SIAM J. FINANCIAL MATH.
Vol. 6, pp. 1171–1194
© 2015 Society for Industrial and Applied Mathematics

Linking Vanillas and VIX Options: A Constrained Martingale Optimal Transport Problem*

Stefano De Marco† and Pierre Henry-Labordère‡

Dedicated to the memory of Prof. Peter Laurence

Abstract. VIX options traded on the CBOE have become popular volatility derivatives. As S&P 500 vanilla options and VIX both depend on S&P 500 volatility dynamics, it is important to understand the link between these products. In this paper, we bound VIX options from vanilla options and VIX futures. This leads us to introduce a new *martingale optimal transportation* problem that we solve numerically. Analytical lower and upper bounds are also provided which already highlight some (potential) arbitrage opportunities. We fully characterize the class of marginal distributions for which these explicit bounds are optimal, and illustrate numerically that they seem to be optimal for the market-implied marginal distributions.

Key words. VIX options, robust sub/super-replication, duality, martingale optimal transport

AMS subject classifications. 91G20, 91G80, 90C46, 90C05, 90C34

DOI. 10.1137/140960724

1. Introduction.

VIX futures and VIX options, traded on the CBOE, have become popular volatility derivatives. The VIX index at a future expiry t_1 is by definition the 30-day variance swap volatility, computed by replication using market prices of listed S&P 500 options (at t_1):

$$
\begin{aligned}
(1.1)\qquad \mathrm{VIX}_{t_1}^2 &\equiv -\frac{2}{\Delta}\mathbb{E}_{t_1}\left[\ln\left(\frac{S_{t_2}}{F_{t_1}^{t_2}}\right)\right] \\
&= \frac{2}{\Delta}\left(\int_0^{F_{t_1}^{t_2}} \frac{P(t_1,t_2,K)}{K^2}dK + \int_{F_{t_1}^{t_2}}^{\infty} \frac{C(t_1,t_2,K)}{K^2}dK\right)
\end{aligned}
$$

with $\Delta = t_2 - t_1 = 30$ days, $F_{t_1}^{t_2}$ the forward price at t_1 with maturity t_2, and $P(t_1,t_2,K)$ (resp., $C(t_1,t_2,K)$) the undiscounted market price at t_1 of a put (resp., call) option with strike K and maturity t_2. In the formula used by CBOE (see [8]), the integrals are discretized using the trapezoidal rule, the integration is cut off for small and large strikes wherever a zero bid price is encountered for two consecutive prices, and the two shortest available market maturities following t_1 are linearly interpolated in order to attain $t_2 = t_1 + 30$ days. The payoff of a call option on VIX expiring at t_1 with strike K is $(\mathrm{VIX}_{t_1} - K)^+$. Its fair value

*Received by the editors March 12, 2014; accepted for publication (in revised form) September 2, 2015; published electronically December 8, 2015.
http://www.siam.org/journals/sifin/6/96072.html
†CMAP, École Polytechnique, 91128 Palaiseau, France (demarco@cmap.polytechnique.fr).
‡Global Markets Quantitative Research, Société Générale, 92800 Puteaux, France (pierre.henry-labordere@sgcib.com).

C_{VIX} is usually quoted in implied VIX volatility, i.e., the constant volatility that must be put into the Black–Scholes formula with the spot value equal to the VIX future (i.e., $K = 0$) in order to match C_{VIX}. Below, the market value (at $t = 0$) for the VIX future is denoted VIX, and the market value (at $t = 0$) of a forward log-contract, which pays $\mathrm{VIX}^2_{t_1}$ at t_1, is denoted $\sigma^2_{1,2}$. In terms of call and put prices, we have

$$\sigma^2_{1,2} = \frac{2}{\Delta}\left(\int_0^{F_0^{t_2}} \frac{P(0,t_2,K)}{K^2}dK + \int_{F_0^{t_2}}^{\infty} \frac{C(0,t_2,K)}{K^2}dK\right) - \frac{2}{\Delta}\left(\int_0^{F_0^{t_1}} \frac{P(0,t_1,K)}{K^2}dK + \int_{F_0^{t_1}}^{\infty} \frac{C(0,t_1,K)}{K^2}dK\right).$$

In the no-arbitrage framework, the pricing of VIX options can be achieved by postulating a model, that is, a probability measure $\mathbb{P}$ on $\mathbb{R}^2_+$ under which the coordinate process $(S_{t_i})_{i=1,2}$ is required to be a discrete martingale in its own filtration. For the sake of simplicity, we assume here zero rate, dividend, and repo. This can be easily relaxed by considering the process f_{t_i} introduced in [18] (see equation 14), which has the property of being a martingale. Additionally, we may impose that the model reproduces the market value of VIX futures and a continuum of S&P 500 call options for each expiry t_i. The construction of such a model is not at all obvious. For example, the calibration of vanillas and the VIX future can be achieved in principle with the so-called local stochastic volatility models (LSVMs); see, e.g., [17, 25, 27]. However, the computation of the VIX index at t_1, as defined by (1.1), in such models requires the evaluation of conditional expectations: this task is typically not straightforward and requires performing a time-consuming Monte-Carlo of Monte-Carlo method (with nested simulations) or relying on a numerical PDE implementation for low-dimensional models. Moreover, within this class of continuous stochastic volatility models calibrated to vanillas, the VIX future is bounded from above by the Dupire local volatility model (LVM) [15] as Δ is small:

$$\mathrm{VIX} = \mathbb{E}\left[\sqrt{\frac{1}{\Delta}\int_{t_1}^{t_2} \mathbb{E}_{t_1}\sigma_s^2\, ds}\right] \simeq \mathbb{E}\Big[\sqrt{\sigma^2_{t_1}}\Big] \overset{\text{Jensen}}{\leq} \mathbb{E}\Big[\sqrt{\mathbb{E}[\sigma^2_{t_1}|S_{t_1}]}\Big] = \mathbb{E}\Big[\sigma_{\mathrm{loc}}(t_1, S_{t_1})\Big] \simeq \mathrm{VIX}_{\mathrm{LV}},$$

where $\sigma_t^2 \equiv d\langle \ln \frac{S}{S_0}\rangle_t/dt$ and σ_{loc} is the Dupire local volatility. As market prices of VIX futures typically lie in practice above prices given by the LVM, LSVMs, calibrated on the S&P 500 smile, are unable to calibrate VIX futures, hence VIX options. An alternative approach is to use jump-diffusion models (see, e.g., [1, 28, 10]).

Disregarding the information of vanillas, the calibration of VIX futures (and also some VIX options) can be easily obtained using a multifactor variance swap curve model as introduced in [3]. However, this model is by construction unable to address the (structural) link between vanillas and VIX smiles. In this paper, we follow a different route. Instead of postulating a model, we focus on the computation of model-independent bounds (and super-replicating strategies) for VIX options consistent with t_1, t_2 vanillas and the t_1 VIX future. By duality, we will show that there exist arbitrage-free models that attain these bounds.

The rest of the paper is organized as follows. In section 2, we reframe the computation of these bounds in terms of robust sub/super-replication strategies. In section 3, we establish a dual version which is connected to a martingale optimal transport problem as introduced in [2] and in [16]. It corresponds to the maximization of the expectation of a VIX payoff with respect to a martingale measure $\mathbb{P}$ with marginals μ, ν and with the constraint on the VIX future $\mathrm{VIX} = \mathbb{E}^{\mathbb{P}}\left[\sqrt{\mathbb{E}^{\mathbb{P}}_{t_1}[-(2/\Delta)\ln S_{t_2}/S_{t_1}]}\right]$. Note that as this additional constraint, which is not present in the original martingale optimal transport, is nonlinear with respect to the (martingale) measure $\mathbb{P}$, this optimal transport problem is more involved. Then, we derive analytical (a priori nonoptimal) lower and upper bounds, and we characterize, in terms of an order condition between measures, the class of marginals μ, ν for which these bounds are actually optimal. Finally, we compare our bounds against market values of VIX options and highlight some arbitrage opportunities. Moreover, our numerical experiments suggest that the analytical bound is optimal for a large class of marginals μ, ν.

2. Robust sub/super-replications.

For technical reasons, we will assume that the random variables $(S_{t_1}, S_{t_2}, \mathrm{VIX}_{t_1})$ are supported on a compact interval $I_1 \times I_2 \times I_X \subset (\mathbb{R}_+^*)^3$. In section 4 we will allow for random variables with support on the positive real line.

$M_+(X)$ (resp., $\mathcal{P}(X)$) denotes the set of positive (resp., probability) measures on a space X. For further reference, we denote by $\mathcal{M}(\mu, \nu, \mathrm{VIX})$ the set of all martingale measures $\mathbb{P}$ on the (path space) $I_1 \times I_2 \times I_X$ having first two marginals μ, ν with mean S_0 and such that $\mathrm{VIX} = \mathbb{E}^{\mathbb{P}}[\mathrm{VIX}_{t_1}]$, that is,

$$
\begin{aligned}
\mathcal{M}(\mu, \nu, \mathrm{VIX}) = \Big\{ \mathbb{P} \in \mathcal{P}(I_1 \times I_2 \times I_X) : {} & S_{t_1} \overset{\mathbb{P}}{\sim} \mu, \; S_{t_2} \overset{\mathbb{P}}{\sim} \nu, \\
& \mathbb{E}^{\mathbb{P}}[\mathrm{VIX}_{t_1}] = \mathrm{VIX}, \\
& \mathbb{E}^{\mathbb{P}}[S_{t_2} | S_{t_1}, \mathrm{VIX}_{t_1}] = S_{t_1}, \\
& \mathbb{E}^{\mathbb{P}}\left[-\frac{2}{\Delta} \log \frac{S_{t_2}}{S_{t_1}} \Big| S_{t_1}, \mathrm{VIX}_{t_1} \right] = \mathrm{VIX}_{t_1}^2 \Big\}.
\end{aligned} \tag{2.1}
$$

The price (at $t = 0$) of the forward log-contract reads $\sigma_{1,2}^2 = -\frac{2}{\Delta}(\mathbb{E}^{\nu}[\log(S_{t_2})] - \mathbb{E}^{\mu}[\log(S_{t_1})])$.

We define the robust seller's price UB of a VIX call option expiring at t_1 with strike K as follows.

Definition 2.1 (seller's price).

$$
\mathrm{UB} \equiv \inf_{u_1 \in \mathrm{L}^1(\mu), u_2 \in \mathrm{L}^1(\nu), \lambda \in \mathbb{R}, \Delta_S, \Delta_X} \mathbb{E}^{\mu}[u_1(S_{t_1})] + \mathbb{E}^{\nu}[u_2(S_{t_2})] + \lambda \mathrm{VIX}
$$

such that

$$
u_1(s_1) + u_2(s_2) + \lambda\sqrt{x} + \Delta_S(s_1, x)(s_2 - s_1) + \Delta_X(s_1, x)\left(-\frac{2}{\Delta}\ln\left(\frac{s_2}{s_1}\right) - x\right) \geq \left(\sqrt{x} - K\right)^+ \quad \forall\, (s_1, s_2, \sqrt{x}) \in I_1 \times I_2 \times I_X, \tag{2.2}
$$

where the functions $\Delta_S, \Delta_X : I_1 \times I_X \to \mathbb{R}$ *are assumed to be bounded continuous functions on* $I_1 \times I_X$, $u_1 \in \mathrm{L}^1(\mu)$ *and* $u_2 \in \mathrm{L}^1(\nu)$.

Note that this defines a *linear* semi-infinite infinite-dimensional programming problem. The variable x should be interpreted as the t_1-value of a log-contract $-2/\Delta \ln \frac{s_2}{s_1}$, i.e., the square of the VIX index $\mathrm{VIX}^2_{t_1}$. Note that we have ensured that the above inequality is valid for all s_1 and s_2, not just $\mu \times \nu$ a.s. As explained in [30] (see section 1.1), this can be achieved by allowing u_1 and u_2 to take values in $\mathbb{R} \cup \{+\infty\}$. Indeed, we introduce negligible sets N_1 and N_2 such that the inequality holds true for all $(s_1, s_2) \in N_1^c \times N_2^c$, and redefine the values of u_1 and u_2 to be $+\infty$ on N_1, N_2, respectively.

Remark 2.2. Our analysis applies to the exact definition of the CBOE VIX index by replacing in inequality (2.2) the function $-(2/\Delta)\ln(s_2/s_1)$ by its trapezoidal approximation.

Remark 2.3. Our results can be easily extended to the case of a general payoff on the VIX, $\psi(x)$, instead of $(\sqrt{x} - K)^+$. For example, we never use the convexity of the payoff $(\mathrm{VIX}_{t_1} - K)^+$.

Similarly, we define the robust buyer's price LB as follows.

Definition 2.4 (buyer's price).

$$\mathrm{LB} \equiv \sup_{u_1 \in \mathrm{L}^1(\mu), u_2 \in \mathrm{L}^1(\nu), \lambda \in \mathbb{R}, \Delta_S, \Delta_X} \mathbb{E}^\mu[u_1(S_{t_1})] + \mathbb{E}^\nu[u_2(S_{t_2})] + \lambda \mathrm{VIX}$$

such that

$$u_1(s_1) + u_2(s_2) + \lambda\sqrt{x} + \Delta_S(s_1, x)(s_2 - s_1) + \Delta_X(s_1, x)\left(-\frac{2}{\Delta}\ln\left(\frac{s_2}{s_1}\right) - x\right) \leq \left(\sqrt{x} - K\right)^+ \quad \forall\, (s_1, s_2, x) \in I_1 \times I_2 \times I_X.$$

Interpretation at t_2. This semistatic super-replication consists in holding statically t_1- and t_2-European payoffs with market prices $\mathbb{E}^\mu[u_1(S_{t_1})]$ and $\mathbb{E}^\nu[u_2(S_{t_2})]$, a VIX future with market price VIX, and delta hedging at t_1 on the spot and on a forward log-contract with price x. The intrinsic value of this portfolio is greater than the payoff $(\sqrt{x} - K)^+$. If somebody offers this VIX option at a price p above UB, the arbitrage can be locked in by selling this option and going long in the above super-replication:

$$u_1(s_1) + u_2(s_2) + \lambda\sqrt{x} + \Delta_S(s_1, x)(s_2 - s_1) + \Delta_X(s_1, x)\left(-\frac{2}{\Delta}\ln\left(\frac{s_2}{s_1}\right) - x\right) - \left(\sqrt{x} - K\right)^+ + (p - \mathrm{UB}) \geq 0 \quad \forall\, (s_1, s_2, x) \in I_1 \times I_2 \times I_X.$$

As an illustration, note that from Jensen's inequality, we have the no-arbitrage condition

$$\mathrm{VIX} \leq \sigma_{12}. \tag{2.3}$$

This inequality corresponds to the pathwise super-hedging

$$-\frac{1}{2\sigma_{1,2}}\left(-\frac{2}{\Delta}\ln\frac{s_2}{s_1} - x\right) - \frac{1}{2\sigma_{1,2}}\frac{2}{\Delta}\ln\frac{s_2}{s_1} + \frac{\sigma_{1,2}}{2} \geq \sqrt{x} \quad \forall\, (s_1, s_2, x) \in (\mathbb{R}_+)^3,$$

which involves a constant delta hedging $-1/2\sigma_{1,2}$ in the forward log-contract, a static hedging in t_1 and t_2 log-contracts with payoffs $\ln s_1/\Delta\sigma_{1,2}$, $-\ln s_2/\Delta\sigma_{1,2}$, and a cash strategy $\sigma_{1,2}/2$. By taking the expectation of the left-hand side with respect to $\mathbb{P} \in \mathcal{M}(\mu, \nu, \mathrm{VIX})$, the fair value of this super-replicating portfolio is $\sigma_{1,2}$.

Interpretation at t_1. As the VIX option expires at t_1, it may sound strange to unwind our super-hedging strategy only at t_2. In particular, by setting $x = \mathrm{VIX}_{t_1}^2$, $s_2 = S_{t_2}$, $s_1 = S_{t_1}$ in inequality (2.2), and by taking the conditional expectation with respect to the random variables $S_{t_1} = s_1$ and $\mathrm{VIX}_{t_1} = \mathrm{vix}_{t_1}$, we obtain

$$u_1(s_1) + \mathbb{E}^{\mathbb{P}}\left[u_2(S_{t_2})|(S_{t_1}, \mathrm{VIX}_{t_1}) = (s_1, \mathrm{vix}_{t_1})\right] + \lambda \mathrm{vix}_{t_1} \geq (\mathrm{vix}_{t_1} - K)^+$$

for all measures $\mathbb{P} \in \mathcal{P}(I_1 \times I_2 \times I_X)$ such that $\mathbb{E}^{\mathbb{P}}[S_{t_2}|(S_{t_1}, \mathrm{VIX}_{t_1}) = (s_1, \mathrm{vix}_{t_1})] = s_1$ and $\mathbb{E}^{\mathbb{P}}[-(2/\Delta)\ln(S_{t_2}/S_{t_1})|(S_{t_1}, \mathrm{VIX}_{t_1}) = (s_1, \mathrm{vix}_{t_1})] = \mathrm{vix}_{t_1}^2$. This means that the arbitrage can be locked at t_1 by exercising at t_1 both the European option with payoff $u_1(s_1)$ and the VIX future and selling the European option with payoff $u_2(s_2)$ and maturity t_2 at the market price $\mathbb{E}^{\mathbb{P}}[u_2(S_{t_2})|S_{t_1} = s_1, \mathrm{VIX}_{t_1} = \mathrm{vix}_{t_1}]$. This second interpretation indicates that the seller's price can be defined as follows.

Definition 2.5 (seller's price).

$$\mathrm{UB}' \equiv \inf_{u_1, u_2, \lambda} \mathbb{E}^{\mu}[u_1(S_{t_1})] + \mathbb{E}^{\nu}[u_2(S_{t_2})] + \lambda \mathrm{VIX}$$

such that for all $s_1 \in I_1$, for all $\mathrm{vix}_{t_1} \in I_X$,

$$u_1(s_1) + \mathbb{E}^{\mathbb{P}}[u_2(S_{t_2})] + \lambda \mathrm{vix}_{t_1} \geq (\mathrm{vix}_{t_1} - K)^+ \quad \forall\, \mathbb{P} \in \mathcal{P}_{s_1, \mathrm{vix}_{t_1}}, \tag{2.4}$$

where $\mathcal{P}_{s_1,\mathrm{vix}_{t_1}}$ is the set of probability measures on I_2 such that $\mathbb{E}^{\mathbb{P}}[S_{t_2}] = s_1$ and $-\frac{2}{\Delta}\mathbb{E}^{\mathbb{P}}\left[\ln\left(\frac{S_{t_2}}{s_1}\right)\right] = \mathrm{vix}_{t_1}^2$.

Proposition 2.6. *The two definitions of seller's price coincide:* $\mathrm{UB}' = \mathrm{UB}$.

Similarly, $\mathrm{LB}' = \mathrm{LB}$.

Proof. (i) $\mathrm{UB}' \leq \mathrm{UB}$. By setting $x = \mathrm{vix}_{t_1}^2$, $s_2 = S_{t_2}$ in inequality (2.2) and by taking the expectation with respect to $\mathbb{P} \in \mathcal{P}_{s_1,\mathrm{vix}_{t_1}}$, we obtain

$$u_1(s_1) + \mathbb{E}^{\mathbb{P}}[u_2(S_{t_2})] + \lambda \mathrm{vix}_{t_1} \geq (\mathrm{vix}_{t_1} - K)^+$$

from which we conclude the following.

(ii) $\mathrm{UB}' \geq \mathrm{UB}$. Let us take a feasible solution u_1, u_2, λ of (2.4) (a solution exists, taking $u_1 = u_2 = 0$, $\lambda = 1$) such that $\mathbb{E}^{\mu}[u_1(S_{t_1})] + \mathbb{E}^{\nu}[u_2(S_{t_2})] + \lambda \mathrm{VIX} \leq \mathrm{UB}' + \delta$. By taking the infimum over $\mathbb{P} \in \mathcal{P}_{s_1,\mathrm{vix}_{t_1}}$, we get

$$u_1(s_1) + \inf_{\mathbb{P} \in \mathcal{P}_{\mathrm{vix}_{t_1}, s_1}} \mathbb{E}^{\mathbb{P}}[u_2(S_{t_2})] + \lambda \mathrm{vix}_{t_1} \geq (\mathrm{vix}_{t_1} - K)^+ \quad \forall\, (s_1, \mathrm{vix}_{t_1}) \in I_1 \times I_X. \tag{2.5}$$

We have for all $s_1 \in I_1$, for all $\mathrm{vix}_{t_1} \in I_X$, that

$$\inf_{\mathbb{P}\in\mathcal{P}_{\mathrm{vix}_{t_1},s_1}} \mathbb{E}^{\mathbb{P}}\Big[u_2(S_{t_2})\Big] = \inf_{\mathbb{P}\in M_+(I_2)} \sup_{\Delta_S,\Delta_X} \mathbb{E}^{\mathbb{P}}\Big[u_2(S_{t_2}) + \Delta_X(s_1,\mathrm{vix}_{t_1})\left(-\frac{2}{\Delta}\ln\frac{S_{t_2}}{s_1} - \mathrm{vix}_{t_1}^2\right) + \Delta_S(s_1,\mathrm{vix}_{t_1})(S_{t_2}-s_1)\Big]$$
$$= \sup_{\Delta_S,\Delta_X} \inf_{\mathbb{P}\in M_+(I_2)} \mathbb{E}^{\mathbb{P}}\Big[u_2(S_{t_2}) + \Delta_X(s_1,\mathrm{vix}_{t_1})\left(-\frac{2}{\Delta}\ln\frac{S_{t_2}}{s_1} - \mathrm{vix}_{t_1}^2\right) + \Delta_S(s_1,\mathrm{vix}_{t_1})(S_{t_2}-s_1)\Big]$$
$$(2.6) \qquad = \sup_{\Delta_S,\Delta_X}\left\{\Delta_X(s_1,\mathrm{vix}_{t_1})\left(\frac{2}{\Delta}\ln s_1 - \mathrm{vix}_{t_1}^2\right) - \Delta_S(s_1,\mathrm{vix}_{t_1})s_1\right\}$$

such that

$$(2.7) \qquad u_2(s_2) + \Delta_X(s_1,\mathrm{vix}_{t_1})\left(-\frac{2}{\Delta}\ln s_2\right) + \Delta_S(s_1,\mathrm{vix}_{t_1})s_2 \geq 0 \quad \forall\ s_2 \in I_2.$$

In the second equality, we have used a strong duality result (which is proved in Theorem 3.1 below as an application of the Fenchel–Rockafellar duality). By taking a maximizing sequence $(\Delta_S^{(k)}, \Delta_X^{(k)})$ of the supremum in (2.6) satisfying (2.7), we obtain from (2.5) that

$$u_1(s_1) + \Delta_X^{(k)}(s_1,\mathrm{vix}_{t_1})\left(\frac{2}{\Delta}\ln s_1 - \mathrm{vix}_{t_1}\right) - \Delta_S^{(k)}(s_1,\mathrm{vix}_{t_1})s_1 + \epsilon_k + \lambda\mathrm{vix}_{t_1}$$
$$= u_1(s_1) + \inf_{\mathbb{P}\in\mathcal{P}_{\mathrm{vix}_{t_1},s_1}} \mathbb{E}^{\mathbb{P}}\Big[u_2(S_{t_2})\Big] + \lambda\mathrm{vix}_{t_1} \geq (\mathrm{vix}_{t_1} - K)^+$$

with the sequence $(\epsilon_k)_k \in \mathbb{R}_+$ and $\epsilon_k \underset{k\to\infty}{\to} 0$. By adding (2.7), we get

$$u_1(s_1) + u_2(s_2) + \Delta_X^{(k)}(s_1,\mathrm{vix}_{t_1})\left(-\frac{2}{\Delta}\ln\frac{s_2}{s_1} - \mathrm{vix}_{t_1}\right) + \Delta_S^{(k)}(s_1,\mathrm{vix}_{t_1})(s_2-s_1) + \lambda\mathrm{vix}_{t_1} + \epsilon_k \geq (\mathrm{vix}_{t_1} - K)^+.$$

Then $(u_1^k \equiv u_1 + \epsilon_k, u_2, \Delta_S^{(k)}, \Delta_X^{(k)}, \lambda)$ is a feasible solution of UB, and $\mathrm{UB} \leq \mathrm{UB}' + \delta + \epsilon_k$. Since δ was arbitrary, we conclude by taking the limit $\epsilon_k \underset{k\to\infty}{\to} 0$. ■

3. Duality and a new martingale optimal transport.

Theorem 3.1 (duality). *Assume that μ, ν are probability measures on I_1 and I_2, respectively, such that $\mathcal{M}(\mu,\nu,\mathrm{VIX})$ is nonempty. Then,*

$$\mathrm{UB} = \max_{\mathbb{P}\in\mathcal{M}(\mu,\nu,\mathrm{VIX})} \mathbb{E}^{\mathbb{P}}[(\mathrm{VIX}_{t_1} - K)^+],$$
$$\mathrm{LB} = \min_{\mathbb{P}\in\mathcal{M}(\mu,\nu,\mathrm{VIX})} \mathbb{E}^{\mathbb{P}}[(\mathrm{VIX}_{t_1} - K)^+].$$

This is part of the proof that we have a max and not only a sup, meaning that the seller's price (resp., buyer's price) is attained by a martingale measure, i.e., a model, calibrated to the t_1 and t_2 vanilla smiles and to the VIX future. It will be seen in section 4 that this property holds also in the unbounded setting $I_1 = I_2 = I_X = \mathbb{R}_+$. As highlighted in the introduction, the construction of such a model is not at all obvious. This result justifies the interpretation of UB (resp., LB) as the upper (resp., lower) bound of a VIX call option within the class of arbitrage-free models consistent with t_1, t_2 vanillas and the VIX future.

A similar duality result was established in [2] in a discrete and in [16] in a continuous-time setting (see also [14]). Dropping the martingale constraint and the constraint on the VIX future, this optimization problem degenerates into a classical optimal transportation problem. Recently, some variants in optimal transport have been studied. In [23], the authors introduce an optimal transportation with capacity constraints, which consists in minimizing a cost among joint densities $\mathbb{P}$ with marginals μ and ν and under the *capacity* constraint

$$\mathbb{P}(s_1, s_2) \leq \bar{\mathbb{P}}(s_1, s_2)$$

for some prior joint density $\bar{\mathbb{P}}(s_1, s_2)$. In [7], optimal transports with congestion are considered. Our variant involves a martingale constraint as in [2, 16] but also a *nonlinear* constraint $\mathrm{VIX} = \mathbb{E}^{\mathbb{P}}\left[\sqrt{\mathbb{E}^{\mathbb{P}}_{t_1}[-(2/\Delta)\ln S_{t_2}/S_{t_1}]}\right]$. As a crucial step, by introducing a delta hedging on a forward log-contract, this problem has been converted into a linear programming problem that can be solved with a simplex algorithm (see section 5). Note that a similar trick is used in [5] for converting a quantile hedging approach into a super-replication problem.

Proof. We follow closely the proof of Kantorovich duality in [30]. Let $E = C_b(I_1 \times I_2 \times I_X)$ be the set of all bounded continuous functions on $I_1 \times I_2 \times I_X$. By Riesz's theorem, its dual can be identified with the space of regular Radon measures: $E^* = M(I_1 \times I_2 \times I_X)$. Then, we introduce two convex functionals Θ, Ξ on E:

$$\Theta(u) = \begin{cases} 0 & \text{if } u(s_1, s_2, x) \geq (\sqrt{x} - K)^+, \\ +\infty & \text{else,} \end{cases}$$

$$\Xi(u) = \begin{cases} \mathbb{E}^{\mu}[u_1] + \mathbb{E}^{\nu}[u_2] + \lambda \mathrm{VIX} & \text{if } u(s_1, s_2, x) = U(s_1, s_2, x), \\ +\infty & \text{else,} \end{cases}$$

with $U(s_1, s_2, x) \equiv u_1(s_1) + u_2(s_2) + \lambda\sqrt{x} + \Delta_S(s_1, x)(s_2 - s_1) + \Delta_X(s_1, x)(-\frac{2}{\Delta}\ln(\frac{s_2}{s_1}) - x)$.

For $u_0 \equiv 1 + \max_{x \in I_x} (\sqrt{x} - K)^+$, we have $\Xi(u_0) < \infty$, $\Theta(u_0) = 0$, and Θ is continuous at u_0. The assumptions in the Fenchel–Rockafellar duality theorem are then satisfied, and therefore (see, e.g., Theorem 1.9 in [30])

$$\inf_{u \in E} [\Theta(u) + \Xi(u)] = \max_{\pi \in E^*} [-\Theta^*(-\pi) - \Xi^*(\pi)], \tag{3.1}$$

where Θ^* and Ξ^* are the Legendre–Fenchel transforms of Θ, Ξ, respectively. Let us compute both sides of (3.1). The left-hand side is $\inf_{u \in E}[\Theta(u) + \Xi(u)] = \widetilde{\mathrm{UB}} \equiv \inf_{u_1, u_2, \Delta_S, \Delta_X \in \Phi, \lambda}$ $\{\mathbb{E}^{\mu}[u_1(S_{t_1})] + \mathbb{E}^{\nu}[u_2(S_{t_2})] + \lambda \mathrm{VIX}\}$, where Φ is the space of functions $(u_1, u_2, \Delta_S, \Delta_X)$ that satisfy the inequality (2.2) and are bounded and continuous. The Legendre transform of Θ is, for all $\pi \in E^*$,

$$\Theta^*(-\pi) \equiv \sup_{u \in E} \left\{ -\int u d\pi - \Theta(u) \right\} = - \inf_{u(s_1, s_2, x) \geq (\sqrt{x} - K)^+} \int u d\pi;$$

therefore

$$\Theta^*(-\pi) = \begin{cases} -\int(\sqrt{x}-K)^+ d\pi & \text{if } \pi \in M_+(I_1 \times I_2 \times I_X), \\ +\infty & \text{else.} \end{cases}$$

Similarly, the Legendre transform of Ξ is, for all $\pi \in E^*$,

$$\Xi^*(\pi) \equiv \sup_{u\in E} \int u d\pi - \Xi(u) = \sup_{u_1,u_2,\lambda,\Delta_S,\Delta_X} \left\{ \int U(s_1,s_2,x)d\pi - \mathbb{E}^\mu[u_1] - \mathbb{E}^\nu[u_2] - \lambda \mathrm{VIX} \right\};$$

therefore

$$\Xi^*(\pi) = \begin{cases} 0 & \text{if } \pi \in \mathcal{M}(\mu,\nu,\mathrm{VIX}), \\ +\infty & \text{else.} \end{cases}$$

Putting everything together in (3.1), we get $\widetilde{\mathrm{UB}} = \max_{\mathbb{P}\in\mathcal{M}(\mu,\nu,\mathrm{VIX})} \mathbb{E}^{\mathbb{P}}[(\mathrm{VIX}_{t_1}-K)^+]$. As $\widetilde{\mathrm{UB}} \geq \mathrm{UB}$ and $\mathrm{UB} \geq \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu,\mathrm{VIX})} \mathbb{E}^{\mathbb{P}}[(\mathrm{VIX}_{t_1}-K)^+]$ by weak duality, we can conclude. The lower bound can be obtained similarly by replacing $(\sqrt{x}-K)^+$ above by $-(\sqrt{x}-K)^+$. ∎

4. Analytical bound. In this section, we will give an (a priori) nonoptimal explicit upper bound. Define the following set of probability measures on $\mathbb{R}_+$:

$$\mathcal{P}_{1,2} = \mathcal{P}_{1,2}(\mathrm{VIX},\sigma_{1,2}^2) \equiv \{\mathbb{P} \in \mathcal{P}_{\mathbb{R}_+} : \ \mathbb{E}^{\mathbb{P}}[X] = \mathrm{VIX}, \mathbb{E}^{\mathbb{P}}[X^2] = \sigma_{1,2}^2\},$$

where X is the identity on $\mathbb{R}_+$. Since for any $\mathbb{P} \in \mathcal{M}(\mu,\nu,\mathrm{VIX})$ we have $\mathbb{E}^{\mathbb{P}}[\mathrm{VIX}_{t_1}] = \mathrm{VIX}$ and $\mathbb{E}^{\mathbb{P}}[\mathrm{VIX}_{t_1}^2] = \mathbb{E}^{\mathbb{P}}[-\frac{2}{\Delta}\log\frac{S_{t_2}}{S_{t_2}}] = \sigma_{1,2}^2$, it is clear that

$$\overline{\mathrm{UB}} \equiv \sup_{\mathbb{P}\in\mathcal{P}_{1,2}} \mathbb{E}^{\mathbb{P}}[(X-K)^+] \tag{4.1}$$

defines an upper bound $\overline{\mathrm{UB}} \geq \mathrm{UB}$.

Problem (4.1) is a typical example of an "extremal moment problem": maximizing the expectation of a certain payoff under the constraint that some (here, the first and second) moments of the underlying distribution are fixed. This type of problem has been widely studied in the actuarial science literature; for the case of piecewise affine payoffs, see [6, 12, 13, 22] and the comprehensive review by Hürlimann [21]. Problem (4.1) can be solved with the so-called *majorant polynomial method* (see [21]), which consists in looking for a quadratic polynomial $q(x) = ax^2+bx+c \geq (x-k)^+$ for all $x \geq 0$, and a finitely supported distribution $\overline{\mathbb{P}}$ such that all the atoms of $(X-K)^+$ under $\overline{\mathbb{P}}$ are simultaneously atoms of $q(X)$ (that is, $(X-K)^+ = q(X)$, $\overline{\mathbb{P}}$-a.s.). If such q and $\overline{\mathbb{P}}$ are found, then it is clear that

$$\begin{aligned} E^{\mathbb{P}}[q(X)] &\geq E^{\mathbb{P}}[(X-K)^+] \qquad \forall\, \mathbb{P} \in \mathcal{P}_{1,2}, \\ E^{\overline{\mathbb{P}}}[q(X)] &= E^{\overline{\mathbb{P}}}[(X-K)^+]. \end{aligned} \tag{4.2}$$

Since the left-hand side $E^{\mathbb{P}}[q(X)] = a\sigma_{1,2}^2 + b\mathrm{VIX} + c$ depends only on the fixed values VIX and $\sigma_{1,2}$, then $\overline{\mathbb{P}}$ necessarily maximizes $\mathbb{E}[(X-K)^+]$ over $\mathcal{P}_{1,2}$.

It is well known that the solution to (4.1) is given by a biatomic distribution $\overline{\mathbb{P}}(dx) = p\delta_{x_0}(dx) + (1-p)\delta_{x_1}(dx)$. The first part of the following proposition reports the explicit solution for p, x_0, x_1, as it can be found in, e.g., Jansen, Haezendonck, and Goovaerts [22]; the second part translates the dual solution (a, b, c) in terms of the super-replicating strategy in (2.2). We provide a short proof for the sake of completeness.

Proposition 4.1 (analytical upper bound). *A maximizer for problem* (4.1) *is given by the biatomic measure*

$$\overline{\mathbb{P}}(dx) = p\delta_{x_0}(dx) + (1-p)\delta_{x_1}(dx) \tag{4.3}$$

with

$$\begin{cases} x_0 = K - I; \quad x_1 = K + I; \quad p = \frac{K - \mathrm{VIX} + I}{2I(K)} & \text{if } K \geq K^*, \\ x_0 = 0; \quad x_1 = \frac{\sigma_{1,2}^2}{\mathrm{VIX}}; \quad p = \frac{\sigma_{1,2}^2 - \mathrm{VIX}^2}{\sigma_{1,2}^2} & \text{if } K < K^*, \end{cases}$$

where $K^* = \frac{\sigma_{1,2}^2}{2\mathrm{VIX}}$ *and* $I = I(K) = \sqrt{\sigma_{1,2}^2 - \mathrm{VIX}^2 + (\mathrm{VIX} - K)^2}$. *The value of the problem is then*

$$\overline{\mathrm{UB}}(\mathrm{VIX}, \sigma_{1,2}^2) = \mathbb{E}^{\overline{\mathbb{P}}}[(X-K)^+] = \begin{cases} \frac{1}{2}(\mathrm{VIX} - K + I), & K \geq K^*, \\ \mathrm{VIX} - K\frac{\mathrm{VIX}^2}{\sigma_{1,2}^2}, & K < K^*. \end{cases} \tag{4.4}$$

This bound is attained by the semistatic super-replication

$$\overline{u}_1(s_1) = -\frac{2}{\Delta}\Delta_X \ln\frac{s_1}{S_0} + \nu, \quad \overline{u}_2(s_2) = \frac{2}{\Delta}\Delta_X \ln\frac{s_2}{S_0}, \quad \overline{\Delta}_S = 0, \quad \overline{\Delta}_X(s_1, x) = \Delta_X,$$

where for $K \geq K^*$,

$$\Delta_X = -\frac{1}{4I}, \quad \nu = \frac{-2KI + \sigma_{1,2}^2 - 2K(\mathrm{VIX} - K)}{4I}, \quad \overline{\lambda} = \frac{1}{2} - \frac{K}{2I},$$

and for $K \leq K^*$,

$$\Delta_X = -K\left(\frac{\mathrm{VIX}}{\sigma_{1,2}^2}\right)^2, \quad \nu = 0, \quad \overline{\lambda} = 1 - 2K\frac{\mathrm{VIX}}{\sigma_{1,2}^2}.$$

This implies that the pathwise super-replication

$$\overline{u}_1(s_1) + \overline{u}_2(s_2) + \overline{\lambda}\sqrt{x} + \overline{\Delta}_S(s_1, x)(s_2 - s_1) + \overline{\Delta}_X(s_1, x)\left(-\frac{2}{\Delta}\ln\left(\frac{s_2}{s_1}\right) - x\right) \geq \left(\sqrt{x} - K\right)^+ \tag{4.5}$$

holds for all $(s_1, s_2, \sqrt{x}) \in \mathbb{R}_+^3$.

The measure $\overline{\mathbb{P}}$ in (4.3) is known to be the unique solution to (4.1) among the finitely supported distributions contained in $\mathcal{P}_{1,2}$; see [21]. Note that the bound $\overline{\mathrm{UB}}$ depends only on the $(t = 0)$ market values VIX, $\sigma_{1,2}$, and it is well defined if and only if the arbitrage-free condition $\mathrm{VIX} \leq \sigma_{1,2}$ holds.

Proof. The conditions $(X - K)^+ = q(X)$, $\overline{\mathbb{P}}$-a.s., and $q(x) \geq (x - K)^+$ for all $x \geq 0$ (cf. the discussion preceding (4.2)) translate into $ax^2 + bx + c = (x - K)^+$, $\overline{\mathbb{P}}$-a.s., and

$\inf_{x\in\mathbb{R}_+}\{ax^2+bx+c-(x-K)^+\}=0$ is attained at x_0, x_1. This implies the algebraic equations

$$px_0+(1-p)x_1=m_1, \quad px_0^2+(1-p)x_1^2=m_2, \tag{4.6}$$

$$ax_0^2+bx_0+c=(x_0-K)^+, \quad ax_1^2+bx_1+c=(x_1-K)^+, \tag{4.7}$$

$$2ax_0+b=1_{x_0\geq K} \text{ if } x_0\neq 0, \quad 2ax_1+b=1_{x_1\geq K} \tag{4.8}$$

with $a>0$. The solutions (p,x_0,x_1,a,b,c) are reported below.

Solutions of (4.6), (4.7), (4.8):

(i) $x_1\geq K$ and $x_0\leq K$. We get

$$p=\frac{K-m_1+I}{2I}, \quad x_0=K-I, \quad x_1=K+I,$$
$$a=\frac{1}{4I}, \quad b=\frac{1}{2}-\frac{K}{2I}, \quad c=\frac{-2KI+m_2-2K(m_1-K)}{4I}$$

with $I\equiv\sqrt{(K-m_1)^2+m_2-m_1^2}$. Finally, $\mathbb{E}^{\overline{\mathbb{P}}}[(X-K)^+]=\frac{1}{2}(m_1-K+I)$. This solution is valid if and only if $x_0\geq 0$, which is equivalent to our condition $K\geq K^*\equiv\frac{m_2}{2m_1}$.

(ii) $x_1\geq K$ and $x_0=0$. We get

$$p=\frac{m_2-m_1^2}{m_2}, \quad x_0=0, \quad x_1=\frac{m_2}{m_1},$$
$$a=\frac{Km_1^2}{m_2^2}, \quad b=\frac{m_2-2Km_1}{m_2}, \quad c=0.$$

Moreover, $\mathbb{E}^{\overline{\mathbb{P}}}[(x-K)^+]=m_1-K\frac{m_1^2}{m_2}$. This solution is valid (i.e., $ax^2+bx+c\geq(x-K)^+$ for all $x\in\mathbb{R}_+$) if $K\leq K^*$.

Finally, the coefficients Δ_X, λ, and ν in the super-replication strategy are deduced from the values of a,b,c. It is easy to check that inequality (4.5) holds for all $(s_1,s_2,\sqrt{x})\in\mathbb{R}_+^3$ (then in particular for all $(s_1,s_2,\sqrt{x})\in I_1\times I_2\times I_X$) and that

$$\mathbb{E}^{\mu}[u_1(S_{t_1})]+\mathbb{E}^{\nu}[u_2(S_{t_2})]+\lambda\text{VIX}=a\,\sigma_{1,2}^2+b\,\text{VIX}+c=\mathbb{E}^{\overline{\mathbb{P}}}[(X-K)^+]=\overline{\text{UB}}. \qquad ■$$

Remark 4.2. Our solution has explicitly used that the VIX index at t_1 can be written as the difference of two log-contracts, i.e., $\ln s_2-\ln s_1$. This is not the case if we use the (market) trapezoidal approximation appearing in the definition of the VIX index.

Analogously, it is possible to formulate an explicit lower bound.

Proposition 4.3 (analytical lower bound). *A lower bound* $\overline{\text{LB}}$ *is given by*

$$\overline{\text{LB}}=(\text{VIX}-K)^+.$$

This bound is attained by the sub-replication for $K\geq\text{VIX}$,

$$u_1(s_1)=0, \quad u_2(s_2)=0, \quad \Delta_S=0, \quad \Delta_X=0, \quad \lambda=0, \tag{4.9}$$

and for $K<\text{VIX}$,

$$u_1(s_1)=-K, \quad u_2(s_2)=0, \quad \Delta_S=0, \quad \Delta_X=0, \quad \lambda=1. \tag{4.10}$$

Proof. As the payoff $(\mathrm{VIX}_{t_1} - K)^+$ is convex, this result is obvious from Jensen's inequality. This lower bound corresponds to a null VIX implied volatility. The determination of $\overline{\mathbb{P}}$ (as a Bernoulli distribution), defined as $\overline{\mathrm{LB}} = \mathbb{E}^{\overline{\mathbb{P}}}[(X-K)^+]$, copycats the proof of the upper bound and consists in looking at the solutions of system (4.6)–(4.8) with $a \leq 0$. We will obtain an explicit characterization only in the case $K < m_1$ and $K \geq \frac{m_2}{m_1}$.

(i) $x_0 = 0$ and $x_1 \leq K$:

$$p = \frac{m_2 - m_1^2}{m_2}, \quad x_0 = 0, \quad x_1 = \frac{m_2}{m_1},$$
$$a = 0, \quad b = 0, \quad c = 0.$$

Finally, the lower bound is $\mathbb{E}^{\overline{\mathbb{P}}}[(X-K)^+] = 0$. This solution is valid if and only if $K \geq \frac{m_2}{m_1}$.

(ii) $x_0 = K$ and $x_1 \geq K$:

$$p = \frac{m_2 - m_1^2}{I^2}, \quad x_0 = K, \quad x_1 = m_1 \frac{K - \frac{m_2}{m_1}}{K - m_1},$$
$$a = 0, \quad b = 1, \quad c = -K.$$

Finally, the lower bound is $\mathbb{E}^{\overline{\mathbb{P}}}[(X-K)^+] = m_1 - K$. This solution is valid if and only if $K < m_1$. ■

Remark 4.4. Note that $\overline{\mathbb{P}}$ is not unique. Below, we list some additional solutions.

(iii) $0 \leq x_0 \leq K$ and $x_1 = K$. We get

$$p = \frac{(K - m_1)^2}{I^2}, \quad x_0 = m_1 \frac{K - \frac{m_2}{m_1}}{K - m_1}, \quad x_1 = K,$$
$$a = 0, \quad b = 0, \quad c = 0.$$

Finally, the lower bound is $\mathbb{E}^{\overline{\mathbb{P}}}[(X-K)^+] = 0$. This solution is valid if and only if $K \geq \frac{m_2}{m_1}$ when $m_2 > m_1^2$, and $K > m_1$ when $m_2 = m_1^2$.

(iv) $x_0 \geq K$ and $x_1 \geq K$:

$$p = \frac{(m_1 - x_1)^2}{m_2 - 2m_1 x_1 + x_1^2}, \quad x_0 = \frac{m_2 - m_1 x_1}{m_1 - x_1},$$
$$a = 0, \quad b = 1, \quad c = -K.$$

This solution is valid if and only if $K < m_1$ when $m_2 > m_1^2$, and $m_1 \leq K$ when $m_2 = m_1^2$.

(v) $x_0 \leq K$ and $x_1 \leq K$:

$$p = \frac{(m_1 - x_1)^2}{m_2 - 2m_1 x_1 + x_1^2}, \quad x_0 = \frac{m_2 - m_1 x_1}{m_1 - x_1},$$
$$a = 0, \quad b = 0, \quad c = 0.$$

This solution is valid if and only if $K \geq \frac{m_2}{m_1}$ when $m_2 > m_1^2$, and $m_1 \leq K$ when $m_2 = m_1^2$.

4.1. The extreme points of $\mathcal{P}_{1,2}$. It is clear that $\mathcal{P}_{1,2}$ is a convex set. We know from [31] the following result (where $\#A$ denotes the cardinality of the set A).

Theorem 4.5 (see Winkler [31]). *$\mathbb{P} \in \mathcal{P}_{1,2}$ is an extreme point of $\mathcal{P}_{1,2}$ if and only if $\#supp(\mathbb{P}) \leq 3$.*

In particular, the measure $\overline{\mathbb{P}}$ in Proposition 4.1 is an extreme point of $\mathcal{P}_{1,2}$ that maximizes the linear function $\mathbb{P} \mapsto \mathbb{E}^{\mathbb{P}}[(X-K)^+]$, and we know from [21] that it is the unique maximizer among finitely supported distributions. Therefore, by Theorem 4.5, $\overline{\mathbb{P}}$ is also the unique maximizer among the extreme points of $\mathcal{P}_{1,2}$.

The main result of this section is the following.

Proposition 4.6. *$\overline{\mathbb{P}}$ is the unique maximizer of $\mathbb{E}^{\mathbb{P}}[(X-K)^+]$ over $\mathcal{P}_{1,2}$.*

We are interested in describing the set $\mathcal{P}_{1,2}$ in terms of its extreme points (which we denote by $ext(\mathcal{P}_{1,2})$), since this will allow us to exploit the uniqueness of the maximizer $\mathbb{P}$ among the elements of $ext(\mathcal{P}_{1,2})$. In the finite-dimensional setting, a classical theorem by Minkowski (see [26, section 1]) asserts that every point of a convex compact subset of $\mathbb{R}^n$ can be written as a convex combination of a finite number of its extreme points (furthermore, a sharper version of the same theorem due to Carathéodory establishes that these points can be chosen to be at most $n+1$; see again [26]). The analogue of this result for general (infinite-dimensional) topological vector spaces is given by Choquet's theorem, which we recall here in the setting of metrizable spaces (see [9]; see also [26, section 3]).

Theorem 4.7 (see Choquet [9]). *Let K be a metrizable compact convex subset of some locally convex topological vector space E. Then for every $x \in K$, there exists a Borel probability measure τ_x on K with $supp(\tau_x) \subseteq ext(K)$ such that for every continuous linear functional f on E,*

$$f(x) = \int f(y)\tau_x(dy).$$

The set $\mathcal{P}_{1,2}$ is not closed with respect to the weak topology: we will need to consider its closure (but this will not introduce any substantial difficulty). Note that Markov's inequality implies $\mathbb{P}(X > R) \leq \frac{\mathrm{VIX}}{R}$ for all $\mathbb{P} \in \mathcal{P}_{1,2}$ and $R > 0$. It follows that the set $\mathcal{P}_{1,2}$ is tight, hence relatively compact from Prohorov's theorem, and its closure $\overline{\mathcal{P}_{1,2}}$ is a compact set.

Let us quickly inspect the set $\overline{\mathcal{P}_{1,2}}$. The following classical result (see, e.g., Billingsley [4, Chap. 1]) will be useful.

Lemma 4.8. *Let Y_n be a sequence of ($\mathbb{R}^n$-valued) random variables such that $Y_n \overset{D}{\to} Y$ for some random variable Y. Then the following hold:*

(i) $\mathbb{E}[|Y|] \leq \liminf_n \mathbb{E}[|Y_n|]$.

(ii) *If the Y_n are uniformly integrable, then $\mathbb{E}[Y_n] \to \mathbb{E}[Y]$.*

We note that Lemma 4.8 contains Fatou's lemma and Vitali's theorem (when the Y_n and Y are defined on the same probability space).

It follows from Lemma 4.8(i) that the limit $\mathbb{P}$ of a weakly converging subsequence $\mathbb{P}_n$ in $\mathcal{P}_{1,2}$ is such that $\mathbb{E}^{\mathbb{P}}[X^2] \leq \sigma_{1,2}^2$. Using integration to the limit for uniformly integrable sequences as in Lemma 4.8(ii), we see that the limit $\mathbb{P}$ also satisfies the condition $\mathbb{E}^{\mathbb{P}}[X] = \mathrm{VIX}$. Therefore

$$\overline{\mathcal{P}_{1,2}} \subseteq \{\mathbb{P} \in \mathcal{P}_{\mathbb{R}_+} : \ \mathbb{E}^{\mathbb{P}}[X] = \mathrm{VIX}, \ \mathbb{E}^{\mathbb{P}}[X^2] \leq \sigma_{1,2}^2\}. \tag{4.11}$$

Using again the uniform integrability of the sequence $((X_n - K)^+)_n$, where the X_n are random variables with law $\mathbb{P}_n$, one sees that the map $\mathbb{P} \mapsto \mathbb{E}^{\mathbb{P}}[(X-K)^+]$ is continuous over $\overline{\mathcal{P}_{1,2}}$. It is then clear that

$$\overline{\text{UB}} = \sup_{\mathbb{P}\in\overline{\mathcal{P}_{1,2}}} \mathbb{E}^{\mathbb{P}}[(X-K)^+].$$

The proof of Proposition 4.6 relies on the following two lemmas.

Lemma 4.9. *For every $K > 0$ and $\text{VIX} > 0$, the map $\sigma^2 \mapsto \overline{\text{UB}}(\sigma^2)$ defined by (4.4) is strictly increasing on $[\text{VIX}, \infty)$.*

The proof of Lemma 4.9 is elementary and is postponed to the appendix. The property of the upper bound $\overline{\text{UB}}$ stated in Lemma 4.9 is illustrated by the curves in Figure 2.

Let us now consider the set $ext(\overline{\mathcal{P}_{1,2}})$: it is obvious from the definition of extreme points that $ext(\overline{\mathcal{P}_{1,2}}) \cap \mathcal{P}_{1,2} \subseteq ext(\mathcal{P}_{1,2})$.

Lemma 4.10. *The measure $\overline{\mathbb{P}}$ in Proposition 4.1 is contained in $ext(\overline{\mathcal{P}_{1,2}})$, and it is the unique global maximizer of $\mathbb{E}^{\mathbb{P}}[(X-K)^+]$ among the elements of $ext(\overline{\mathcal{P}_{1,2}})$.*

Proof. We claim (and prove later) that

$$ext(\overline{\mathcal{P}_{1,2}}) = ext(\mathcal{P}_{1,2}) \cup \left(ext(\overline{\mathcal{P}_{1,2}}) \cap \mathcal{P}^c_{1,2}\right). \tag{4.12}$$

In other words, if $\mathbb{P} \in ext(\overline{\mathcal{P}_{1,2}})$, then either $\mathbb{P} \in ext(\mathcal{P}_{1,2})$ or, because of (4.11), $\mathbb{P}(x^2) = \int_{\mathbb{R}_+} x^2 \mathbb{P}(dx) < \sigma^2$ (where we denote $\sigma^2 = \sigma^2_{1,2}$ for simplicity in the rest of the proof). Moreover, it follows from (4.12) that $ext(\overline{\mathcal{P}_{1,2}})$ contains the global maximizer $\overline{\mathbb{P}}$ defined in (4.3).

Now consider $\tilde{\mathbb{P}} \in ext(\overline{\mathcal{P}_{1,2}})$ that maximizes $\mathbb{E}^{\mathbb{P}}[(X-K)^+]$, so that $\mathbb{E}^{\tilde{\mathbb{P}}}[(X-K)^+] = \sup_{\mathbb{P}\in\overline{\mathcal{P}_{1,2}}} \mathbb{E}^{\mathbb{P}}[(X-K)^+] = \overline{\text{UB}}(\sigma^2)$. Let us assume $\tilde{\mathbb{P}} \notin \mathcal{P}_{1,2}$. This entails $\tilde{\mathbb{P}}(x^2) =: a^2 < \sigma^2$; therefore it follows from Lemma 4.9 that

$$\mathbb{E}^{\tilde{\mathbb{P}}}[(X-K)^+] \leq \sup_{\mathbb{P}\in\mathcal{P}_{1,2}(\text{VIX},a^2)} \mathbb{E}^{\mathbb{P}}[(X-K)^+] = \overline{\text{UB}}(a^2) < \overline{\text{UB}}(\sigma^2),$$

contradicting the optimality of $\tilde{\mathbb{P}}$. Therefore, according to the decomposition (4.12), we must have $\tilde{\mathbb{P}} \in ext(\mathcal{P}_{1,2})$. By the uniqueness of the maximizer $\overline{\mathbb{P}}$ inside $ext(\mathcal{P}_{1,2})$, we conclude that $\tilde{\mathbb{P}} = \overline{\mathbb{P}}$.

Proof of (4.12). We need to prove that $ext(\overline{\mathcal{P}_{1,2}}) \cap \mathcal{P}_{1,2} = ext(\mathcal{P}_{1,2})$. Let us prove the inclusion $ext(\mathcal{P}_{1,2}) \subseteq ext(\overline{\mathcal{P}_{1,2}}) \cap \mathcal{P}_{1,2}$. Consider $\mathbb{P} \in ext(\mathcal{P}_{1,2})$, $\mathbb{P}_1$ and $\mathbb{P}_2 \in \overline{\mathcal{P}_{1,2}}$, and $\theta \in (0,1)$ such that

$$\theta\mathbb{P}_1 + (1-\theta)\mathbb{P}_2 = \mathbb{P}. \tag{4.13}$$

By (4.11), we have $\mathbb{P}_i(x^2) \leq \sigma^2$ for $i = 1, 2$. If $\mathbb{P}_i(x^2) < \sigma^2$ for some i, then $(\theta\mathbb{P}_1 + (1-\theta)\mathbb{P}_2)(x^2) < \sigma^2$, which is impossible because $\mathbb{P} = \theta\mathbb{P}_1 + (1-\theta)\mathbb{P}_2 \in \mathcal{P}_{1,2}$. Thus we must have $\mathbb{P}_i(x^2) = \sigma^2$, which implies $\mathbb{P}_i \in \mathcal{P}_{1,2}$ for $i = 1, 2$. Since $\mathbb{P} \in ext(\mathcal{P}_{1,2})$, by the definition of extreme point, the identity (4.13) implies $\mathbb{P}_1 = \mathbb{P}_2 = \mathbb{P}$, which in turn entails $\mathbb{P} \in ext(\overline{\mathcal{P}_{1,2}})$, hence proving the claimed inclusion. The converse inclusion $ext(\overline{\mathcal{P}_{1,2}}) \cap \mathcal{P}_{1,2} \subseteq ext(\mathcal{P}_{1,2})$ is obvious, and (4.12) is proved. ■

Proof of Proposition 4.6. Let $f(\mathbb{P}) = \mathbb{E}^{\mathbb{P}}[(X-K)^+]$. Assume that $\mathbb{P}_0 \in \overline{\mathcal{P}_{1,2}}$ maximizes f over $\overline{\mathcal{P}_{1,2}}$, so that $f(\mathbb{P}_0) = \overline{\text{UB}}$. By Choquet's theorem (4.7), there exists a probability measure τ_0 supported by $ext(\overline{\mathcal{P}_{1,2}})$ such that

$$\begin{aligned}\overline{\mathrm{UB}} = f(\mathbb{P}_0) &= \int_{ext(\overline{\mathcal{P}_{1,2}})} f(\mathbb{P})\tau_0(d\mathbb{P}) \\ &= \int_{ext(\overline{\mathcal{P}_{1,2}})\cap\{\mathbb{P}:f(\mathbb{P})=\overline{\mathrm{UB}}\}} f(\mathbb{P})\tau_0(d\mathbb{P}) + \int_{ext(\overline{\mathcal{P}_{1,2}})\cap\{\mathbb{P}:f(\mathbb{P})<\overline{\mathrm{UB}}\}} f(\mathbb{P})\tau_0(d\mathbb{P}).\end{aligned} \tag{4.14}$$

It follows from (4.14) that the measure τ_0 cannot charge the set $\{f(\mathbb{P}) < \overline{\mathrm{UB}}\}$, for otherwise we would get the contradiction $\overline{\mathrm{UB}} < \overline{\mathrm{UB}}$. On the other hand, Lemma 4.10 implies that $ext(\overline{\mathcal{P}_{1,2}}) \cap \{f(\mathbb{P}) = \overline{\mathrm{UB}}\} = \{\overline{\mathbb{P}}\}$, where $\overline{\mathbb{P}}$ is defined in Proposition 4.1, and it follows that τ_0 is the Dirac mass on the singleton $\{\overline{\mathbb{P}}\}$. Applying Theorem 4.7 again, we have

$$\int \varphi \, d\mathbb{P}_0 = \int \varphi \, d\overline{\mathbb{P}}$$

for all continuous bounded functions $\varphi : \mathbb{R}_+ \to \mathbb{R}$; therefore we conclude $\mathbb{P}_0 = \overline{\mathbb{P}}$. Note that we have shown that $\overline{\mathbb{P}}$ is the unique maximizer over the set $\overline{\mathcal{P}_{1,2}}$; in particular, Proposition 4.6 follows. ■

4.2. Optimality. It is natural to ask whether there are cases in which the (a priori non-optimal) analytical upper bound $\overline{\mathrm{UB}}$ is actually optimal, that is, we have $\overline{\mathrm{UB}} = \mathrm{UB}$. The main contribution of this section is to formulate conditions on the (market) data μ, ν, VIX that are equivalent to $\overline{\mathrm{UB}} = \mathrm{UB}$. The resulting condition on the marginals μ, ν is an order relationship. We will state this result in the more general setting of unbounded intervals $I_1 = I_2 = I_X = \mathbb{R}_+$.

With a slight abuse of notation, we still denote $(S_{t_1}, S_{t_2}, \mathrm{VIX}_{t_1})$ the canonical process on $\mathbb{R}^3_+$. We assume that μ and ν are two probability measures on $\mathbb{R}_+$ having the same finite mean $\int_{\mathbb{R}_+} x\mu(dx) = \int_{\mathbb{R}_+} y\nu(dy) = S_0$ and satisfying the following condition:

$$\log(S_1) \in L^q(\mu), \qquad \log(S_2) \in L^q(\nu) \qquad \text{for some } q > 1. \tag{4.15}$$

We define the set of martingale measures $\mathcal{M}(\mu,\nu,\mathrm{VIX}) \subset \mathcal{P}(\mathbb{R}^3_+)$ as in (2.1), and we set $\mathrm{UB}(\mu,\nu,\mathrm{VIX}) \equiv \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu,\mathrm{VIX})} \mathbb{E}^{\mathbb{P}}[(\mathrm{VIX}_{t_1} - K)^+]$, which is now the upper bound for the price of the VIX option over all models $\mathbb{P}$ with support on $\mathbb{R}^3_+$ compatible with market data. We denote $\overline{\mathrm{UB}}\left(\mathrm{VIX}, \sigma^2_{1,2}(\mu,\nu)\right)$ the value of the analytical upper bound defined in (4.4), with $\sigma^2_{1,2} = \sigma^2_{1,2}(\mu,\nu) = -\frac{2}{\Delta}(\int_{\mathbb{R}_+} \log(s_2)d\nu(s_2) - \int_{\mathbb{R}_+} \log(s_1)d\mu(s_1))$.

Using the conditional Jensen's inequality, the second conditional expectation constraint in (2.1) implies that for every $\mathbb{P} \in \mathcal{M}(\mu,\nu,\mathrm{VIX})$,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[(\mathrm{VIX}^2_{t_1})^q\right] &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}\left[-\frac{2}{\Delta}\log\frac{S_2}{S_1}\middle| S_1, \mathrm{VIX}_{t_1}\right]^q\right] \\ &\le \mathbb{E}^{\mathbb{P}}\left[\left|-\frac{2}{\Delta}\log\frac{S_2}{S_1}\right|^q\right] \le \frac{2^{2q-1}}{\Delta^q}\left(\mathbb{E}^{\nu}[|\log S_2|^q] + \mathbb{E}^{\mu}[|\log S_1|^q]\right) =: m_q(\mu,\nu),\end{aligned} \tag{4.16}$$

where the last term is finite under condition (4.15). Using the estimate above, one can prove that the set $\mathcal{M}(\mu,\nu,\mathrm{VIX})$ is compact with respect to the weak topology on $\mathcal{P}(\mathbb{R}^3_+)$. As a consequence, the existence of a maximizer for $\mathrm{UB}(\mu,\nu,\mathrm{VIX})$ (which we already know in the compact case from Theorem 3.1) also holds in the unbounded setting.

Proposition 4.11. *Assume that $\mu,\nu\in\mathcal{P}(\mathbb{R}_+)$ have the same finite mean and satisfy condition (4.15). Then, the set $\mathcal{M}(\mu,\nu,\mathrm{VIX})$ is compact with respect to the weak topology. Consequently, if $\mathcal{M}(\mu,\nu,\mathrm{VIX})$ is nonempty, the supremum is attained: there exists $\mathbb{P}^*\in\mathcal{M}(\mu,\nu,\mathrm{VIX})$ such that $\mathrm{UB}(\mu,\nu,\mathrm{VIX})=\mathbb{E}^{\mathbb{P}^*}[(\mathrm{VIX}_{t_1}-K)^+]$.*

Proof. We denote $\mathcal{M}=\mathcal{M}(\mu,\nu,\mathrm{VIX})$ for simplicity. If $\mathcal{M}$ is empty, there is nothing to prove. Then, let us assume that $\mathcal{M}\neq\emptyset$.

Step 1 *($\mathcal{M}$ is relatively compact).* Fix $\varepsilon>0$, and consider compact sets $K_1,K_2\subset\mathbb{R}_+$ such that $\mu(K_1^c)\le\varepsilon/3$ and $\nu(K_2^c)\le\varepsilon/3$. For every $\mathbb{P}\in\mathcal{M}$ and $R>0$, one has

$$\begin{aligned}\mathbb{P}((K_1\times K_2\times[0,R])^c)&\le\mathbb{P}\left(\{S_1\in K_1^c\}\cup\{S_2\in K_2^c\}\cup\{\mathrm{VIX}_{t_1}>R\}\right)\\&\le\mu(K_1^c)+\nu(K_2^c)+\mathbb{P}(\mathrm{VIX}_{t_1}>R)\le\frac{2}{3}\varepsilon+\frac{\mathrm{VIX}}{R},\end{aligned}$$

where the last step follows from Markov's inequality. It follows that $\mathcal{M}$ is tight, hence relatively compact, from Prohorov's theorem.

Step 2 *($\mathcal{M}$ is closed).* Let $\mathbb{P}_n$ be a sequence in $\mathcal{M}$ converging weakly to some $\tilde{\mathbb{P}}\in\mathcal{P}(\mathbb{R}_+^3)$. It is clear that the marginal laws of S_1 and S_2 under $\tilde{\mathbb{P}}$ are given by μ and ν (since expectations of continuous bounded functions are preserved by weak convergence). Consider a sequence (S_1^n,S_2^n,X^n) of random variables (resp., a random variable $(\tilde S_1,\tilde S_2,\tilde X)$) with laws $\mathbb{P}_n$ (resp., $\tilde{\mathbb{P}}$). It follows from the definition of $\mathcal{M}$ that $\sup_n\mathbb{E}[(X^n)^2]=\sup_n\mathbb{E}^{\mathbb{P}_n}[\mathrm{VIX}_{t_1}^2]=\sigma_{1,2}^2(\mu,\nu)$. Therefore, the X^n are uniformly integrable, and Lemma 4.8 implies that $\mathbb{E}^{\tilde{\mathbb{P}}}[\mathrm{VIX}_{t_1}]=\mathbb{E}[\tilde X]=\lim_{n\to\infty}\mathbb{E}[X^n]=\mathrm{VIX}$. Note that using the sublinear growth of the payoff function $x\mapsto(x-K)^+$, the same argument also shows that the map $\mathbb{P}\mapsto\mathbb{E}^{\mathbb{P}}[(\mathrm{VIX}_{t_1}-K)^+]$ is continuous on $\mathcal{M}$ with respect to weak topology.

Let us now discuss the constraints on conditional expectations. First of all, the estimate (4.16) implies $\sup_n\mathbb{E}[(X^n)^{2q}]\le m_q(\mu,\nu)$; therefore the $(X^n)^2$ are uniformly integrable. We are going to use the fact that if $c:\mathbb{R}_+^3\to\mathbb{R}$ is a function satisfying

$$c(s_1,s_2,x)\le C(1+s_1+s_2+|\log(s_1)|+|\log(s_2)|+x^2)\tag{4.17}$$

for some constant $C>0$ and all $(s_1,s_2,x)\in\mathbb{R}_+^3$, the sequence $c(S_1^n,S_2^n,X^n)$ is also uniformly integrable. This follows from the finite mean condition for μ and ν, the integrability of the function $s\mapsto\log(s)$ under the same two measures, and the uniform integrability of the $(X^n)^2$.

Recall that $\mathbb{P}\in\mathcal{P}(\mathbb{R}_+^3)$ is such that $\mathbb{E}^{\mathbb{P}}[S_2|S_1,\mathrm{VIX}_{t_1}]=S_1$ and $\mathbb{E}^{\mathbb{P}}[-\frac{2}{\Delta}\log\frac{S_2}{S_1}|S_1,\mathrm{VIX}_{t_1}]=\mathrm{VIX}_{t_1}^2$ if and only if

$$\mathbb{E}^{\mathbb{P}}[(S_2-S_1)\phi(S_1,\mathrm{VIX}_{t_1})]=0\tag{4.18}$$

and

$$\mathbb{E}^{\mathbb{P}}\left[\left(-\frac{2}{\Delta}\log\frac{S_2}{S_1}-\mathrm{VIX}_{t_1}^2\right)\phi(S_1,\mathrm{VIX}_{t_1})\right]=0\tag{4.19}$$

for every bounded continuous function ϕ. For every such ϕ, the functions $c_1(s_1,s_2,x)=(s_2-s_1)\phi(s_1,x)$ and $c_2(s_1,s_2,x)=(-\frac{2}{\Delta}\log\frac{s_2}{s_1}-x^2)\phi(s_1,x)$ fulfill condition (4.17). Then, using Lemma 4.8, we see that the conditions (4.18) and (4.19) for the measure $\tilde{\mathbb{P}}$ follow from $\mathbb{E}^{\tilde{\mathbb{P}}}[c_i(S_1, S_2, \mathrm{VIX}_{t_1})] = \mathbb{E}[c_i(S_1^n, S_2^n, X^n)] = \lim_{n\to\infty} \mathbb{E}[c_i(S^n, S_2^n, X^n)] = 0$.

Finally, the existence of a maximizer $\mathbb{P}^*$ follows from the compactness of $\mathcal{M}$ and the continuity of $\mathbb{P} \mapsto \mathbb{E}^{\mathbb{P}}[(\mathrm{VIX}t_1 - K)^+]$. ■

We now state the main result for this section.

Theorem 4.12. *Let μ, ν be probability measures in $\mathcal{P}(\mathbb{R}_+)$ with the same finite mean, satisfying condition (4.15), and such that $\sigma_{1,2}^2(\mu,\nu) > 0$. Let $\mathrm{VIX} \in (0, \sigma_{1,2}(\mu,\nu)]$. The following are equivalent:*

(i) $\mathrm{UB}(\mu, \nu, \mathrm{VIX}) = \overline{\mathrm{UB}}(\mathrm{VIX}, \sigma_{1,2}^2(\mu,\nu))$.

(ii) *There exist two couples of measures (μ_0, ν_0) and (μ_1, ν_1) on $\mathbb{R}_+$ such that*

$$\mu = p\mu_0 + (1-p)\mu_1, \qquad \nu = p\nu_0 + (1-p)\nu_1, \tag{4.20}$$

and

$$\int f(z, \log(z))\, \nu_i(dz) \geq \int f\left(y, \log(y) - \frac{\Delta}{2} x_i^2\right) \mu_i(dy), \qquad i = 0, 1, \tag{4.21}$$

for every convex function $f : \mathbb{R}^2 \to \mathbb{R}$, where the coefficients p, x_0, x_1 and the explicit value of $\overline{\mathrm{UB}}\left(\mathrm{VIX}, \sigma_{1,2}^2\right)$ have been given in Proposition 4.1.

Note that condition (4.21) entails that the measures (μ_i, ν_i) increase in the convex order (take $f(y,z) = \tilde{f}(y)$ with $\tilde{f}$ convex in (4.21)); hence by (4.20) so do the measures (μ, ν). An example of a class of measures μ, ν satisfying conditions (4.20), (4.21) is given in Remark 4.13 below (and a concrete instance within this class in Example 4.14).

Proof. We prove the implication (i) ⇒ (ii).

Let $\mathbb{P}^* \in \mathcal{M}(\mu, \nu, \mathrm{VIX})$ be a maximizer in (3.1), so that $\mathbb{E}^{\mathbb{P}^*}[(\mathrm{VIX}_{t_1} - K)^+] = \mathrm{UB} = \overline{\mathrm{UB}}$ (where we denote the two bounds by UB and $\overline{\mathrm{UB}}$ for simplicity). Denote λ the marginal law of the third component VIX_{t_1} under $\mathbb{P}^*$. It follows from $\mathbb{E}^{\mathbb{P}^*}[(\mathrm{VIX}_{t_1} - K)^+] = \mathbb{E}^{\lambda}[(\mathrm{VIX}_{t_1} - K)^+]$ and the uniqueness result in Proposition 4.6 that λ is equal to $\overline{\mathbb{P}}$, the biatomic distribution defined in Proposition 4.1. Then, denote $\mathbb{Q}_i$ the law of (S_{t_1}, S_{t_2}) conditional on $\mathrm{VIX}_{t_1} = x_i$ for $i = 0, 1$. Similarly, denote μ_i (resp., ν_i) the law of S_{t_1} (resp., S_{t_2}) conditional on $\mathrm{VIX}_{t_1} = x_i$ for $i = 0, 1$. The condition $S_{t_1} \overset{\mathbb{P}^*}{\sim} \mu$ is equivalent to

$$p\mu_0 + (1-p)\mu_1 = \mu,$$

and correspondingly, $S_{t_2} \overset{\mathbb{P}^*}{\sim} \nu$ gives $p\nu_0 + (1-p)\nu_1 = \nu$. On the other hand, the two conditions $\mathbb{E}^{\mathbb{P}^*}[S_{t_2} | S_{t_1}, \mathrm{VIX}_{t_1}] = S_{t_1}$ and $\mathbb{E}^{\mathbb{P}^*}\left[-\frac{2}{\Delta} \ln \frac{S_{t_2}}{S_{t_1}} \middle| S_{t_1}, \mathrm{VIX}_{t_1}\right] = (\mathrm{VIX}_{t_1})^2$ in (2.1) are equivalent to the (four) conditions $\mathbb{E}^{\mathbb{Q}_i}[S_{t_2} | S_{t_1}] = S_{t_1}$ and $\mathbb{E}^{\mathbb{Q}_i}[\ln S_{t_2} | S_{t_1}] = \ln S_{t_1} - \frac{\Delta}{2} x_i^2$ for $i = 0, 1$. It is immediate to see that the latter conditions are equivalent to the following: for every $i \in \{0, 1\}$, the $\mathbb{R}^2$-valued process

$$(Z_2, Z_1) = \left(\begin{pmatrix} S_{t_2} \\ \ln S_{t_2} \end{pmatrix}, \begin{pmatrix} S_{t_1} \\ \ln S_{t_1} - \frac{\Delta}{2} x_i^2 \end{pmatrix} \right) \tag{4.22}$$

is a martingale under $\mathbb{Q}_i$, that is, $\mathbb{E}^{\mathbb{Q}_i}[Z_2|Z_1] = Z_1$. Using Jensen's inequality for conditional expectation, one obtains that for every convex function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathbb{E}^{\mathbb{Q}_i}[f(Z_1)] = \int_{\mathbb{R}^2} f(z_1)\mu_i^{Z_1}(dz_1) \le \int_{\mathbb{R}^2} f(z_2)\mu_i^{Z_2}(dz_2) = \mathbb{E}^{\mathbb{Q}_i}[f(Z_2)]. \tag{4.23}$$

Since the components of each vector Z_j, $j = 1, 2$, are deterministic functions of S_j, one can rewrite condition (4.23) in terms of the laws μ_i and ν_i defined above, finally obtaining (4.21).

(ii) $\Rightarrow$ (i). Let μ_i, ν_i, $i = 0, 1$, be measures in $\mathcal{P}(\mathbb{R}_+)$ satisfying condition (4.21). In terms of the process (Z_1, Z_2) defined above, condition (4.21) can be rewritten as $\mathbb{E}^{\mu_i}[f(Z_1)] = \mathbb{E}^{\mu_i}[f(S_{t_1}, \ln S_{t_1} - \frac{\Delta}{2}x_i^2)] \le \mathbb{E}^{\nu_i}[f(S_{t_2}, \ln S_{t_2})] = \mathbb{E}^{\nu_i}[f(Z_2)]$ for every convex function f on $\mathbb{R}^2$ and $i = 0, 1$. Strassen's theorem [29] then ensures that the set $\mathcal{M}_i = \{\mathbb{Q} \in \mathcal{P}(\mathbb{R}_+^2) : S_{t_1} \overset{\mathbb{Q}}{\sim} \mu_i, S_{t_2} \overset{\mathbb{Q}}{\sim} \nu_i, \mathbb{E}^{\mathbb{Q}}[Z_2|Z_1] = Z_1\}$ is not empty. Then, pick an element $\mathbb{Q}_i$ in $\mathcal{M}_i$, and define $\mathbb{P}^* \in \mathcal{P}(\mathbb{R}_+^3)$ by

$$\mathbb{P}^*(ds_1, ds_2, dx) \equiv \sum_{i=0,1} p_i\, \delta_{x_i}(dx)\, \mathbb{Q}_i(ds_1, ds_2), \tag{4.24}$$

where we have set $p_0 = p$ and $p_1 = 1 - p$. By construction, we have $\mathbb{P}^* \in \mathcal{M}(\mu, \nu, \mathrm{VIX})$. Indeed, the marginal law of S_{t_1} under $\mathbb{P}^*$ is $\sum_{i=0,1} p_i\mu_i = \mu$, where the last equality follows from (4.20). Analogously, the marginal law of S_{t_2} under $\mathbb{P}^*$ is ν. Having chosen $\mathbb{Q}_i \in \mathcal{M}_i$, it is immediate to verify that $\mathbb{P}^*$ also satisfies the conditional expectation constraints in (2.1). Finally, we can check that we have $\mathrm{UB} \ge \mathbb{E}^{\mathbb{P}^*}[(\mathrm{VIX}_{t_1} - K)^+] = \mathbb{E}^{\overline{\mathbb{P}}}[(\mathrm{VIX}_{t_1} - K)^+] = \overline{\mathrm{UB}}$. Recalling that $\mathrm{UB} \le \overline{\mathrm{UB}}$, this concludes the proof. ■

Remark 4.13. Let T_u^0 and T_u^1 be two strictly increasing C^1 functions on $\mathbb{R}_+$ such that $T_u^i(x) > x$ and $x\partial_x T_u^i(x) \ge T_u^i(x)$, $i = 0, 1$. It can be shown that the equations

$$\frac{T_u^i(x) - x}{T_u^i(x) - T_d^i(x)} \ln \frac{T_d^i(x)}{x} + \frac{x - T_d^i(x)}{T_u^i(x) - T_d^i(x)} \ln \frac{T_u^i(x)}{x} = -\Delta \frac{x_i^2}{2}, \quad i = 0, 1 \tag{4.25}$$

(where x_0, x_1 are reported in the proof of Proposition 4.1), uniquely define two strictly increasing C^1 functions T_d^0 and T_d^1 on $\mathbb{R}_+$ such that $T_d^i(x) \le x$. Setting $q^i(s) = \frac{s - T_d^i(s)}{T_u^i(s) - T_d^i(s)}$, denote

$$\tilde{\mathbb{P}}(ds_1, ds_2, dx) = \sum_{i=0,1} p_i\, \delta_{x_i}(dx) \left(q^i(s_1)\delta_{T_u^i(s_1)}(ds_2) + (1 - q^i(s_1))\delta_{T_d^i(s_1)}(ds_2) \right) \mu(ds_1).$$

It is easy to verify that $\tilde{\mathbb{P}}$ is a probability measure on $\mathbb{R}_+^3$ such that $S_{t_1} \overset{\tilde{\mathbb{P}}}{\sim} \mu$, $\mathbb{E}^{\tilde{\mathbb{P}}}[\mathrm{VIX}_{t_1}] = \mathrm{VIX}$, $\mathbb{E}^{\tilde{\mathbb{P}}}[S_{t_2}|S_{t_1}, \mathrm{VIX}_{t_1}] = S_{t_1}$, and $\mathbb{E}^{\tilde{\mathbb{P}}}[-\frac{2}{\Delta} \log \frac{S_{t_2}}{S_{t_1}} | S_{t_1}, \mathrm{VIX}_{t_1}] = \mathrm{VIX}_{t_1}^2$. The marginal law $\tilde{\nu}$ of S_{t_2} under $\tilde{\mathbb{P}}$ is the push-forward of μ via the maps T_u^0, T_u^1 and the corresponding T_d^0, T_d^1; more precisely,

$$d\tilde{\nu}(x) = \sum_{i=0}^{1} p_i \left(d\mu^1(T_d^{i,-1}(x)) \frac{T_u^i(s) - s}{(T_u^i(s) - T_d^i(s))\partial_s T_d^i(s)} \Big|_{s = T_d^{i,-1}(x)} \right. \\ \left. + d\mu(T_u^{i,-1}(x)) \frac{s - T_d^i(s)}{(T_u^i(s) - T_d^i(s))\partial_s T_u^i(x)} \Big|_{s = T_u^{i,-1}(x)} \right). \tag{4.26}$$

Denote $\mathcal{P}(\mu)$ the class of probability measures on $\mathbb{R}_+$ that can be written as $\tilde{\nu}$ for some strictly increasing C^1 maps T_u^0 and T_u^1 such that $T_u^i(x) > x$ and $x\partial_x T_u^i(x) \geq T_u^i(x)$, $i = 0, 1$. If $\nu \in \mathcal{P}(\mu)$, it is clear that $\tilde{\mathbb{P}}$ belongs to $\mathcal{M}(\mu, \nu, \mathrm{VIX})$ and that $\mathbb{E}^{\tilde{\mathbb{P}}}[(\mathrm{VIX}_{t_1} - K)^+] = \overline{\mathrm{UB}}$, so that $\tilde{\mathbb{P}}$ is a maximizer and we have $\mathrm{UB} = \overline{\mathrm{UB}}$. Of course, this situation falls within the scope of Theorem 4.12: in this case, the measures μ and ν satisfy condition (ii) in Theorem 4.12, where $\mu_0 = \mu_1 = \mu$ and the $\{\nu_i\}_{i=0,1}$ are given by the term inside the brackets on the right-hand side of (4.26).

Note that the class of measures $\tilde{\mathbb{P}}$ defined above is a specific instance of the more general construction considered in the proof of Theorem 4.12. While it seems difficult, for given μ and ν, to show existence of two maps T_u^0 and T_u^1 which ensure that $\nu \in \mathcal{P}(\mu)$,[1] Theorem 4.12 provides a condition on μ, ν for the optimality of $\overline{\mathrm{UB}}$ that is both explicitly checkable and more general (for example, the joint law $p\,\mathbb{Q}_0 + (1-p)\mathbb{Q}_1$ of (S_{t_1}, S_{t_2}) induced by $\mathbb{P}^*$ in (4.24) may admit a density with respect to the Lebesgue measure on $\mathbb{R}_+^2$, which is not the case for the conditional law of (S_{t_1}, S_{t_2}) under $\tilde{\mathbb{P}}$).

Example 4.14 (example of $\nu \in \mathcal{P}(\mu)$). As an example of density in $\mathcal{P}(\mu)$, we take a log-normal distribution with mean 1 and volatility 0.2 (maturity = 1 year) for μ and $T_u^i(x) = \alpha_u^i x$ with $\alpha_u^i > 1$. Then,

$$\nu(\alpha_u^0, \alpha_u^1)(s) = \sum_{i=0}^{1} p_i \left(\mu^1 \left(\frac{s}{\alpha_d^i} \right) \frac{\alpha_u^i - 1}{(\alpha_u^i - \alpha_d^i)\alpha_d^i} + \mu \left(\frac{s}{\alpha_u^i} \right) \frac{1 - \alpha_d^i}{(\alpha_u^i - \alpha_d^i)\alpha_u^i} \right),$$

where $\alpha_d^i < 1$ is uniquely fixed by

$$\frac{\alpha_u^i - 1}{\alpha_u^i - \alpha_d^i} \ln \alpha_d^i + \frac{1 - \alpha_d^i}{\alpha_u^i - \alpha_d^i} \ln \alpha_u^i = -\Delta \frac{x_i^2}{2}, \quad i = 0, 1.$$

We have chosen VIX $= 0.3$, $\sigma_{1,2} = 0.4$. For $K = \mathrm{VIX}$ and for the upper bound, we have $x_0 = 0.035$, $x_1 = 0.567$, $p = 0.5$ (recall that the expressions for p, x_0, and x_1 can be found in Proposition 4.1). We have chosen $\alpha_u^0 = 1.1$, $\alpha_u^1 = 1.2$ (giving $\alpha_d^0 = 0.998$, $\alpha_d^1 = 0.863$). In Figure 1, we have plotted the densities μ and $\nu(\alpha_u^0, \alpha_u^1)$.

We conclude this section with an extension of the duality result of section 3 to the present setting of unbounded intervals. The super-replication strategies of the VIX option are now defined for all $(s_1, s_2, x) \in \mathbb{R}_+^3$.

Proposition 4.15. *Under the hypotheses and condition* (ii) *of Theorem* 4.12*, the duality result of Theorem* 3.1 *holds; that is,*

$$\sup_{\mathbb{P} \in \mathcal{M}(\mu,\nu,\mathrm{VIX})} \mathbb{E}^{\mathbb{P}}[(\mathrm{VIX}_{t_1} - K)^+] = \inf_{u_1, u_2, \lambda, \Delta_S, \Delta_X} \mathbb{E}^{\mu}[u_1(S_{t_1})] + \mathbb{E}^{\nu}[u_2(S_{t_2})] + \lambda \mathrm{VIX},$$

where inf *runs over all* $u_1 \in L^1(\mathbb{R}_+, \mu)$, $u_2 \in L^1(\mathbb{R}_+, \nu)$, $\lambda \in \mathbb{R}$, *and* $\Delta_S, \Delta_X \in C_b(\mathbb{R}_+^2)$ *satisfying inequality* (2.2) *for all* $(s_1, s_2, \sqrt{x}) \in \mathbb{R}_+^3$.

[1] A similar question is considered in Hobson and Neuberger [20]. Note that in [20] only the martingale constraint $q(s)T_u(s) + (1 - q(s))T_d(s) = s$ is in force, while here the maps T_u and T_d also need to satisfy condition (4.25).

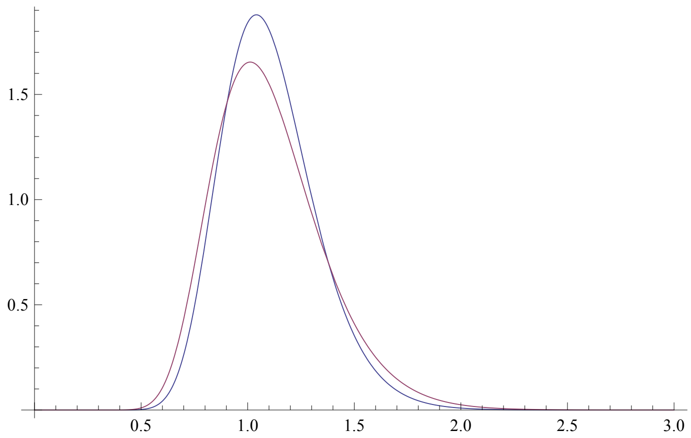

Figure 1. *Example of density ν in $\mathcal{P}(\mu)$. We have chosen $\alpha_u^0 = 1.1$, $\alpha_u^1 = 1.2$ (hence $\alpha_d^0 = 0.998$, $\alpha_d^1 = 0.863$). The blue (resp., red) curve is μ (resp., ν).*

Proof. Denote $J(u_1, u_2, \lambda) = \mathbb{E}^\mu[u_1(S_{t_1})] + \mathbb{E}^\nu[u_2(S_{t_2})] + \lambda\mathrm{VIX} = \int_{\mathbb{R}_+} u_1(s_1)d\mu(s_1) + \int_{\mathbb{R}_+} u_2(s_2)d\nu(s_2) + \lambda\mathrm{VIX}$. We have already shown in Proposition 4.1 (see (4.5)) that there exists an admissible super-replication strategy $\overline{u}_1, \overline{u}_2, \overline{\lambda}, \overline{\Delta}_S, \overline{\Delta}_X$ defined on $\mathbb{R}_+^3$ whose value is $J(\overline{u}_1, \overline{u}_2, \overline{\lambda}) = \overline{\mathrm{UB}}$. From Theorem 4.12, under condition (ii) we have $\sup \mathbb{E}^{\mathbb{P}}[(\mathrm{VIX}_{t_1} - K)^+] = \overline{\mathrm{UB}} \geq \inf_{u_1, u_2, \lambda, \Delta_S, \Delta_X} J(u_1, u_2, \lambda)$. Since by weak duality $\inf J(u_1, u_2, \lambda) \geq \sup \mathbb{E}^{\mathbb{P}}[(\mathrm{VIX}_{t_1} - K)^+]$, the equality is proved. ∎

4.3. Finite number of strikes. In our previous discussion, we assumed a continuum of t_1 and t_2 S&P 500 call options. In this section, we consider briefly the practical situation where we have a finite number of strikes. In Definitions 2.1 and 2.4, the infimum over u_1 (a similar change applies to u_2) is replaced by an infimum over variables $\{\omega_1^i\}_{i=1,\dots,n} \in \mathbb{R}$ with $u_1(s_1) = \sum_{i=1}^n \omega_1^i (s_1 - K_i)^+$. $\mathbb{E}^\mu[u_1]$ is defined by $\mathbb{E}^\mu[u_1] = \sum_{i=1}^n \omega_1^i C(t_1, K_i)$ with $C(t_1, K_i)$ the market value of the S&P 500 call option with strike K_i and maturity t_1. We define the robust seller's price of a log-contract expiring at t_2 as follows.

Definition 4.16 (seller's price of the t_2 log-contract).

$$\overline{\sigma}_2^2 = \inf_{\omega_2^i \in \mathbb{R}} \sum_{i=1}^n \omega_2^i C(t_2, K_i)$$

such that $\sum_{i=1}^n \omega_2^i (s_2 - K_i)^+ \geq -\frac{2}{t_2} \ln s_2$ *for all* $s_2 \in I_2$.

Similarly, we define the robust buyer's price of a log-contract expiring at t_1 as follows.

Definition 4.17 (buyer's price of the t_1 log-contract).

$$\underline{\sigma}_1^2 = \sup_{\omega_1^i \in \mathbb{R}} \sum_{i=1}^n \omega_1^i C(t_1, K_i)$$

such that $\sum_{i=1}^n \omega_1^i (s_1 - K_i)^+ \leq -\frac{2}{t_1} \ln s_1$ *for all* $s_1 \in I_1$.

These sub/super-replication problems have been studied in detail in [11], and one can show that these model-independent bounds are attained by arbitrage-free models, i.e., measures $\mathbb{P}_*^1$

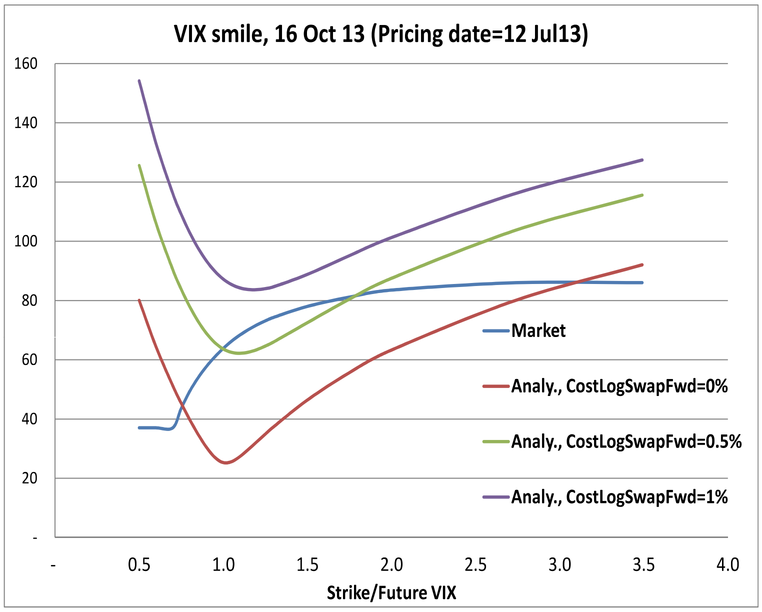

Figure 2. *Analytical upper bound versus market values for VIX smile* $t_1 = 16$ *Oct.* 13 *(pricing date* = 12 *Jul.* 13*).* $\sigma_{1,2} = 18.15\%$ $(\mathrm{VIX} = 18.05\%)$*.*

and $\mathbb{P}^2_*$ such that $\overline{\sigma}_2^2 = \mathbb{E}^{\mathbb{P}^2_*}[-\frac{2}{t_2}\ln S_{t_2}]$ and $\overline{\sigma}_1^2 = \mathbb{E}^{\mathbb{P}^1_*}[-\frac{2}{t_1}\ln S_{t_1}]$. Then Propositions 4.1 and 4.3 still hold if we replace the market value of the forward log-contract σ_{12}^2 by

$$\sigma_{12}^2 \to \frac{t_2\overline{\sigma_2}^2 - t_1\underline{\sigma_1}^2}{\Delta}.$$

5. Numerical experiments.

5.1. Analytical upper bound. We have compared the analytical upper bound $\overline{\mathrm{UB}}$ against market prices for VIX options with expiry 16 Oct. 13, pricing date = 12 Jul. 13 (see Figure 2). Surprisingly, market prices of VIX options are well above our analytical upper bound $\overline{\mathrm{UB}}$ for $K/\mathrm{VIX} > 1$. This upper bound depends on the $t = 0$ market values of the VIX future and the forward log-contract $\sigma_{1,2}$. As log-contracts are not sold on the market, they must be replicated using a strip of vanillas and therefore depend strongly on transaction costs (for low and high strikes). Note that the value of $\sigma_{1,2}$ has been properly computed using rate, dividend, and repo. In order to see the impact of transaction costs, we have added +0.5% and +1.0% to the market value of the forward log-contract $\sigma_{1.2} = 18.15\%$ (VIX = 18.05% here). As the transaction cost increases, the arbitrage opportunities evaporate as expected.

5.2. Optimal bound = analytical upper bound. In Table 1, we have computed the optimal bound for VIX options with expiry 16 Oct. 13 (pricing date = 12 Jul. 13) by numerically solving our semi-infinite linear program. This is achieved using a simplex method within a cutting-plane algorithm as described in [19]. More precisely, the European payoffs u_1 and u_2 are decomposed over a basis of call options with payoffs $(s_j - K_i)^+$ and a log-contract. This

Table 1
Numerical versus analytical upper bounds.

K/VIX	$\mathrm{UB}_{\mathrm{num}} \times 100$	$\overline{\mathrm{UB}} \times 100$
0.90	30.56	30.56
0.95	27.13	27.13
1.00	25.31	25.31
1.05	25.59	25.59
1.10	27.33	27.33
1.15	29.70	29.70

is not really an approximation in practice, as the range of liquid strikes quoted on the market is finite. We have used a fourth-order polynomial approximation for Δ_S and Δ_X in s_1 and x. Our optimal bound reads as (note that $\overline{\mathrm{UB}} \geq \mathrm{UB}_{\mathrm{num}} \geq \mathrm{UB}$)

$$\mathrm{UB}_{\mathrm{num}} \equiv \inf \sum_{j=1}^{2} \sum_{i=1}^{n} \omega_j^i C(t_j, K_i) + \sum_{i=1}^{2} \beta_i \mathbb{E}^{\mathbb{P}_i}[\ln S_{t_i}] + \lambda \mathrm{VIX}$$

such that

$$(5.1) \quad \sum_{j=1}^{2} \sum_{i=1}^{n} \omega_j^i (s_j - K_i)^+ + \sum_{i=1}^{2} \beta_i \ln s_i + \lambda \sqrt{x} + \Delta_{s,\mathrm{Poly}}(s_1, x)(s_2 - s_1) + \Delta_{X,\mathrm{Poly}}(s_1, x) \left(-\frac{2}{\Delta} \ln \left(\frac{s_2}{s_1} \right) - x \right) \geq (\sqrt{x} - K)^+ \quad \forall\, (s_1, s_2, \sqrt{x}) \in I_1 \times I_2 \times I_X$$

with $\mathbb{P}_0 = \mu$, $\mathbb{P}_1 = \nu$, $I_1 = I_2 = [0.3S_0, 2S_0]$, and $I_X = [0.1, 1]$. The infimum is taken over the variables $(\omega_j^i)_{j=1,2}^{1 \leq i \leq n}$, λ, and $(\beta_i)_{i=1,2}$ in $\mathbb{R}$. The upper bound $\mathrm{UB}_{\mathrm{num}}$ for VIX options $t_1 =$ 16 Oct. 13 (pricing date = 12 Jul. 13) with different strikes K are reported in Table 1 and compared with $\overline{\mathrm{UB}}$. $\mathrm{UB}_{\mathrm{num}}$ and $\overline{\mathrm{UB}}$ are quoted in implied volatility ($\times 100$). We have that $\overline{\mathrm{UB}} = \mathrm{UB}_{\mathrm{num}}$ are identical, indicating that the bound $\overline{\mathrm{UB}}$ seems optimal. We have checked that the numerical bound $\mathrm{UB}_{\mathrm{num}}$ is not improved by increasing the range of liquid strikes and the rank of the polynomial approximations for Δ_S and Δ_X. This result suggests that all "reasonable" market marginals $\nu \geq \mu$ (convex order) imply $\mathrm{UB} = \overline{\mathrm{UB}}$, and the optimality result proved in Theorem 4.12 still seems to be valid in this larger setting.

We have plotted in Figure 3 the optimal super-replication strategy reported in Proposition 4.1 as a function of s_1/S_0 and s_2/S_0 for $\sqrt{x} = \mathrm{VIX} = 18.05\%$ and $K/\mathrm{VIX} = 1$ (precisely, we have plotted the left-hand side of (4.5) minus $(\sqrt{x} - K)^+$, which has to remain positive for all (s_1, s_2)).

6. Summary. In this work, we have derived model-independent lower and upper bounds for VIX options. These bounds are attained by a sub/super-replicating strategy involving log-contracts and the VIX future. As a consequence, log-contracts seem to be the most relevant vanilla instruments for sub/super-hedging VIX options. We have characterized the class of

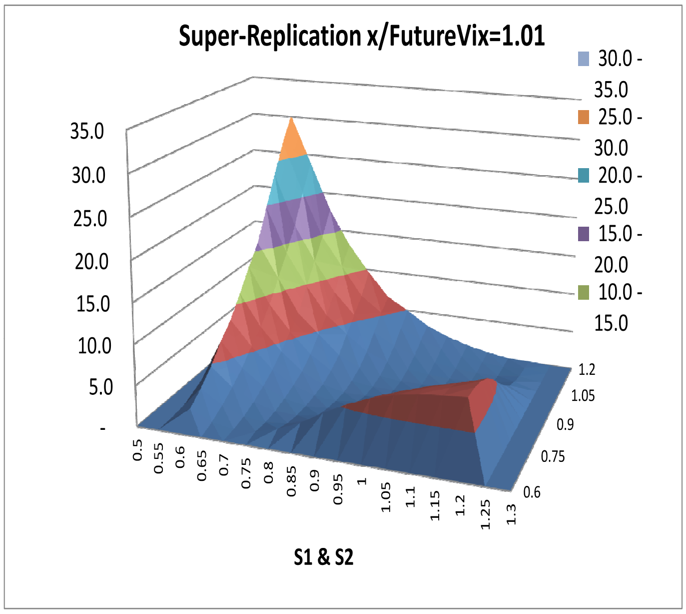

Figure 3. *Super-replication strategy for VIX smile* $t_1 = 16$ *Oct.* 13 *with strike* $K/\mathrm{VIX} = 1$ *(pricing date* $= 12$ *Jul.* 13*) as a function of* s_1/S_0 *and* s_2/S_0 *for* $\sqrt{x} = \mathrm{VIX} = 18.05\%$*. We have plotted the left-hand side of* (4.5) *minus* $(\sqrt{x} - K)^+$*, which has to remain positive for all* (s_1, s_2)*.*

marginals μ, ν for which these bounds are optimal, and we have illustrated numerically that this optimality result holds for the marginal laws implied by market data for options on the S&P 500 and the VIX future quote.

Appendix.

Proof of Lemma 4.9. Let K and VIX be fixed. Set $f(v) = \mathrm{VIX} - K\frac{\mathrm{VIX}^2}{v}$ and $g(v) = \frac{1}{2}(\mathrm{VIX} - K + \sqrt{v - \mathrm{VIX}^2 + (\mathrm{VIX} - K)^2})$, so that

$$\overline{\mathrm{UB}}(\sigma^2) = f(\sigma^2) 1_{K < K^*_\sigma} + g(\sigma^2) 1_{K \geq K^*_\sigma}$$

for every $\sigma > 0$, where $K^*_\sigma = \frac{\sigma^2}{2\mathrm{VIX}}$. Clearly, the two maps f and g are strictly increasing and satisfy $f(\sigma^2) = g(\sigma^2)$ if $K = K^*_\sigma$. Consider $\sigma_1, \sigma_2 \in \mathbb{R}_+$ with $\sigma_1 < \sigma_2$: we want to show that $\overline{\mathrm{UB}}(\sigma_1^2) < \overline{\mathrm{UB}}(\sigma_2^2)$. Obviously, $K^*_{\sigma_1} < K^*_{\sigma_2}$. Assume $K \leq K^*_{\sigma_1}$: then $\overline{\mathrm{UB}}(\sigma_1^2) = f(\sigma_1^2) < f(\sigma_2^2) = \overline{\mathrm{UB}}(\sigma_2^2)$. Analogously, if $K \geq K^*_{\sigma_2}$, we have $\overline{\mathrm{UB}}(\sigma_1^2) = g(\sigma_1^2) < g(\sigma_2^2) = \overline{\mathrm{UB}}(\sigma_2^2)$. It remains to study the case $K^*_{\sigma_1} < K < K^*_{\sigma_2}$: in this case, $\overline{\mathrm{UB}}(\sigma_1^2) = g(\sigma_1^2)$, while $\overline{\mathrm{UB}}(\sigma_2^2) = f(\sigma_2^2)$; therefore we need to show that

$$f(\sigma_2^2) = \mathrm{VIX} - K\frac{\mathrm{VIX}^2}{\sigma_2^2} > \frac{1}{2}\left(\mathrm{VIX} - K + \sqrt{\sigma_1^2 - \mathrm{VIX}^2 + (\mathrm{VIX} - K)^2}\right) = g(\sigma_1^2).$$

The previous inequality is equivalent to

$$\mathrm{VIX} + K - 2K\frac{\mathrm{VIX}^2}{\sigma_2^2} > \sqrt{\sigma_1^2 + K^2 - 2K\,\mathrm{VIX}}.$$

Using $K < \frac{\sigma_2^2}{2\mathrm{VIX}}$, it is immediate to see that the left-hand side is strictly greater than K. On the other hand, using $K > \frac{\sigma_1^2}{2\mathrm{VIX}}$, one sees that the right-hand side is strictly smaller than K, and the proof is concluded. ■

Acknowledgments. Stefano De Marco would like to thank Ismail Laachir, Claude Martini, and Giovanni Puccetti for useful insights. He is indebted to Alexandre Zhou for his interest in this topic and for stimulating discussions during an early phase of this work.

This paper is dedicated to the memory of Prof. Peter Laurence, who initiated our interest in robust model-independent bounds (see, e.g., [24]) and offered invaluable comments on Pierre Henry-Labordère's research on heat kernel expansion in mathematical finance and martingale optimal transport.

REFERENCES

[1] J. Baldeaux and A. Badran, *Consistent modelling of VIX and equity derivatives using a 3/2 plus jumps model*, Appl. Math. Finance, 21 (2014), pp. 299–312.
[2] M. Beiglböck, P. Henry-Labordère, and F. Penkner, *Model-independent bounds for option prices: A mass-transport approach*, Finance Stoch., 17 (2013), pp. 477–501.
[3] L. Bergomi, *Smile dynamics* III, Risk magazine, Oct. (2008).
[4] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley-Interscience, John Wiley & Sons, New York, 1995.
[5] B. Bouchard, R. Elie, and N. Touzi, *Stochastic target problems with controlled loss*, SIAM J. Control Optim., 48 (2009), pp. 3123–3150.
[6] H. Bühlmann, B. Gagliardi, H. Gerber, and E. Straub, *Some inequalities for stop-loss premiums*, ASTIN Bull., 9 (1977), pp. 75–83.
[7] G. Carlier, C. Jimenez, and F. Santambrogio, *Optimal transportation with traffic congestion and Wardrop equilibria*, SIAM J. Control Optim., 47 (2008), pp. 1330–1350.
[8] *The CBOE Volatility Index-VIX*, white paper, www.cboe.com/micro/vix/vixwhite.pdf.
[9] G. Choquet, *Existence et unicité des represéntations intégrales au moyen des points extrémaux dans les cônes convexes*, in Séminaire Bourbaki, Vol. 4, Société Mathématique de France, Paris, 1995, Exp. No. 139, pp. 33–47.
[10] R. Cont and T. Kokholm, *A consistent pricing model for index options and volatility derivatives*, Math. Finance, 23 (2013), pp. 248–274.
[11] M. Davis, J. Obłój, and V. Raval, *Arbitrage bounds for weighted variance swap prices*, Math. Finance, 24 (2014), pp. 821–854.
[12] F. De Vylder, *Best upper bounds for integrals with respect to measures allowed to vary under conical and integral constraints*, Insurance Math. Econom., 1 (1982), pp. 109–130.
[13] F. De Vylder and M. Goovaerts, *Upper and lower bounds on stop-loss premiums in case of known expectation and variance of the risk variable*, Mitt. Verein. Schweiz. Versicherungsmath., no. 1 (1982), pp. 149–164.
[14] Y. Dolinsky and H. M. Soner, *Martingale optimal transport and robust hedging in continuous time*, Probab. Theory Related Fields, 160 (2014), pp. 391–427.
[15] B. Dupire, *Pricing with a smile*, Risk magazine, 7 (1994), pp. 18–20.
[16] A. Galichon, P. Henry-Labordère, and N. Touzi, *A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options*, Ann. Appl. Probab., 24 (2014), pp. 312–336.
[17] J. Guyon and P. Henry-Labordère, *Being particular about smile calibration*, Risk magazine, Jan. (2012).
[18] P. Henry-Labordère, *Calibration of local stochastic volatility models to market smiles: A Monte-Carlo approach*, Risk magazine, Sept. (2009).

[19] P. Henry-Labordère, *Automated option pricing: Numerical methods*, Int. J. Theor. Appl. Finance, 16 (2013), 1350042.
[20] D. Hobson and A. Neuberger, *Robust bounds for forward start options*, Math. Finance, 22 (2012), pp. 31–56.
[21] W. Hürlimann, *Extremal moment methods and stochastic orders: Application in actuarial science*, Bol. Asoc. Mat. Venez., 15 (2008), pp. 5–110, 153–301.
[22] K. Jansen, J. Haezendonck, and M. Goovaerts, *Upper bounds on stop-loss premiums in case of known moments up to the fourth order*, Insurance Math. Econom., 5 (1986), pp. 315–334.
[23] J. Korman and R. J. McCann, *Optimal transportation with capacity constraints*, Trans. Amer. Math. Soc., 367 (2015), pp. 1501–1521.
[24] P. Laurence and T. H. Wang, *What's a basket worth?*, Risk magazine, Feb. (2004).
[25] A. Lipton, *The vol smile problem*, Risk magazine, Feb. (2002).
[26] R. R. Phelps, *Lectures on Choquet's Theorem*, 2nd ed., Springer, Berlin, Heidelberg, 2011.
[27] Y. Ren, D. Madan, and M. Qian Qian, *Calibrating and pricing with embedded local volatility models*, Risk magazine, July (2009).
[28] A. Sepp, *VIX option pricing in a jump-diffusion model*, Risk magazine, April (2008).
[29] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist., 36 (1965), pp. 423–439.
[30] C. Villani, *Topics in optimal transportation*, Grad. Stud. Math. 58, AMS, Providence, RI, 2003.
[31] G. Winkler, *Extreme points of moment sets*, Math. Oper. Res., 13 (1988), pp. 581–587.