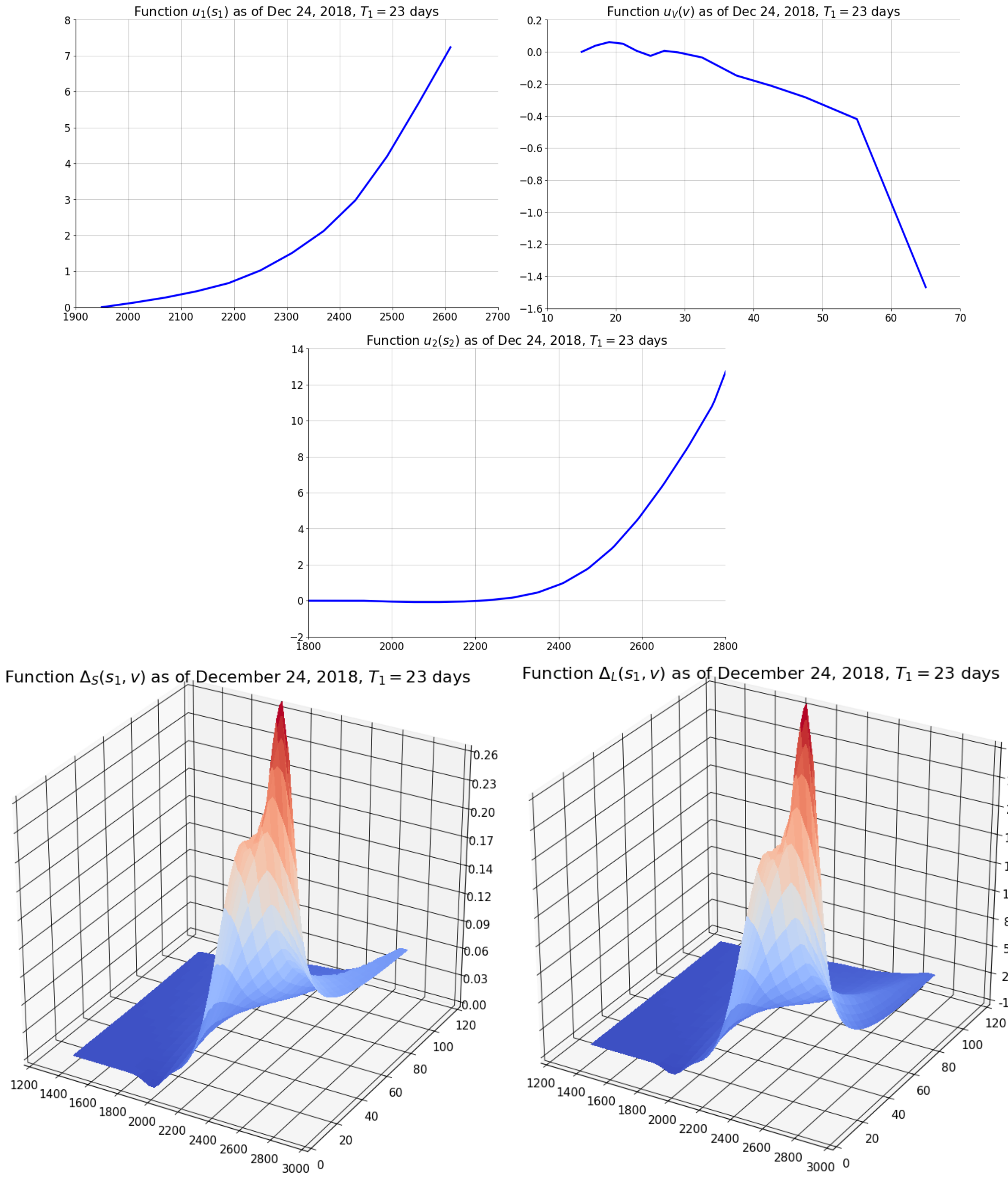

FIGURE 9.9. Description of the calibrated model model $\mu^*_{\mathcal{K}}$. Top: optimal functions u^*_1, u^*_V and u^*_2. Bottom: optimal functions $\Delta^*_S(s_1, v)$ and $\Delta^*_L(s_1, v)$ for (s_1, v) in the quadrature grid. Calibration as of December 24, 2018; $T_1 = 23$ days

10. Appendix: Functionally generated portfolios superreplicating zero

In this appendix, in the spirit of [24, Section 6], we introduce a new family of portfolios that superreplicate zero. Their main merits are their simple functional form and that their superreplication property is guaranteed by construction for all values of the underlyings—in contrast to numerical solutions to the linear programming problems. When a portfolio in this family has a negative price, it reveals the existence of an (S_1, S_2, V)-arbitrage. However, there might exist an (S_1, S_2, V)-arbitrage even if all portfolios in the family have nonnegative prices.

Our portfolios are based on convex payoffs of both the SPX and its logarithm. Let us consider a convex function $\varphi : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}$. For $u_1 : \mathbb{R}_{>0} \to \mathbb{R} \cup \{+\infty\}$ and $v \geq 0$, we denote by

$$u_1^{\varphi}(v) := \sup_{s_1>0} \left\{ -\varphi(s_1, L(s_1) + v^2) - u_1(s_1) \right\}$$

the φ-transform of u_1, i.e., the smallest function $u_V : \mathbb{R}_{\geq 0} \to \mathbb{R} \cup \{+\infty\}$ such that for all $s_1 > 0$ and $v \geq 0$, $u_1(s_1) + u_V(v) + \varphi(s_1, L(s_1) + v^2) \geq 0$. Similarly, for $u_V : \mathbb{R}_{\geq 0} \to \mathbb{R} \cup \{+\infty\}$ and $s_1 > 0$, we denote by

$$u_V^{\varphi}(s_1) := \sup_{v \geq 0} \left\{ -\varphi(s_1, L(s_1) + v^2) - u_V(v) \right\}$$

the φ-transform of u_V, i.e., the smallest function $u_1 : \mathbb{R}_{>0} \to \mathbb{R} \cup \{+\infty\}$ such that for all $s_1 > 0$ and $v \geq 0$, $u_1(s_1) + u_V(v) + \varphi(s_1, L(s_1) + v^2) \geq 0$. We denote $u_1^{\varphi\varphi} := (u_1^{\varphi})^{\varphi}$ and $u_V^{\varphi\varphi} := (u_V^{\varphi})^{\varphi}$. The pair $(u_1^{\varphi\varphi}, u_1^{\varphi})$ is called a pair of φ-conjugate functions. It is easy to check that $(u_1^{\varphi\varphi})^{\varphi} = u_1^{\varphi}$, $(u_V^{\varphi\varphi})^{\varphi} = u_V^{\varphi}$, $u_1^{\varphi\varphi} \leq u_1$, and $u_V^{\varphi\varphi} \leq u_V$, see for instance [41, Exercise 2.35]. Moreover, we denote by $\partial_{i,r}\varphi$ the right derivative of φ with respect to its i-th argument.

Proposition 11. *Let $\varphi : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}$ be convex. Let*

$$u_{2,\varphi}(s_2) = \varphi(s_2, L(s_2)), \quad \Delta_{S,\varphi}(s_1, v) = -\partial_{1,r}\varphi(s_1, L(s_1) + v^2), \quad \Delta_{L,\varphi}(s_1, v) = -\partial_{2,r}\varphi(s_1, L(s_1) + v^2).$$

Then

(i) *For any $u_1 : \mathbb{R}_{>0} \to \mathbb{R}$ the following portfolios superreplicate zero:*

$$\Pi^1_{\varphi,u_1} := (u_1, u_1^{\varphi}, u_{2,\varphi}, \Delta_{S,\varphi}, \Delta_{L,\varphi}), \qquad \bar{\Pi}^1_{\varphi,u_1} := (u_1^{\varphi\varphi}, u_1^{\varphi}, u_{2,\varphi}, \Delta_{S,\varphi}, \Delta_{L,\varphi}).$$

(ii) *For any $u_V : \mathbb{R}_{\geq 0} \to \mathbb{R}$ the following portfolios superreplicate zero:*

$$\Pi^V_{\varphi,u_V} := (u_V^{\varphi}, u_V, u_{2,\varphi}, \Delta_{S,\varphi}, \Delta_{L,\varphi}), \qquad \bar{\Pi}^V_{\varphi,u_V} := (u_V^{\varphi}, u_V^{\varphi\varphi}, u_{2,\varphi}, \Delta_{S,\varphi}, \Delta_{L,\varphi}).$$

Proof. Since φ is convex,

$$\begin{aligned}&\varphi(s_2, L(s_2)) - \varphi(s_1, L(s_1) + v^2) \\ &\qquad \geq \partial_{1,r}\varphi(s_1, L(s_1) + v^2)(s_2 - s_1) + \partial_{2,r}\varphi(s_1, L(s_1) + v^2)(L(s_2) - L(s_1) - v^2)\end{aligned}$$

and as a consequence,

$$\begin{aligned} & u_{2,\varphi}(s_2) + \Delta_{S,\varphi}(s_1, v)(s_2 - s_1) + \Delta_{L,\varphi}(s_1, v)\left(L\left(\frac{s_2}{s_1}\right) - v^2\right) \\ = \; & \varphi(s_2, L(s_2)) - \partial_{1,r}\varphi(s_1, L(s_1) + v^2)(s_2 - s_1) - \partial_{2,r}\varphi(s_1, L(s_1) + v^2)(L(s_2) - L(s_1) - v^2) \\ \geq \; & \varphi(s_1, L(s_1) + v^2). \end{aligned}$$

Therefore

$$\begin{aligned}&u_1(s_1) + u_{2,\varphi}(s_2) + u_1^{\varphi}(v) + \Delta_{S,\varphi}(s_1, v)(s_2 - s_1) + \Delta_{L,\varphi}(s_1, v)\left(L\left(\frac{s_2}{s_1}\right) - v^2\right) \\ &\qquad \geq u_1(s_1) + u_1^{\varphi}(v) + \varphi(s_1, L(s_1) + v^2) \geq 0\end{aligned}$$

(by the definition of $u_1^{\varphi}(v)$) and

$$\begin{aligned}&u_V^{\varphi}(s_1) + u_{2,\varphi}(s_2) + u_V(v) + \Delta_{S,\varphi}(s_1, v)(s_2 - s_1) + \Delta_{L,\varphi}(s_1, v)\left(L\left(\frac{s_2}{s_1}\right) - v^2\right) \\ &\qquad \geq u_V^{\varphi}(s_1) + u_V(v) + \varphi(s_1, L(s_1) + v^2) \geq 0\end{aligned}$$

(by the definition of $u_V^{\varphi}(s_1)$). This proves the first parts of (i) and (ii). Applying them with $u_1 = u_V^{\varphi}$ and $u_V = u_1^{\varphi}$ gives the second parts of (ii) and (i), respectively. □

Corollary 12. *Let* $\varphi : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}$ *be a convex function such that* $\varphi(S_2, L(S_2)) \in L^1(\mu_2)$.

(i) *Let* $u_1 : \mathbb{R}_{>0} \to \mathbb{R}$ *such that* $u_1 \in L^1(\mu_1)$ *and* $u_1^\varphi \in L^1(\mu_V)$. *If* $\mathbb{E}^1[u_1(S_1)] + \mathbb{E}^V[u_1^\varphi(V)] + \mathbb{E}^2[\varphi(S_2, L(S_2))] < 0$ *then the portfolio* Π^1_{φ,u_1} *is an* (S_1, S_2, V)*-arbitrage.*

(ii) *Let* $u_1 : \mathbb{R}_{>0} \to \mathbb{R}$ *such that* $u_1^{\varphi\varphi} \in L^1(\mu_1)$ *and* $u_1^\varphi \in L^1(\mu_V)$. *If* $\mathbb{E}^1[u_1^{\varphi\varphi}(S_1)] + \mathbb{E}^V[u_1^\varphi(V)] + \mathbb{E}^2[\varphi(S_2, L(S_2))] < 0$ *then the portfolio* $\bar{\Pi}^1_{\varphi,u_1}$ *is an* (S_1, S_2, V)*-arbitrage.*

(iii) *Let* $u_V : \mathbb{R}_{\geq 0} \to \mathbb{R}$ *such that* $u_V^\varphi \in L^1(\mu_1)$ *and* $u_V \in L^1(\mu_V)$. *If* $\mathbb{E}^1[u_V^\varphi(S_1)] + \mathbb{E}^V[u_V(V)] + \mathbb{E}^2[\varphi(S_2, L(S_2))] < 0$ *then the portfolio* Π^V_{φ,u_V} *is an* (S_1, S_2, V)*-arbitrage.*

(iv) *Let* $u_V : \mathbb{R}_{\geq 0} \to \mathbb{R}$ *such that* $u_V^\varphi \in L^1(\mu_1)$ *and* $u_V^{\varphi\varphi} \in L^1(\mu_V)$. *If* $\mathbb{E}^1[u_V^\varphi(S_1)] + \mathbb{E}^V[u_V^{\varphi\varphi}(V)] + \mathbb{E}^2[\varphi(S_2, L(S_2))] < 0$ *then the portfolio* $\bar{\Pi}^V_{\varphi,u_V}$ *is an* (S_1, S_2, V)*-arbitrage.*

From (4.1), in order to identify an (S_1, S_2, V)-arbitrage, one should optimize over the set of constrained functions $(u_1, u_2, u_V, \Delta_S, \Delta_V) \in \mathcal{U}_0$. Corollary 12 tells us that it might be enough to simply optimize over the set of *unconstrained* functions (φ, u_1) or (φ, u_V), where $\varphi : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}$ is convex. Since testing all convex functions in dimension two is impractical [33], one may only test functions φ of the form $\varphi(x, y) = \psi(ax + y)$ where ψ is a convex function in dimension one and $a \in \mathbb{R}$.

In dual form, this means that in order to identify an (S_1, S_2, V)-arbitrage, instead of solving a dispersion-constrained martingale optimal transport problem with two dates (T_1 and T_2) and three underlyings (S_1 and V at date T_1, and S_2 at date T_2), it might be enough to solve a family (indexed by the convex functions φ) of much simpler transport problems, namely, classical unconstrained optimal transport problems with only one date (T_1) and two underlyings (S_1 and V). Indeed, for each convex function $\varphi : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}$, let us define

$$\mathcal{U}^1_\varphi = \{(u_1, u_V) \in L^1(\mu_1) \times L^1(\mu_V) \mid \forall s_1 > 0, \forall v \geq 0,\ u_1(s_1) + u_V(v) \geq -\varphi(s_1, L(s_1) + v^2)\},$$

the set of vanilla payoffs on S_1 and V whose sum superreplicates the payoff $-\varphi(S_1, L(S_1) + V^2)$. Then by the classical Kantorovich duality we have

Proposition 13. *Let* $\varphi : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}$ *be convex. Then*

$$P^1_\varphi := \inf_{(u_1,u_V)\in\mathcal{U}^1_\varphi} \left\{\mathbb{E}^1[u_1(S_1)] + \mathbb{E}^V[u_V(V)]\right\} = \sup_{\nu\in\mathcal{T}(\mu_1,\mu_V)} \mathbb{E}^\nu\left[-\varphi(S_1, L(S_1) + V^2)\right] =: D^1_\varphi.$$

If $D^1_\varphi < -\mathbb{E}^2[\varphi(S_2, L(S_2))]$*, then there exists an* (S_1, S_2, V)*-arbitrage.*

Bloomberg L.P., Quantitative Research, 731 Lexington Ave, New York, NY 10022, USA.