Quantitative Finance

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

Inversion of convex ordering in the VIX market

Julien Guyon

To cite this article: Julien Guyon (2020) Inversion of convex ordering in the VIX market, Quantitative Finance, 20:10, 1597-1623, DOI: 10.1080/14697688.2020.1753885

To link to this article: https://doi.org/10.1080/14697688.2020.1753885

Published online: 18 May 2020.

Quantitative Finance, 2020
Vol. 20, No. 10, 1597–1623, https://doi.org/10.1080/14697688.2020.1753885

Inversion of convex ordering in the VIX market

JULIEN GUYON*

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

(*Received 14 December 2019; accepted 3 April 2020; published online 18 May 2020*)

We investigate conditions for the existence of a *continuous* model on the S&P 500 index (SPX) that jointly calibrates to a *full surface* of SPX implied volatilities and to the VIX smiles. We present a novel approach based on the SPX smile calibration condition $\mathbb{E}[\sigma_t^2|S_t] = \sigma_{\text{lv}}^2(t, S_t)$. In the limiting case of instantaneous VIX, a novel application of martingale transport to finance shows that such model exists if and only if, for each time t, the local variance $\sigma_{\text{lv}}^2(t, S_t)$ is smaller than the instantaneous variance σ_t^2 in convex order. The real case of a 30-day VIX is more involved, as averaging over 30 days and projecting onto a filtration can undo convex ordering.

We show that in usual market conditions, and for reasonable smile extrapolations, the distribution of VIX_T^2 in the market local volatility model is *larger* than the market-implied distribution of VIX_T^2 in convex order for short maturities T, and that the two distributions are not rankable in convex order for intermediate maturities. In particular, a *necessary* condition for continuous models to jointly calibrate to the SPX and VIX markets is the *inversion of convex ordering* property: the fact that, even though associated local variances are smaller than instantaneous variances in convex order, the VIX squared is larger in convex order in the associated local volatility model than in the original model for short maturities. We argue and numerically demonstrate that, when the (typically negative) spot–vol correlation is large enough in absolute value, (a) traditional stochastic volatility models with large mean reversion, and (b) rough volatility models with small Hurst exponent, satisfy the inversion of convex ordering property, and more generally can reproduce the market term-structure of convex ordering of the local and stochastic squared VIX.

Keywords: VIX; Convex order; Inversion of convex ordering; Martingale transport; Local volatility; Stochastic volatility; Mean reversion; Rough volatility; Smile calibration

1. Introduction

Volatility indices, such as the VIX index (CBOE 2017), do not only serve as market-implied indicators of volatility. Futures and options on these indices are also widely used as risk-management tools to hedge the volatility exposure of options portfolios. The existence of a liquid market for these futures and options has led to the need for models that jointly calibrate to the prices of options on the underlying asset and the prices of volatility derivatives. Without such models, financial institutions could possibly arbitrage each other, and even desks within the same institution could do so, e.g. the VIX desk could arbitrage the SPX desk.

In particular, since VIX options started trading on the CBOE in 2006, many researchers and practitioners have attempted to build a model for the SPX that is consistent with market data on both SPX options and VIX futures and options. The first attempt, by Gatheral (2008, 2013), used a diffusive (double mean reverting) model. Interestingly, the numerical results show that, in usual market conditions, this model, though it is very flexible, cannot fit both the negative at-the-money (ATM) SPX skew (not large enough in absolute value) and the ATM VIX volatility (too large) for short maturities (up to 5 months). One should decrease the volatility of volatility ('vol-of-vol') to decrease the latter, but this would also decrease the former, which is already too small.

Guyon's experiments (Guyon 2018a, 2018b) using very flexible models such as the skewed two-factor Bergomi model (Bergomi 2008), the skewed rough Bergomi model, independently introduced by Guyon (2018a) and De Marco (2018), and their stochastic local volatility versions, are also suggesting that joint calibration seems out of the reach of classical continuous-time models with continuous SPX paths ('continuous models' for short): either the SPX smile is well fitted, but then the model ATM VIX implied volatility is too large; or the VIX smile is well calibrated, but then the model ATM SPX skew is too small in absolute value. Song and Xiu (2012) argued that 'the state-of-the-art stochastic volatility models in the literature cannot capture the S&P 500 and VIX option prices simultaneously'. Jacquier *et al.* (2018) investigated the rough Bergomi model and reached a similar conclusion:

*Email: jguyon2@bloomberg.net

> Interestingly, we observe a 20% difference between the [vol-of-vol] parameter obtained through VIX calibration and the one obtained through SPX. This suggests that the volatility of volatility in the SPX market is 20% higher when compared to VIX, revealing potential data inconsistencies (arbitrage?).

Note that the fact that a class of models is not able to jointly fit SPX options, VIX futures and VIX options does not reveal any joint SPX/VIX arbitrage. It simply means that this class of models is inconsistent with market data.

Fouque and Saporito (2018) have tried to jointly calibrate using a Heston stochastic volatility model with stochastic vol-of-vol, but their approach does not apply to short maturities (below 4 months), for which VIX derivatives are most liquid and the joint calibration is most difficult. Goutte *et al.* (2017) used a Heston model whose parameters are driven by a hidden Markov chain, but the SPX smile used in their calibration tests is erroneous—and the reported fit is only approximate.

To try to achieve joint calibration, many authors have incorporated jumps in the SPX (Sepp 2011, Cont and Kokholm 2013, Baldeaux and Badran 2014, Papanicolaou and Sircar 2014, Kokholm and Stisen 2015, Pacati *et al.* 2018). Indeed, jumps offer extra degrees of freedom to partly decouple the ATM SPX skew and the ATM VIX implied volatility. Some of the suggested models, in particular those in Cont and Kokholm (2013) and Pacati *et al.* (2018), achieve a quite accurate fit, though not perfect.†

By taking a radically different approach, Guyon (2020) was the first to build a model that perfectly fits SPX options, VIX futures, and VIX options. Instead of postulating a parametric (jump-)diffusion model, he built a nonparametric, discrete-time model, cast as a solution of a dispersion-constrained martingale transport problem which is solved using (an extension of) the Sinkhorn algorithm. In particular, Guyon's result shows that in usual market conditions the market is free of joint SPX/VIX arbitrage. Should a joint SPX/VIX arbitrage arise in the market, his algorithm would detect it.

In this article, we go back to *continuous* models on the SPX, that is, continuous-time models on the SPX with continuous SPX paths. We investigate conditions for a continuous model on the SPX to jointly fit the *full surface* of SPX implied volatilities and the VIX smiles. We first investigate the limiting case of an 'instantaneous VIX', replacing the VIX by the instantaneous volatility. In this case, as a first main contribution of this paper, first reported in Guyon (2017), we show that there exists a jointly calibrating continuous model on the SPX if and only if, for each time t, the market local variance, computed using Dupire's formula (Dupire 1994), is smaller than the instantaneous variance in convex order. Interestingly, our proof uses a novel application of martingale transport to quantitative finance.

We then move to the case of the (stylized) real VIX, which is not the instantaneous volatility of the SPX, but the implied volatility of the 30-day log-contract on the SPX. In a continuous model on the SPX, the VIX squared at T is equal to the risk-neutral conditional expectation of the realized variance of the SPX over the next 30 days, given the information $\mathcal{F}_T$ available at time T. Since averaging over 30 days and projecting onto $\mathcal{F}_T$ can undo convex ordering, characterizing the existence of a jointly calibrating continuous model is not as simple as in the case of the instantaneous VIX. We carefully investigate market data and show, for the first time, that in usual market conditions, and for reasonable smile extrapolations,‡ the distribution of VIX_T^2 in the market local volatility model and the market-implied distribution of VIX_T^2 exhibit a remarkable, surprising *term-structure of convex ordering*: the distribution of VIX_T^2 in the market local volatility model is *larger* than the market-implied distribution of VIX_T^2 in convex order for short maturities T (typically, up to 2–3 months); the two distributions are not rankable in convex order for intermediate maturities (typically, 3–6 months); maturity extrapolation suggests that the distribution of VIX_T^2 in the market local volatility model is smaller than the market-implied distribution of VIX_T^2 in convex order for longer maturities T (VIX options only trade for maturities up to 6 months, therefore extrapolation is needed beyond that maturity). This observation, first made in Guyon (2017), is the second main contribution of this paper.

In particular, this shows that a *necessary* (but not sufficient) condition for a continuous model to jointly calibrate to the full surface of SPX implied volatilities and to the (reasonably extrapolated) VIX smiles is that it possesses what we call the *inversion of convex ordering* property: the fact that, even though associated local variances are smaller than instantaneous variances in convex order, the VIX squared in the associated local volatility model is *larger* in convex order than the VIX squared in the original model for short VIX future maturities. A question then naturally arises: Does there exist a continuous model that has the inversion of convex ordering property? It turns out that the answer is yes. We argue that, when the (typically negative) spot–vol correlation is large enough in absolute value, (a) stochastic volatility models with large mean reversion, and (b) rough volatility models with small Hurst exponent, not only possess the inversion of convex ordering property,§ but also can reproduce the term-structure of convex ordering of the local and stochastic squared VIX observed in (or extrapolated from) the market—and we check it numerically. This is the third main contribution of this paper: we prove (numerically at least) that the market term-structure of convex ordering can be produced by familiar continuous stochastic volatility models, for certain values of the model parameters that we identify. For these familiar models, it seems difficult to obtain a mathematical proof. In a companion paper, Acciao and Guyon (2020) build more extreme stochastic volatility models and mathematically prove that these models do generate the term-structure of convex ordering observed in (or extrapolated from) the VIX market.

The rest of this article is structured as follows. After introducing the setting and notations in Section 2, we analyze the

† Note, however, that in Cont and Kokholm (2013) the authors compute the model prices of VIX futures and VIX options based on an approximation of the VIX in the model, see equation (41).

‡ Strictly speaking, market data, which is finite, is usually not enough to conclude that two random variables X and Y are in convex order. In particular, smile extrapolations might make calls on Y cheaper than calls on X for extreme strikes, even if for all traded strikes calls on Y are more expensive than calls on X. See Remark 1 and Appendix 1.

§ At least when it is restricted to non-extreme strikes.

case of the instantaneous VIX in Section 3. We turn to the case of the real, 30-day VIX in Section 4, where we carefully analyze market data and reveal the term-structure of convex ordering in the VIX market. In Section 5, we investigate one-factor lognormal forward instantaneous variance models and explain what ingredients are needed for them to satisfy the inversion of convex ordering property, e.g. (a) large mean reversion, or (b) small Hurst exponent, together with a large spot–vol correlation. We also argue that these models can reproduce the term-structure of convex ordering observed in the VIX market. We numerically check our claims in Section 6. Finally, Section 7 concludes.

2. Setting and notation

For simplicity, let us assume zero interest rates, repos, and dividends. Let $\mathcal{F}_t$ denote the market information available up to time t. We consider continuous models on the SPX index of the form

$$\frac{\mathrm{d}S_t}{S_t} = \sigma_t\,\mathrm{d}W_t, \quad S_0 = x, \tag{1}$$

where W denotes a standard one-dimensional $(\mathcal{F}_t)$-Brownian motion, (σ_t) is an $(\mathcal{F}_t)$-adapted process such that for all $t \geq 0$, $\int_0^t \sigma_s^2\,\mathrm{d}s < \infty$ a.s., and $x > 0$ is the initial SPX price. By continuous model, we mean that the SPX has no jump, but the volatility process (σ_t) may be discontinuous. The local volatility function associated to model (1) is the function σ_{loc} defined by

$$\sigma_{\mathrm{loc}}^2(t, S_t) := \mathbb{E}[\sigma_t^2 | S_t]. \tag{2}$$

The associated local volatility model is defined by

$$\frac{\mathrm{d}S_t^{\mathrm{loc}}}{S_t^{\mathrm{loc}}} = \sigma_{\mathrm{loc}}(t, S_t^{\mathrm{loc}})\,\mathrm{d}W_t, \quad S_0^{\mathrm{loc}} = x.$$

From Gyöngy (1986), the marginal distributions of the processes $(S_t, t \geq 0)$ and $(S_t^{\mathrm{loc}}, t \geq 0)$ agree:

$$\forall t \geq 0, \quad S_t^{\mathrm{loc}} \overset{(d)}{=} S_t. \tag{3}$$

Using Dupire (1994), we conclude that model (1) is calibrated to the full SPX smile if and only if

$$\sigma_{\mathrm{loc}} = \sigma_{\mathrm{lv}}, \tag{4}$$

where σ_{lv} is the local volatility function derived from market prices of vanilla options on the SPX using Dupire's formula. We denote by S^{lv} the market local volatility model, defined by

$$\frac{\mathrm{d}S_t^{\mathrm{lv}}}{S_t^{\mathrm{lv}}} = \sigma_{\mathrm{lv}}(t, S_t^{\mathrm{lv}})\,\mathrm{d}W_t, \quad S_0^{\mathrm{lv}} = x.$$

Let $T \geq 0$. By definition, the (idealized) VIX at time T is the implied volatility of a 30-day log-contract on the SPX index starting at T. For continuous models (1), applying the Itô formula to the log payoff, this translates into (Dupire 1993, Neuberger 1994)

$$\mathrm{VIX}_T^2 = \mathbb{E}\left[\frac{1}{\tau}\int_T^{T+\tau} \sigma_t^2\,\mathrm{d}t \,\middle|\, \mathcal{F}_T\right] = \frac{1}{\tau}\int_T^{T+\tau} \mathbb{E}[\sigma_t^2 \,|\, \mathcal{F}_T]\,\mathrm{d}t,$$

where $\tau = \frac{30}{365}$ (30 days). In the associated local volatility model, since by the Markov property of $(S_t^{\mathrm{loc}}, t \geq 0)$, $\mathbb{E}[\sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}})|\mathcal{F}_T] = \mathbb{E}[\sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}})|S_T^{\mathrm{loc}}]$, the VIX, denoted by $\mathrm{VIX}_{\mathrm{loc},T}$, satisfies

$$\begin{aligned}\mathrm{VIX}_{\mathrm{loc},T}^2 &= \frac{1}{\tau}\int_T^{T+\tau} \mathbb{E}[\sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}}) \,|\, S_T^{\mathrm{loc}}]\,\mathrm{d}t \\ &= \mathbb{E}\left[\frac{1}{\tau}\int_T^{T+\tau} \sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}})\,\mathrm{d}t \,\middle|\, S_T^{\mathrm{loc}}\right].\end{aligned} \tag{5}$$

Similarly,

$$\begin{aligned}\mathrm{VIX}_{\mathrm{lv},T}^2 &= \frac{1}{\tau}\int_T^{T+\tau} \mathbb{E}[\sigma_{\mathrm{lv}}^2(t, S_t^{\mathrm{lv}}) \,|\, S_T^{\mathrm{lv}}]\,\mathrm{d}t \\ &= \mathbb{E}\left[\frac{1}{\tau}\int_T^{T+\tau} \sigma_{\mathrm{lv}}^2(t, S_t^{\mathrm{lv}})\,\mathrm{d}t \,\middle|\, S_T^{\mathrm{lv}}\right].\end{aligned} \tag{6}$$

The prices at time 0 of the VIX future and the VIX call options with common maturity T in model (1) are respectively given by

$$\mathbb{E}[\mathrm{VIX}_T] = \mathbb{E}\left[\sqrt{\mathbb{E}\left[\frac{1}{\tau}\int_T^{T+\tau} \sigma_t^2\,\mathrm{d}t \,\middle|\, \mathcal{F}_T\right]}\right],$$

$$\mathbb{E}[(\mathrm{VIX}_T - K)_+] = \mathbb{E}\left[\left(\sqrt{\mathbb{E}\left[\frac{1}{\tau}\int_T^{T+\tau} \sigma_t^2\,\mathrm{d}t \,\middle|\, \mathcal{F}_T\right]} - K\right)_+\right].$$

We observe market prices for those instruments, for a list of liquid monthly VIX future maturities T_i, with $T_i \leq 6$ months.† The main question we investigate in this article is the following: Under what conditions does there exist a continuous model on the SPX that calibrates to the *full surface* of SPX implied volatilities and to the market prices of liquid VIX futures and VIX options? In mathematical terms: Under what conditions does there exist a model satisfying (1) and (4) and such that for all T_i and K, the observed market price of the T_i-VIX future coincides with $\mathbb{E}[\mathrm{VIX}_{T_i}]$ and the observed market price of the (T_i, K)-VIX call coincides with $\mathbb{E}[(\mathrm{VIX}_{T_i} - K)_+]$?

3. The case of instantaneous VIX

To simplify the problem, let us first consider the case of an instantaneous VIX: $\tau \to 0$. The realized variance over 30

† The largest quoted maturity of VIX options is 6 months, while for VIX futures it is 9 months. We disregard weekly VIX futures data, as those futures are not very liquid.

days is then simply replaced by the instantaneous variance: $\mathrm{VIX}_T \to \mathrm{instVIX}_T := \sigma_T$.

We recall that (the distributions of) two random variables X and Y are said to be in convex order, which we denote by $X \leq_c Y$, if and only if for any convex function f, $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$. Loosely speaking, this means that both distributions have the same mean, but the distribution of Y is more 'spread' than that of X. In dimension 1, this is equivalent to the fact that $\mathbb{E}[X] = \mathbb{E}[Y]$ and for all strike $K \in \mathbb{R}$, $\mathbb{E}[(X-K)_+] \leq \mathbb{E}[(Y-K)_+]$. In financial terms, this means that the assets X and Y share the same forward price, but that calls (resp. puts) on Y are more expensive than calls (resp. puts) on X. From the definition (2) of the model local volatility σ_{loc}, conditional Jensen's inequality implies that $\sigma^2_{\mathrm{loc}}(t, S_t) \leq_c \sigma_t^2$. Using (3), we get

$$\forall t \geq 0, \quad \sigma^2_{\mathrm{loc}}(t, S_t^{\mathrm{loc}}) \leq_c \sigma_t^2. \tag{7}$$

Assume that model (1) jointly calibrates to the SPX smile and the instantaneous VIX smile. Then (7) and (4) imply that

$$\forall t \geq 0, \quad \sigma^2_{\mathrm{lv}}(t, S_t^{\mathrm{lv}}) \leq_c \mathrm{instVIX}_t^2. \tag{8}$$

Conversely, assume (8). Can we build a model of the type (1) that jointly calibrates to the SPX smile and the instantaneous VIX smile? Let us discretize time, assume that we were able to simulate S_{t_i} and that it follows the risk-neutral distribution of the SPX at date t_i. From (8), $\sigma^2_{\mathrm{lv}}(t_i, S_{t_i}) \overset{(d)}{=} \sigma^2_{\mathrm{lv}}(t_i, S^{\mathrm{lv}}_{t_i}) \leq_c \mathrm{instVIX}^2_{t_i}$. Strassen's theorem (Strassen 1965) then implies that we can draw a random variable $\sigma^2_{t_i}$ having the same distribution as $\mathrm{instVIX}^2_{t_i}$ and such that $\mathbb{E}[\sigma^2_{t_i} \,|\, \sigma^2_{\mathrm{lv}}(t_i, S_{t_i})] = \sigma^2_{\mathrm{lv}}(t_i, S_{t_i})$. Let us define the distribution of $\sigma^2_{t_i}$ given S_{t_i} as the distribution of $\sigma^2_{t_i}$ given $\sigma^2_{\mathrm{lv}}(t_i, S_{t_i})$. Then we have $\mathbb{E}[\sigma^2_{t_i} \,|\, S_{t_i}] = \sigma^2_{\mathrm{lv}}(t_i, S_{t_i})$, and we simulate $S_{t_{i+1}}$ using

$$\frac{\mathrm{d}S_t}{S_t} = \sigma_{t_i}\,\mathrm{d}W_t, \quad t \in [t_i, t_{i+1}].$$

Assuming we can pass to the continuous time limit (which may require some regularity in time of the martingale transport from $\sigma^2_{\mathrm{lv}}(t, S_t)$ to σ_t^2), we have built a model of the type (1) that jointly calibrates the SPX smile and the instantaneous VIX smile:

$$\forall t \geq 0, \quad \sigma_t^2 \overset{(d)}{=} \mathrm{instVIX}_t^2 \quad \text{and} \quad \mathbb{E}[\sigma_t^2 | S_t] = \sigma^2_{\mathrm{lv}}(t, S_t).$$

As a consequence, (8) is a necessary and sufficient condition for the existence of a continuous model (1) on the SPX that jointly fits the SPX smile and the instantaneous VIX smile.

If the continuum of call prices on $\mathrm{instVIX}_t$ were accessible in the market, we could imply from it the distribution of $\mathrm{instVIX}_t^2$, and compare it to the distribution of $\sigma^2_{\mathrm{lv}}(t, S_t^{\mathrm{lv}})$ and check whether these two market-implied distributions satisfy (8).

Note that this general construction does not address the issue of the dynamics of the instantaneous volatility: σ_t and $\sigma_{t'}$ could be very loosely related for arbitrarily close t and t'. Passing to the continuous time limit may actually require some kind of continuity for the process (σ_t). In practice, if (8) holds, in order to build a calibrating process, one would discretize time and recursively solve martingale transport problems:

$$\begin{aligned} &\mathrm{Law}(\sigma^2_{\mathrm{lv}}(t_i, S^{\mathrm{lv}}_{t_i})) \text{ and } \mathrm{Law}(\sigma^2_{t_i}) \text{ given,} \\ &\mathbb{E}[\sigma^2_{t_i} | \sigma^2_{\mathrm{lv}}(t_i, S^{\mathrm{lv}}_{t_i})] = \sigma^2_{\mathrm{lv}}(t_i, S^{\mathrm{lv}}_{t_i}). \end{aligned} \tag{9}$$

Solutions to these martingale transport problems include left- and right-curtains (Beiglböck and Juillet 2016) (see also Henry-Labordère 2017, Section 2.2.3, which is the reference monograph on martingale optimal transport), forward-starting solutions to the Skorokhod embedding problem (Michon 2016), and the local variance gamma model of Carr and Nadtochiy (2017). Recent advances in computational martingale optimal transport include Guo and Obłój (2019) and De March (2018). Note that (9) is a new type of application of martingale transport to finance: usually, the martingality constraint applies to the underlying at two different dates (see Henry-Labordère 2017 and references therein); here it applies to two types of instantaneous variance—local and stochastic—at a single date, ensuring that the full surface of SPX implied volatilities is matched.

Even though the limiting case of the instantaneous VIX is only an approximation, it is interesting as it shows that it might be impossible to build a continuous model on the SPX that jointly calibrates to SPX and VIX options. In the limit, this happens if (and only if) for some t the market-implied distribution of $\sigma^2_{\mathrm{lv}}(t, S_t^{\mathrm{lv}})$ is 'more spread' than that of the instantaneous VIX squared.

4. The real case

In reality, the VIX is not the instantaneous volatility but the implied volatility of the 30-day log-contract on the SPX. Assume that we observe a continuum of market prices of VIX options maturing at T for all strikes $K \geq 0$. The VIX future corresponds to $K = 0$. Let $\mathrm{VIX}^2_{\mathrm{mkt},T}$ denote the distribution of VIX_T^2 implied from those market prices. In absence of arbitrage, we must have

$$\mathbb{E}[\mathrm{VIX}^2_{\mathrm{mkt},T}] = \mathbb{E}[\mathrm{VIX}^2_{\mathrm{lv},T}] \tag{10}$$

since the payoff VIX_T^2 can be replicated:

- either using the VIX future and VIX options maturing at T, which gives the price on the l.h.s. of (10):

$$\begin{aligned} &\mathrm{Price}[\mathrm{VIX}_T]^2 \\ &\quad + 2\int_0^{\mathrm{Price}[\mathrm{VIX}_T]} \mathrm{Price}[(K - \mathrm{VIX}_T)_+]\,\mathrm{d}K \\ &\quad + 2\int_{\mathrm{Price}[\mathrm{VIX}_T]}^{\infty} \mathrm{Price}[(\mathrm{VIX}_T - K)_+]\,\mathrm{d}K \\ &= \mathrm{Price}[\mathrm{VIX}_T^2] = \mathbb{E}[\mathrm{VIX}^2_{\mathrm{mkt},T}], \end{aligned}$$

- or using a calendar spread of SPX log-contracts maturing at T and $T + \tau$, which yields the price on the r.h.s. of (10):

$$
\begin{aligned}
&-\frac{2}{\tau}\{\mathrm{Price}(\ln(S_{T+\tau}))-\mathrm{Price}(\ln(S_T))\}\\
&=-\frac{2}{\tau}(\mathbb{E}[\ln(S^{\mathrm{lv}}_{T+\tau})]-\mathbb{E}[\ln(S^{\mathrm{lv}}_{T})])\\
&=-\frac{2}{\tau}\mathbb{E}\left[\ln\frac{S^{\mathrm{lv}}_{T+\tau}}{S^{\mathrm{lv}}_{T}}\right]=\mathbb{E}\left[\frac{1}{\tau}\int_T^{T+\tau}\sigma^2_{\mathrm{lv}}(t,S^{\mathrm{lv}}_t)\,\mathrm{d}t\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\frac{1}{\tau}\int_T^{T+\tau}\sigma^2_{\mathrm{lv}}(t,S^{\mathrm{lv}}_t)\,\mathrm{d}t\,\middle|\,S^{\mathrm{lv}}_T\right]\right]=\mathbb{E}[\mathrm{VIX}^2_{\mathrm{lv},T}].
\end{aligned}
$$

Violations of (10) in the market have been reported, suggesting arbitrage opportunities, see, e.g. Bergomi (2016, Section 7.7.4). However, the two quantities in (10) do not purely depend on market data. The l.h.s. depends on an (arbitrage-free) extrapolation of the smile of VIX_T beyond the smallest and largest quoted strikes, while the r.h.s. depends on (arbitrage-free) extrapolations of the SPX smile at maturities T and $T+\tau$.† Both sides can actually be made arbitrarily large by modifying the smile extrapolations (Guyon 2018b). The reported violations of (10) actually rely on some arbitrary smile extrapolations. In our numerical experiments, we always first build *consistent* (and reasonable) extrapolations of the VIX and SPX smiles so that (10) holds. Therefore, throughout the rest of this article, we assume that (10) holds true. In particular, since they have the same mean, the distributions $\mathrm{VIX}^2_{\mathrm{mkt},T}$ and $\mathrm{VIX}^2_{\mathrm{lv},T}$ may be rankable in convex order.

4.1. Market data: term-structure of convex ordering of $\mathrm{VIX}^2_{\mathrm{lv},T}$ and $\mathrm{VIX}^2_{\mathrm{mkt},T}$

A careful analysis of market data (reasonably interpolated and extrapolated) reveals, for the first time, a remarkable, surprising *term-structure of convex ordering* of $\mathrm{VIX}^2_{\mathrm{lv},T}$ and $\mathrm{VIX}^2_{\mathrm{mkt},T}$:

$$
\begin{cases}
\mathrm{VIX}^2_{\mathrm{lv},T}\geq_c \mathrm{VIX}^2_{\mathrm{mkt},T} \\
\quad\text{for short maturities } T\lesssim 2\text{–}3\text{ months,}\\
\mathrm{VIX}^2_{\mathrm{lv},T}\text{ and }\mathrm{VIX}^2_{\mathrm{mkt},T}\quad\text{are not rankable}\\
\quad\text{in convex order for } 2\text{–}3\text{ months}\lesssim T\lesssim 6\text{ months.}
\end{cases}
\tag{11}
$$

VIX options only trade for maturities up to 6 months. Extrapolating market data to longer maturities suggests that $\mathrm{VIX}^2_{\mathrm{lv},T}\leq_c \mathrm{VIX}^2_{\mathrm{mkt},T}$ for large enough T.

For short maturities, the market local volatility (LV) model yields a distribution of the squared VIX that is 'more spread' than the distribution of the squared VIX implied from VIX option prices, whereas for intermediate maturities the two distributions cannot be ranked in convex order. This is illustrated in figures 1 ($T=21$ days) and 2 ($T=77$ days) where we plot the market-implied distributions of $\mathrm{VIX}^2_{\mathrm{mkt},T}$ and $\mathrm{VIX}^2_{\mathrm{lv},T}$ as of August 1, 2018‡ (top left). Note that the density of $\mathrm{VIX}^2_{\mathrm{lv},T}$, which is infinite at the minimum possible value and then decreases, does not at all look like the bell-shaped density (with positive skewness) of $\mathrm{VIX}^2_{\mathrm{mkt},T}$. To verify if these two distributions are rankable in convex order, we plot the prices of call options on these quantities as a function of the strike (top right). The two price curves coincide at strike 0, since from (10) the two distributions have the same mean. If a price curve is always below the other, it means that the corresponding distribution is smaller in convex order. For example, figure 1 (top right) shows that $\mathrm{VIX}^2_{\mathrm{mkt},T}\leq_c \mathrm{VIX}^2_{\mathrm{lv},T}$ for $T=21$ days. If the price curves cross each other, the two corresponding distributions are not rankable in convex order, which is the case for $T=77$ days. The corresponding VIX futures and VIX smiles are also reported (bottom).

Let us look at the market data (as of August 1, 2018) in more detail. The positive slope of market VIX smiles (the red smile in the bottom graphs) means that the distribution $\mathrm{VIX}^2_{\mathrm{mkt},T}$ (the red distribution in the top left graphs) has positive skewness, i.e. a fat right tail. Let us look at the shortest VIX future and options maturity ($T=21$ days, figure 1). The steep negative ATM slope of market SPX smiles at short maturities yields steep negative slopes of market SPX local volatilities $\sigma_{\mathrm{lv}}(t,\cdot)$ for $t\in[T,T+\tau]$ for SPX values around the money and below. As a consequence, the function

$$
\psi : S\mapsto \mathbb{E}\left[\frac{1}{\tau}\int_T^{T+\tau}\sigma^2_{\mathrm{lv}}(t,S^{\mathrm{lv}}_t)\,\mathrm{d}t\,\middle|\,S^{\mathrm{lv}}_T=S\right],
$$

which is an average in time and space of SPX local variances over $[T,T+\tau]$, has a low minimum value $\psi_{\min}\approx(8.3\%)^2$ for $S\approx 2875$ and takes very large values for SPX values that are not so far from the money, e.g. $(48.5\%)^2$ for $S=2500$; see the blue curve on the middle left graph of figure 1. The distribution of $\mathrm{VIX}^2_{\mathrm{lv},T}=\psi(S^{\mathrm{lv}}_T)$ is the image of the risk-neutral distribution $\mathrm{Law}(S^{\mathrm{lv}}_T)$ of the SPX at maturity T by the function ψ; see the orange and blue distributions on the middle graphs of figure 1. Since ψ reaches its minimum in the bulk of $\mathrm{Law}(S^{\mathrm{lv}}_T)$, $\mathrm{Law}(\mathrm{VIX}^2_{\mathrm{lv},T})$ is very peaked at the low value $\psi_{\min}$.

In particular, the LV model predicts that the VIX can take low values, say less than 9%, with a much larger probability than the one implied by VIX options: $\mathbb{P}(\mathrm{VIX}_{\mathrm{lv},T}\leq 9\%)$ is much larger than $\mathbb{P}(\mathrm{VIX}_{\mathrm{mkt},T}\leq 9\%)$, computed with any reasonable arbitrage-free extrapolation of the VIX smile for low strikes;§ and for small strikes around 9%, put (hence call) prices on $\mathrm{Law}(\mathrm{VIX}^2_{\mathrm{lv},T})$ are more expensive than put (hence call) prices on $\mathrm{Law}(\mathrm{VIX}^2_{\mathrm{mkt},T})$.¶ Moreover, for reasonable SPX and VIX smile extrapolations, $\mathrm{Law}(\mathrm{VIX}^2_{\mathrm{lv},T})$ has a

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

‡ The same patterns were observed for the many other dates that we have tested.

§ Note that, at the time of writing, the all-time lowest VIX close was 9.14% on November 3, 2017.

¶ Note, however, that if we extrapolate the market VIX smile so that VIX implied volatilities are positive for strikes smaller than $\sqrt{\psi_{\min}}$, then for those strikes the puts on $\mathrm{VIX}^2_{\mathrm{lv},T}$ are worthless while the puts on $\mathrm{VIX}^2_{\mathrm{mkt},T}$ have positive values. Yet, given the quoted VIX implied volatilities, the value of $\psi_{\min}$, and the value of the all-time lowest VIX close (9.14%), it is reasonable to extrapolate the VIX smile so that VIX implied volatilities are zero below $\sqrt{\psi_{\min}}$ (see figure 1 and remark 1). That is what we do.

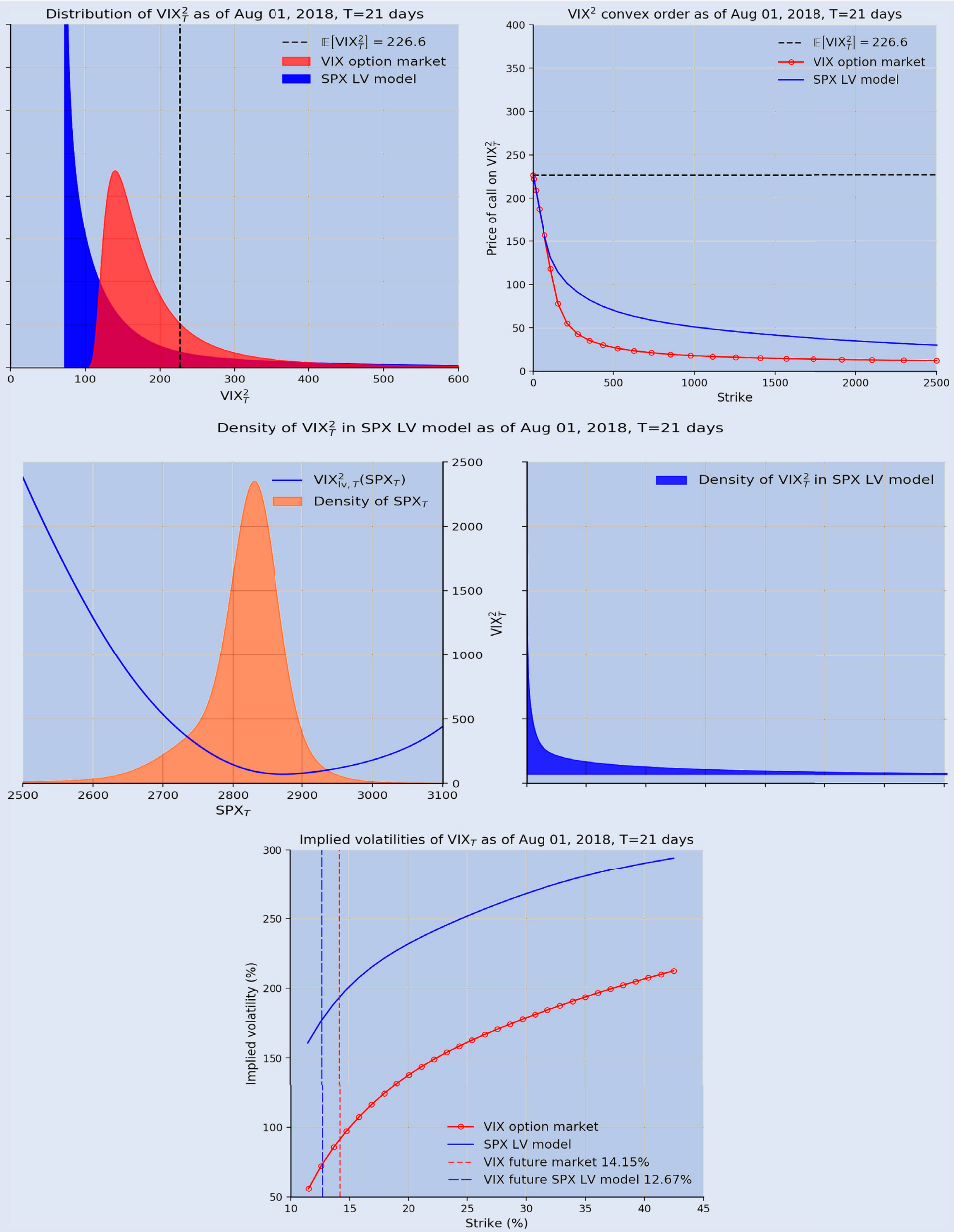

Figure 1. Top left: Market-implied distributions of $\mathrm{VIX}^2_{\mathrm{lv},T}$ (blue) and $\mathrm{VIX}^2_{\mathrm{mkt},T}$ (red). Top right: Prices of call options on $\mathrm{VIX}^2_{\mathrm{lv},T}$ (blue) and $\mathrm{VIX}^2_{\mathrm{mkt},T}$ (red) as a function of the strike. Middle: Market-implied density of S_T^{lv} (left) and its transform $\mathrm{VIX}^2_{\mathrm{lv},T}$ via the function $\psi(S_T^{\mathrm{lv}}) = \mathrm{VIX}^2_{\mathrm{lv},T}(S_T^{\mathrm{lv}})$ (right). Bottom: Implied volatilities of $\mathrm{VIX}_{\mathrm{lv},T}$ (blue) and $\mathrm{VIX}_{\mathrm{mkt},T}$ (red) as a function of the strike. $T = 21$ days, market data as of August 1, 2018. Forward SPX value: 2811.68.

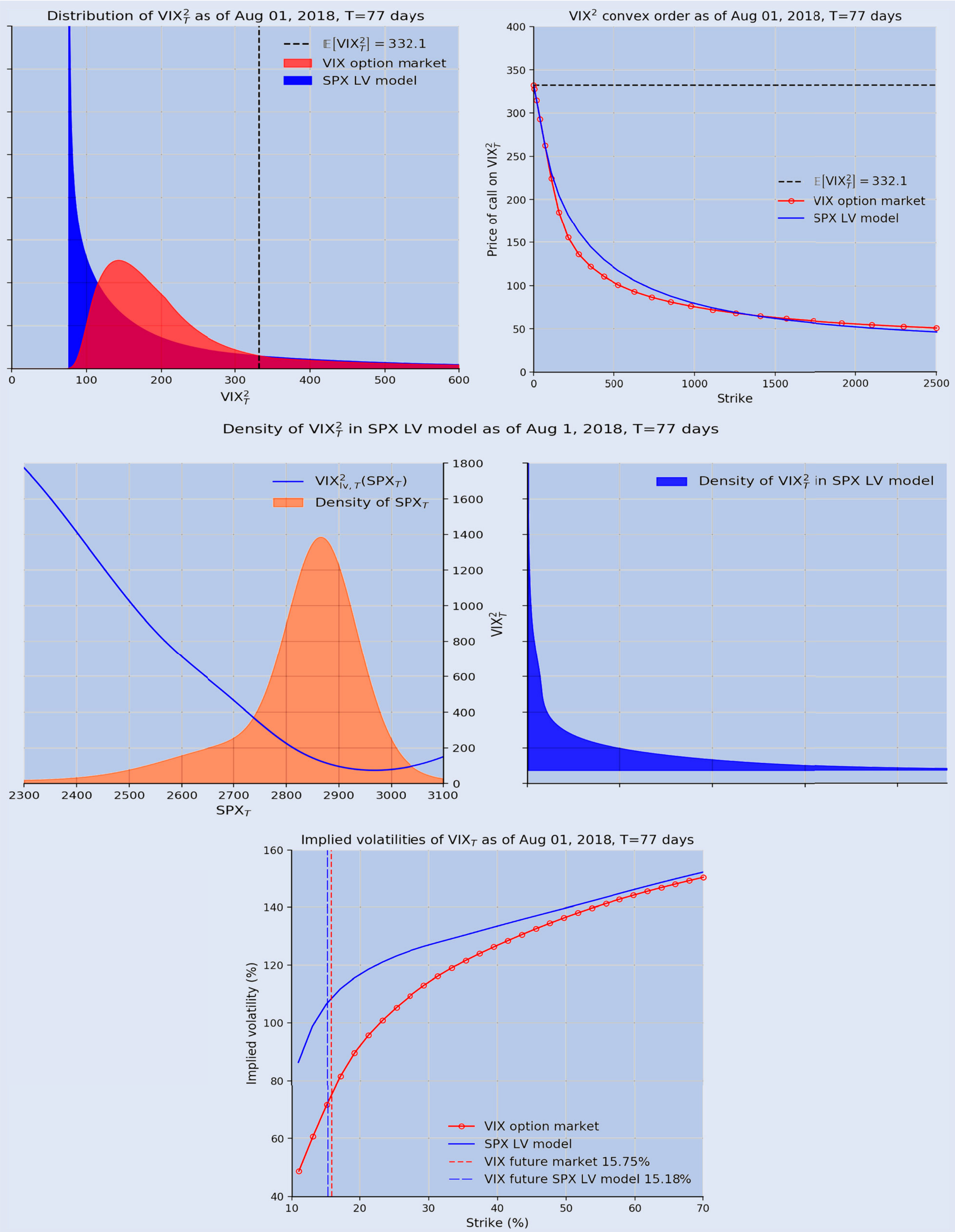

Figure 2. Top left: Market-implied distributions of $VIX^2_{lv,T}$ (blue) and $VIX^2_{mkt,T}$ (red). Top right: Prices of call options on $VIX^2_{lv,T}$ (blue) and $VIX^2_{mkt,T}$ (red) as a function of the strike. Middle: Market-implied density of S_T^{lv} (left) and its transform $VIX^2_{lv,T}$ via the function $\psi(S_T^{lv}) = VIX^2_{lv,T}(S_T^{lv})$ (right). Bottom: Implied volatilities of $VIX_{lv,T}$ (blue) and $VIX_{mkt,T}$ (red) as a function of the strike. $T = 77$ days, market data as of August 1, 2018. Forward SPX value: 2814.34.

larger right tail than $\mathrm{Law}(\mathrm{VIX}^2_{\mathrm{mkt},T})$, due to the large left tail of $\mathrm{Law}(S^{\mathrm{lv}}_T)$ (e.g. $\mathbb{P}(S^{\mathrm{lv}}_T \le 2500) \approx 1\%$) and the large values taken by ψ on the left-hand side. This materializes in the fact that for large strikes calls on $\mathrm{VIX}^2_{\mathrm{lv},T}$ are more expensive than calls on $\mathrm{VIX}^2_{\mathrm{mkt},T}$.† Eventually, call prices on $\mathrm{Law}(\mathrm{VIX}^2_{\mathrm{lv},T})$ are more expensive than call prices on $\mathrm{Law}(\mathrm{VIX}^2_{\mathrm{mkt},T})$ for all strikes (top right graph), i.e. $\mathrm{VIX}^2_{\mathrm{mkt},T} \le_c \mathrm{VIX}^2_{\mathrm{lv},T}$.

As a result, for this maturity, the VIX future is much cheaper in the LV model than in the market (by concavity of the square root; bottom graph), and the VIX puts are more expensive in the LV model (by convexity of $x \mapsto (K - \sqrt{x})_+$). Moreover, it can be numerically checked that the implied volatilities of VIX options produced by the LV model are much larger (approximately 100 volatility points larger across all listed strikes!) than the market ones (bottom graph). This huge gap between LV and market VIX implied volatilities is a sign that it may be difficult to simultaneously calibrate a continuous stochastic volatility model to short term SPX and VIX option prices.

Let us now look at a larger maturity, $T = 77$ days (figure 2). Since market local volatilities are not as steep for these larger maturities $t \in [T, T + \tau]$, the function ψ is 'flatter'. As a consequence, the minimum $\psi_{\min}$ of ψ is slightly larger than for $T = 21$ days, and the right tail of $\mathrm{Law}(\mathrm{VIX}^2_{\mathrm{lv},T})$ is not as fat. The top right graph shows that the two distributions are not rankable in convex order: calls on VIX^2_T are more expensive in the LV model for small strikes, but they are cheaper for large strikes. For this maturity, the differences between LV and market implied volatilities of VIX options are smaller, as well as the difference between LV and market values of the VIX future (bottom graph), a sign that it may be easier to jointly calibrate a continuous stochastic volatility model to SPX and VIX option prices for these maturities.

VIX options only trade for maturities up to 6 months. Extrapolating market data to maturities longer than 6 months suggests that calls on VIX^2_T would eventually become cheaper in the LV model for all strikes, i.e. $\mathrm{VIX}^2_{\mathrm{lv},T} \le_c \mathrm{VIX}^2_{\mathrm{mkt},T}$.

Remark 1 (Convex ordering of $\mathrm{VIX}^2_{\mathrm{mkt},T}$ and $\mathrm{VIX}^2_{\mathrm{lv},T}$ typically depend on smile extrapolations) The fact that, say, $\mathrm{VIX}^2_{\mathrm{lv},T} \ge_c \mathrm{VIX}^2_{\mathrm{mkt},T}$ typically does not purely depend on market data. Both SPX and VIX smiles are usually interpolated and extrapolated (in an arbitrage-free way). As explained at the beginning of this section, in our numerical tests, we perform a joint extrapolation of SPX and VIX smiles so that (10) holds. Whether $\mathrm{VIX}^2_{\mathrm{lv},T} \ge_c \mathrm{VIX}^2_{\mathrm{mkt},T}$ depends on the chosen extrapolations. In our numerical experiments, we have used realistic extrapolations based on cubic splines. For example, for $T = 21$ days (figure 1), given the quoted VIX implied volatilities, the value of $\psi_{\min}$, and the value of the all-time lowest VIX close (9.14% at the time of writing), we extrapolated the VIX smile so that VIX implied volatilities are zero below $\sqrt{\psi_{\min}}$. With this reasonable choice, $\mathrm{VIX}^2_{\mathrm{lv},T} \ge_c \mathrm{VIX}^2_{\mathrm{mkt},T}$ does hold. But one could also choose a quite unrealistic small strike extrapolation of the VIX smile such that $\mathbb{P}(\mathrm{VIX}_{\mathrm{mkt},T} < \sqrt{\psi_{\min}}) > 0$; in this case, since $\psi_{\min}$ is the smallest value taken by $\mathrm{VIX}^2_{\mathrm{lv},T}$, the VIX put with strike $\sqrt{\psi_{\min}}$ would be strictly more expensive in the (extrapolated) market than in the LV model, therefore $\mathrm{VIX}^2_{\mathrm{lv},T} \ge_c \mathrm{VIX}^2_{\mathrm{mkt},T}$ would not hold.

A well-posed mathematical problem is: Does there exist jointly arbitrage-free interpolations and extrapolations of both SPX and VIX market data such that $\mathrm{VIX}^2_{\mathrm{lv},T} \ge_c \mathrm{VIX}^2_{\mathrm{mkt},T}$? such that $\mathrm{VIX}^2_{\mathrm{lv},T} \le_c \mathrm{VIX}^2_{\mathrm{mkt},T}$? such that $\mathrm{VIX}^2_{\mathrm{lv},T}$ and $\mathrm{VIX}^2_{\mathrm{mkt},T}$ are not rankable in convex order? These are difficult questions. See Appendix 1 for partial results. Note that the problem of determining whether SPX and VIX market data are jointly arbitrage-free is already a very difficult one, which was only recently solved in Guyon (2020).

4.2. *Inversion of convex ordering: a necessary condition for joint calibration*

From the case of instantaneous VIX (see section 3), since 30 days is a relatively short maturity for SPX options, one may be tempted to believe that there exists a model of the form (1) calibrated to the full surface of SPX implied volatilities and to the VIX smiles if and only if for all T, $\mathrm{VIX}^2_{\mathrm{lv},T} \le_c \mathrm{VIX}^2_{\mathrm{mkt},T}$—and then conclude that such a model does not exist, since for short maturities T this inequality does not hold. However, the timewise convex ordering of $\sigma^2_{\mathrm{lv}}(t, S_t)$ and σ^2_t does *not* necessarily imply that the random variables‡ $\mathbb{E}[(1/\tau)\int_T^{T+\tau} \sigma^2_{\mathrm{lv}}(u, S_u)\,\mathrm{d}u \,|\, \mathcal{F}_T]$ and $\mathbb{E}[(1/\tau)\int_T^{T+\tau} \sigma^2_u\,\mathrm{d}u \,|\, \mathcal{F}_T]$ are in convex order. Indeed, time-averaging over 30 days and conditioning on $\mathcal{F}_T$ may undo convex ordering (see appendix 2).

From the first line of (11), a necessary condition for model (1) to jointly fit the full surface of SPX implied volatilities and the (reasonably extrapolated) VIX smiles is that it satisfies what we call the *inversion of convex ordering* property:

Definition 2 *We say that model* (1) *satisfies the inversion of convex ordering property if and only if, for small maturities* T, $\mathrm{VIX}^2_T \le_c \mathrm{VIX}^2_{\mathrm{loc},T}$, *i.e.*

$$\mathbb{E}\left[\frac{1}{\tau}\int_T^{T+\tau} \sigma^2_u\,\mathrm{d}u \,\middle|\, \mathcal{F}_T\right] \le_c \mathbb{E}\left[\frac{1}{\tau}\int_T^{T+\tau} \sigma^2_{\mathrm{loc}}(u, S^{\mathrm{loc}}_u)\,\mathrm{d}u \,\middle|\, S^{\mathrm{loc}}_T\right]. \tag{12}$$

Typically, 'small T' means $T \le \bar{T}$, with $\bar{T}$ equal to a few months, i.e. a few times τ. Inversion of convex ordering means that, despite the fact that for all $u \ge 0$, $\sigma^2_{\mathrm{loc}}(u, S^{\mathrm{loc}}_u) \le_c \sigma^2_u$ (recall (7)), $\mathrm{VIX}^2_{\mathrm{loc},T} \ge_c \mathrm{VIX}^2_T$ for small T. The convex ordering between local variances and stochastic variances is reversed, when we move from instantaneous variance to the squared VIX.

Statement (11) actually gives a more precise information about the term-structure of convex ordering of $\mathrm{VIX}^2_{\mathrm{lv},T}$ and

† Note, however, that one could possibly build unrealistic arbitrage-free extrapolations of the SPX and VIX smiles such that, for very large strikes, calls on $\mathrm{VIX}^2_{\mathrm{mkt},T}$ are more expensive than calls on $\mathrm{VIX}^2_{\mathrm{lv},T}$.

‡ From now on, we denote by u, instead of t, a generic time between T and $T + \tau$. This will allow us to use the classical notation ξ^u_t for instantaneous forward variances, $t \le u$.

$\mathrm{VIX}^2_{\mathrm{mkt},T}$. In particular, a more restrictive necessary condition for model (1) to jointly fit the full surface of SPX implied volatilities and the (reasonably extrapolated) VIX smiles is that it satisfies what we call the *market term-structure of VIX convex ordering* property:

DEFINITION 3 *We say that model* (1) *satisfies the market term-structure of VIX convex ordering property if and only if there exist* $0 < T_1^* < T_2^*$ *such that*

$$\begin{cases} \mathrm{VIX}^2_{\mathrm{loc},T} \geq_c \mathrm{VIX}^2_T \ \textit{for}\ T \leq T_1^*, \\ \mathrm{VIX}^2_{\mathrm{loc},T} \ \textit{and}\ \mathrm{VIX}^2_T \ \textit{are not rankable} \\ \quad \textit{in convex order for}\ T_1^* < T < T_2^*. \end{cases} \tag{13}$$

We say that model (1) *satisfies the extrapolated market term-structure of VIX convex ordering property if and only if there exist* $0 < T_1^* < T_2^*$ *such that*

$$\begin{aligned} \mathrm{VIX}^2_{\mathrm{loc},T} &\geq_c \mathrm{VIX}^2_T \quad \textit{for}\ T \leq T_1^*, \\ \mathrm{VIX}^2_{\mathrm{loc},T} &\leq_c \mathrm{VIX}^2_T \quad \textit{for}\ T \geq T_2^*. \end{aligned} \tag{14}$$

The definition of the *extrapolated market term-structure of VIX convex ordering* property (14) is based on the maturity extrapolation explained below (11). Obviously, if model (1) satisfies the extrapolated market term-structure of VIX convex ordering property (14), it also likely possesses the market term-structure of VIX convex ordering property (13).

In a companion paper, Acciaio and Guyon (2020) have recently proved that (strict) inversion of convex ordering can be produced by continuous stochastic volatility models (1). Their example can also capture that $\mathrm{VIX}^2_{\mathrm{loc},T} <_c \mathrm{VIX}^2_T$ for all large enough T.† However, the mathematical proof uses an *ad hoc*, unrealistic stochastic volatility model. In the following sections, we follow a different route: we investigate realistic stochastic volatility models and *numerically* show that for model parameters lying in a certain domain they are able to generate the desired term-structure of convex ordering (14).

Note that $\mathbb{E}[\sigma_u^2|\mathcal{F}_T]$ and $\mathbb{E}[\sigma^2_{\mathrm{loc}}(u, S_u^{\mathrm{loc}})|S_T^{\mathrm{loc}}]$ have the same mean, $\mathbb{E}[\sigma_u^2]$. One natural way to achieve (12) is to require that

$$\mathbb{E}[\sigma_u^2|\mathcal{F}_T] \leq_c \mathbb{E}[\sigma^2_{\mathrm{loc}}(u, S_u^{\mathrm{loc}})|S_T^{\mathrm{loc}}] \tag{15}$$

for many $u \in (T, T+\tau]$ and check that this convex ordering of forward instantaneous variances is preserved when we sum over u. That is, we produce the inversion of convex ordering by projecting instantaneous variances onto $\mathcal{F}_T$ rather than by averaging them over 30 days. When $\mathcal{F}_T$ contains little information on σ_u^2, (15) may hold. In Acciaio and Guyon's extreme example (Acciaio and Guyon 2020), for small T, $(\sigma_u)_{u\in[T,T+\tau]}$ is chosen to be independent of $\mathcal{F}_T$ so that $\mathbb{E}[\sigma_u^2|\mathcal{F}_T]$ is a constant and satisfies (15). Note that in general (15) is unlikely to hold for u close to T, as when $u = T$, $\mathbb{E}[\sigma_u^2|\mathcal{F}_T] = \sigma_T^2 \geq_c \sigma^2_{\mathrm{loc}}(T, S_T^{\mathrm{loc}}) = \mathbb{E}[\sigma^2_{\mathrm{loc}}(u, S_u^{\mathrm{loc}})|S_T^{\mathrm{loc}}]$. Therefore we will require that (15) holds for all $T \leq T_1^*$ and $u \in [T+\varepsilon, T+\tau]$, for some small $\varepsilon > 0$. When (15) holds, the convex ordering $\sigma^2_{\mathrm{loc}}(u, S_u^{\mathrm{loc}}) \leq_c \sigma_u^2$ is actually reversed after conditioning on $\mathcal{F}_T$:

$$\mathbb{E}[\sigma_u^2|\mathcal{F}_T] \leq_c \mathbb{E}[\sigma^2_{\mathrm{loc}}(u, S_u^{\mathrm{loc}})|\mathcal{F}_T] \leq_c \sigma^2_{\mathrm{loc}}(u, S_u^{\mathrm{loc}}) \leq_c \sigma_u^2. \tag{16}$$

In the next section, we explain why we expect realistic stochastic volatility models with large spot–vol correlation (in absolute value) and (a) fast mean reversion of volatility, or (b) small Hurst exponent, to satisfy the market term-structure of VIX convex ordering property (at least when extremely small strikes are ignored when we compare call prices on the VIX squared, see remark 6). Then we numerically check it in the following section.

5. One-factor lognormal forward instantaneous variance models

5.1. Setting

The quantity $\mathbb{E}[\sigma_u^2 \,|\, \mathcal{F}_T]$ in (15) is a forward instantaneous variance. We denote it by

$$\xi_T^u := \mathbb{E}[\sigma_u^2 \,|\, \mathcal{F}_T], \qquad u \geq T.$$

It is the instantaneous lognormal variance of S at $u \geq T$ seen from T. It is well known (Dupire 1993, Bergomi 2005) that forward instantaneous variances are driftless. Second-generation stochastic volatility models directly model the dynamics of $(\xi_t^u)_{t\in[0,u]}$ under a risk-neutral measure. The only requirement is that these processes, indexed by u, be nonnegative and driftless. For simplicity, let us assume in this section that forward instantaneous variances are lognormal and all driven by a single standard one-dimensional $(\mathcal{F}_t)$-Brownian motion Z, correlated with the Brownian motion W that drives the SPX dynamics:

$$\frac{\mathrm{d}\xi_t^u}{\xi_t^u} = K(t,u)\,\mathrm{d}Z_t. \tag{17}$$

The SPX dynamics simply reads as (1) with $\sigma_t^2 := \xi_t^t$:

$$\frac{\mathrm{d}S_t}{S_t} = \sigma_t\,\mathrm{d}W_t, \quad \sigma_t^2 := \xi_t^t, \quad \mathrm{d}\langle W, Z\rangle_t = \rho\,\mathrm{d}t. \tag{18}$$

To reproduce the negative link between SPX returns and SPX volatility, a negative correlation ρ is used. The so-called kernel K is a deterministic function of current time t and forward instantaneous variance maturity $u \geq t$. The initial condition ξ_0^u is computed from the market prices $\mathrm{VS}(T)$ of variances swaps on the SPX: $\xi_0^u = (\mathrm{d}/\mathrm{d}u)(u\mathrm{VS}(u))$. The solution to (17) is simply $\xi_t^u = \xi_0^u \exp(\int_0^t K(s,u)\,\mathrm{d}Z_s - \frac{1}{2}\int_0^t K(s,u)^2\,\mathrm{d}s)$, which yields

$$\sigma_u^2 = \xi_0^u \exp\left(\int_0^u K(s,u)\,\mathrm{d}Z_s - \frac{1}{2}\int_0^u K(s,u)^2\,\mathrm{d}s\right). \tag{19}$$

† The notation $X <_c Y$ means that X is strictly smaller than Y in convex order, i.e. $X \leq_c Y$ and $\mathrm{Law}(X) \neq \mathrm{Law}(Y)$. In appendix 3, we prove that this is essentially equivalent to the fact that for any strictly convex function f such that $\mathbb{E}[f(X)]$ and $\mathbb{E}[f(Y)]$ exist and are finite, $\mathbb{E}[f(X)] < \mathbb{E}[f(Y)]$ (see theorem A.2).

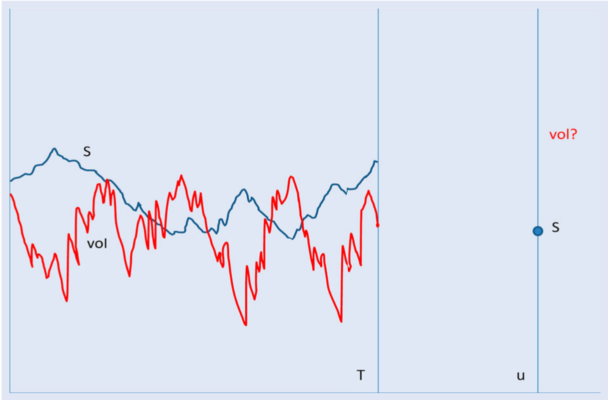

Figure 3. Ingredients required for inversion of convex ordering: loosely speaking, S_u should give more information than $\mathcal{F}_T$ on σ_u^2.

For simplicity, let us choose a time-homogeneous kernel,

$$K(s,u) = K(u-s).$$

This is a natural stationarity assumption: the (lognormal) volatility of ξ_t^u depends on t and u only through $u-t$. In this case, K acts as a convolution kernel on the Brownian increments, through $\int_0^t K(u-s)\,\mathrm{d}Z_s$. Financially, we expect the kernel $K : \mathbb{R}_+ := [0,+\infty) \to \mathbb{R}_+$ to be decreasing: the further the instantaneous forward variance maturity u, the less volatile the instantaneous forward variance.

5.2. *Ingredients required for the inversion of convex ordering*

Can we choose a kernel K such that (15) holds for all $T \le T_1^*$ and $u \in [T+\varepsilon, T+\tau]$? Such a kernel K should make the distribution of $\mathbb{E}[\sigma_u^2|\mathcal{F}_T]$ 'more narrow' than that of $\mathbb{E}[\sigma_{\mathrm{loc}}^2(u,S_u^{\mathrm{loc}})\,|\,S_T^{\mathrm{loc}}]$ for those T and u. For this to hold, two ingredients are needed:

- **I1**: The knowledge of $\mathcal{F}_T := \sigma((W_s,Z_s)_{0\le s\le T})$ should give little information on σ_u^2, so that the distribution of $\mathbb{E}[\sigma_u^2\,|\,\mathcal{F}_T]$ is narrow.
- **I2**: The knowledge of S_u should give significant enough information on σ_u^2, so that $S \mapsto \sigma_{\mathrm{loc}}^2(u,S) = \mathbb{E}[\sigma_u^2\,|\,S_u = S]$ varies enough with S and the distribution of $\mathbb{E}[\sigma_{\mathrm{loc}}^2(u,S_u^{\mathrm{loc}})\,|\,S_T^{\mathrm{loc}}]$ is not as narrow as that of $\mathbb{E}[\sigma_u^2|\mathcal{F}_T]$.

Loosely speaking, S_u should give more information than $\mathcal{F}_T$ on σ_u^2. This may look surprising as S_u is just one real number while $\mathcal{F}_T$ represents two paths (the SPX and its instantaneous volatility), but only the paths information up to time T is taken into account while S_u gives an information at the same time $u > T$ as the time at which the instantaneous variance σ_u^2 is evaluated. Figure 3 illustrates this situation.

Ingredients **I1** and **I2** are somewhat antagonistic. From (19), for given T and $u \in (T, T+\tau]$, **I1** requires that $K(u-s)$ be small for all $s \in [0,T]$. Conversely, **I2** requires that (a) $K(u-s)$ be (comparatively) large for at least some $s \in [0,u]$, and (b) ρ, which is typically negative, be large in absolute value. Indeed, the knowledge of S_u gives partial information on $(W_s, 0 \le s \le u)$, which is passed to $(Z_s, 0 \le s \le u)$ through the correlation ρ; this information can impact σ_u^2 only if $K(u-s)$ is large for at least some $s \in [0,u]$.† For **I1** and **I2** to hold jointly, it is thus required that on average $K(u-s)$ be smaller for $s \in [0,T]$ than for $s \in [T,u]$, i.e. that $K(\theta)$ be on average smaller for $\theta \in [u-T,u]$ than for $\theta \in [0,u-T]$. Since this should hold for many $u \in (T, T+\tau]$, $\theta \mapsto K(\theta)$ should be fast decreasing. Such a fast decreasing kernel K is reminiscent of fast mean reversion, with short characteristic time around $\tau = 30$ days or less.

REMARK 4 (Term-structure of convex ordering) One nice feature of the models (17)–(18) with fast decreasing kernel K and large absolute spot–vol correlation $|\rho|$ is that we expect them to not only generate inversion of convex ordering (12) at short maturities but also produce the term-structure of convex ordering (13) and/or (14). Indeed, the smaller T, the more information S_u gives on $(\mathrm{d}W_s, u-\varepsilon \le s \le u)$, hence on $(\mathrm{d}Z_s, u-\varepsilon \le s \le u)$, and finally on σ_u^2, for $u \in [T, T+\tau]$. This is because the smaller T, the smaller u, and the larger the portion of time ε/u covered by $[u-\varepsilon, u]$ out of $[0,u]$. Stated otherwise, in those models, the local volatility $S \mapsto \sigma_{\mathrm{loc}}(t,S)$ flattens over time, as typically observed in the market. As a consequence, the distribution of $\mathbb{E}[\sigma_{\mathrm{loc}}^2(u,S_u^{\mathrm{loc}})|S_T^{\mathrm{loc}}]$ narrows over time for fixed $u-T$. Comparatively, the distribution of

$$\mathbb{E}[\sigma_u^2\,|\,\mathcal{F}_T] = \xi_T^u = \xi_0^u \exp\left(\int_0^T K(u-s)\,\mathrm{d}Z_s - \frac{1}{2}\int_0^T K(u-s)^2\,\mathrm{d}s\right)$$

does not depend much on T for fixed $u-T$, if $(\xi_0^u, u \ge 0)$ does not vary much, since the kernel K is fast decreasing. Therefore, in models (17)–(18), (15) is likely to hold for small T and not to hold for large T. This is precisely in line with market data (11).

REMARK 5 (Local volatility does not mean revert) The fact that $S \mapsto \sigma_{\mathrm{loc}}^2(u,S)$ varies a lot with S does not necessarily mean that the distribution of $\mathbb{E}[\sigma_{\mathrm{loc}}^2(u,S_u^{\mathrm{loc}})\,|\,S_T^{\mathrm{loc}}]$ is spread. Precisely, the above procedure describes a model where $\mathbb{E}[\sigma_u^2\,|\,\mathcal{F}_T]$ is much smaller than σ_u^2 in convex order, and the cautious reader may wonder how much smaller in convex order $\mathbb{E}[\sigma_{\mathrm{loc}}^2(u,S_u^{\mathrm{loc}})\,|\,S_T^{\mathrm{loc}}]$ is, compared to $\sigma_{\mathrm{loc}}^2(u,S_u^{\mathrm{loc}})$. Since $\sigma_{\mathrm{loc}}(t,S_t^{\mathrm{loc}})$ does not mean revert, contrary to σ_t, we expect $\mathbb{E}[\sigma_{\mathrm{loc}}^2(u,S_u^{\mathrm{loc}})\,|\,S_T^{\mathrm{loc}}]$ to be only slightly smaller than $\sigma_{\mathrm{loc}}^2(u,S_u^{\mathrm{loc}})$ in convex order. More precisely, since $(S_t^{\mathrm{loc}})_{t\ge0}$ is a martingale, the knowledge of S_T^{loc} gives a lot of information on S_u^{loc}, $u \in [T, T+\tau]$, hence on $\sigma_{\mathrm{loc}}^2(u,S_u^{\mathrm{loc}})$. This is indeed verified numerically, e.g. by comparing figures 4 and 6 in section 6.

Two well-known examples of deterministic kernels K are the exponential kernel and the power-law kernel, which we briefly recall below.

† The knowledge of S_u may also give partial direct information on $(\sigma_s, 0 \le s \le u)$. For instance, if S_u is extremely large, then many σ_s, $0 \le s \le u$, must have been very large, and σ_u is likely to be large. This explains why the smile has a positive slope at large strikes in stochastic volatility models even if $\rho < 0$. However, for values of S_u not too far from S_0, the knowledge of S_u is transferred to σ_u^2 mostly through the paths of W and Z up to u, via the spot–vol correlation ρ.

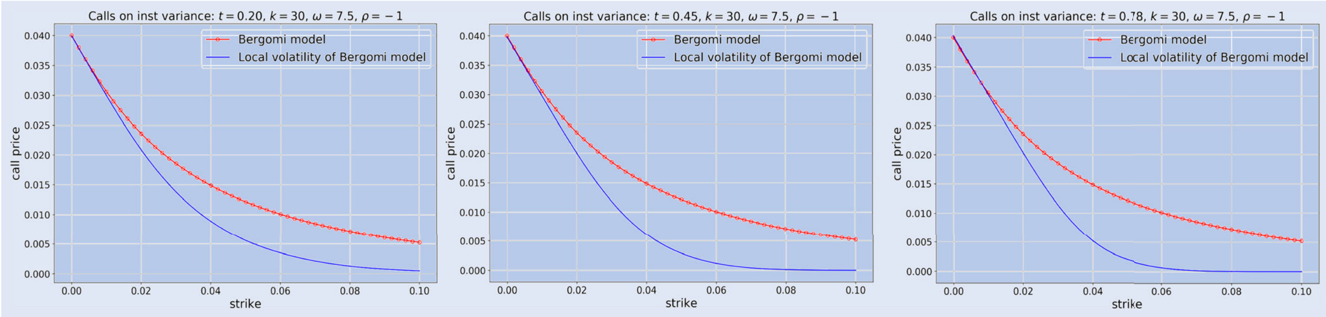

Figure 4. Call on instantaneous variance in the Bergomi model and its associated local volatility model; left: $t = 0.2$; middle: $t = 0.45$; right: $t = 0.78$; $k = 30$, $\omega = 7.5$, $\rho = -1$.

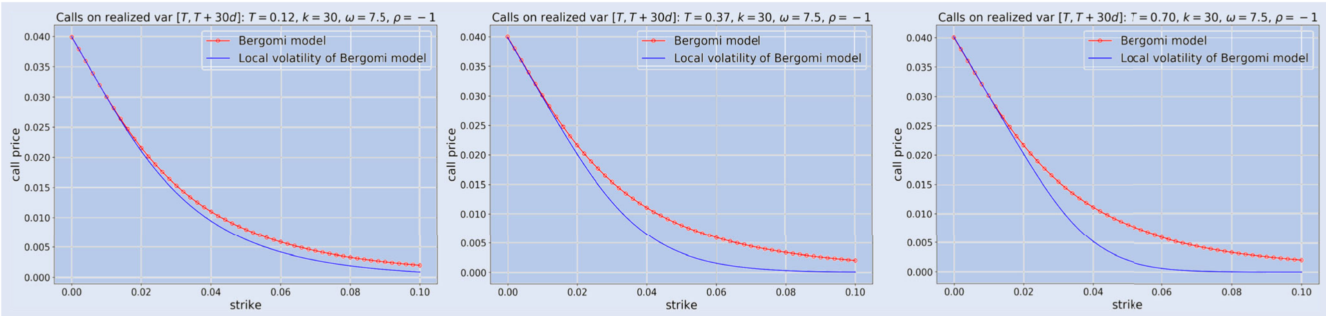

Figure 5. Call on realized variance over $[T, T + \tau]$ in the Bergomi model and its associated local volatility model; left: $T = 0.12$; middle: $T = 0.37$; right: $T = 0.70$; $k = 30$, $\omega = 7.5$, $\rho = -1$.

5.3. *Exponential kernel: one-factor Bergomi model*

The exponential kernel reads (Dupire 1993, Bergomi 2005):

$$K(\theta) = \omega \exp(-k\theta), \quad \omega \geq 0,\ k > 0.$$

In this case, ξ_t^u admits a one-dimensional Markov representation: $\xi_t^u = \xi_0^u f^u(t, X_t)$ where

$$f^u(t,x) = \exp\left(\omega\, \mathrm{e}^{-k(u-t)} X_t - \frac{\omega^2}{2}\, \mathrm{e}^{-2k(u-t)} \operatorname{Var}(X_t)\right),$$

$$\operatorname{Var}(X_t) = \frac{1 - \mathrm{e}^{-2kt}}{2k}$$

and the Ornstein–Uhlenbeck process $X_t = \int_0^t \mathrm{e}^{-k(t-s)}\, \mathrm{d}Z_s$ follows the Markov dynamics:

$$\mathrm{d}X_t = -kX_t\, \mathrm{d}t + \mathrm{d}Z_t, \quad X_0 = 0. \tag{20}$$

From (20), the parameter k should be interpreted as a parameter of mean reversion: it is the rate at which the factor X that drives all the forward instantaneous variances mean reverts to zero. As for ω, which has the dimension of a volatility, it is a parameter of volatility of volatility: it is the instantaneous (lognormal) volatility of the instantaneous variance $\sigma_t^2 := \xi_t^t$; the instantaneous (lognormal) volatility of the instantaneous volatility σ_t is simply $\omega/2$. This model, first introduced by Dupire (1993), is known as the (one-factor) Bergomi model (Bergomi 2005). For the inversion of convex ordering to hold, we choose the kernel K to be fast decreasing, i.e. we impose that k be large, of the order of magnitude of $1/\tau$ or larger. We are then in a fast mean reversion regime, where the characteristic time of mean reversion $1/k$ is similar to or smaller than the (already small) length τ of the time window used in the VIX computation (30 days).

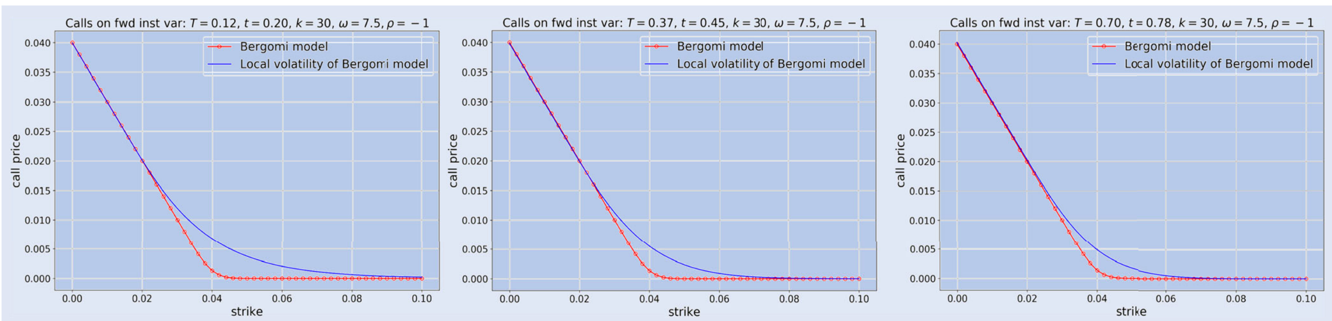

Figure 6. Call on forward instantaneous variance ξ_T^t in the Bergomi model and its associated local volatility model; left: $T = 0.12$; middle: $T = 0.37$; right: $T = 0.70$; $t = T + \tau$; $k = 30$, $\omega = 7.5$, $\rho = -1$.

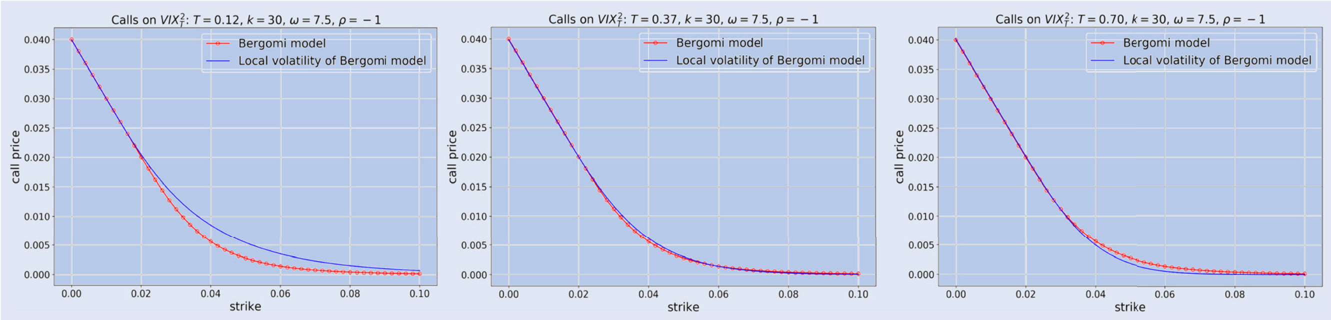

Figure 7. Call on VIX_T^2 in the Bergomi model and its associated local volatility model; left: $T = 0.12$; middle: $T = 0.37$; right: $T = 0.70$; $k = 30$, $\omega = 7.5$, $\rho = -1$.

Remark 6 (Small strikes) Strictly speaking, the inversion of convex ordering may fail to hold due to the left tails of the two distributions $\mathrm{VIX}^2_{\mathrm{loc},T}$ and VIX_T^2. Indeed,

$$\mathrm{VIX}_T^2 = \frac{1}{\tau}\int_T^{T+\tau} \xi_T^u \,\mathrm{d}u = \frac{1}{\tau}\int_T^{T+\tau} \xi_0^u \times \exp\left(\omega\, \mathrm{e}^{-k(u-T)} X_T - \frac{\omega^2}{2}\mathrm{e}^{-2k(u-T)}\mathrm{Var}(X_T)\right)\mathrm{d}u$$

has support $(0,+\infty)$ since X_T is Gaussian and is thus supported by the whole real line. Now, the support of $\mathrm{VIX}^2_{\mathrm{loc},T}$ is likely bounded away from zero. As a consequence, for small enough strikes, puts on $\mathrm{VIX}^2_{\mathrm{loc},T}$ are worthless while puts on VIX_T^2 have positive price. However, X_T being Gaussian, the left tail of VIX_T^2 is extremely thin and the issue is immaterial: those puts on VIX_T^2 are numerically worthless (see figures 7 and 10).

Remark 7 (Limiting regimes) The limiting regime where k tends to $+\infty$ while ω is kept constant is simply the Black–Scholes model with deterministic, time-dependent variance $\sigma_t^2 = \xi_0^t$. In this regime, both realized variances over $[T, T+\tau]$, $X := (1/\tau)\int_T^{T+\tau}\sigma_t^2\,\mathrm{d}t$ and $Y := (1/\tau)\int_T^{T+\tau}\sigma^2_{\mathrm{loc}}(t, S_t^{\mathrm{loc}})\,\mathrm{d}t$, are constant, and they are equal, so there is no strict inversion of convex ordering.

Another limiting regime, in which the variance of σ_t^2 has a finite limit, makes more financial sense: when k and ω tend to $+\infty$ while ω^2/k is kept constant. This corresponds to an ergodic limit where $(\omega X_t)_{t\ge 0}$ quickly reaches its stationary distribution $\mathcal{N}(0, \omega^2/2k)$. It has been deeply studied in Fouque *et al.* (2000). In this limit, $K(\theta) \sim \sqrt{k}\exp(-k\theta) \longrightarrow 0$ for all $\theta > 0$, while $K(0) \longrightarrow +\infty$. The same holds in any limiting regime where ω grows as a power of k. However, the case where $\omega \sim \sqrt{k}$ corresponds to the only regime where the variance of σ_t^2 has a finite limit, which is a natural regime in finance.

Note that from Fouque *et al.* (2000), in the ergodic limit, X is constant, and the associated local volatility is flat, so again both X and Y are constant and equal, therefore in this limit too there is no strict inversion of convex ordering. Strict inversion of convex ordering can only be seen in the intermediate regimes, not in the limit. This is one of the reasons why it is difficult to mathematically prove the inversion of convex ordering (at least the restricted version where very small strikes are ignored when we compare call prices on the VIX squared, see Remark 6) in these models.

5.4. *Power-law kernel: rough Bergomi model*

The power-law kernel reads (Bayer *et al.* 2016):

$$K(\theta) = \nu\theta^{H-1/2}, \quad \nu > 0,\ 0 < H < \tfrac{1}{2}.$$

H is called the Hurst exponent. In this case, $\lim_{\theta\to 0^+} K(\theta) = +\infty$, and

$$\xi_t^u = \xi_0^u \exp\left(\nu X_t^u - \frac{\nu^2}{2}\mathrm{Var}(X_t^u)\right), \tag{21}$$

where

$$X_t^u = \int_0^t (u-s)^{H-1/2}\,\mathrm{d}Z_s, \quad \mathrm{Var}(X_t^u) = \frac{u^{2H} - (u-t)^{2H}}{2H}.$$

This model is known as the rough Bergomi model (Bayer *et al.* 2016, Gatheral *et al.* 2018). For the inversion of convex ordering to hold, we choose K to be fast decreasing, i.e. we impose that H be very small. In the limit where H tends to zero, for fixed ν, $\nu^2(t^{2H}/2H)$ tends to $+\infty$ for any $t>0$. In order for $\mathrm{Var}(\sigma_t^2)$ to tend to a finite limit, we must impose that $\nu^2/2H$ tend to a finite limit. As a consequence, a natural limiting regime, analogous to the ergodic regime described in remark 7, consists of letting H and ν tend to zero, while keeping $\nu^2/2H$ constant. In this limit $K(\theta) \sim \sqrt{H}\theta^{H-1/2} \longrightarrow 0$ for all $\theta > 0$, while $\lim_{\theta\to 0^+} K(\theta) = +\infty$ for any $H > 0$.

In the next section, we numerically check that, when the (typically negative) spot–vol correlation is large enough in absolute value, (a) exponential kernels with fast mean reversion of volatility, and (b) power-law kernels with small Hurst exponent, do indeed produce the desired term-structure of convex ordering, i.e. satisfy the extrapolated market term-structure of VIX convex ordering property (14).

6. Numerical results

In this section, we check that the intuition developed in the previous section is correct. We compare distributions with respect to the convex order: the distribution of a quantity X

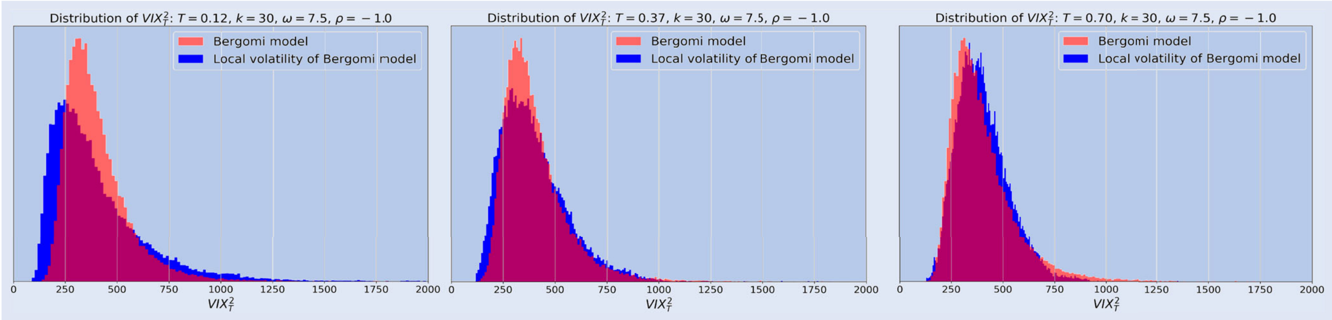

Figure 8. Empirical distribution of the VIX squared in the Bergomi model and its associated local volatility model; $N = 40{,}000$ paths; left: $T = 0.12$; middle: $T = 0.37$; right: $T = 0.70$; $k = 30$, $\omega = 7.5$, $\rho = -1$.

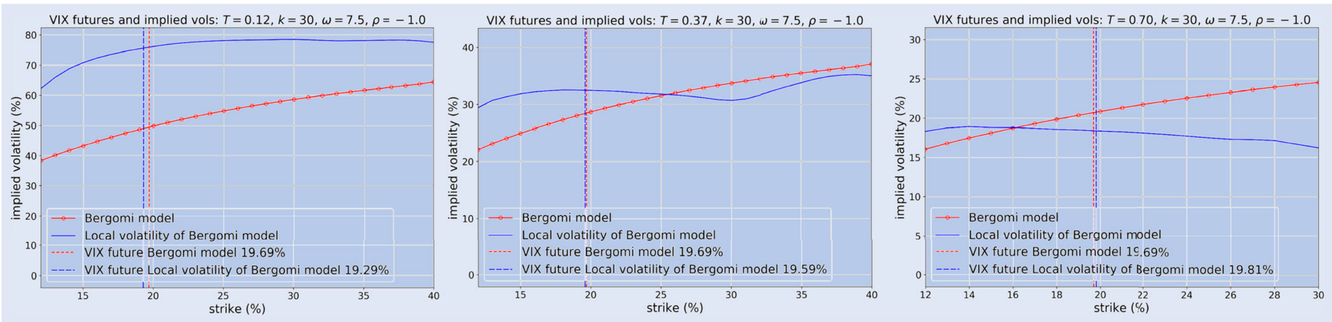

Figure 9. VIX implied volatilities and VIX future in the Bergomi model and its associated local volatility model; left: $T = 0.12$; middle: $T = 0.37$; right: $T = 0.70$; $k = 30$, $\omega = 7.5$, $\rho = -1$.

in the stochastic volatility (SV) model (one-factor Bergomi model or rough Bergomi model), and the distribution of its counterpart Y in the associated local volatility model. We recall that (the distributions of) two nonnegative integrable scalar random variables X and Y are in convex order if and only if $\mathbb{E}[X] = \mathbb{E}[Y]$ and for all strike $K \geq 0$, $\mathbb{E}[(X - K)_+] \leq \mathbb{E}[(Y - K)_+]$. In financial terms, X and Y have the same forward value, and calls on X are cheaper than calls on Y. Therefore we will compare the graphs of $K \mapsto \mathbb{E}[(X - K)_+]$ (SV model, red curve with circle markers) and $K \mapsto \mathbb{E}[(Y - K)_+]$ (associated LV model, solid blue curve); the values for $K = 0$ give the expectations of X and Y, which are equal.

The simulation of the Bergomi model is very easy. Since it is a Markov model, several quantities, such as ξ_T^u or VIX_T^2, are simply functions of X_T, the Ornstein–Uhlenbeck process at time T, and the price of call options on them can be computed using one-dimensional Gaussian quadrature. It is easy, though slow, to simulate exactly the ξ_t^us in the rough Bergomi model, since it suffices to simulate Gaussian vectors whose covariance matrices are known. Instead, we use the recent hybrid scheme by Bennedsen *et al.* (2017) to approximately simulate σ_t^2 quickly. We choose a constant curve of instantaneous forward variances $\xi_0(t) = 0.04 = (20\%)^2$. For the simulation of the associated local volatilty models, we use kernel regressions to estimate the conditional expectations (given S_t and given S_T^{loc}) that are needed to compute $S \mapsto \sigma_{\mathrm{loc}}^2(t, S)$ and $\mathrm{VIX}_{\mathrm{loc},T}^2$, see (2) and (5). We always numerically verify that the associated local volatility model produces the same smile of SPX implied volatilities as the original stochastic volatility model.

6.1. *Exponential kernel: one-factor Bergomi model*

We first consider the one-factor Bergomi model with a very large mean reversion $k = 30$ (characteristic time of mean reversion $1/\tau = \frac{1}{30}$ years, i.e. approximately 12 days) and a very large volatility of variance $\omega = 7.5$. With those values, $(\omega X_t)_{t\geq 0}$ quickly reaches its stationary distribution $\mathcal{N}(0, \omega^2/2k = 0.9375)$ after just a few weeks. Later we will consider other values of (k, ω), see section 6.1.6. We enforce the maximum negative spot–vol correlation, $\rho = -1$, so as to generate the maximum negative SPX ATM skew and, from section 5.2, maximize the inversion of convex ordering at short maturities.

6.1.1. Instantaneous variance. Here we pick $X = \sigma_t^2$ and $Y = \sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}})$. From (7), we know that $Y \leq_c X$, which we check numerically in figure 4 for three dates, $t = 0.2$, $t = 0.45$, and $t = 0.78$. Note that the distributions of σ_t^2 are (almost) the same at these three dates, since all these dates are much larger than $1/k$ and therefore at these dates $(\omega X_t)_{t\geq 0}$ has already reached its stationary distribution. This materializes in the fact that the calls on the instantaneous variances (and on all the other quantities that we investigate below) have the same price at those three dates (and for all dates $t \gg 1/k$) in the Bergomi model. By contrast, the associated local volatility process is not stationary. Over time, the local volatility flattens, since the information content of S_t on σ_t^2 decreases (see Remark 4). As a consequence, the distribution of $Y = \sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}})$ gets less ‘spread’ over time and the calls on Y become cheaper, as observed in figure 4.

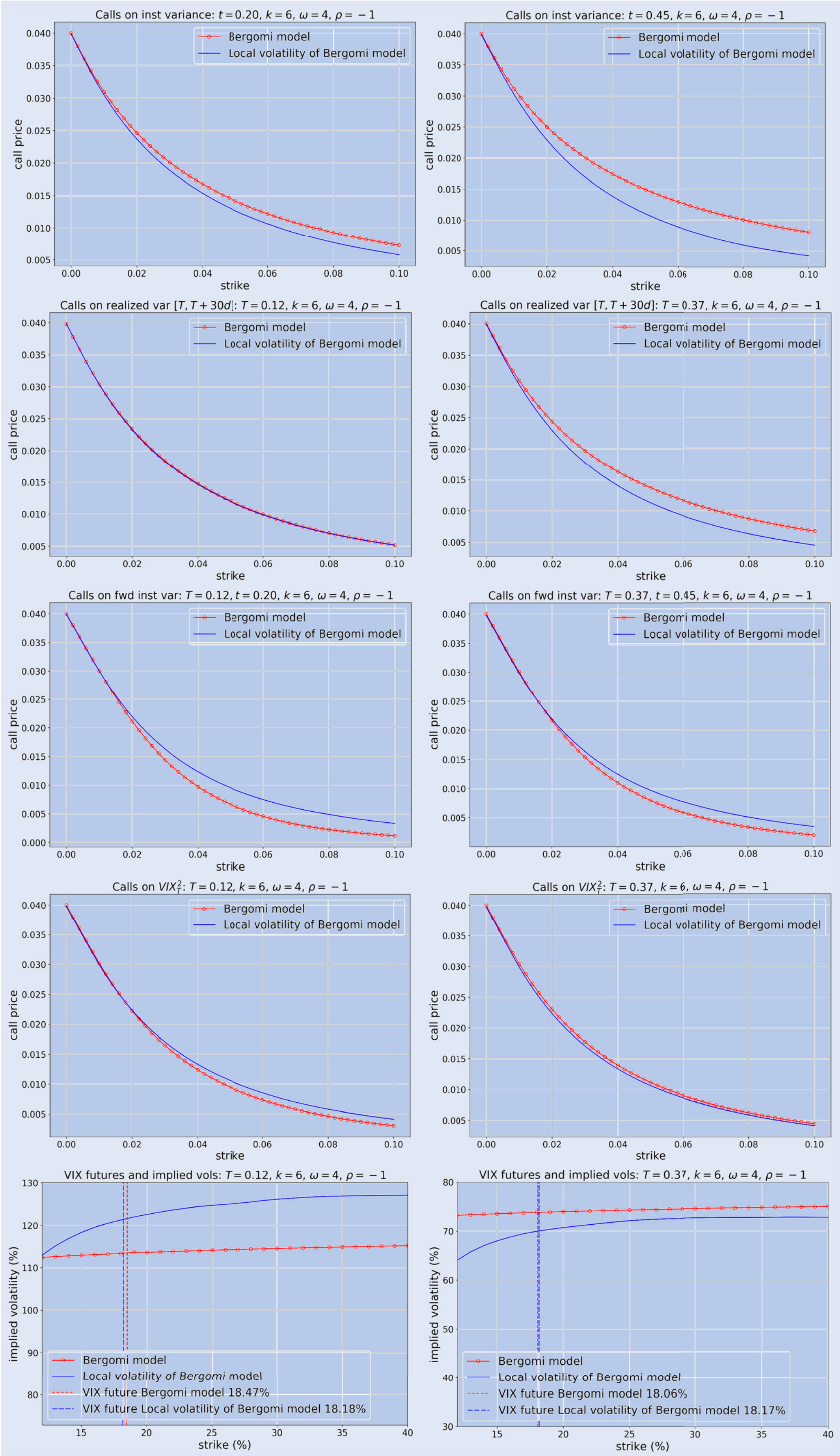

Figure 10. Row 1: call on instantaneous variance at t; row 2: call on realized variance over $[T, T+\tau]$; row 3: call on forward instantaneous variance ξ_T^t; row 4: call on VIX_T^2; row 5: VIX implied volatilities and VIX future for maturity T in the Bergomi model and its associated local volatility model; $t = T + \tau$; left: $T = 0.12$, $t = 0.20$; right: $T = 0.37$, $t = 0.45$; $k = 6$, $\omega = 4$, $\rho = -1$.

REMARK 8 (Term-structure of prices of calls on variance) For the asset price $(S_t)_{t\geq 0}$, we know that calls get more expensive when the maturity increases, by absence of arbitrage. Mathematically, this results from the fact that under any risk-neutral measure, $(S_t)_{t\geq 0}$ is a martingale. By contrast, the volatility process has a nonzero drift. Due to mean reversion, it behaves very differently from the asset price process. In particular, in our case, calls on instantaneous variance become constant (Bergomi model) or get *less* expensive (associated local volatility) over time. Calls on other types of variance also become constant (Bergomi model) or get *less* expensive (associated local volatility) over time, see figures 4–7.

6.1.2. Realized variance over $[T, T+\tau]$**.** Here we average the variances over 30 days, i.e. we pick $X = (1/\tau)\int_T^{T+\tau} \sigma_t^2\,\mathrm{d}t$ and $Y = (1/\tau)\int_T^{T+\tau} \sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}})\,\mathrm{d}t$. Note that we do not apply the conditional expectation given $\mathcal{F}_T$, therefore X and Y represent realized variances, not the VIX squared. Figure 5 shows that averaging instantaneous variances over $[T, T+\tau]$ does *not* undo the timewise convex ordering observed in figure 4. This does not contradict our argument of Section 5.2. Indeed, we argued that in this model the inversion of convex ordering property results from projecting instantaneous variances onto $\mathcal{F}_T$ rather than from averaging them over 30 days.

By comparing figures 4 and 5, we note however that averaging instantaneous variances over 30 days decreases the 'gap' between the two distributions in convex order. To be precise, averaging over 30 days significantly decreases the variance call prices in the Bergomi model, because of the strong mean reversion with characteristic time much shorter than 30 days—in the ergodic limit described in section 5.3, the realized variance becomes constant. By contrast, averaging over 30 days has almost no impact on the variance call prices in the associated local volatility model, in which the volatility $\sigma_{\mathrm{loc}}(t, S_t^{\mathrm{loc}})$ is not mean reverting—it is a function of the martingale $(S_t^{\mathrm{loc}})_{t\geq 0}$.

REMARK 9 Note that if we use the approximations $\mathbb{E}[\sigma_t^2|S_t] \approx \mathbb{E}[\sigma_t^2|S_T]$ for $t \in (T, T+\tau]$, as well as $\mathrm{Law}((S_t)_{t\in[T,T+\tau]}) \approx \mathrm{Law}((S_t^{\mathrm{loc}})_{t\in[T,T+\tau]})$, we get

$$\begin{aligned}\mathbb{E}&\left[\frac{1}{\tau}\int_T^{T+\tau}\sigma_t^2\,\mathrm{d}t\,\middle|\,S_T\right]\\ &= \frac{1}{\tau}\int_T^{T+\tau}\mathbb{E}[\sigma_t^2\,|\,S_T]\,\mathrm{d}t \approx \frac{1}{\tau}\int_T^{T+\tau}\mathbb{E}[\sigma_t^2\,|\,S_t]\,\mathrm{d}t\\ &= \frac{1}{\tau}\int_T^{T+\tau}\sigma_{\mathrm{loc}}^2(t,S_t)\,\mathrm{d}t \overset{(\mathrm{Law})}{\approx} \frac{1}{\tau}\int_T^{T+\tau}\sigma_{\mathrm{loc}}^2(t,S_t^{\mathrm{loc}})\,\mathrm{d}t,\end{aligned}$$

which implies that the local realized variance Y is approximately smaller in convex order than the realized variance X. When these approximations are correct, the timewise convex ordering $\sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}}) \leq_c \sigma_t^2$ is preserved under summation over $t \in [T, T+\tau]$, which is what we observe in figures 5 and 10 to 16.

6.1.3. Forward instantaneous variance. Here we check that projecting instantaneous variances σ_t^2 onto $\mathcal{F}_T$, $t \in (T, T+\tau]$, reverses their convex ordering, provided that $t-T$ is not too small. We pick $X = \mathbb{E}[\sigma_t^2|\mathcal{F}_T]$ and $Y = \mathbb{E}[\sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}})\,|\,\mathcal{F}_T] = \mathbb{E}[\sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}})\,|\,S_T^{\mathrm{loc}}]$, with $t = T+\tau$. From section 5.2, we expect $X \leq_c Y$ even if $\sigma_t^2 \geq_c \sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}})$. Figure 6 shows that indeed, at least for short enough maturities T, projecting onto $\mathcal{F}_T$ undoes the timewise convex ordering observed in figure 4: (16) holds with $u = t = T+\tau$. Due to the very large mean reversion of the volatility, X is almost constant ($\mathcal{F}_T$ gives almost no information on σ_t^2) and the call prices on X are almost as small as they can be—they look like a hockey stick. By contrast, by the arguments explained in section 5.2, the associated local volatility $\sigma_{\mathrm{loc}}(t, S_t^{\mathrm{loc}})$ is not constant, since we are still far from the ergodic limit. Moreover, as explained in remark 5, $\sigma_{\mathrm{loc}}(t, S_t^{\mathrm{loc}})$ is not mean reverting, the knowledge of S_T^{loc} gives a lot of information on $\sigma_{\mathrm{loc}}^2(t, S_t^{\mathrm{loc}})$, so Y is far from being constant, which explains why, at least numerically, we observe that $X <_c Y$.

Note however that $X = \mathbb{E}[\sigma_t^2\,|\,\mathcal{F}_T] = \xi_T^t = \xi_0^t f(T, X_T) = \exp(\omega\,\mathrm{e}^{-k(t-T)}X_T - (\omega^2/2)\,\mathrm{e}^{-2k(t-T)}\mathrm{Var}(X_T))$ actually has support $(0, +\infty)$ while the support of Y, like associated local volatilities, is likely bounded away from zero. In such a case, for small enough strikes, puts on X are more expensive than puts on Y: the convex ordering $X <_c Y$ only holds if we ignore very small strikes. However, figure 6 shows that this issue is immaterial. See also remark 6.

6.1.4. The VIX squared. Here we put everything together: averaging over $[T, T+\tau]$ and projecting onto $\mathcal{F}_T$. That is, we pick $X = \mathrm{VIX}_T^2$, the forward realized variance over $[T, T+\tau]$, and $Y = \mathrm{VIX}_{\mathrm{loc},T}^2$. Figure 7 shows that the Bergomi model with large mean reversion, large vol-of-vol, and large spot–vol correlation (in absolute value) is able to reproduce the extrapolated term-structure of convex ordering (14).† The shorter the maturity, the stronger the inversion of convex ordering (16) for forward instantaneous variances (see figure 6), and the 'closer' (in convex order) the local and stochastic realized variances over 30 days (see figure 5). As a combined result, inversion of convex ordering holds for the VIX squared for short maturities, but not for longer maturities. For long enough maturities, $\mathrm{VIX}_{\mathrm{loc},T}^2 \leq_c \mathrm{VIX}_T^2$, just as for instantaneous variances.

To shed more light on the model's behavior, in figure 8, we have graphed the empirical distributions of VIX_T^2 and $\mathrm{VIX}_{\mathrm{loc},T}^2$ for the three maturities considered in figure 7, based on 40,000 Monte Carlo simulations. The distribution of VIX_T^2 is constant in T for $T \gg 1/k$, since the Ornstein–Uhlenbeck process $(X_t)_{t\geq 0}$ has then reached its stationary distribution. By contrast, the distribution of $\mathrm{VIX}_{\mathrm{loc},T}^2$ gets less spread when the maturity grows, as the local volatility flattens. This helps understand how the model reproduces the extrapolated term-structure of convex ordering (14). Since we use $\rho = -1$, the local volatility function $S \mapsto \sigma_{\mathrm{loc}}(t, S)$ does not reach its minimum in the bulk of the distribution of S_t, $t \in [T, T+\tau]$, therefore in figure 8 we do not observe the infinite density that is seen in the market for the local VIX squared (see figures 1 and 2). When we use $\rho = -0.6$, we do observe the infinite density at the minimum possible value of $\mathrm{VIX}_{\mathrm{loc},T}^2$ for short

† as long as very small strikes are ignored, see remark 6.

maturities (see the bottom graphs in figure 11).† However, in this case, we do not see an inversion of convex ordering, even for very short maturities, e.g. $T = 0.04$.

6.1.5. VIX implied volatility. Finally, figure 9 shows the corresponding VIX implied volatilities. We observe that for short maturities the VIX implied volatilities are much larger in the associated local volatility model, while the reverse is true, at least for most strikes, for long maturities.

Note that since $x \mapsto (K - \sqrt{x})_+$ is convex, the convex ordering of $X = \mathrm{VIX}_T^2$ and $Y = \mathrm{VIX}_{\mathrm{loc},T}^2$ implies that the corresponding VIX put prices are ordered. (By contrast, $x \mapsto (\sqrt{x} - K)_+$ is neither convex nor concave, and the VIX call prices are not necessarily ordered.) However, the ordering of the VIX put prices does not imply any ordering of the VIX implied volatilities, since the Black–Scholes price of the VIX put is a decreasing function of the VIX future price and an increasing function of the implied volatility, and the ordering of VIX future prices is the opposite of the ordering of the VIX put prices, since the square root function is concave while $x \mapsto (K - \sqrt{x})_+$ is convex.

Remark 10 (Positive VIX skew in the one-factor Bergomi model with large mean reversion) Note that with these parameters the Bergomi model, which is generally known to produce an almost flat VIX smile, does actually generate a quite strong positive VIX skew, due to the very large mean reversion $k \gg 1/\tau$. The general argument is that in the Bergomi model $\mathrm{VIX}_T^2 = (1/\tau) \int_T^{T+\tau} \xi_T^u \, \mathrm{d}u$ is almost lognormal, as an average of highly correlated lognormal random variables ξ_T^u. However, when $k \gg 1/\tau$, some of the lognormal variables ξ_T^u are not highly correlated. Consider for example

$$\xi_T^T = \xi_0^T \exp\left(\omega X_T - \frac{\omega^2}{2} \mathrm{Var}(X_T)\right),$$

$$\xi_T^{T+\tau} = \xi_0^{T+\tau} \exp\left(\omega \, \mathrm{e}^{-k\tau} X_T - \frac{\omega^2}{2} \mathrm{e}^{-2k\tau} \mathrm{Var}(X_T)\right).$$

Since $k\tau \gg 1$, $\xi_T^{T+\tau}$ is almost constant at $\xi_0^{T+\tau}$, and those two lognormal random variables are almost independent. If we approximate the VIX squared by $\mathrm{VIX}_T^2 \approx \frac{1}{2}(\xi_T^T + \xi_T^{T+\tau})$, it reads as a mixture of two exponentials of the Gaussian random variable X_T, with different lognormal volatilities. This is exactly the ingredient used by Bergomi (2008) to generate positive VIX skew.

6.1.6. Other sets of parameters. The parameters $k = 30$, $\omega = 7.5$ used in the above example are plausible, in that they correspond to a characteristic time of mean reversion of 12 days for the volatility, and a long-term variance of the instantaneous variance ωX_t around unit. However, they correspond to an instantaneous volatility of a very short-term volatility of $\omega/2 = 375\%$, which is probably too large, even for equity markets. In figure 10, we report what figures 4– 9 look like when we use $k = 6$, $\omega = 4$. These values correspond to a characteristic time of mean reversion of 2 months for the volatility, a long-term variance of the instantaneous variance ωX_t of $\omega^2/2k \approx 1.33$, and a more reasonable instantaneous volatility of a very short-term volatility of $\omega/2 = 200\%$. The results are qualitatively similar to figures 4– 9, thus proving that one does not need to use a very large mean reversion or a very large vol-of-vol in the one-factor Bergomi model to produce the term structure of convex ordering observed in the VIX market. We still observe the inversion of convex ordering when we use smaller vols-of-vol (not reported here), but then the corresponding VIX implied volatilities are too small compared to the typical market values (see figures 1 and 2, bottom graphs). Note that, unlike in figure 9, the model VIX smile is almost flat, since we are now using a much smaller mean reversion k ($k\tau \approx 0.5$; see remark 10).

Figures 11 and 12 respectively show that when the spot–vol correlation is too small in absolute value, or when the mean reversion is too small, we do *not* observe the inversion of convex ordering, even for short maturities, in the one-factor Bergomi model. This was expected from section 5.2, as in both cases one ingredient needed for the inversion of convex ordering is missing.

6.2. *Power-law kernel: rough Bergomi model*

We now repeat the same experiment with the rough Bergomi model. Figure 13 shows that when the Hurst exponent H is small enough and when the spot–vol correlation ρ is large enough in absolute value, then as expected from Section 5.2, the rough Bergomi model possesses the inversion of convex ordering property (12), as well as the extrapolated market term-structure of VIX convex ordering property (14).‡ Here we use $H = 0.1$, a typical value for the Hurst exponent in equity markets (Bayer *et al.* 2016, Gatheral *et al.* 2018), and $\rho = -1$.

Figures 14 and 15 show that we do *not* observe the inversion of convex ordering for large values of H ($H = 0.4$ in figure 14) or for small $|\rho|$ ($\rho = -0.6$ in figure 15), thus confirming the intuition developed in Section 5.2.

6.3. *Skewed one-factor Bergomi model and skewed rough Bergomi model*

Finally let us mention that the same qualitative conclusions hold for the skewed one-factor Bergomi model and skewed rough Bergomi model. Those two models were introduced to produce positive VIX skew. Figures 9–15 show that the plain, unskewed versions have an almost flat VIX skew, except if $k \gg 1/\tau$ in the one-factor Bergomi model (see remark 10). The skewed Bergomi model was introduced by Bergomi (2008) (originally for the two-factor version of the model):

$$\sigma_u^2 = \xi_0^u \left((1-\lambda)\mathcal{E}\left(\omega_0 \int_0^u \mathrm{e}^{-k(u-s)} \, \mathrm{d}Z_s\right) + \lambda \mathcal{E}\left(\omega_1 \int_0^u \mathrm{e}^{-k(u-s)} \, \mathrm{d}Z_s\right)\right) \tag{22}$$

† Of course, we cannot directly observe the infinite density on the empirical distribution, but we clearly see the minimal value of the local VIX squared.

‡ At least as long as extreme strikes are ignored, see remark 6.

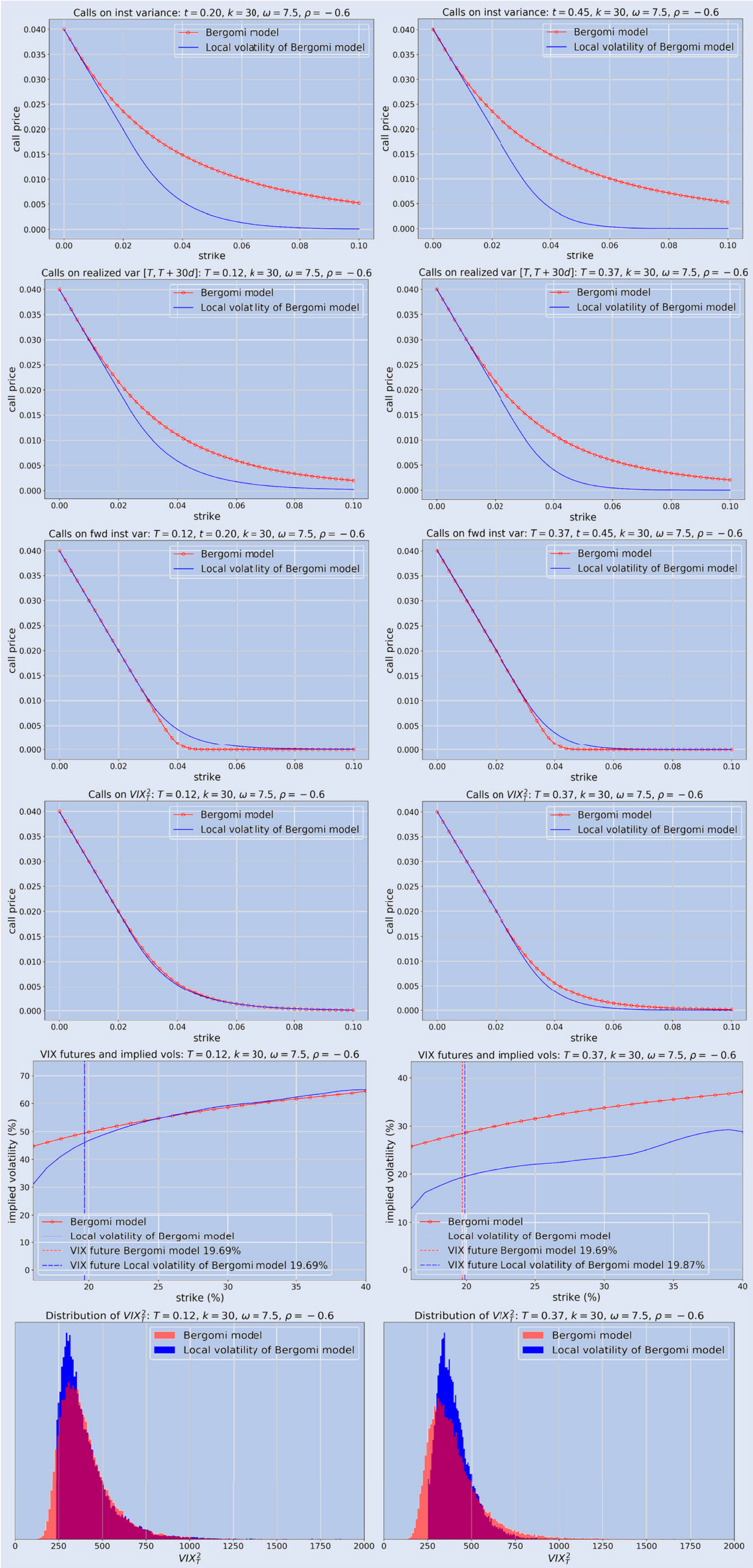

Figure 11. Row 1: call on instantaneous variance at t; row 2: call on realized variance over $[T, T+\tau]$; row 3: call on forward instantaneous variance ξ_T^t; row 4: call on VIX_T^2; row 5: VIX implied volatilities and VIX future; row 6: empirical distribution of the VIX squared for maturity T in the Bergomi model and its associated local volatility model; $t = T + \tau$; left: $T = 0.12$, $t = 0.20$; right: $T = 0.37$, $t = 0.45$; $N = 40{,}000$ paths; $k = 30$, $\omega = 7.5$, $\rho = -0.6$.

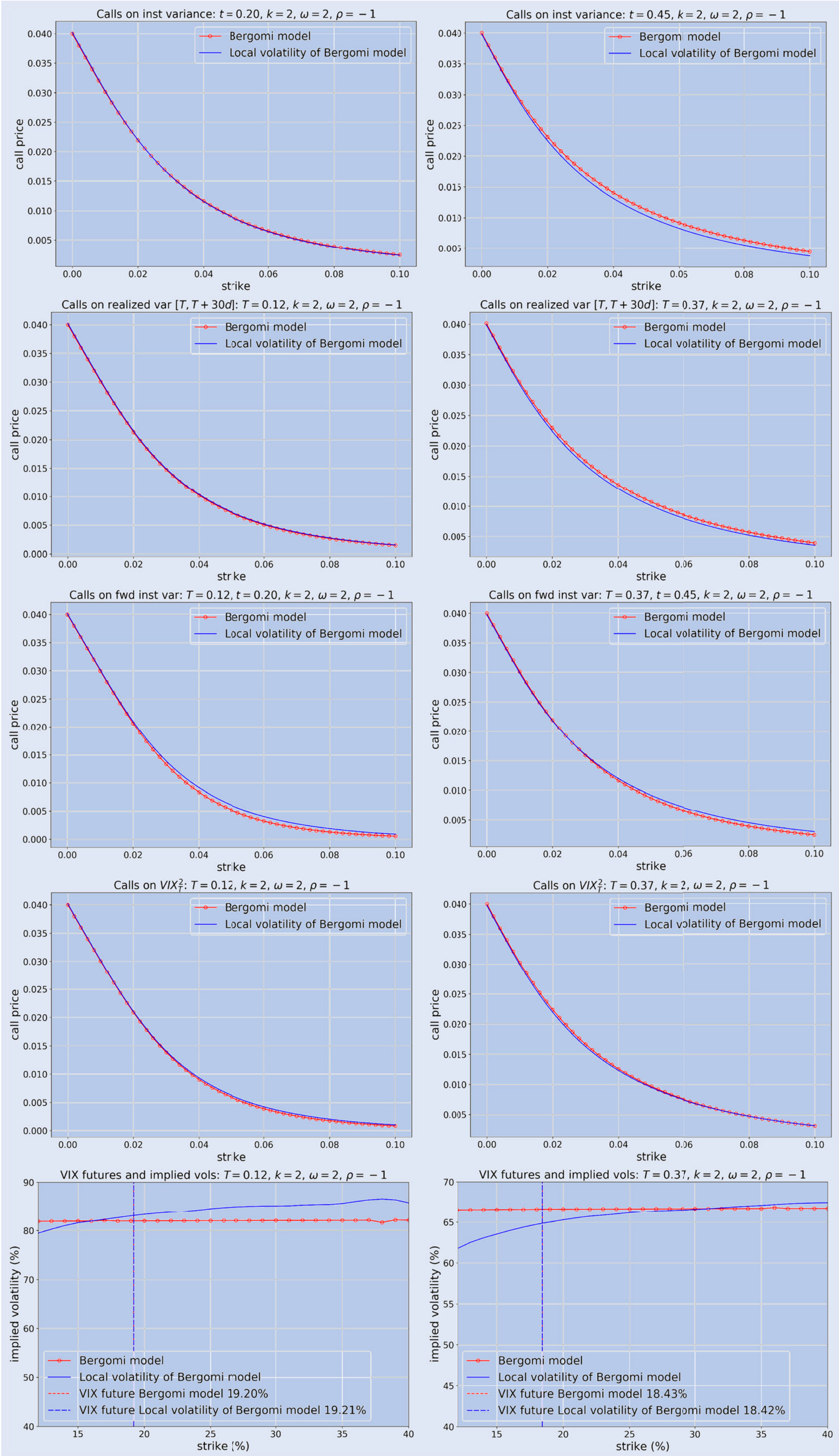

Figure 12. Row 1: call on instantaneous variance at t; row 2: call on realized variance over $[T, T + \tau]$; row 3: call on forward instantaneous variance ξ_T^t; row 4: call on VIX_T^2; row 5: VIX implied volatilities and VIX future for maturity T in the Bergomi model and its associated local volatility model; $t = T + \tau$; left: $T = 0.12$, $t = 0.20$; right: $T = 0.37$, $t = 0.45$; $k = 2$, $\omega = 2$, $\rho = -1$.

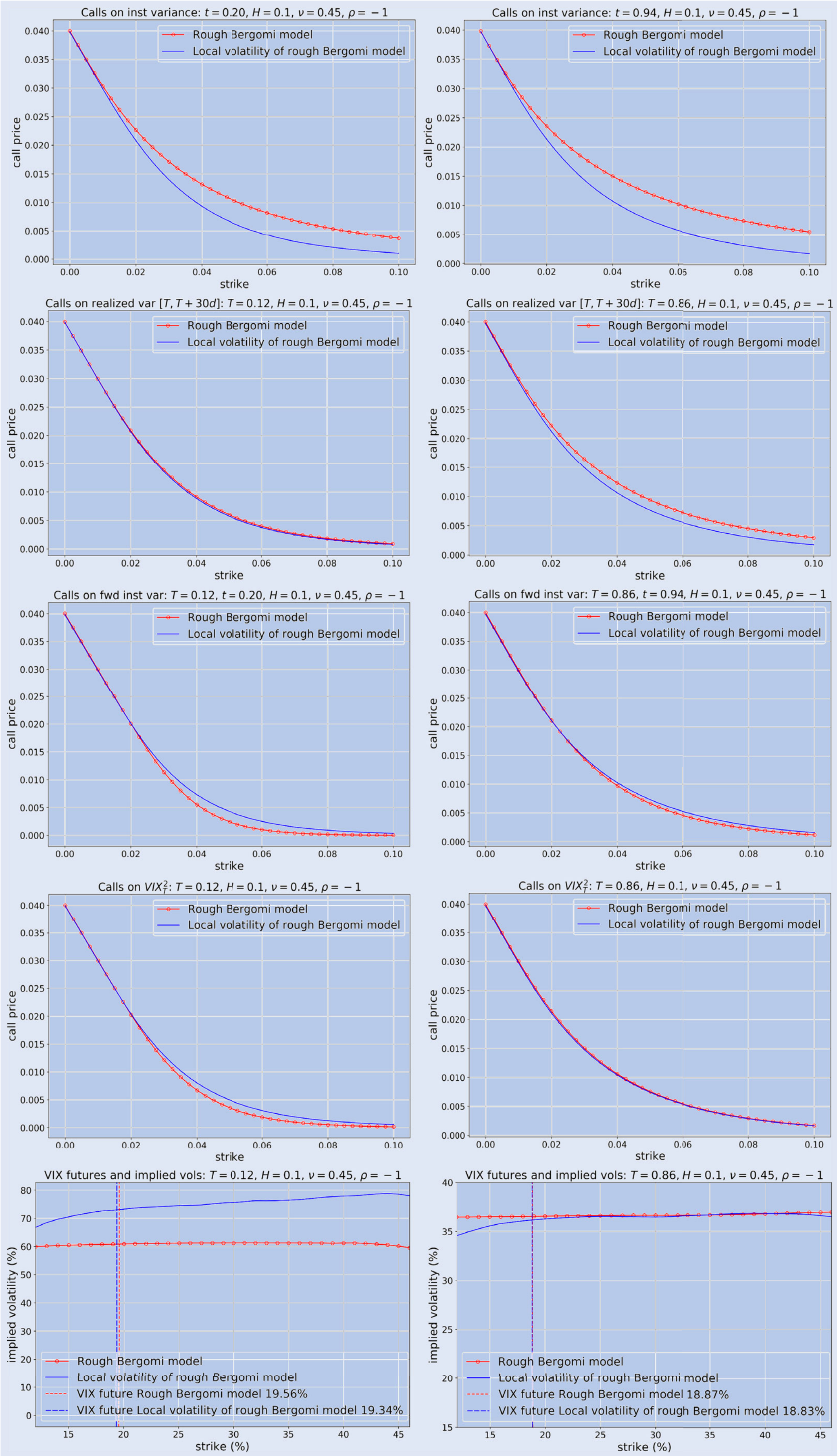

Figure 13. Row 1: call on instantaneous variance at t; row 2: call on realized variance over $[T, T+\tau]$; row 3: call on forward instantaneous variance ξ_T^t; row 4: call on VIX_T^2; row 5: VIX implied volatilities and VIX future for maturity T in the rough Bergomi model (21) and its associated local volatility model; $t = T + \tau$; left: $T = 0.12$, $t = 0.20$; right: $T = 0.86$, $t = 0.94$; $H = 0.1$, $\nu = 0.45$, $\rho = -1$.

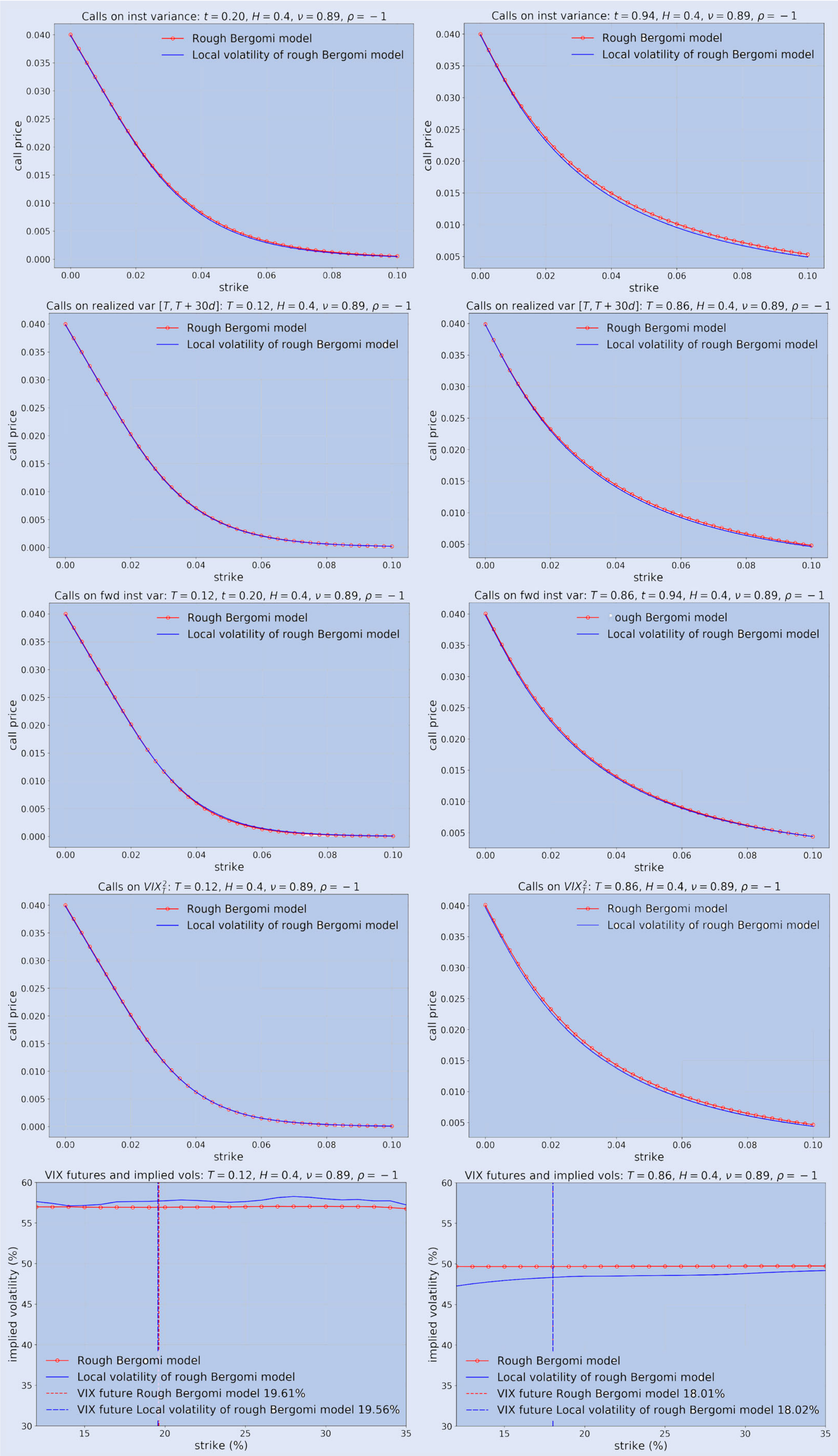

Figure 14. Row 1: call on instantaneous variance at t; row 2: call on realized variance over $[T, T+\tau]$; row 3: call on forward instantaneous variance ξ_T^t; row 4: call on VIX_T^2; row 5: VIX implied volatilities and VIX future for maturity T in the rough Bergomi model (21) and its associated local volatility model; $t = T + \tau$; left: $T = 0.12$, $t = 0.20$; right: $T = 0.86$, $t = 0.94$; $H = 0.4$, $\nu = 0.89$, $\rho = -1$.

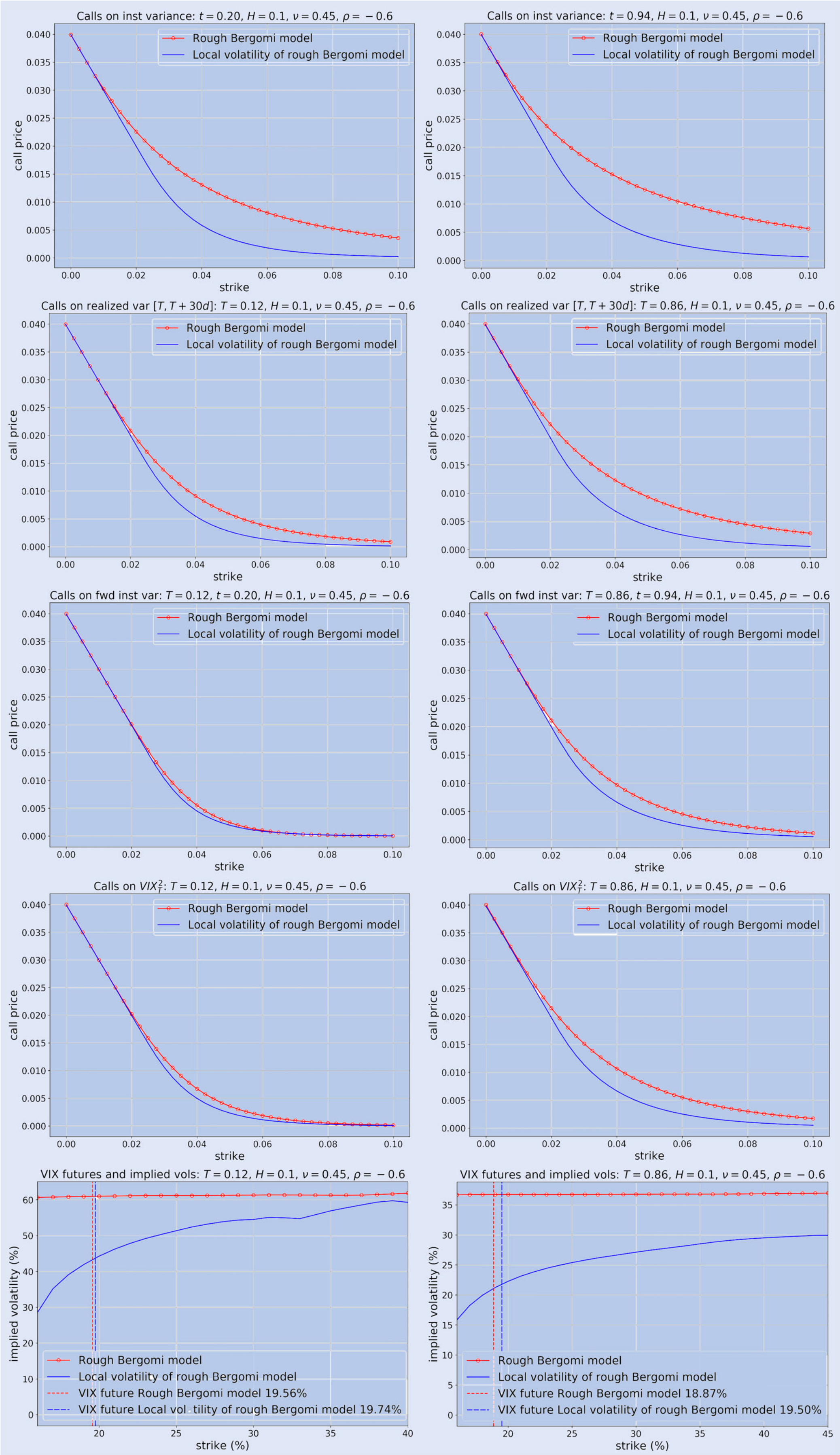

Figure 15. Row 1: call on instantaneous variance at t; row 2: call on realized variance over $[T, T+\tau]$; row 3: call on forward instantaneous variance ξ_T^t; row 4: call on VIX_T^2; row 5: VIX implied volatilities and VIX future for maturity T in the rough Bergomi model (21) and its associated local volatility model; $t = T + \tau$; left: $T = 0.12$, $t = 0.20$; right: $T = 0.86$, $t = 0.94$; $H = 0.1$, $\nu = 0.45$, $\rho = -0.6$.

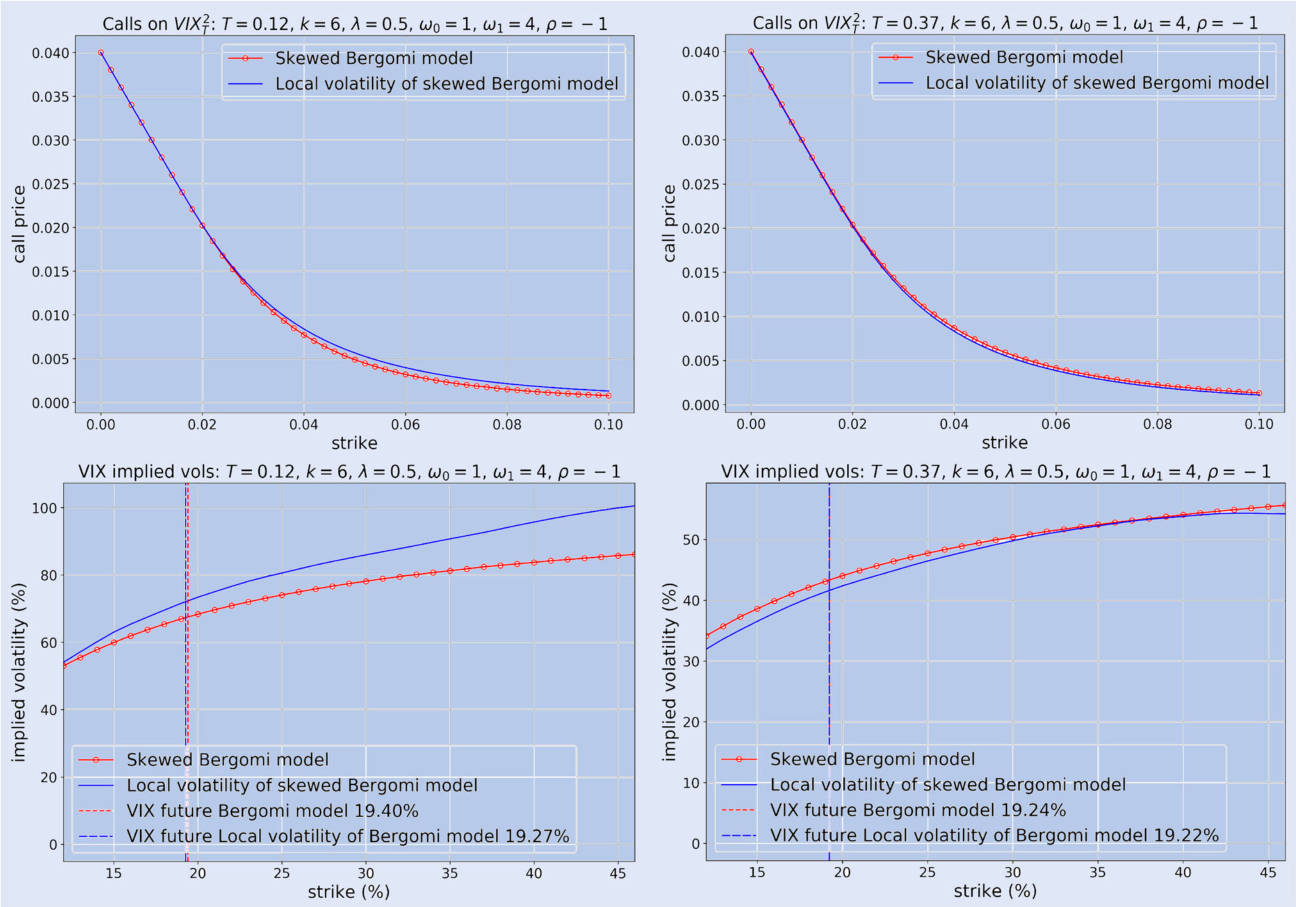

Figure 16. Top: call on VIX_T^2; bottom: VIX implied volatilities and VIX future for maturity T in the skewed Bergomi model (22) and its associated local volatility model; left: $T = 0.12$; right: $T = 0.37$; $k = 6$, $\lambda = 0.5$, $\omega_0 = 1$, $\omega_1 = 4$, $\rho = -1$.

while the skewed rough Bergomi model was independently introduced by Guyon (2018a) and De Marco (2018) as an extension of Bergomi's idea to the rough volatility case:

$$\sigma_u^2 = \xi_0^u \left((1-\lambda)\mathcal{E}\left(\nu_0 \int_0^u (u-s)^{H-1/2}\,\mathrm{d}Z_s \right) + \lambda \mathcal{E}\left(\nu_1 \int_0^u (u-s)^{H-1/2}\,\mathrm{d}Z_s \right) \right). \tag{23}$$

$\mathcal{E}(X)$ is simply a shorthand notation for $\exp(X - \frac{1}{2}\mathrm{Var}(X))$. Two exponentials of the type (19), with two different vol-of-vol parameters, are mixed; the mixing parameter λ lies in $[0, 1]$.

Figures 16 and 17 show that skewed Bergomi models behave similarly as their plain, unskewed versions w.r.t. inversion of convex ordering; and that they do indeed generate positive VIX skew, even when $k < 1/\tau$ in the one-factor Bergomi model (compare figures 10 and 16 on the one hand, and figures 13 and 17 on the other hand).

7. Conclusion

In this article, we have shown that, even though 30 days is a relatively short expiry, the VIX may not behave like the instantaneous volatility in continuous stochastic volatility models on the SPX. In particular, our numerical experiments show that when the spot–vol correlation is large in absolute value, models with large mean reversion or small Hurst exponent have an intriguing property that we call *inversion of convex ordering*: despite the fact that the associated local variance is always smaller in convex order than the instantaneous variance, the VIX squared in the associated local volatility model is *larger* than the VIX squared in the original model for short VIX future maturities (at least if we ignore the very end of the left tail, see remark 6). For longer maturities, the two VIX squared distributions are not rankable in convex order. For even longer maturities, the two VIX squared distributions are in the same convex order as the instantaneous variance distributions.

Most important, we have shown that this term-structure of convex ordering of VIX squared distributions (the market-implied distribution of the VIX squared, and the distribution of the VIX squared in the market local volatility model) is the one that is actually observed in the market. Models with small spot–vol correlation (in absolute value), or with small mean reversion, or with large Hurst exponent, do not produce this term-structure of convex ordering and therefore cannot jointly calibrate SPX options, VIX futures, and VIX options. Conversely, models with large negative value of spot–vol correlation and large mean reversion (or small Hurst exponent) are good candidates for solving this joint calibration problem—in particular the skewed versions of section 6.3,

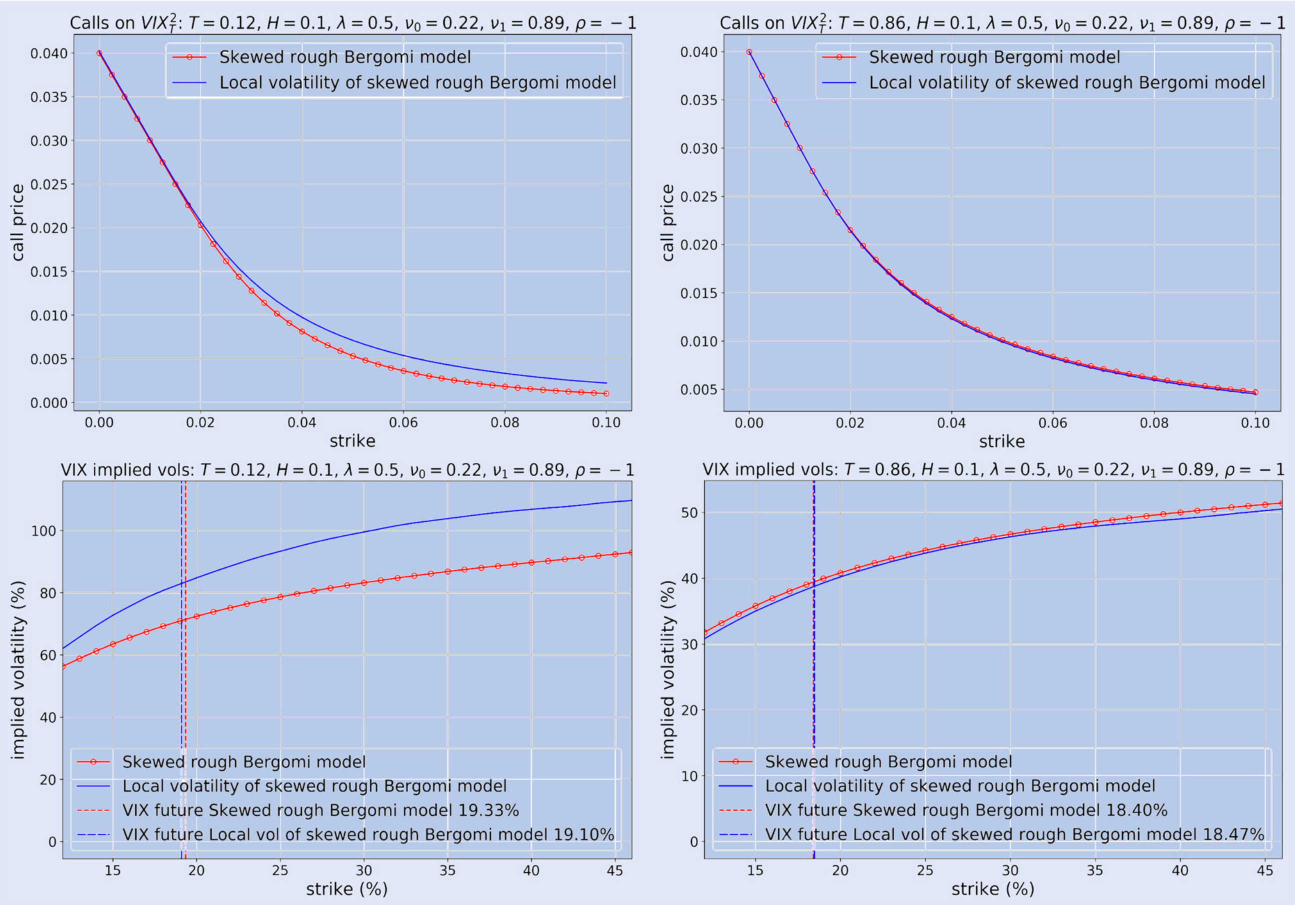

Figure 17. Top: call on VIX_T^2; bottom: VIX implied volatilities and VIX future for maturity T in the skewed rough Bergomi model (23) and its associated local volatility model; left: $T = 0.12$; right: $T = 0.86$; $H = 0.1$, $\lambda = 0.5$, $\nu_0 = 0.22$, $\nu_1 = 0.89$, $\rho = -1$.

which can reproduce positive VIX skew. However, so far we have not been able to jointly fit SPX options, VIX futures, and VIX options using these 'classical' models, even when they are enriched with time-dependent parameters, a second factor driving the volatility curve, and a local volatility component (stochastic local volatility). While this article was under review, Gatheral *et al.* (2020) have shown that a quadratic rough Heston model can jointly fit SPX and VIX smiles simultaneously with good accuracy for very short maturities, even though, given its low number of parameters, it cannot perfectly fit market data.† While we have recently solved the joint calibration problem exactly using a discrete-time model (Guyon 2020), it is still an open question whether one can solve it exactly using continuous models on the SPX. We leave this task for future research.

Acknowledgments

We would like to thank the Managing Editor and the anonymous referees for their useful comments. We would also like to thank Beatrice Acciao and Bryan Liang for helpful discussions, and Bryan Liang for helping with figures 1 and 2.

† While the VIX smile is well calibrated, the model does not seem to produce enough SPX ATM skew in absolute value, a general phenomenon that we have explained in the introduction.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Acciaio, B. and Guyon, J., Inversion of convex ordering: Local volatility does not maximize the price of VIX futures. *SIAM J. Finan. Math.*, 2020, **11**(1), SC1–SC13.
Baldeaux, J and Badran, A., Consistent modelling of VIX and equity derivatives using a 3/2 plus jumps model. *Appl. Math. Finance*, 2014, **21**(4), 299–312.
Bayer, C., Friz, P. and Gatheral, J., Pricing under rough volatility. *Quant. Finance*, 2016, **16**(6), 887–904.
Beiglböck, M. and Juillet, N., On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 2016, **44**(1), 42–106.
Bennedsen, M., Lunde, A. and Pakkanen, M., Hybrid scheme for Brownian semistationary processes. *Finance Stoch.*, 2017, **21**(4), 931–965.

Bergomi, L., Smile dynamics II. *Risk*, October 2005.
Bergomi, L., Smile Dynamics III. *Risk*, March 2008.
Bergomi, L., *Stochastic Volatility Modeling*, Financial Mathematics Series, 2016 (Chapman & Hall/CRC: Boca Raton).
Carr, P. and Madan, D., A note on sufficient conditions for no arbitrage. *Finance Res. Lett.*, 2005, **2**, 125–130.
Carr, P. and Nadtochiy, S., Local variance gamma and explicit calibration to option prices. *Math. Finance*, 2017, **27**(1), 151–193.
Cheung, K.C., Dhaene, J., Kukush, A. and Linders, D., Ordered random vectors and equality in distribution. *Scand. Actuar. J.*, 2015, **3**, 221–244.
Cont, R. and Kokholm, T., A consistent pricing model for index options and volatility derivatives. *Math. Finance*, 2013, **23**(2), 248–274.
Cousot, L., Conditions on option prices for absence of arbitrage and exact calibration. *J. Bank. Finance*, 2007, **31**, 3377–3397.
Davis, M. and Hobson, D, The range of traded option prices. *Math. Finance*, 2007, **17**(1), 1–14.
De March, A., Entropic approximation for multi-dimensional martingale optimal transport, preprint, 2018. Available at: arxiv.org/abs/1812.11104.
De March, A. and Henry-Labordère, P., Building arbitrage-free implied volatility: Sinkhorn's algorithm and variants, preprint, 2019. Available at: ssrn.com/abstract = 3326486.
De Marco, S., Volatility derivatives in (rough) forward variance models. Paper presented at the Bachelier Congress, Dublin, July, 2018.
Dupire, B., Arbitrage pricing with stochastic volatility. Preprint, 1993.
Dupire, B., Pricing with a smile. *Risk*, January, 1994.
Fouque, J.-P. and Saporito, Y., Heston stochastic vol-of-vol model for joint calibration of VIX and S&P 500 options. *Quant. Finance*, 2018, **18**(6), 1003–1016.
Fouque, J.-P., Papanicolaou, G. and Sircar, R., *Derivatives in Financial Markets with Stochastic Volatility*, 2000 (Cambridge University Press: Cambridge, UK).
Gatheral, J., Consistent modeling of SPX and VIX options. Paper presented at Bachelier Congress, 2008.
Gatheral, J., Joint modeling of SPX and VIX. Paper presented at Peking University, 2013.
Gatheral, J., Jaisson, T. and Rosenbaum, M., Volatility is rough. *Quant. Finance*, 2018, **18**(6).
Gatheral, J., Jusselin, P. and Rosenbaum, M., The quadratic rough Heston model and the joint calibration problem. Preprint, 2020. Available at arxiv.org/abs/2001.01789.
Goutte, S, Amine, I. and Pham, H., Regime-switching stochastic volatility model: Estimation and calibration to VIX options. *Appl. Math. Finance*, 2017, **24**(1), 38–75.
Guo, G. and Obłój, J., Computational methods for martingale optimal transport problems. *Ann. Appl. Probab.*, 2019, **29**(6), 3311–3347.
Guyon, J., On the joint calibration of SPX and VIX options. Paper presented at Jim Gatheral's 60th Birthday Conference, New York, October 14, 2017.
Guyon, J., On the joint calibration of SPX and VIX options. Paper presented at Imperial College London, March 28, 2018a.
Guyon, J., On the joint calibration of SPX and VIX options. Paper presented at QuantMinds 2018, Lisbon, May 16, 2018b.
Guyon, J., The joint S&P 500/VIX smile calibration puzzle solved. *Risk*, April 2020.
Gyöngy, I., Mimicking the one-dimensional marginal distributions of processes having an itô differential. *Probab. Theory Relat. Fields*, 1986, **71**, 501–516.
Henry-Labordère, P., *Model-free Hedging: A Martingale Optimal Transport Viewpoint*, Financial Mathematics Series, 2017 (Chapman & Hall/CRC: Boca Raton).
Jacquier, A., Martini, C. and Muguruza, A., On the VIX futures in the rough Bergomi model. *Quant. Finance*, 2018, **18**(1), 45–61.
Kokholm, T. and Stisen, M., Joint pricing of VIX and SPX options with stochastic volatility and jump models. *J. Risk Finance*, 2015, **16**(1), 27–48.
Michon, A., Skorokhod embedding problem and subreplication of variance options. Research internship at Bloomberg LP, 2016.
Neuberger, A., The log contract. *J. Portf. Manag.*, 1994, **20**(2), 74–80.
Pacati, C., Pompa, P. and Renò, R., Smiling twice: The Heston + + model. *J. Bank. Finance*, 2018, **96**(C), 185–206.
Papanicolaou, A. and Sircar, R., A regime-switching Heston model for VIX and S&P 500 implied volatilities. *Quant. Finance*, 2014, **14**(10), 1811–1827.
Sepp, A., Achieving consistent modeling of VIX and equity derivatives. In *Imperial College Mathematical Finance Seminar*, November 2, 2011.
Shaked, M. and Shantikumar, J.G., *Stochastic Orders*, 2007 (Springer: Berlin).
Song, Z. and Xiu, D., A tale of two option markets: State-price densities implied from S&P 500, VIX option prices. Preprint, 2012. Available at: ssrn.com/abstract = 2976966.
Strassen, V., The existence of probability measures with given marginals. *Ann. Math. Stat.*, 1965, **36**, 423–439.
The CBOE volatility index-VIX. Available online at: www.cboe.com/micro/vix/vixwhite.pdf (accessed 18 August 2017).

Appendix 1. Existence of distributions in convex order given price constraints

In this appendix, we further investigate the issue raised in remark 1: the fact that the convex ordering of $\mathrm{VIX}^2_{\mathrm{mkt},T}$ and $\mathrm{VIX}^2_{\mathrm{lv},T}$ typically depends on smile extrapolations.

Let us start with a simpler problem. We consider a financial asset and for simplicity we assume zero interest rates, repo, and dividends. Let $0 < T_1 < T_2$ be two maturities. Assume we are given call prices $C_1(K_i) \geq 0$ and $C_2(K_i) \geq 0$ on this asset, corresponding to a finite set of strikes $0 < K_1 < \cdots < K_n$, as well as forward prices denoted by $C_1(0) \geq 0$ and $C_2(0) \geq 0$. The indices 1 and 2 refer to the maturities T_1 and T_2. We set $K_0 := 0$, $\mathcal{K} := \{K_0, \ldots, K_n\}$ and, from no strike arbitrage constraints (Davis and Hobson 2007), we assume that for $j \in \{1,2\}$, $K_i \in \mathcal{K} \mapsto C_j(K_i)$ is (discretely) convex and nonincreasing, and that $(C_j(K_1) - C_j(K_0))/(K_1 - K_0) \geq -1$. Moreover we assume that $C_j(K_i) = C_j(K_{i+1}) \Longrightarrow C_j(K_i) = 0$. We denote by Π the set of pairs of probability measures on $\mathbb{R}_+$, (μ_1, μ_2), satisfying the constraints

$$\forall j \in \{1,2\},\ \forall i \in \{0,\ldots,n\}, \qquad \int (x - K_i)_+ \mu_j(\mathrm{d}x) = C_j(K_i). \quad \text{(A1)}$$

Π represents the set of pairs of measures that are consistent with market data. We address the following questions:

(1) Does there exist $(\mu_1, \mu_2) \in \Pi$ such that $\mu_1 \leq_c \mu_2$?
(2) When the answer is positive, does there also exist $(\mu_1', \mu_2') \in \Pi$ such that μ_1' and μ_2' are not rankable in convex order?

In terms of random variables, we consider the set $\mathcal{R}$ of all nonnegative random variables (S_1, S_2) such that $\mathbb{E}[(S_j - K_i)_+] = C_j(K_i)$ and we ask whether there exist $(S_1, S_2) \in \mathcal{R}$ such that $S_1 \leq_c S_2$, and in that case if there also exist $(S_1', S_2') \in \mathcal{R}$ such that S_1' and S_2' are not rankable in convex order. By Strassen's theorem (Strassen 1965), there exist $(\mu_1, \mu_2) \in \Pi$ such that $\mu_1 \leq_c \mu_2$ if and only if there exist $(S_1, S_2) \in \mathcal{R}$ such that $\mathbb{E}[S_2|S_1] = S_1$; in that case (S_1, S_2) is a martingale and the model (S_1, S_2) is free of calendar arbitrage.

We recall that two nonnegative integrable scalar random variables S_1 and S_2 are in convex order if and only if $\mathbb{E}[S_1] = \mathbb{E}[S_2]$ and for all strike $K \geq 0$, $\mathbb{E}[(S_1 - K)_+] \leq \mathbb{E}[(S_2 - K)_+]$. From this characterization of convex order, it is easy to prove the following result.

Proposition A.1 *There exist* $(\mu_1, \mu_2) \in \Pi$ *such that* $\mu_1 \leq_c \mu_2$ *if and only if* $C_1(0) = C_2(0)$ *and for all* $i \in \{1, \ldots, n\}$, $C_1(K_i) \leq C_2(K_i)$. *In this case, if at least one of the following conditions holds*:

(i) $C_1(K_n) > 0$,
(ii) $(C_1(K_1) - C_1(K_0))/(K_1 - K_0) > -1$ *and* $(C_2(K_2) - C_2(K_1))/(K_2 - K_1) > (C_2(K_1) - C_2(K_0))/(K_1 - K_0)$,

then there also exist $(\mu_1', \mu_2') \in \Pi$ *such that* μ_1' *and* μ_2' *are not rankable in convex order.*

Note in particular that if the answer to (1) is positive, then there also exists $(\mu_1'', \mu_2'') \in \Pi$ such that $\mu_2'' \leq_c \mu_1''$ if and only if for all $i \in \{1, \dots, n\}$, $C_1(K_i) = C_2(K_i)$. If condition (i) is satisfied, then it is possible to build random variables S_1 and S_2 such that $\mathbb{E}[(S_1 - K)_+] > \mathbb{E}[(S_2 - K)_+]$ for large enough K; see an example in figure 1 (left). If condition (ii) is satisfied, then it is possible to build random variables S_1 and S_2 such that $\mathbb{E}[(S_1 - K)_+] > \mathbb{E}[(S_2 - K)_+]$ for small enough $K > 0$. Note that in general the sufficient condition (i) or (ii) is not necessary, even in the case where $C_1(K_i) < C_2(K_i)$ for all $i \in \{1, \dots, n\}$; see a counterexample in figure 1 (right).

The first part of proposition A.1 is classical and was proved in Carr and Madan (2005) and Davis and Hobson (2007). Davis and Hobson (2007) consider the general case where the strikes may depend on the maturity. Cousot (2007) even included bid/ask prices. In those works, the probability distributions μ_1, μ_2 are built as discrete probabilities. In order to build more realistic distributions that are close in the entropic sense to some reference measure, De March and Henry-Labordère (2019) use an extension of Sinkhorn's algorithm.

Let us now turn to the VIX squared. Assume that the market local volatility σ_{lv} is given, so that the distribution μ_{lv} of $\mathrm{VIX}^2_{\mathrm{lv},T}$ (defined in (6)) is known. We observe prices $C_{\mathrm{mkt}}(K_i)$ of VIX calls at strikes $K_1 < \cdots < K_n$ and the price $C_{\mathrm{mkt}}(0)$ of the VIX future, for maturity T. Again we set $K_0 := 0$, $\mathcal{K} := \{K_0, \dots, K_n\}$ and, from no arbitrage constraints, we assume that $K_i \in \mathcal{K} \mapsto C_{\mathrm{mkt}}(K_i)$ is (discretely) convex and nonincreasing, and that $(C_{\mathrm{mkt}}(K_1) - C_{\mathrm{mkt}}(K_0))/(K_1 - K_0) \geq -1$; we also assume that $C(K_i) = C(K_{i+1}) \Longrightarrow C(K_i) = 0$. We consider the set Π of probability distributions μ_{mkt} on the VIX squared which are consistent with VIX market data, i.e. such that

$$\forall i \in \{0, \dots, n\}, \qquad \int (\sqrt{x} - K_i)_+ \mu_{\mathrm{mkt}}(\mathrm{d}x) = C_{\mathrm{mkt}}(K_i).$$

In the rest of this appendix, we are interested in the following questions: Does there exist $\mu_{\mathrm{mkt}} \in \Pi$ such that $\mu_{\mathrm{mkt}} \leq_c \mu_{\mathrm{lv}}$? such that $\mu_{\mathrm{mkt}} \geq_c \mu_{\mathrm{lv}}$? such that μ_{mkt} and μ_{lv} are not rankable in convex order?

Let us denote $P_{\mathrm{lv}}(K) := \int (K - \sqrt{x})_+ \mu_{\mathrm{lv}}(\mathrm{d}x)$ the price of the put on the local VIX with strike K, and $P_{\mathrm{mkt}}(K_i) = C_{\mathrm{mkt}}(K_i) - C_{\mathrm{mkt}}(0) + K_i$ the market price of the VIX put with strike K_i. The following proposition gives necessary conditions for the existence of $\mu_{\mathrm{mkt}} \in \Pi$ such that $\mu_{\mathrm{mkt}} \leq_c \mu_{\mathrm{lv}}$ (resp. $\mu_{\mathrm{mkt}} \geq_c \mu_{\mathrm{lv}}$). It also gives sufficient conditions for the existence of $\mu_{\mathrm{mkt}} \in \Pi$ such that μ_{mkt} and μ_{lv} are not rankable in convex order.

PROPOSITION A.2 (i) *If there exists* $\mu_{\mathrm{mkt}} \in \Pi$ *such that* $\mu_{\mathrm{mkt}} \leq_c \mu_{\mathrm{lv}}$ *(resp.* $\mu_{\mathrm{mkt}} \geq_c \mu_{\mathrm{lv}}$*), then for all* $i \in \{1, \dots, n\}$, $P_{\mathrm{mkt}}(K_i) \leq P_{\mathrm{lv}}(K_i)$ *(resp.* $P_{\mathrm{mkt}}(K_i) \geq P_{\mathrm{lv}}(K_i)$*) and* $C_{\mathrm{mkt}}(0) \geq C_{\mathrm{lv}}(0)$ *(resp.* $C_{\mathrm{mkt}}(0) \leq C_{\mathrm{lv}}(0)$*).*

(ii) *Let* $\psi_{\min} \geq 0$ *be the lower bound of the support of* μ_{lv} *and assume that* $0 < \sqrt{\psi_{\min}} < K_1$. *Assume that* $0 < P_{\mathrm{mkt}}(K_1) < P_{\mathrm{lv}}(K_1)$. *Then there exists* $\mu_{\mathrm{mkt}} \in \Pi$ *such that* μ_{mkt} *and* μ_{lv} *are not rankable in convex order.*

Note that the sufficient conditions of (ii) are met in the market data of figures 1 and 2. In the case of figure 1 ($T = 21$ days), even though we can extrapolate the VIX smile on the low strike side so that $\mathrm{VIX}^2_{\mathrm{mkt},T}$ and $\mathrm{VIX}^2_{\mathrm{lv},T}$ are not rankable in convex order, natural, reasonable extrapolations of the VIX smile give a zero VIX implied volatility for strikes smaller than $\sqrt{\psi_{\min}} \approx 8.3\%$ (at the time of writing, the all-time lowest VIX close was 9.14% on November 3, 2017) and result in $\mathrm{VIX}^2_{\mathrm{mkt},T} \leq_c \mathrm{VIX}^2_{\mathrm{lv},T}$.

Proof (i) Directly follows from the fact that for all $K \geq 0$, $f : x \mapsto (K - \sqrt{x})_+$ is convex, and $x \mapsto \sqrt{x}$ is concave. (ii) Let $K \in \mathbb{R}_+ \mapsto C(K) \in \mathbb{R}_+$ be a convex nonincreasing function, with right derivatives larger than or equal to -1, such that for all $i \in \{0, \dots, n\}$, $C(K_i) = C_{\mathrm{mkt}}(K_i)$. Since $P_{\mathrm{mkt}}(K_1) > 0$, $C_{\mathrm{mkt}}(K_1) > C_{\mathrm{mkt}}(0) - K_1$ and we can pick C such that for all $K > 0$, $C_{\mathrm{mkt}}(K) > C_{\mathrm{mkt}}(0) - K$. The continuum of call prices C defines a risk-neutral distribution $\mathbb{P}$ for $\mathrm{VIX}_{\mathrm{mkt},T}$ that is consistent with market data, and such that $\mathbb{P}(\mathrm{VIX}_{\mathrm{mkt},T} < \sqrt{\psi_{\min}}) > 0$. Let μ_{mkt} be the distribution of $\mathrm{VIX}^2_{\mathrm{mkt},T}$ under $\mathbb{P}$. It is easy to check that $\mu_{\mathrm{mkt}} \in \Pi$ and that $0 = \int (\sqrt{\psi_{\min}} - \sqrt{x})_+ \mu_{\mathrm{lv}}(\mathrm{d}x) < \int (\sqrt{\psi_{\min}} - \sqrt{x})_+ \mu_{\mathrm{mkt}}(\mathrm{d}x)$. Moreover, by assumption, $\int (K_1 - \sqrt{x})_+ \mu_{\mathrm{lv}}(\mathrm{d}x) > \int (K_1 - \sqrt{x})_+ \mu_{\mathrm{mkt}}(\mathrm{d}x)$. By the convexity of $x \mapsto (K - \sqrt{x})_+$, this proves that μ_{mkt} and μ_{lv} are not rankable in convex order. ∎

Appendix 2. Convex order is not preserved under sum or conditioning

EXAMPLE A.1 (Convex order is not preserved under sum) Let W be a standard Brownian motion. Let us denote $X_0 = W_{t_1}$, $X_1 = -W_{t_2}$, $Y_0 = W_{t_3}$, and $Y_1 = -W_{t_3}$, with $0 < t_1 < t_2 < t_3$. We have $\mathbb{E}[Y_0 \,|\, X_0] = X_0$, $\mathbb{E}[Y_1 \,|\, X_1] = X_1$, hence $X_0 \leq_c Y_0$ and $X_1 \leq_c Y_1$, yet $0 = Y_0 + Y_1 <_c X_0 + X_1$. In this example, Y_0 and Y_1 (resp. X_0 and X_1) are negatively correlated, and convex order is not preserved under the sum.

EXAMPLE A.2 (Convex order is not preserved under sum) We generalize the previous example by considering a four-dimensional Gaussian vector $G = (X_0, Y_0, X_1, Y_1)$. We assume that $\mathbb{E}[Y_0 \,|\, X_0] = X_0$ and $\mathbb{E}[Y_1 \,|\, X_1] = X_1$, and look for necessary and sufficient conditions under which $X_0 + X_1 \leq_c Y_0 + Y_1$. We ignore trivial cases by assuming that all components of G have positive variance. Let us denote by m_X the expectation of a random variable X, by σ_X its standard deviation, and by ρ_{XY} the correlation between two random variables X and Y, which is well defined if and only if σ_X and σ_Y are positive. Since G is Gaussian, for $i \in \{0, 1\}$, $\mathbb{E}[Y_i \,|\, X_i] = m_{Y_i} + \rho_{X_i Y_i}(\sigma_{Y_i}/\sigma_{X_i})(X_i - m_{X_i})$ so $\mathbb{E}[Y_i \,|\, X_i] = X_i$ yields

$$m_{X_i} = m_{Y_i} \quad \text{and} \quad \sigma_{X_i} = \rho_{X_i Y_i} \sigma_{Y_i}. \tag{A2}$$

In particular, this implies that $\rho_{X_i Y_i} > 0$. As a consequence, $m_{X_0+X_1} = m_{Y_0+Y_1}$, and since $X_0 + X_1$ and $Y_0 + Y_1$ are Gaussian,

$$X_0 + X_1 \leq_c Y_0 + Y_1 \Longleftrightarrow \mathrm{Var}(X_0 + X_1) \leq \mathrm{Var}(Y_0 + Y_1).$$

Now, using the second equation in (A2), we have

$$\begin{aligned} \mathrm{Var}(X_0 + X_1) &= \sigma^2_{X_0} + \sigma^2_{X_1} + 2\rho_{X_0 X_1}\sigma_{X_0}\sigma_{X_1} \\ &= \rho^2_{X_0 Y_0}\sigma^2_{Y_0} + \rho^2_{X_1 Y_1}\sigma^2_{Y_i} + 2\rho_{X_0 X_1}\rho_{X_0 Y_0}\sigma_{Y_0}\rho_{X_1 Y_1}\sigma_{Y_1} \end{aligned}$$

so $X_0 + X_1 \leq_c Y_0 + Y_1$ if and only if

$$\begin{aligned} &\sigma^2_{Y_0}(1 - \rho^2_{X_0 Y_0}) + \sigma^2_{Y_1}(1 - \rho^2_{X_1 Y_1}) \\ &\quad + 2\sigma_{Y_0}\sigma_{Y_1}(\rho_{Y_0 Y_1} - \rho_{X_0 X_1}\rho_{X_0 Y_0}\rho_{X_1 Y_1}) \geq 0. \end{aligned}$$

In particular, if $\sigma_{Y_0} = \sigma_{Y_1}$, $\rho_{X_i Y_i} \neq 1$ for $i \in \{0, 1\}$, and

$$\chi := \frac{\rho_{Y_0 Y_1} - \rho_{X_0 X_1}\rho_{X_0 Y_0}\rho_{X_1 Y_1}}{\sqrt{1 - \rho^2_{X_0 Y_0}}\sqrt{1 - \rho^2_{X_1 Y_1}}} < -1$$

then $X_0 + X_1 \not\leq_c Y_0 + Y_1$. Example A.1 corresponds to $\chi = -\sqrt{(t_3 - t_1)/(t_3 - t_2)} < -1$.

EXAMPLE A.3 (Convex order is not preserved under conditioning) Assume that $X \leq_c Y$ with X $\mathcal{F}$-measurable and not constant, and Y independent of $\mathcal{F}$. Then $\mathbb{E}[Y] = \mathbb{E}[Y \,|\, \mathcal{F}] <_c \mathbb{E}[X \,|\, \mathcal{F}] = X$. Since $\mathcal{F}$ contains no information on Y, $\mathbb{E}[Y \,|\, \mathcal{F}]$ is as small as it can be in convex order, i.e. a constant. This is the idea used by Acciaio and Guyon (2020) in their counterexample. In Section 5.2, we explain that using stochastic volatility models with fast mean reversion or small Hurst exponent lead to a similar behavior, in the sense that $\mathcal{F}_T$ contains little information on σ^2_u, $u > T$.

Appendix 3. Strict convex ordering of probability measures on the real line

Let $d \in \mathbb{N}^*$. Let μ be a nonnegative measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $f : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. We recall that $\int f \, d\mu$ always exists, i.e. is always well defined, when f is nonnegative, in which case $\int f \, d\mu$ may take the value $+\infty$; and that in general, $\int f \, d\mu$ exists if and only if $\inf(\int f^+ \, d\mu, \int f^- \, d\mu) < +\infty$, in which case $\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu$. We also recall the convention that for any $a \in \mathbb{R}$, $-\infty < a < +\infty$, and that $-\infty \le -\infty$, $-\infty < +\infty$, and $+\infty \le +\infty$. A convex function is continuous, hence measurable.

Definition A.1 (Convex order and strict convex order) *Let μ and ν be two probability measures on $\mathbb{R}^d$.*

(i) *We say that μ is smaller than ν in convex order, or that μ and ν are in convex order, denoted by $\mu \le_c \nu$, if and only if for any convex function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\int f \, d\mu$ and $\int f \, d\nu$ exist, $\int f \, d\mu \le \int f \, d\nu$.*
(ii) *We say that μ is strictly smaller than ν in convex order, or that μ and ν are in strict convex order, denoted by $\mu <_c \nu$, if and only if $\mu \le_c \nu$ and $\mu \ne \nu$.*

Remark A.1 Since a convex function f is bounded from below by an affine function, if $\int |x| \mu(dx) < +\infty$, then $\int f^- \, d\mu < +\infty$, so $\int f \, d\mu$ is well defined and takes values in $\mathbb{R} \cup \{+\infty\}$.

As is customary with stochastic orders, we also say that two random variables X and Y are in *strict convex order*, or that X is strictly smaller than Y in convex order, if and only if Law(X) and Law(Y) are in strict convex order. We recall that a function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be strictly convex if and only if for all $x, y \in \mathbb{R}^d$ such that $x \ne y$, and $\lambda \in (0, 1)$,

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda) f(y).$$

For μ a nonnegative measure on $\mathbb{R}^d$, we also recall that a function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be μ-integrable if and only if $\int |f| \, d\mu < +\infty$. If f is μ-integrable, then $\int f \, d\mu$ exists, since both $\int f^+ \, d\mu$ and $\int f^- \, d\mu$ are finite, their sum being equal to $\int |f| \, d\mu < +\infty$.

Proposition A.3 (Characterization of convex order) *Let μ and ν be two probability measures on $\mathbb{R}^d$. The following assertions are equivalent*:

(i) *The distributions μ and ν are in convex order.*
(ii) *For any convex function $f : \mathbb{R}^d \to \mathbb{R}$ which is nonnegative or $(\mu + \nu)$-integrable, $\int f \, d\mu \le \int f \, d\nu$.*
(iii) *For any $(\mu + \nu)$-integrable convex function $f : \mathbb{R}^d \to \mathbb{R}$, $\int f \, d\mu \le \int f \, d\nu$. Moreover, for any nonnegative convex function $f : \mathbb{R}^d \to \mathbb{R}_+$, if $\int f \, d\mu = +\infty$ then $\int f \, d\nu = +\infty$.*

Proof (i) $\Rightarrow$ (ii): Follows directly from Definition A.1. (ii) $\Rightarrow$ (i): Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function such that $\int f \, d\mu$ and $\int f \, d\nu$ exist. Let us show that $\int f \, d\mu \le \int f \, d\nu$. This obviously holds if $\int f \, d\mu$ and $\int f \, d\nu$ are finite (from (ii)), or if $\int f \, d\mu = -\infty$, or if $\int f \, d\nu = +\infty$. As a consequence, it is enough to prove (a) $\int f^+ \, d\nu < +\infty \Longrightarrow \int f^+ \, d\mu < +\infty$, and (b) $\int f^- \, d\nu = +\infty \Longrightarrow \int f^- \, d\mu = +\infty$. Claim (a) is clear, as f^+ is a nonnegative convex function, and therefore (ii) implies that $\int f^+ \, d\mu \le \int f^+ \, d\nu < +\infty$. To complete the proof, let us derive Claim (b). Assume that $\int f^- \, d\nu = +\infty$. Then $\int f^+ \, d\nu < +\infty$, $\int f \, d\nu = -\infty$, and by monotone convergence, $\lim_{N \to +\infty} \int f_N \, d\nu = \int f \, d\nu = -\infty$, where $f_N := \max(f, -N)$, $N \in \mathbb{N}$. Since $f_N^+ = f^+$ is convex and nonnegative, (ii) implies that $\int f_N^+ \, d\mu \le \int f_N^+ \, d\nu = \int f^+ \, d\nu < +\infty$. Moreover, f_N^- is bounded by N, so f_N is $(\mu + \nu)$-integrable. Since f_N is convex, as the maximum of two convex functions, we can apply (ii) again and deduce that $\int f_N \, d\mu \le \int f_N \, d\nu$. By monotone convergence, this yields $\int f \, d\mu = \lim_{N \to +\infty} \int f_N \, d\mu \le \lim_{N \to +\infty} \int f_N \, d\nu = \int f \, d\nu = -\infty$, i.e. $\int f^- \, d\mu = +\infty$. Finally, note that (ii) $\iff$ (iii) is obvious. ■

The following theorem was stated without proof in Shaked and Shantikumar (2007, theorem 3.A.43). It was proved in Cheung *et al.* (2015, theorem 7) under the extra assumption that f' is absolutely continuous. It holds for probability measures on the real line ($d = 1$).

Theorem A.1 *Let μ and ν be two probability measures on $\mathbb{R}$ in convex order. Assume that there exists a strictly convex function $f : \mathbb{R} \to \mathbb{R}$ such that $\int f \, d\mu$ and $\int f \, d\nu$ exist and are finite and equal. Then $\mu = \nu$.*

Proof The extra assumption in Cheung *et al.* (2015, Theorem 7) is unnecessary, since by Alexandrov's theorem, a convex function has a second derivative almost everywhere. In dimension 1, this follows from the differentiability almost everywhere of monotone functions of one variable. ■

Remark A.2 Note that the converse of theorem A.1 is not true. Indeed, there might not exist a μ-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$. For instance, let μ denote the probability distribution on $\mathbb{R}$ with density $p_\alpha(x) = c_\alpha / (1 + |x|^\alpha)$, $\alpha \in (1, 2]$, where $c_\alpha > 0$ is the normalization constant. Note that μ does not admit a first moment: $\int x p_\alpha(x) \, dx$ is not defined, as $\int x^+ p_\alpha(x) \, dx = \int x^- p_\alpha(x) \, dx = +\infty$. Let us prove that the only μ-integrable convex functions $f : \mathbb{R} \to \mathbb{R}$ are the constant functions. Let f be a non-constant convex function. There exists $a \ne 0$ and $b \in \mathbb{R}$ such that for all $x \in \mathbb{R}$, $f(x) \ge ax + b$. Since $f^+(x) \ge (ax + b)^+$, $\int f^+ \, d\mu = +\infty$, so f is not μ-integrable. Therefore the only μ-integrable convex functions are the constant functions. In particular, there exists no μ-integrable strictly convex function.

When we assume the existence of a μ-integrable strictly convex function, we get the following corollary.

Corollary A.1 *Let μ and ν be two probability measures on $\mathbb{R}$ in convex order. Assume that there exists a μ-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$. Then the following conditions are equivalent*:

(i) *$\mu = \nu$.*
(ii) *There exists a $(\mu + \nu)$-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$ such that $\int f \, d\mu = \int f \, d\nu$.*
(iii) *For all $(\mu + \nu)$-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$, $\int f \, d\mu = \int f \, d\nu$.*

This corollary can also read as follows.

Corollary A.2 *Let μ and ν be two probability measures on $\mathbb{R}$ in convex order. Assume that there exists a $(\mu + \nu)$-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$. Then the following conditions are equivalent*:

(i) *$\mu <_c \nu$.*
(ii) *For any $(\mu + \nu)$-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$, $\int f \, d\mu < \int f \, d\nu$.*
(iii) *There exists a $(\mu + \nu)$-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$ such that $\int f \, d\mu < \int f \, d\nu$.*

The next proposition gives sufficient conditions for the existence of a μ-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$.

Proposition A.4 (Existence of a $\boldsymbol{\mu}$-integrable strictly convex function) *Let μ be a probability measure on the real line. Assume one of the following assumptions*:

(i) *There exists $a \in \mathbb{R}$ such that $\mu((a, +\infty)) = 1$ or $\mu((-\infty, a)) = 1$.*
(ii) *There exists $\alpha > 0$ such that $\int |x|^{1+\alpha} \mu(dx) < +\infty$.*

Then there exists a μ-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$.

Proof In each case, it is enough to identify an element g of the set E of μ-integrable strictly convex functions $f : \mathbb{R} \to \mathbb{R}$. If $\mu((a, +\infty)) = 1$, one can pick $g(x) = e^{-x}$. If $\mu((-\infty, a)) = 1$, one may choose $g(x) = e^x$. Finally, in the case of Assumption (ii), one can simply pick $g(x) = |x|^{1+\alpha}$. ■

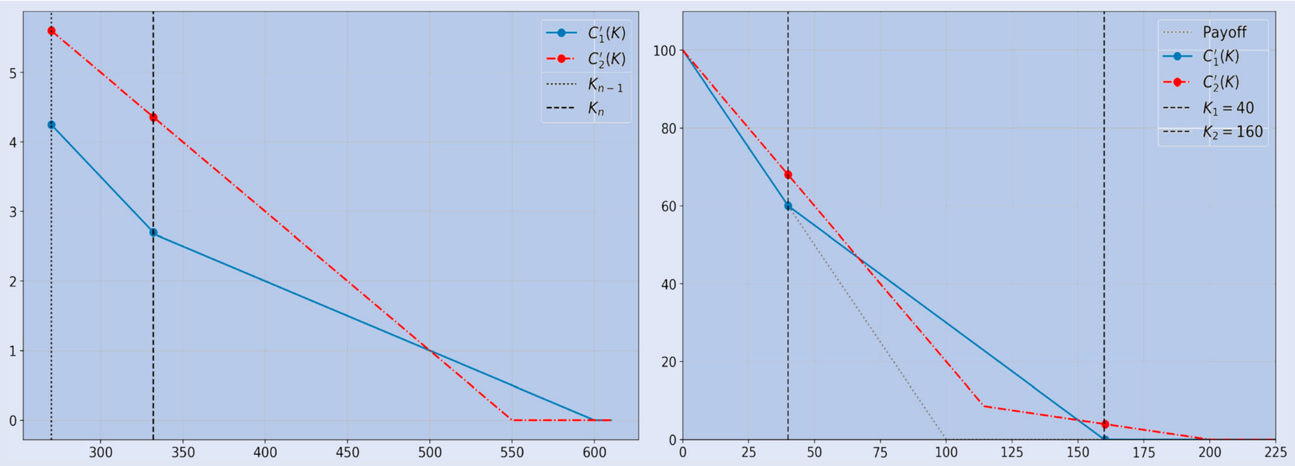

Figure A1. Left: If $C_1(K_n) > 0$, then it is possible to build random variables S_1 and S_2 such that $\mathbb{E}[(S_1 - K)_+] > \mathbb{E}[(S_2 - K)_+]$ for large enough K. Right: An example where there exist $(\mu_1, \mu_2) \in \Pi$ such that $\mu_1 \leq_c \mu_2$, condition {(i) or (ii)} is not satisfied, yet there also exist $(\mu'_1, \mu'_2) \in \Pi$ such that μ'_1 and μ'_2 are not rankable in convex order. Here $n = 2$; $C_1(0) = C_2(0) = 100$; $K_1 = 40$, $K_2 = 160$; $C_1(K_1) < C_2(K_1)$ and $C_1(K_2) < C_2(K_2)$. In both graphs, $C'_1(K)$ and $C'_2(K)$ are the call prices corresponding to μ'_1 and μ'_2.

Corollary A.2 leads to the following characterization of strict convex ordering of probability measures on the real line, which says that $\mu <_c \nu$ is essentially equivalent to the fact that for any $(\mu + \nu)$-integrable strictly convex function f, $\int f \, \mathrm{d}\mu < \int f \, \mathrm{d}\nu$.

THEOREM A.2 (Characterization of strict convex ordering of probability measures on the real line) *Let μ and ν be two probability measures on the real line. Assume that there exists a $(\mu + \nu)$-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$. Then the following assertions are equivalent*:

(i) *The distributions μ and ν are in strict convex order.*
(ii) *For any $(\mu + \nu)$-integrable strictly convex function $f : \mathbb{R} \to \mathbb{R}$, $\int f \, \mathrm{d}\mu < \int f \, \mathrm{d}\nu$. Moreover, for any nonnegative convex function $f : \mathbb{R} \to \mathbb{R}_+$, if $\int f \, \mathrm{d}\mu = +\infty$ then $\int f \, \mathrm{d}\nu = +\infty$.*

Proof (i) $\Rightarrow$ (ii): See corollary A.2. (ii) $\Rightarrow$ (i): Let E denote the set of all $(\mu + \nu)$-integrable strictly convex functions, and let $g \in E$. Let us show that $\mu \leq_c \nu$. By proposition A.3, it is enough to show that for any convex function $f : \mathbb{R} \to \mathbb{R}$ which is nonnegative or $(\mu + \nu)$-integrable, $\int f \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\nu$. Let f be such a function. Assume that f is $(\mu + \nu)$-integrable. Note that for all $\varepsilon > 0$, $f_\varepsilon := f + \varepsilon g \in E$. Assumption (ii) thus yields $\int f_\varepsilon \, \mathrm{d}\mu < \int f_\varepsilon \, \mathrm{d}\nu$. Letting ε go to zero, by dominated convergence, we get $\int f \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\nu$. If f is not $(\mu + \nu)$-integrable, then f is nonnegative and at least one of the two integrals $\int f \, \mathrm{d}\mu$, $\int f \, \mathrm{d}\nu$ is infinite. From assumption (ii), if $\int f \, \mathrm{d}\mu = +\infty$, then $\int f \, \mathrm{d}\nu = +\infty$. As a consequence $\int f \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\nu$ also holds. We have thus proved that $\mu \leq_c \nu$. Moreover, $\int g \, \mathrm{d}\mu < \int g \, \mathrm{d}\nu$, so $\mu \neq \nu$. Finally, $\mu <_c \nu$. ■