Optimality properties of Dupire and Brunick-Shreve diffusion-type processes - extremal models for vol swaps and VIX futures given full marginals

Martin Forde

27th Oct 2025

Abstract: For a general class of $\mathcal{F}_t^X$-adapted path-dependent functionals Y_t of the canonical sample path $(X_t)_{t\in[0,T]}$ and a given family of marginals $\mu_t(dx,dy)$ for $t\in[0,T]$, we identify the process which maximizes (resp. minimizes) $\mathbb{E}(\int_0^T f(\sigma_t^2)g(X_t,Y_t)\nu(dt))$ over the space of continuous martingale Itô processes X such that $(X_t,Y_t)\sim\mu_t$ for all $t\in[0,T]$ for a concave (convex) function f, a non-negative function g and a non-negative Borel measure ν. Using a simple conditional Jensen argument we show that the extremal model is a Brunick-Shreve[BS13] diffusion-type process of the form $d\hat{X}_t=\hat{\sigma}(\hat{X}_t,\hat{Y}_t,t)d\hat{W}_t$, where $\hat{\sigma}(x,y,t)^2=\mathbb{E}(\sigma_t^2|X_t=x,Y_t=y)$ and $\hat{W}_t$ is a Brownian motion on some filtered probability space (and this model is clearly complete). As a special case this includes discrete and continuously monitored volatility swap-type payoffs of the form $\int_0^{T_1}|\sigma_t|\nu(dt)$ for any maturity $T_1\in[0,T]$. Similarly, the process which maximizes or minimizes $\mathbb{E}(\int_0^T f(\sigma_t^2)g(X_t)\nu(dt))$ over continuous martingale Itô processes X such that $X_t\sim\mu_t$ for all $t\in[0,T]$ is the associated Dupire local volatility model, and we construct the appropriate super(sub)-hedge using dynamic trading in X and European-type contracts on X. We also give a positive answer to the problem discussed in Acciaio&Guyon[AG20], namely that an ε-optimal model for the largest price of a VIX future with maturity T given marginals for S at all maturities in $[0,T+\Delta]$ (or a finite set of tradeable options) is given by a Brunick-Shreve type-mimicking model of the form $dS_t=S_t\sigma_{loc}(S_t,t)dW_t$ for $t\in[0,T]$ and $dS_t=S_t\sigma_{loc}(S_t,S_t/S_T,t)dW_t$ for $t\in(T,T+\Delta]$, where T is the maturity of the VIX future (and similarly for forward-starting options). This sheds further light on some long-standing questions about when a calibrated Dupire local volatility model gives rise to extremal prices for certain types of volatility derivatives, and complements the counterexample result in [AG20] which shows that the usual Dupire model does not maximize the price of a VIX future.*

1. Background and literature review

The mimicking problem involves constructing a process that mimics certain properties of a given Itô process, but is simpler in the sense that the mimicking process solves a stochastic differential equation, or more generally a stochastic functional differential equation, while the original Itô process may have drift and diffusion terms that are themselves adapted stochastic processes. The classical [Gyö86] article considers a multi-dimensional Itô process, and constructs a weak solution to an SDE which mimics the marginals of the original Itô process at each fixed time. The drift and covariance coefficient for the mimicking process can be interpreted as the expected value of the instantaneous drift and covariance of original Itô process, conditioned on its terminal level.

Brunick&Shreve[BS13] relax the conditions of non-degeneracy and boundedness on the covariance of the Itô process imposed in [Gyö86], and they also significantly extend the Gyöngy result. More specifically, the main result Theorem 3.6 in [BS13] proves that we can match the *joint* distribution at each fixed time of various functionals of the Itô process, including the maximum-to-date or the running average of one component of the Itô process. The mimicking process now takes the form of a stochastic functional differential equation (SFDE) and the diffusion coefficient for the SFDE is given by the so-called *Markovian projection*; in the case when we are mimicking the law of the terminal value of the process X_t and another path-dependent functional Y_t, the Markovian projection is given by the conditional expectation $\mathbb{E}(\sigma_t^2\,|X_t,Y_t)$. If in addition $\hat{\sigma}(x,y)^2$ is continuous and strictly positive, then Corollary 3.13 in [BS13] shows that all weak solutions to the SFDE have the same law.

[BS13] do not provide a constructive method for computing $\hat{\sigma}(x,y,t)^2$; however, for the standard problem of just mimicking the law of the terminal value of the process, this can be computed from the well known Dupire forward equation for continuous semimartingales, in terms of infinitesimal calendar and butterfly spreads of put or call options. This equation was derived heuristically in [Dup96] and can be proved rigorously using the Tanaka-Meyer formula for continuous semimartingales, see Klebaner[Kle02].

*The author would like to thank Benjamin Smith for helpful comments

The other main technical obstacle in establishing fitting and mimicking results of this nature is establishing *uniqueness* for the associated forward Kolmogorov equation (or associated partial integro-differential equation when there is a jump component), in the sense of distributions. This can be done when Y_t is an a.s. absolutely continuous functional using standard existence and uniqueness theorems for the forward Kolmogorov equation associated with the mimicking diffusion process, which is degenerate when we are just mimicking the marginals of the two quantities (X_t, Y_t), because there is only one driving Brownian motion. It is less clear how to proceed for a.s. non-absolutely continuous functionals likes the running maximum or local time, because the mimicking process now takes the form of a non-standard stochastic functional differential equation for which the theory is less developed.

[Forde14] considers an Itô process of the form $dX_t = \sigma_t dW_t$, and a general $\mathcal{F}_t^X$-adapted non-decreasing process Y_t (this is the path-dependent functional of interest), and first considers the case when Y_t is a.s. non absolutely continuous and $X_t = g(Y_t)$ for some continuous function $g(.)$, on the *growth set* of Y_t; this condition is satisfied when for example when Y is the running maximum of X with $g(y) = y$, or if $Y_t = L_t^a$ the local time of X at a with $g(y) = a$. In this setup, a general forward equation for the Fourier-Laplace transform of the law of (X_t, Y_t) is established, and the forward equation can be inverted to compute the Markovian projection $\hat{\sigma}(x, y, t)^2$ explicitly via a Fourier-Laplace inversion, without the a priori assumption that (X_t, Y_t) has a density at (x, y) or that $\hat{\sigma}(t, ., .)$ is continuous at (x, y). A mimicking result for the case when Y is quadratic variation is also established under more stringent conditions on $\hat{\sigma}(x, y)$, by establishing uniqueness for the associated martingale problem (in the Stroock-Varadhan sense).

Backhoff et al.[BBHK20] show that the simple Bass martingale maximizes the expectation of the vol-swap type payout $\int_0^T |\sigma_t| dt$, subject to the single-marginal constraint $X_T \sim \mu$, and one can easily verify that the Bass martingale is particular type of local volatility model. Acciaio&Guyon[AG20] build a continuous stochastic volatility model in which a VIX future is strictly more expensive than in its associated local volatility model, which disproves the long-held conjecture that the latter maximizes the price of a VIX future over all continuous martingale models consistent with the given family of marginals (this is closely related to the final section in this article where a positive result in this direction is established), see also Beiglböck et al.[BFS11].

Touzi et al.[RTT18] consider the full marginals limit of the multiple-marginals Root embedding considered in Cox et al.[COT18], and they show that the family of potential functions can be characterized as a viscosity solution to a variational inequality, part of which involves the heat equation. [KTT17] consider the full marginals limit for the multi-maturity Azéma-Yor embedding, which is constructed in Obloj&Spoida[OS17]. Hobson[Hob16] constructs a family of pure-jump martingales consistent with a given family of univariate marginals, and within this family identifies the martingale which minimizes the expected total variation over all martingales consistent with these marginals (see also Madan&Yor[MY02] for another discontinuous martingale which is also consistent with marginals at all maturities based on a family of Azema-Yor stopping times). We also mention the Local Variance Gamma (LVG) model of Carr&Nadtochiy[CN17], which consists of a local volatility model evaluated at an independent gamma subordinator, where the volatility function can be chosen so as to be consistent with a finite number of implied volatility smiles at multiple maturities, and [GJMN16] who derive explicit arbitrage-free SVI parametrizations for the whole implied volatility surface.

1.1. The Brunick-Shreve mimicking result

We now briefly summarize the main result in Brunick&Shreve[BS13] for the special case when the dimension $n = 1$ and the process under consideration is driftless.

For an Itô process of the form in (2.1), [BS13] consider a certain class of path-dependent functionals Y of X which can be associated with an *updating function* $\Phi_t(e_1, e_2; x)$, which can include $Y_t = \bar{X}_t$ (the running maximum of X), $Y_t = \underline{X}_t$ (the running minimum of X) or an additive functional of the form $Y_t = \int_0^t g(X_s) ds$, but cannot include $\langle X \rangle_t$ the quadratic variation of X or L_t^a the local time of X at $x = a$ because these functionals are not continuous in the sup norm topology (see [Forde14] for mimicking theorems which deal with these two cases). For continuous functionals Y in the class of updating functions, Theorem 3.6 in [BS13] proves that there exists a filtered probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathcal{F}}_t, \hat{\mathbb{P}})$ that supports a continuous adapted process $\hat{X}$ on $\mathbb{R}$ and a one-dimensional Brownian motion $\hat{W}$ satisfying

$$\hat{X}_t = \int_0^t \hat{\sigma}(\hat{X}_s, \hat{Y}_s, s) d\hat{W}_s \tag{1.1}$$

where $\hat{\sigma}(\hat{X}_t, \hat{Y}_t, t)^2 = \mathbb{E}(\sigma_t^2 | X_t, Y_t)$ $\mathbb{P}$-a.s. $t \in N^c$ for some Lebesgue null set $N \subset [0, \infty)$, such that the distribution of (X_t, Y_t) under $\mathbb{P}$ agrees with the distribution of $(\hat{X}_t, \hat{Y}_t)$ under $\hat{\mathbb{P}}$ for all $t \in [0, T]$. If in addition $\hat{\sigma}(x,y)^2$ is continuous and strictly positive, then Corollary 3.13 in [BS13] shows that all weak solutions to (1.1) have the same law.

2. The modelling set up

We will say that X is an Itô process if there exists a filtered probability space $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbb{P})$ satisfying the usual conditions such that

$$X_t \;=\; \int_0^t \sigma_s dW_s \tag{2.1}$$

where W is a standard one-dimensional Brownian motion adapted to $\mathcal{F}_t$, and σ_t is an $\mathcal{F}_t$-adapted process with $\mathbb{E}(\int_0^t \sigma_s^2 ds) < \infty$ for all $t \le T$, so X is an $\mathcal{F}_t$-martingale. Let $\mathcal{F}_t^X$ denote the natural filtration of X, and let Y_t^X be an $\mathcal{F}_t^X$-adapted process which can be associated to an updating function $\Phi(e_1,e_2,x)$ of X in the sense in section 3 in [BS13].

We henceforth use Y_t as shorthand for Y_t^X.

We begin with an elementary lemma:

Lemma 2.1. *For each $t \in \times(0,T]$, there exist Borel functions $\hat{\sigma}^2(.,.,t) : \mathbb{R}\times\mathbb{R} \mapsto \mathbb{R}^+$ and $\bar{\sigma} : \mathbb{R}\times\mathbb{R}\times\mathbb{R}^+ \mapsto \mathbb{R}^+$ such that*

$$\begin{aligned}
\mathbb{E}(\sigma_t^2 \,|\, X_t, Y_t) &= \hat{\sigma}^2(X_t,Y_t,t) \qquad p_t - a.s. \\
\mathbb{E}(|\sigma_t| \,|\, X_t, Y_t) &= \bar{\sigma}(X_t,Y_t,t) \qquad p_t - a.s.
\end{aligned}$$

where $p_t(dx,dy)$ denotes the law of (X_t,Y_t), with $\hat{\sigma}^2(.,.,t)$ and $\bar{\sigma}(.,.,t)$ unique up to a set of $p_t(.,.)$-measure zero.

Proof. See Appendix. □

Remark 2.2. *$\hat{\sigma}^2(x,y,t)$ is known as the Markovian projection of σ_t^2 on (X_t,Y_t).*

From the conditional Jensen inequality, we know that

$$\mathbb{E}(\sigma_t^2 |\, X_t, Y_t) \;\ge\; \mathbb{E}(|\sigma_t| \,|\, X_t,Y_t)^2 \;=\; \bar{\sigma}(X_t,Y_t)^2$$

p_t-a.s., so

$$\hat{\sigma}(x,y,t)^2 \;\ge\; \bar{\sigma}(x,y,t)^2 \qquad p_t\text{-a.s.} \tag{2.2}$$

2.1. Extremal properties of Brunick-Shreve diffusion-type processes

Now consider a family of probability measures $(\mu_t)_{t\in[0,T]}$ on $\mathbb{R}\times\mathbb{R}$ such that $\int x\mu_t(dx,dy) = 0$ and $\int |x|\mu_t(dx,dy) < \infty$ and $\mu_0(dx,dy) = \delta_{(0,0)}$. Let $\mathcal{M}(\mu_{(.)})$ denote the collection of all Itô processes such that $(X_t,Y_t) \sim \mu_t$ for all $t \in [0,T]$, and we make the following assumption throughout:

Assumption 2.3. *There exists an Itô process such that $(X_t,Y_t) \sim \mu_t$ for all $t \in [0,T]$.*

Remark 2.4. *A necessary condition for this Assumption to be satisfied is that the $\hat{\sigma}(x,y)^2$ calculated from $(\mu_t)_{t\in[0,T]}$ using the Fourier inversion method in [Forde14] is real-valued. An important open question is what is a sufficient condition (this would require establishing uniqueness for the forward equation which links $\mathbb{E}(e^{ikX_t-\lambda Y_t})$ and $\mathbb{E}(\sigma_t^2 e^{ikX_t-\lambda Y_t}) = \mathbb{E}(\hat{\sigma}(X_t,Y_t,t)^2 e^{ikX_t-\lambda Y_t})$ in Theorem 3.1 for the non-absolutely continuous case in [Forde14]).*

$\hat{X}$ attains equality in (2.2) Moreover, by Theorem 3.1 in [Forde14], the Fourier–Laplace transform

$$U(k,\lambda,t) \;:=\; \mathbb{E}\big(\hat{\sigma}(X_t,Y_t,t)^2\, e^{ikX_t-\lambda Y_t}\big) \;=\; \iint \hat{\sigma}(x,y,t)^2 e^{ikx-\lambda y}\, p_t(dx,dy)$$

is determined explicitly from the family of joint marginals $\{\mu_t\}_{t\in[0,T]}$. Hence the measure $q_t(dx,dy) := \hat{\sigma}(x,y,t)^2\, p_t(dx,dy)$ is uniquely determined by $\{\mu_t\}$, and therefore $\hat{\sigma}(x,y,t)^2 = \frac{dq_t}{dp_t}(x,y)$ is unique p_t-a.s. It follows that the left-hand side of (2.2) depends only on $\{\mu_t\}$ and is the same for all $X \in \mathcal{M}(\mu_{(.)})$ (for fixed (x,y)). Hence $\hat{X}$ also maximizes $\bar{\sigma}(x,y,t)^2$ (and $|\bar{\sigma}(x,y,t)|$) over $X \in \mathcal{M}(\mu_{(.)})$ and also maximizes the measure

$$|\bar{\sigma}(x,y,t)|\, p_t(dx,dy) \;=\; \mathbb{E}(|\sigma_t| 1_{X_t\in dx, Y_t\in dy})$$

over all $X \in \mathcal{M}(\mu_{(.)})$ (where $p_t(dx,dy)$ denotes the law of (X_t, Y_t)) because $p_t(dx,dy) = \mu_t(dx,dy)$ (i.e. is unchanging) for each $X \in \mathcal{M}(\mu_{(.)})$, from the definition of $\mathcal{M}(\mu_{(.)})$. Thus $\hat{X}$ also maximizes $\mathbb{E}(\int_0^T |\sigma_t| g(X_t, Y_t, t)dt)$ for any non-negative Borel function g, and more generally maximizes

$$\mathbb{E}(\int_0^T f(|\sigma_t|)g(X_t, Y_t)\nu(dt)) \tag{2.3}$$

for any non-negative concave function f, and non-negative Borel measure ν. By trivial modifications, we see that $\hat{X}$ minimizes (2.3) if f is convex. As a special case this includes a volatility swap-type payoff of the form $\int_0^{T_1} |\sigma_t| dt$ for any $T_1 \in [0,T]$ (note that some authors define a volatility swap as paying $(\int_0^T \sigma_t^2 dt)^{\frac{1}{2}}$ not $\int_0^T |\sigma_t| dt$).

Remark 2.5. *If $Y_t = [X]_t$ (the quadratic variation of X), then section 5 of [Forde14] gives us the mimicking result we need, albeit under the more stringent condition that $\hat{\sigma}(x,y,t)$ is bounded and continuous and has two bounded continuous spatial derivatives.*

2.2. The univariate and multi-dimensional cases

From Theorem 3.6 in [BS13] we can also forget about the Y variable, and just maximize $\mathbb{E}(\int_0^T f(\sigma_t^2)g(X_t)\nu(dt))$ over all Itô processes X such that $X_t \sim \mu_t$ for all $t \in [0,T]$ for a given family of univariate marginals $(\mu_t)_{t\in[0,T]}$, assuming such an Itô process exists (for which a necessary condition is that μ_t be be non-decreasing in the convex order, see e.g. Lowther[Low07],[Low08], Beiglböck et al.[BHS17] and Strassen's theorem (Theorem 8 in [Str65]) for more on this). By an almost identical analysis to above, the extremal model is now given by a Dupire-type local volatility model of the form $d\hat{X}_t = \hat{\sigma}(X_t,t)d\hat{W}_t$, where $\hat{\sigma}(X_t,t) = \mathbb{E}(\sigma_t^2 | X_t)$ a.s.. Dupire's model is still a plausible and (historically) popular model which has been used in the industry for 25 years, unlike most extremal Skorkohod embeddings which (despite their mathematical beauty) typically give rise to unrealistic models which stop moving before the final maturity T and upper/lower bounds which are too wide to be of practical use. Or we can work with a d-dimensional Y functional as in [BS13], and/or mimick marginals of an n-dimensional X process (we do not pursue this here).

2.3. Sub/super hedging

For simplicity we concentrate on the univariate case discussed in the previous subsection, and assume that $f \in C^1$ and concave and that ν has a density. Then using that $f(x) \le f(x^*) + f'(x^*)(x - x^*)$ with $x^* = \mathbb{E}(\sigma_t^2|X_t)$ we get

$$\begin{aligned} f(\sigma_t^2) &\le f(\mathbb{E}(\sigma_t^2|X_t)) + f'(\mathbb{E}(\sigma_t^2|X_t))(\sigma_t^2 - \mathbb{E}(\sigma_t^2|X_t)) \\ &= f(\hat{\sigma}(X_t,t)^2) + f'(\hat{\sigma}(X_t,t)^2)(\sigma_t^2 - \hat{\sigma}(X_t,t)^2) \end{aligned}$$

and hence

$$\int_0^T f(\sigma_t^2)g(X_t)\nu(t)dt \le \int_0^T f(\hat{\sigma}(X_t,t)^2)g(X_t)\nu(t)dt + \int_0^T f'(\hat{\sigma}(X_t,t)^2)(\sigma_t^2 - \hat{\sigma}(X_t,t)^2)g(X_t)\nu(t)dt\,.$$

The first and the third terms on the right hand side are just the payoffs of European-type payoffs over the continuum of maturities $t \in [0,T]$. To replicate the middle term $f'(\hat{\sigma}(X_t,t)^2)\sigma_t^2$, we note that for any $\psi \in C^{2,1}$ we have

$$\psi(X_T,T) - \psi(X_0,0) - \int_0^T \psi_x(X_t,t)dX_t - \int_0^T \psi_t(X_t,t)dt = \frac{1}{2}\int_0^T \sigma_t^2 \psi_{xx}(X_t,t)dt$$

and setting $f'(\hat{\sigma}(x,t)^2)g(x)\nu(t) = \psi_{xx}(x,t)$ and solving for ψ, we can also replicate this middle term. Thus we have a model-independent superhedge for $\int_0^T f(|\sigma_t|)g(X_t)\nu(t)dt$, and clearly if we instead assume that f is convex, then this trading strategy becomes a sub-hedge.

We can perform a similar analysis for the case when we have joint marginals for (X_t, Y_t), in this case we require bi-variate European-type contracts of the form $\psi(X_t, Y_t)$ at all maturities.

3. Extremal models for VIX futures given full marginals for X

We now assume S is a strictly positive stock price process of the form

$$dS_t = S_t \sigma_t dW_t \tag{3.1}$$

with $S_0 > 0$ (where σ satisfies the same conditions as in Section 2.1) such that S is an $\mathcal{F}_t$ martingale (for which a sufficient condition is that σ satisfies the Novikov condition). Then $X_t = \log S_t$ of course satisfies $dX_t = -\frac{1}{2}\sigma_t^2 dt + \sigma_t dW_t$. Now let $Y_t := X_t - X_T$ for $t > T$ and zero for $t \in [0,T]$; then from Theorem 3.6 in [BS13], there exists a filtered probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathcal{F}}_t, \hat{\mathbb{P}})$ that supports a continuous adapted process $\hat{X}$ on $\mathbb{R}$ and a one-dimensional Brownian motion $\hat{W}$ satisfying $d\hat{X}_t = -\frac{1}{2}\hat{\sigma}(\hat{X}_t, \hat{Y}_t, t)^2 dt + \hat{\sigma}(\hat{X}_t, \hat{Y}_t, t) d\hat{W}_t$ with $\hat{Y}_t := \hat{X}_t - \hat{X}_T$ for $t > T$ and zero for $t \in [0,T]$ and with $\hat{X}_0 = X_0$, where $\hat{\sigma}(\hat{X}_t, \hat{Y}_t, t)^2 = \mathbb{E}(\sigma_t^2 | X_t, Y_t)$ a.s. and such that $(\hat{X}_t, \hat{Y}_t) \sim (X_t, Y_t)$ for all $t \in [0, T+\Delta]$. Note of course that $\hat{X}$ is just a conventional Dupire-type local volatility model up to time T since $\hat{Y}_t = 0$ is not random on this time interval).

If we now consider a VIX future with maturity T which pays $V := \mathbb{E}(L(S_{T+\Delta}/S_T)|\mathcal{F}_T)^{\frac{1}{2}}$ at time T (where $L(y) := -\frac{2}{\Delta}\log y$, and in practice $\Delta = 30$ days). Then following Proposition 4.10 in [GMN17], using Jensen and the tower property we see that

$$\begin{aligned}
\mathbb{E}(V) \;=\; \mathbb{E}(\mathbb{E}(V|S_T)) &\le \mathbb{E}(\sqrt{(\mathbb{E}(V^2|S_T))}) \\
&= \mathbb{E}(\sqrt{\mathbb{E}(\mathbb{E}(L(S_{T+\Delta}/S_T)|\mathcal{F}_T)|S_T)}) \\
&= \mathbb{E}(\sqrt{\mathbb{E}(L(S_{T+\Delta}/S_T)|S_T)}) \\
&= \int_0^\infty \sqrt{(}\int_0^\infty L(y) q(dy|s))\, p(ds) \\
&\quad \text{(where } q(dy|s) \text{ is conditional law of } S_{T+\Delta}/S_T \text{ given } S_T = s \\
&\quad \text{and } p(ds) \text{ is marginal law of } S_T) \\
&= \mathbb{E}^{\hat{\mathbb{P}}}(\sqrt{\mathbb{E}^{\hat{\mathbb{P}}}(L(\hat{S}_{T+\Delta}/\hat{S}_T)|\hat{S}_T)}) \\
&= \mathbb{E}^{\hat{\mathbb{P}}}(V)
\end{aligned} \tag{3.2}$$

where $\hat{S} = e^{\hat{X}}$, and the penultimate equality follows since $(S_{T_1}, S_{T_2}/S_{T_1}$ and $(\hat{S}_{T_1}, \hat{S}_{T_2}/\hat{S}_{T_1})$ have the same joint laws (from the mimicking property) and the final equality follows since $\hat{X}$ is Markov in $\hat{X}$ on $[0,T]$, and the final line is the price of the VIX future associated with the mimicking process $\hat{S} = e^{\hat{X}}$. In practice we would not have access to joint marginals for S_t and S_t/S_T for all $t \ge T$ (to do so would require tradeable bivariate call contracts with payoff $(S_t - K)^+(S_t/S_T - L)^+$ for all strikes K and L and all maturities $t \in (T, T+\Delta]$ which do not exist in practice), so we cannot compute $\hat{\sigma}(x,y,t)$ from market data if we only have observed European option prices at all strikes and maturities as is often assumed.

To make further progress, we now let $\mathcal{M}$ denote the space of all martingale processes of the form in (3.1) (possibly on different probability spaces) and consider a family $(\mu_t)_{t\in[0,T+\Delta]}$ of probability measures on $(0,\infty)$ with $\mu_0 = \delta_{\{S_0\}}$ (the dirac measure at S_0) and $\int s\mu_t(ds) = S_0$ which are strictly increasing in convex order, and assume $S \in \mathcal{M} : S_t \sim \mu_t, \forall t \in [0, T+\Delta]$ is non-empty [1]. Then for any $\varepsilon > 0$, consider an ε-optimal process S for the problem of maximizing the price of a VIX future over all $S \in \mathcal{M}((\mu_t)_{t\in[0,T+\Delta]})$:

$$P \;:=\; \sup_{S\in\mathcal{M}\,:\,S_t\sim\mu_t,\forall t\in[0,T+\Delta]} \mathbb{E}(\sqrt{\mathbb{E}(L(S_{T+\Delta}/S_T)|\mathcal{F}_T^S)})$$

where (with mild abuse of notation) $\mathcal{F}^S$ here is the filtration being used for the probability space on which each process S is defined (which may be strictly larger than the filtration *generated* by S). Then using (3.2), we know that

$$P - \varepsilon \;\le\; \mathbb{E}^{\hat{\mathbb{P}}}(\sqrt{\mathbb{E}^{\hat{\mathbb{P}}}(L(\hat{S}_{T+\Delta}/\hat{S}_T)|\hat{S}_T)}) \;\le\; P \tag{3.3}$$

where $\hat{\mathbb{P}}$ is the probability measure for the probability space on which $\hat{X}$ is defined, and the second inequality follows since $\hat{S} = e^{\hat{X}}$ is also in $\mathcal{M} : S_t \sim \mu_t, \forall t \in [0, T+\Delta]$. Thus we see that $\hat{S}$ is also ε-optimal. So essentially any extremal (or ε-extremal) model has a Markovian projection in this sense which attains a VIX future price which is at least as large as $P - \varepsilon$.

[1] A sufficient condition for non-emptyness is that the call option price function $C(K,T) := \int (s-K)^+ \mu_t(ds)$ is differentiable in T and the μ_t's admit densities so $C_{KK}(K,T) = \mu_t(K)$ and that the calibrated volatility function obtained from the usual Dupire forward equation for European call options satisfies suitable Lipschitz and growth conditions to ensure that the Dupire forward equation has a unique solution, see also Proposition 4.1 in [Fig08] and Proposition 5.3 in [Forde14] and discussion in the introduction for more on this point. Alternatively one could take the (weaker) approach in Proposition 3.3 in [Forde14] with ik replaced with $-\lambda$ for $\lambda > 0$, and check that the candidate $U(\lambda,t) = \mathbb{E}(\sigma_t^2 e^{-\lambda S_t})$ function is completely monotone in $\lambda > 0$ (for each t) which ensures that it is the Laplace transform of a non-negative Borel measure $q(ds,t)$ and then check that $q(ds,t)$ is absolutely continuous with respect to $p(ds,t) := \mu_t(ds)$, but checking the completely monotone property requires checking the sign of a countably infinite number of derivatives, so this formulation is of more theoretical interest.

Remark 3.1. *Note this does not mean we can easily explicitly compute $\hat{\sigma}(.,.,.)$, because we would need to know the original extremal (or ε-extremal) model to be able to compute $\hat{\sigma}(x,,y,t)^2 = \mathbb{E}(\sigma_t^2 \,|\, X_t = x, Y_t = y)$, so in its current form this tells us the qualitative form of an extremal model rather but not an explicit recipe for constructing it.*

Note we can also replace $S \in \mathcal{M} : S_t \sim \mu_t, \forall t \in [0, T+\Delta]$ with $S \in \mathcal{M} : \mathbb{E}(f_i(S_{T_i}, S_{T_i \vee T}/S_T)) = P_i), i = 1..N$ for n integrable payoff functions f_i, so long as this set is non-empty, which corresponds to n tradeable European and/or forward starting-type contracts.

Remark 3.2. *Trivially, we also see that any ε-optimal price for a forward-starting option with payoff $(S_{T_2} - \lambda S_{T_1})^+$ given marginals for S for all $t \in [0, T_2]$ (with $0 < T_1 < T_2$) can also be attained by the mimicking model $\hat{S}$, since $(S_{T_1}, S_{T_2}/S_{T_1}$ and $(\hat{S}_{T_1}, \hat{S}_{T_2}/\hat{S}_{T_1})$ have the same joint laws, and in fact this we also be true if we replace the sup with an inf in the definition of P.*

References

[AG20] Acciaio, B. and J. Guyon, "Inversion of Convex Ordering: Local Volatility Does Not Maximize the Price of VIX Futures", SIAM Journal on Financial Mathematics 11/1, SC-1-SC-13, 2020.

[BBHK20] Backhoff, J., M.Beiglböck, M.Huesmann and S.Källblad, "Martingale Benamou-Brenier: a probabilistic perspective", *Ann.Probab.*, 48(5), pp 2258-2289.

[BFS11] Beiglböck, M., Friz, P., Sturm, S., "Is the minimum value of an option on variance generated by local volatility?", *SIAM J.Finan. Math.*, 2:213–220, 2011

[BHS17] Beiglböck, M., M.Huesmann and F.Stebegg, "Root to Kellerer", *Séminaire de Probabilités*, 2017.

[BS13] Brunick, G. and S.Shreve, "Matching an Itô Process by a Solution of a Stochastic Differential Equation", 2013, *Ann. Appl. Probab.*, Vol. 23, No. 4, 1584-1628.

[CN17] Carr, P. and S.Nadtochiy, "Local Variance Gamma and Explicit Calibration to Option Prices", *Mathematical Finance*, 27(1), 2017.

[COT18] Cox, A.M.G., J. Obloj, and N.Touzi, , "The root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach", *Probab. Theory Related Fields*, 2018.

[Dup96] Dupire, B., "A unified theory of volatility", Paribas working paper, 1996 (also appears in "Derivatives Pricing: the classic collection", pages 185-196. Risk Books, 2004).

[Fig08] Figalli, A., "Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients", *J.Funct.Anal.*, Vol. 254, 1, 109-153, 2008.

[Forde11] Forde, M., "A diffusion-type process with a given joint law for the terminal level and supremum at an independent exponential time", *Stoch. Proc. Appl.*, 121, 2802-2817, 2011.

[Forde14] Forde, M., "On the Markovian projection in the Brunick-Shreve mimicking result", *Stat.Prob.Lett.*, 85, 98-105, 2014.

[Gyö86] Gyöngy, I., "Mimicking the one-dimensional marginal distributions of processes having an Itô differential", *Prob. Th. Rel. Fields*, Vol. 71, 4, 501-516, 1986.

[GJMN16] Guo, G., Jacquier, A., C. Martini and L. Neufcourt, "Generalised arbitrage-free SVI volatility surfaces", SIAM Journal on Financial Mathematics, 7(1), 619-641, 2016.

[GMN17] Guyon, J., R.Menegaux and M.Nutz, "Bounds for VIX futures given S&P 500 smiles", *Finance and Stochastics*, Volume 21, pages 593-630, 2017.

[Hob16] Hobson, D., "Mimicking martingales", *Ann. Appl. Probab.*, Volume 26, Number 4 (2016), 2273-2303.

[KTT17] Källblad, S., X.Tan and N.Touzi, "Optimal Skorokhod embedding given full marginals and Azéma-Yor peacocks", *Annals of Applied Probability*, 27(2), 686-719, 2017.

[Low07] Lowther, G., "Limits of one dimensional diffusions", preprint, Dec 2007.

[Low08] Lowther, G., "Fitting martingales to given marginals, preprint, Aug 2008.

[MY02] Madan, D. and M.Yor, "Making Markov martingales meet marginals: with explicit constructions", *Bernoulli*, 8,4, 509-536, 2002.

[Kle02] Klebaner, F., "Option pricing when the stock is a semimartingale", *Elect. Comm. in Probab.*, 7, 79-83, 2002.

[OS17] Obloj, J. and P.Spoida, "An iterated Azema-Yor type embedding for finitely many marginals", *Ann. Probab.*, 45(4), 2210-2247, 2017.

[RTT18] Richard, A., X.Tan and N.Touzi, "On the Root solution to the Skorokhod embedding problem given full marginals", preprint.

[Str65] Strassen, V., "The Existence of Probability Measures with Given Marginals", *Ann. Math. Statist.*, Volume 36, Number 2 (1965), 423-439.

Appendix A: Proof of Lemma 2.1

Similar to the proof of Proposition 4.4. in [Gyö86], we now recall the definition of $\hat{\sigma}^2(x,y,t) = \mathbb{E}(\sigma_t^2|X_t = x, Y_t = y)$ via the Radon-Nikodým theorem: we consider the measure q_t defined by the formula

$$q_t(A) \quad = \quad \mathbb{E}(1_{(X_t,Y_t)\in A}\,\sigma_t^2)\,.$$

for every $A \in \mathcal{B}(\mathbb{R}\times\mathbb{R}^+)$. q_t is absolutely continuous with respect to $p_t(dx,dy)$. Thus, by the Radon-Nikodým theorem, there exists a measurable function $\hat{\sigma}^2(,.,.,t)$ such that

$$q_t(A) \quad = \quad \int_{\mathbb{R}\times\mathbb{R}^+} 1_A\,\hat{\sigma}^2(x,y,t)\,p_t(dx,dy)\,.$$

For every t, $\hat{\sigma}^2(.,.,t)$ is unique up to a set of $p_t(.,.)$-measure zero. We then define $\mathbb{E}(\sigma_t^2|X_t = x, Y_t = y) = \hat{\sigma}^2(x,y,t)$ and from the standard Kolmogorov definition of conditional expectation, $\mathbb{E}(\sigma_t^2|X_t,Y_t) = \hat{\sigma}^2(X_t,Y_t,t)$ a.s. We perform a similar analysis for $\mathbb{E}(\sigma_t|X_t,Y_t)$.