Volatility Modeling

Julien Guyon

Ecole des Ponts ParisTech

January 2023

julien.guyon@enpc.fr

Contents

Contents

The different types of volatility

- **Historical volatility** (realized volatility): standard deviation of past log returns.
- **Implied volatility**: a way to quote option prices $\longleftrightarrow$ anticipation of standard deviation of future returns.
- **Instantaneous volatility**: in a diffusive model, weight of dW_t. Never seen in reality; may not exist. Ex: local volatility; stochastic volatility; local stochastic volatility; path-dependent volatility.
- May apply to asset returns, interest rates... or some type of volatility itself! Ex: historical volatility of implied volatility; inst. volatility of inst. volatility.
- Usually annualized. Std dev of random walk $\sim \sqrt{t} \Longrightarrow$ volatility of 16% means 16% per year $=$ 16% $\times\sqrt{\frac{1}{12}}$ per month $=$ 16% $\times\sqrt{\frac{1}{252}} \approx 1\%$ per day.
- $\sigma^2 t$ is dimensionless $\Longrightarrow [\sigma] = \frac{1}{\sqrt{t}}$.

Says nothing about direction!

Historical volatility

$$\sigma_{\text{hist}} = \sqrt{\frac{1}{N\Delta t}\sum_{i=1}^{N}\left(\ln\frac{S_{t_i}}{S_{t_{i-1}}}\right)^2}, \qquad \Delta t = t_i - t_{i-1}$$

- In the Black-Scholes model, $\sigma_{\text{hist}} \approx \sigma_{\text{BS}}$
- Remove empirical average of log-returns or not? Can we estimate the drift?
- Given a continuous sample path of $X_t := \mu t + \sigma W_t$, $t \in [0, T]$, it is easy to very presisely identify σ (compute the quadratic variation, even on a subinterval of length ε), but for usual values of μ, σ, and T in finance, it is difficult to identify the drift μ: $\lim_{t\to\infty} \frac{X_t}{t} = \mu$ a.s. but Std Dev $\left(\frac{X_T}{T}\right) = \frac{\sigma}{\sqrt{T}}$ is large compared to the unknown μ!
- $\Longrightarrow$ Substract risk-neutral drift or nothing.
- Choice of N and Δt? When we compare historical volatility and an implied volatility with maturity T, it is natural to choose $N\Delta t = T$.
- Other estimators based on open, close, high, low: Parkinson, Garman-Klass, Rogers-Satchell, etc.

Black-Scholes model

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t, \qquad S_t = S_0 \exp\left(\sigma W_t + \left(\mu - \frac{\sigma^2}{2}\right) t\right)$$

- $\mu \in \mathbb{R}$, $\sigma > 0$
- Geometric Brownian motion: $\ln(S_t)$ follows a continuous-time random walk with drift $\mu - \frac{\sigma^2}{2}$ and volatility σ:

$$d\ln(S_t) = \frac{dS_t}{S_t} - \frac{1}{2}\frac{d\langle S\rangle_t}{S_t^2} = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma\, dW_t$$

- Replication argument (delta hedging) $\Longrightarrow$ The price of a call option does not depend on μ, only on the volatility parameter σ (and risk-neutral drift $r - q$ and discounting rate r, assumed constant, where q is the dividend yield, inclusive of repo):

$$\begin{aligned} C_{\mathsf{BS}}(T, K; S_0, \sigma, r, q) &= S_0 e^{-qT} N(d_+) - K e^{-rT} N(d_-) \\ d_\pm(T, K; S_0, \sigma, r, q) &= \frac{\ln \frac{S_0}{K e^{-(r-q)T}}}{\sigma\sqrt{T}} \pm \frac{1}{2}\sigma\sqrt{T} \end{aligned}$$

Implied volatility

- In absence of arbitrage in the market:

$$(S_0 - Ke^{-rT})_+ \leq C_{\mathsf{mkt}}(T,K) \leq S_0$$

- $\sigma \mapsto C_{\mathsf{BS}}(T,K;S_0,\sigma,r,q)$ strictly increases from arbitrage lower bound $\left(S_0 - Ke^{-rT}\right)_+$ to arbitrage upper bound S_0.
- By definition of the implied volatility $\sigma_{\mathsf{BS}}(T,K)$ (of a European call option):

$$C_{\mathsf{mkt}}(T,K) = C_{\mathsf{BS}}(T,K;S_0,\sigma_{\mathsf{BS}}(T,K),r,q)$$

- **The implied volatility is the wrong number to input into the wrong formula to get the correct price**.
- Just a way to quote option prices. Much easier to compare implied volatilities than to compare option prices themselves.

Does the implied volatility say something about expected future standard deviation of log-returns?

- If you assume a diffusive dynamics $\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t$, then $\sigma_{\mathrm{BS}}(T, K)$ is the expected standard deviation of log-returns, as implied by the call option price. It reflects the anticipation of the market participants.
- However, $\sigma_{\mathrm{BS}}(T, K)$ depends on (T, K)...
- Moreover, implied volatilities can be computed even within a pure jump model. What is the meaning of the implied volatility in a market where the asset moves only through jumps?

Is the implied volatility of a European put option the same as the implied volatility of a European call option with same strike and maturity?

$$\begin{aligned}
(S_T - K)_+ - (K - S_T)_+ &= S_T - K \\
C_{\mathsf{mkt}}(T,K) - P_{\mathsf{mkt}}(T,K) &= S_0 e^{-qT} - K e^{-rT} \\
C_{\mathsf{BS}}(T,K;\sigma_{\mathsf{BS}}^C(T,K)) - P_{\mathsf{BS}}(T,K;\sigma_{\mathsf{BS}}^P(T,K)) &= S_0 e^{-qT} - K e^{-rT} \\
C_{\mathsf{BS}}(T,K;\sigma) - P_{\mathsf{BS}}(T,K;\sigma) &= S_0 e^{-qT} - K e^{-rT} \qquad \forall \sigma > 0
\end{aligned}$$

$$\begin{aligned}
&C_{\mathsf{BS}}(T,K;\sigma) - P_{\mathsf{BS}}(T,K;\sigma) \\
&\qquad = C_{\mathsf{BS}}(T,K;\sigma_{\mathsf{BS}}^C(T,K)) - P_{\mathsf{BS}}(T,K;\sigma_{\mathsf{BS}}^P(T,K)) \qquad \forall \sigma > 0
\end{aligned}$$

$$\begin{aligned}
&C_{\mathsf{BS}}(T,K;\sigma_{\mathsf{BS}}^P(T,K)) - P_{\mathsf{BS}}(T,K;\sigma_{\mathsf{BS}}^P(T,K)) \\
&\qquad = C_{\mathsf{BS}}(T,K;\sigma_{\mathsf{BS}}^C(T,K)) - P_{\mathsf{BS}}(T,K;\sigma_{\mathsf{BS}}^P(T,K))
\end{aligned}$$

$$\begin{aligned}
C_{\mathsf{BS}}(T,K;\sigma_{\mathsf{BS}}^P(T,K)) &= C_{\mathsf{BS}}(T,K;\sigma_{\mathsf{BS}}^C(T,K)) \\
\sigma_{\mathsf{BS}}^P(T,K) &= \sigma_{\mathsf{BS}}^C(T,K)
\end{aligned}$$

Contents

The different types of volatility derivatives

Payoffs that depend on the future realized variance $\longleftrightarrow$ future historical volatility:

- Variance swaps
- Volatility swaps
- Calls/puts on realized variance

Path-dependent payoffs on the underlying asset.

Payoffs that depend on the future implied volatilities:

- VIX futures
- VIX options
- VXX options

Vanilla payoffs on the prices of options on the underlying asset.

Variance swaps

$$\mathsf{RV}_{t_0,t_N} := \frac{1}{N\Delta t}\sum_{i=1}^{N}\left(\ln\frac{S_{t_i}}{S_{t_{i-1}}}\right)^2, \qquad \Delta t = t_i - t_{i-1} = \frac{1}{252} = 1 \text{ day}$$

$$\text{Payoff at } T = t_N: \quad \mathsf{RV}_{t_0,t_N} - K$$

- Swap: no payment at time $t_0 = 0$. K chosen so that the derivative is costless at time 0.
- $\sigma_{\mathsf{VS}}(T) := \sqrt{K}$ is called the variance swap volatility, or variance swap rate.
- For diffusive processes $\frac{dS_t}{S_t} = \cdots dt + \sigma_t \, dW_t$, in the limit when $\Delta t \to 0$:

$$\mathsf{RV}_{0,T} \longrightarrow \frac{1}{T}\int_0^T \sigma_t^2 \, dt$$

- In this limit, $\sigma_{\mathsf{VS}}(T)$ is the implied volatility of the variance swap.
- Risk-neutral valuation of K: $K = \mathbb{E}[\mathsf{RV}_{t_0,t_N}]$
- Swap contracts are worth zero at date of issuance but can have positive or negative price after issuance.

Volatility swaps

$$\mathsf{RV}_{t_0,t_N} := \frac{1}{N\Delta t}\sum_{i=1}^{N}\left(\ln\frac{S_{t_i}}{S_{t_{i-1}}}\right)^2, \qquad \Delta t = t_i - t_{i-1} = \frac{1}{252} = 1 \text{ day}$$

$$\text{Payoff at } T = t_N : \quad \sqrt{\mathsf{RV}_{t_0,t_N}} - K$$

- Swap: no payment at time $t_0 = 0$. K chosen so that the derivative is costless at time 0.
- $\sigma_{\mathsf{VolSwap}}(T) := K$ is called the volatility swap volatility, or volatility swap rate.
- In the continuous time limit ($\Delta t \to 0$), $\sigma_{\mathsf{VolSwap}}(T)$ is the implied volatility of the volatility swap.
- Risk-neutral valuation of K: $K = \mathbb{E}[\sqrt{\mathsf{RV}_{t_0,t_N}}]$

Calls/puts on realized volatility

$$\mathrm{RV}_{t_0,t_N} := \frac{1}{N\Delta t}\sum_{i=1}^{N}\left(\ln\frac{S_{t_i}}{S_{t_{i-1}}}\right)^2, \qquad \Delta t = t_i - t_{i-1} = \frac{1}{252} = 1 \text{ day}$$

Payoff at $T = t_N$: $\quad (\mathrm{RV}_{t_0,t_N} - K)_+$ or $(K - \mathrm{RV}_{t_0,t_N})_+$

- Premium is paid by buyer at time $t_0 = 0$.
- Buyer cannot lose more than the premium but upside is not limited.
- Risk-neutral valuation at time $t_0 = 0$: $e^{-rT}\mathbb{E}[(\mathrm{RV}_{t_0,t_N} - K)_+]$ or $e^{-rT}\mathbb{E}[(K - \mathrm{RV}_{t_0,t_N})_+]$

Link between variance swap volatility and the implied volatility of the log-contract (Neuberger 1990, Dupire 1994)

Assume the asset price follows a diffusive process $\frac{dS_t}{S_t} = (r-q)\,dt + \sigma_t\,dW_t$.

The T-forward price $F_t^T = \mathbb{E}[S_T|\mathcal{F}_t] = S_t e^{(r-q)(T-t)}$ satisfies $\frac{dF_t^T}{F_t^T} = \sigma_t\,dW_t$.

By Itô's formula,

$$\ln\frac{S_T}{F_0^T} = \ln\frac{F_T^T}{F_0^T} = \int_0^T \frac{dF_t^T}{F_t^T} - \frac{1}{2}\int_0^T \frac{d\langle F^T\rangle_t}{(F_t^T)^2} = \int_0^T \frac{dF_t^T}{F_t^T} - \frac{1}{2}\int_0^T \sigma_t^2\,dt$$

$$\frac{1}{T}\int_0^T \sigma_t^2\,dt = -\frac{2}{T}\ln\frac{S_T}{F_0^T} + \frac{2}{T}\int_0^T \frac{dF_t^T}{F_t^T} = \text{log-contract (vanilla)} + \text{delta-hedge}$$

In the continuous time limit $(\Delta t \to 0)$:

$$\sigma_{\mathsf{VS}}(T)^2 = \mathbb{E}\left[\frac{1}{T}\int_0^T \sigma_t^2\,dt\right] = \mathbb{E}\left[-\frac{2}{T}\ln\frac{S_T}{F_0^T}\right]$$

In the Black-Scholes model, $\mathbb{E}\left[-\frac{2}{T}\ln\frac{S_T}{F_0^T}\right] = \sigma_{\mathsf{BS}}^2$.

For diffusive models on the asset price, $\sigma_{\mathsf{VS}}(T)$ (implied volatility of the variance swap) is also the implied volatility of the log-contract.

VIX index

- VIX $=$ Volatility IndeX.
- Published every 15 seconds by the Chicago Board Options Exchange.
- Indicator of short-term options-implied volatility. Known as "fear factor."
- **Objective: VIX is meant to reflect the 30-day implied volatility of S&P 500 (SPX) options.**
- Problem: implied volatility of S&P 500 call/put options depend on the option strike. VIX should be a strike-free measure of S&P 500 implied vol.
- Natural choice: define VIX as the implied volatility of a 30-day variance swap on S&P 500.
- Problem: Variance swaps are OTC. Not listed on an exchange.
- $\Longrightarrow$ Define VIX as the implied volatility of a 30-day log-contract on S&P 500:

$$(\mathrm{VIX}_t)^2 := -\frac{2}{\tau}\mathrm{Price}_t\left[\ln\left(\frac{S_{t+\tau}}{F_t^{t+\tau}}\right)\right], \qquad \tau = 30 \text{ days}$$

- The log-contract is not listed on an exchange but it can be replicated at t using OTM call and put options on the S&P 500 with maturity $t+\tau$.

Static replication of a vanilla payoff by OTM puts and calls

Taylor-Lagrange:

$$
\begin{aligned}
f(y) &= f(x) + f'(x)(y-x) + \int_x^y f''(K)(y-K)\,dK \\
&= f(x) + f'(x)(y-x) + \int_x^y f''(K)\left((y-K)_+ - (K-y)_+\right)dK \\
&= f(x) + f'(x)(y-x) + \int_y^x f''(K)(K-y)_+\,dK + \int_x^y f''(K)(y-K)_+\,dK \\
&= f(x) + f'(x)(y-x) + \int_0^x f''(K)(K-y)_+\,dK + \int_x^\infty f''(K)(y-K)_+\,dK
\end{aligned}
$$

Take $y = S_T$, $x = F_t^T$ (Carr-Madan formula):

$$
\begin{aligned}
f(S_T) \quad &= \quad f(F_t^T) + f'(F_t^T)(S_T - F_t^T) + \int_0^{F_t^T} f''(K)(K-S_T)_+\,dK \\
&\quad + \int_{F_t^T}^\infty f''(K)(S_T-K)_+\,dK
\end{aligned}
$$

VIX index

$$f(S_T) = f(F_t^T) + f'(F_t^T)(S_T - F_t^T) + \int_0^{F_t^T} f''(K)(K - S_T)_+ \, dK + \int_{F_t^T}^{\infty} f''(K)(S_T - K)_+ \, dK$$

Take **undiscounted** prices at t:

$$\text{Price}_t[f(S_T)] = f(F_t^T) + \int_0^{F_t^T} f''(K) P_t(T,K) \, dK + \int_{F_t^T}^{\infty} f''(K) C_t(T,K) \, dK$$

Take $f = \ln$:

$$\text{Price}_t \left[\ln\left(\frac{S_T}{F_t^T}\right)\right] = -\int_0^{F_t^T} \frac{P_t(T,K)}{K^2} dK - \int_{F_t^T}^{\infty} \frac{C_t(T,K)}{K^2} dK$$

$$(\mathrm{VIX}_t)^2 = -\frac{2}{\tau}\mathrm{Price}_t \left[\ln\left(\frac{S_{t+\tau}}{F_t^{t+\tau}}\right)\right] = \frac{2}{\tau}\int_0^{F_t^{t+\tau}} \frac{P_t(t+\tau,K)}{K^2} dK + \frac{2}{\tau}\int_{F_t^{t+\tau}}^{\infty} \frac{C_t(t+\tau,K)}{K^2} dK$$

VIX index

$$(\mathrm{VIX}_t)^2 = \frac{2}{\tau}\int_0^{F_t^{t+\tau}} \frac{P_t(t+\tau,K)}{K^2}dK + \frac{2}{\tau}\int_{F_t^{t+\tau}}^{\infty} \frac{C_t(t+\tau,K)}{K^2}dK$$

- Requires undiscounting: $P_t(t+\tau,K)$ and $C_t(t+\tau,K)$ are undiscounted put and call prices.
- Requires interpolation/extrapolation in strikes and interpolation in maturities.
- CBOE uses trapezoidal rule for integration over strikes (up to two consecutive zero bid quotes).
- CBOE uses affine interpolation in maturities, using the two closest listed maturities T_- and T_+ s.t. $T_- < t+\tau < T_+$.

VIX futures

- The VIX index cannot be traded, but VIX futures can.
- VIX future expiring at T = the instrument that pays VIX_T at T.
- VIX_T^2 can be replicated using vanilla options on the S&P 500:

$$(\mathrm{VIX}_T)^2 := -\frac{2}{\tau}\mathrm{Price}_T\left[\ln\left(\frac{S_{T+\tau}}{F_T^{T+\tau}}\right)\right]$$

$\Longrightarrow$ To replicate exactly $(\mathrm{VIX}_T)^2$ at time 0: buy $-\frac{2}{\tau}\ln S_{T+\tau}$, sell $-\frac{2}{\tau}\ln F_T^{T+\tau}$.

- But its square root VIX_T cannot.
- Contrary to the price of the future on VIX^2, the price of the VIX future cannot be inferred by arbitrage arguments from the SPX smile.
- T is 30 days before SPX option monthly maturities, i.e., 30 days before the third Friday of each month.

Typical VIX future curves

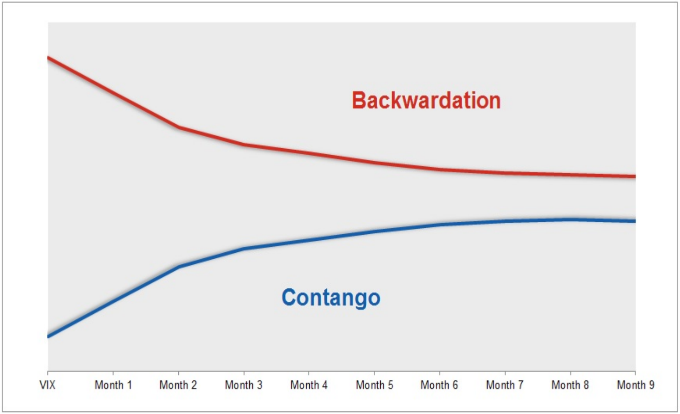


Figure: From CBOE Options Hub

VIX options/VXX options

- Pays $(\text{VIX}_T - K)_+$ or $(K - \text{VIX}_T)_+$ at T; T is a VIX future maturity.
- The underlying asset of VIX options is the VIX future with the same expiry. Only one option maturity per VIX future: the VIX future expiry.

- VXX ETN synthesizes the value of a VIX future with constant 30-day expiry. Linear combination of the front month and second month VIX futures.
- Continuous rolling:
  Contango $\Longrightarrow$ VXX has negative drift.
  Backwardation $\Longrightarrow$ VXX has positive drift.
- VXX options pay a call/put payoff on VXX. Same maturities as SPX options (third Friday of the month).
- They are like mid-term options: option on a VIX future expiring 30 days after the maturity of the option.

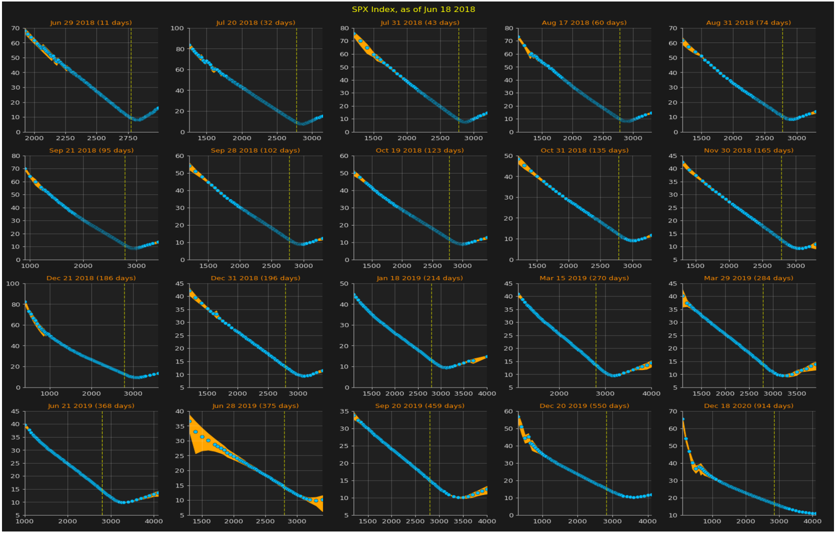
Contents

The volatility smile

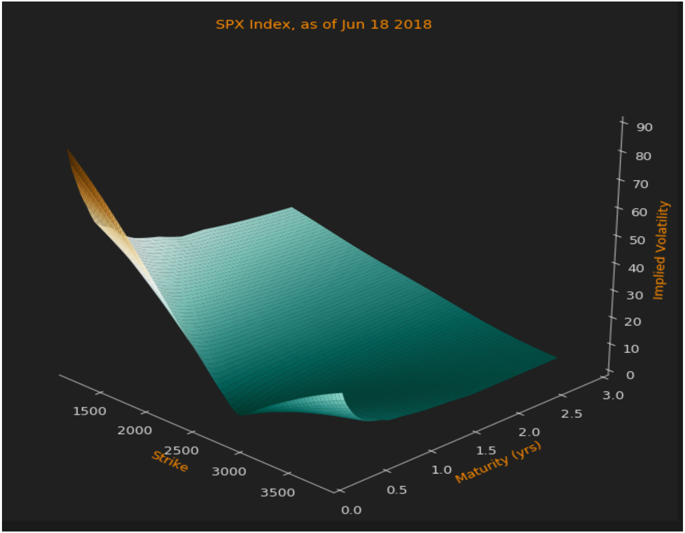
$$C_{\mathrm{mkt}}(T,K) = C_{\mathrm{BS}}(T,K;S_0,\sigma_{\mathrm{BS}}(T,K),r,q)$$

- The implied volatility of European call/put option depends on the maturity T and the strike K of the option.
- The surface of implied volatilities $(T,K) \mapsto \sigma_{\mathrm{BS}}(T,K)$ is referred to as the smile; sometimes smile denotes only a time-slice $K \mapsto \sigma_{\mathrm{BS}}(T,K)$; some people use "skew".
- Smile-like smiles; smirk-like smiles. Stocks, indices. FX.
- No-arbitrage constrains smiles a lot:
  - $K \mapsto C_{\mathrm{BS}}(T,K;S_0,\sigma_{\mathrm{BS}}(T,K),r,q)$ must be convex non-increasing.
  - When the underlying pays no dividend, $T \mapsto C_{\mathrm{BS}}(T,K;S_0,\sigma_{\mathrm{BS}}(T,K),r,q)$ must be non-decreasing.

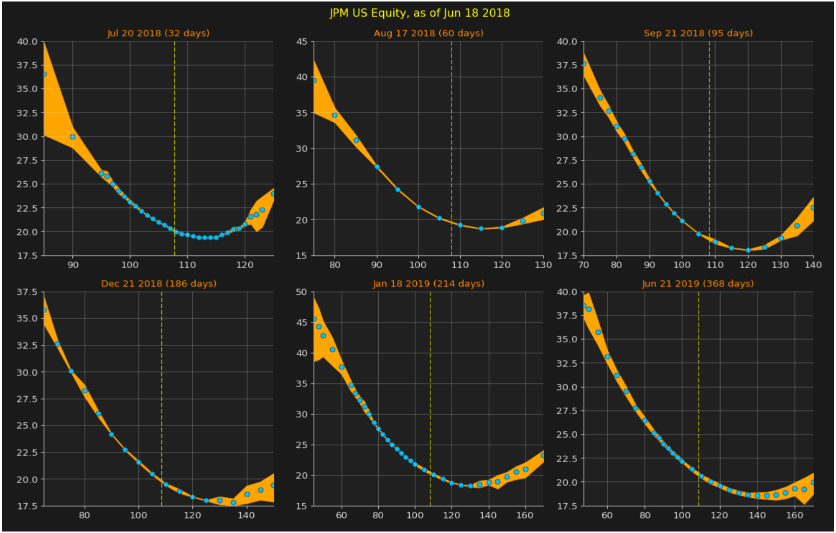
The volatility smile: SPX as of June 18, 2018



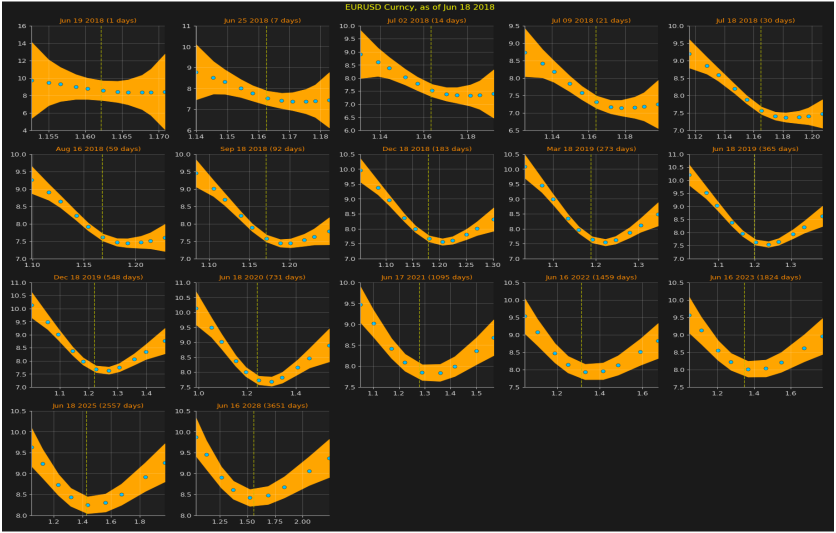
The volatility smile: SPX as of June 18, 2018



The volatility smile: JP Morgan as of June 18, 2018



The volatility smile: EURUSD as of June 18, 2018



Contents

Volatility modeling: a brief history

- **Constant volatility** σ: Black-Scholes (1973)
- **Local volatility** $\sigma(t, S_t)$: Dupire (1994) (see also Bick and Reisman, Derman and Kani)
- **Stochastic volatility**: σ_t is a stochastic process driven by extra random sources, e.g., one or several extra Brownian motions: Hull-White (1987), Stein-Stein (1991), Heston (1993), Dupire (1996), Bergomi (2005)...
- **Local stochastic volatility** $\sigma_t = a_t l(t, S_t)$: Jex (1999), Lipton (2002); see also SABR model (2002)
- **Path-dependent volatility** $\sigma(t, (S_u, u \leq t))$: Hobson-Rogers (1998), Guyon (2014), Guyon-Lekeufack (2022)
- **Rough volatility**: σ_t is a stochastic process driven by one or several fractional Brownian motions (with Hurst exponent $H < \frac{1}{2}$). Gatheral, Jaisson, Rosenbaum, Bayer, Friz... (from 2014)
- **Cross-dependent volatility**. Idea with 2 assets S_t^1, S_t^2: $\sigma_1(t, S_t^1, S_t^2)$, $\sigma_2(t, S_t^1, S_t^2)$. Guyon (2014)
- Some models also include jumps in the asset price and/or the volatility: Merton (1975), Bates (1996), Kou (2002), Lipton (2002), Andersen and Lipton (2012)...

Contents

Static vs dynamic properties of volatility models

Static properties typically refer to the marginal distributions $\mathcal{L}(S_t)$

- Prices of vanilla options
- Volatility smile
- Risk-neutral marginal distribution $e^{rT}\frac{d^2C(T,K)}{dK^2}$

Dynamic properties refer to joint distributions, e.g., $\mathcal{L}(S_{t_1},\ldots,S_{t_N})$:

- Volatilities and correlations of volatilities (spot-starting or forward-starting)
- Joint spot-vol dynamics: How the volatility moves when the spot moves
- Determine the distributions $\mathcal{L}(S_{t_1},\ldots,S_{t_N})$

Models can have the same static properties, and completely different dynamic properties.

Different models can generate the same smile

Theorem (Gyöngy, 1986). Let

$$\frac{dS_t}{S_t} = b_t \, dt + \sigma_t \, dW_t$$

Define

$$\begin{aligned}
\sigma(t,S)^2 &= \mathbb{E}\left[\sigma_t^2 \middle| S_t = S\right] \\
b(t,S) &= \mathbb{E}\left[b_t | S_t = S\right] \\
\frac{d\bar{S}_t}{\bar{S}_t} &= b(t,\bar{S}_t)\, dt + \sigma(t,\bar{S}_t)\, dW_t, \qquad \bar{S}_0 = S_0
\end{aligned}$$

Then for all t, $\mathcal{L}(\bar{S}_t) = \mathcal{L}(S_t)$.

- Different dynamic properties, but same smile (same static properties).
- $(\bar{S}_t)$ is a Markov model. It is called the **Markovian projection** of the original model (S_t)

Proof of the Gyöngy theorem

For simplicity, assume zero drift: $b_t = 0$. By applying Itô-Tanaka's formula on a vanilla call payoff with maturity t and strike K, $\mathcal{P}_t := (S_t - K)_+$, we have:

$$\begin{aligned} d\mathcal{P}_t &= \mathbf{1}_{S_t > K} \sigma_t S_t \, dW_t + \frac{1}{2} S_t^2 \sigma_t^2 \delta(S_t - K) \, dt \\ &= \mathbf{1}_{S_t > K} \sigma_t S_t \, dW_t + \frac{1}{2} K^2 \sigma_t^2 \delta(S_t - K) \, dt \end{aligned}$$

By taking the expectation $\mathbb{E}[\cdot]$ on both sides of the above equation and by assuming that $M_t = \int_0^t \mathbf{1}_{S_s > K} \sigma_s S_s dW_s$ is a true martingale, we get

$$\partial_t \mathcal{C}(t, K) = \frac{1}{2} K^2 \mathbb{E}[\sigma_t^2 \delta(S_t - K)]$$

where $\mathcal{C}(t, K) = \mathbb{E}[\mathcal{P}_t]$ denotes the price of the call option in the initial (stochastic volatility) model.

Proof of the Gyöngy theorem (continued)

Then, by using that $\partial_K \mathcal{C}(t,K) = -\mathbb{E}[\mathbf{1}_{S_t>K}]$ and $\partial_K^2 \mathcal{C}(t,K) = \mathbb{E}[\delta(S_t - K)]$, we deduce that

$$\partial_t \mathcal{C}(t,K) = \frac{1}{2} K^2 \partial_K^2 \mathcal{C}(t,K) \mathbb{E}[\sigma_t^2 | S_t = K] = \frac{1}{2} K^2 \partial_K^2 \mathcal{C}(t,K) \sigma^2(t,K)$$

with the initial condition $\mathcal{C}(0,K) = (S_0 - K)_+$.

Similarly, for the Markov model $(\bar{S}_t)$,

$$\partial_t \bar{\mathcal{C}}(t,K) = \frac{1}{2} K^2 \partial_K^2 \bar{\mathcal{C}}(t,K) \sigma^2(t,K)$$

with the same initial condition.

By uniqueness of the solution to this PDE, $\bar{\mathcal{C}}(t,K) = \mathcal{C}(t,K)$ for all t, K, i.e., $\mathcal{L}(\bar{S}_t) = \mathcal{L}(S_t)$ for all t.

Forward PDE satisfied by call prices - the general case

Proposition. Let us consider the following dynamics for an asset S, where the volatility α_t, the interest rate r_t, and the repo q_t, inclusive of the dividend yield, are all stochastic processes:

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \alpha_t\,dW_t \tag{1}$$

Let $D_{0t} = \exp\left(-\int_0^t r_s ds\right)$ denote the discount factor and $\mathcal{C}(t,K) = \mathbb{E}[D_{0t}(S_t - K)_+]$ denote the price of the call option in the model. Then $\mathcal{C}(t,K)$ satisfies the forward PDE

$$\partial_t \mathcal{C}(t,K) = K\mathbb{E}[D_{0t}(r_t - q_t - (r_t^0 - q_t^0))\mathbf{1}_{S_t > K}] - (r_t^0 - q_t^0)K\partial_K \mathcal{C}(t,K)$$
$$-\,\mathbb{E}[D_{0t}(q_t - q_t^0)(S_t - K)_+] - q_t^0 \mathcal{C}(t,K) + \frac{1}{2}K^2\partial_K^2 \mathcal{C}(t,K)\frac{\mathbb{E}[D_{0t}\alpha_t^2|S_t = K]}{\mathbb{E}[D_{0t}|S_t = K]}$$

with initial condition $\mathcal{C}(0,K) = (S_0 - K)_+$, where r_t^0 and q_t^0 are arbitrary **deterministic** rates and repos.

Forward PDE satisfied by call prices - the case where rates and repo are deterministic and the case of the local volatility model

$$\partial_t \mathcal{C}(t,K) = K\mathbb{E}[D_{0t}(r_t - q_t - (r_t^0 - q_t^0))\mathbf{1}_{S_t>K}] - (r_t^0 - q_t^0)K\partial_K\mathcal{C}(t,K)$$
$$-\mathbb{E}[D_{0t}(q_t - q_t^0)(S_t - K)_+] - q_t^0\mathcal{C}(t,K) + \frac{1}{2}K^2\partial_K^2\mathcal{C}(t,K)\frac{\mathbb{E}[D_{0t}\alpha_t^2|S_t = K]}{\mathbb{E}[D_{0t}|S_t = K]}$$

- In the case where rates r_t and repo/dividend yield q_t are deterministic, we can choose $r_t^0 = r_t$ and $q_t^0 = q_t$. The above PDE then reads

$$\partial_t \mathcal{C}(t,K) = -(r_t - q_t)K\partial_K\mathcal{C}(t,K) - q_t\mathcal{C}(t,K) + \frac{1}{2}K^2\partial_K^2\mathcal{C}(t,K)\mathbb{E}[\alpha_t^2|S_t = K]$$

- If moreover $\alpha_t = \sigma(t, S_t)$ (local volatility model) we get the **Dupire PDE** (1994)

$$\partial_t \mathcal{C}(t,K) = -(r_t - q_t)K\partial_K\mathcal{C}(t,K) - q_t\mathcal{C}(t,K) + \frac{1}{2}K^2\partial_K^2\mathcal{C}(t,K)\sigma(t,K)^2$$

Proof of the general proposition

By applying Itô-Tanaka's formula on a discounted vanilla call payoff with maturity t and strike K, $\mathcal{P}_t := D_{0t}(S_t - K)_+$, we have:

$$\begin{aligned} d\mathcal{P}_t &= -D_{0t}(S_t - K)_+ r_t \, dt + D_{0t}\mathbf{1}_{S_t > K} S_t \left((r_t - q_t)\, dt + \alpha_t dW_t\right) \\ &\quad + \frac{1}{2} S_t^2 \alpha_t^2 D_{0t} \delta(S_t - K)\, dt \\ &= D_{0t}\mathbf{1}_{S_t > K}(r_t - q_t) K \, dt - D_{0t} q_t (S_t - K)_+ \, dt + D_{0t}\mathbf{1}_{S_t > K} \alpha_t S_t dW_t \\ &\quad + \frac{1}{2} K^2 \alpha_t^2 D_{0t} \delta(S_t - K)\, dt \end{aligned}$$

By taking the expectation $\mathbb{E}[\cdot]$ on both sides of the above equation and by assuming that $M_t = \int_0^t D_{0s}\mathbf{1}_{S_s > K} \alpha_s S_s dW_s$ is a true martingale, we get

$$\partial_t \mathcal{C}(t, K) = K\mathbb{E}[D_{0t}(r_t - q_t)\mathbf{1}_{S_t > K}] - \mathbb{E}[D_{0t} q_t (S_t - K)_+] + \frac{1}{2} K^2 \mathbb{E}[D_{0t} \alpha_t^2 \delta(S_t - K)]$$

Proof of the general proposition (continued)

Then, by using that $\partial_K \mathcal{C}(t,K) = -\mathbb{E}[D_{0t}\mathbf{1}_{S_t>K}]$ and $\partial_K^2 \mathcal{C}(t,K) = \mathbb{E}[D_{0t}\,\delta(S_t-K)]$, we deduce that

$$\partial_t \mathcal{C}(t,K) = K\mathbb{E}[D_{0t}(r_t - q_t - (r_t^0 - q_t^0))\mathbf{1}_{S_t>K}] - (r_t^0 - q_t^0)K\partial_K \mathcal{C}(t,K)$$

$$-\mathbb{E}[D_{0t}(q_t - q_t^0)(S_t-K)_+] - q_t^0\mathcal{C}(t,K) + \frac{1}{2}K^2\partial_K^2\mathcal{C}(t,K)\frac{\mathbb{E}[D_{0t}\alpha_t^2|S_t=K]}{\mathbb{E}[D_{0t}|S_t=K]}$$

with the initial condition $\mathcal{C}(0,K) = (S_0-K)_+$.

Contents

Black-Scholes model, P&L analysis

- Delta-hedging argument
- Black-Scholes PDE
- Gamma-theta P&L analysis: What makes the price of an option?
- Black-Scholes robustness
- A quizz

Black-Scholes model

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t$$

- Assume the price of vanilla payoff $g(S_T)$ at time t is a smooth function V of (t, S_t), $V_t = V(t, S_t)$.
- By Itô's lemma,

$$dV_t = \left(\mu S \frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right) dt + \sigma S \frac{\partial V}{\partial S}\, dW_t$$

- The crucial delta-hedging argument of Black and Scholes (and Merton): Consider the so-called the delta-hedge portfolio, short one option and long $\Delta := \partial V / \partial S$ shares at time t. Its value is $\Pi_t = -V_t + \frac{\partial V}{\partial S} S_t$.
- Assume that the portfolio is self-financing: the change of value of Π over $[t, t + dt]$ is only due to the change in the value of S, i.e., Δ is kept constant over $[t, t + dt]$. There is no injection or withdrawal of cash in the portfolio.

The Black-Scholes PDE

- Then the total profit or loss is

$$
\begin{aligned}
d\Pi_t &= -dV_t + \frac{\partial V}{\partial S}(dS_t + qS_t\,dt) \\
&= -\left(\left(\mu S\frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right) dt + \sigma S \frac{\partial V}{\partial S}\, dW_t\right) \\
&\quad + \frac{\partial V}{\partial S}(dS_t + qS_t\,dt) \\
&= \left(-\frac{\partial V}{\partial t} + qS\frac{\partial V}{\partial S} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right) dt
\end{aligned}
$$

- No more dW_t, no more uncertaintly! $\Longrightarrow$ we must have $d\Pi_t = r\Pi_t\,dt$, i.e.,

$$
\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r-q)S\frac{\partial V}{\partial S} - rV = 0, \quad t < T, \quad V(T,S) = g(S).
$$

- This is the famous **Black-Scholes PDE**.

Gamma-theta P&L analysis: What makes the price of an option?

- For simplicity, assume that $r = q = 0$. Then the Black-Scholes PDE reads

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = 0, \quad t < T, \qquad V(T,S) = g(S).$$

- $\frac{\partial V}{\partial t}$ = "Theta", $\frac{\partial^2 V}{\partial S^2} = \Gamma =$ "Gamma", $S^2 \frac{\partial^2 V}{\partial S^2} =$ "dollar Gamma".
- Case of a convex payoff $S \mapsto g(S)$, e.g., a call or put option.
- Between t and $t + \Delta t$, the asset price moves by ΔS with

$$(\Delta S)^2 \approx \sigma^2 S^2 \Delta t.$$

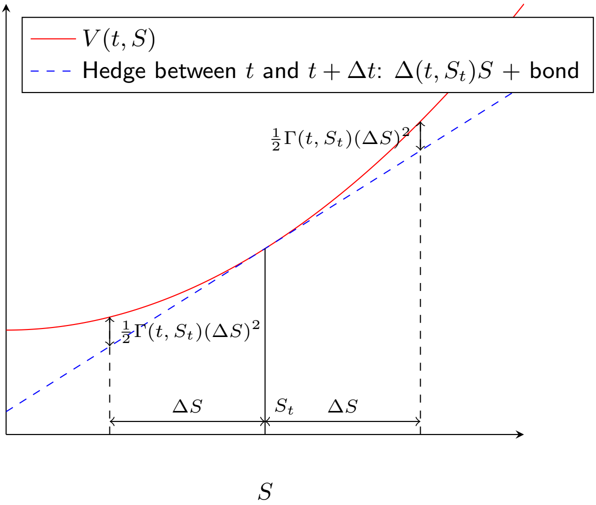
The (delta-hedged) derivative's seller always loses money; she loses

$$\frac{1}{2}\frac{\partial^2 V}{\partial S^2}(\Delta S)^2 \approx \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2 \Delta t.$$

- The derivative's seller thus charges this amount to the buyer: the option at t should be more expensive than the option at $t + \Delta t$ by the amount $\frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2 \Delta t$. That is, $-\frac{\partial V}{\partial t} = \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}$. This is precisely the Black-Scholes PDE:

$$\text{Theta} + \frac{1}{2}\sigma^2 S^2 \text{Gamma} = 0$$

Gamma-theta P&L analysis: What makes the price of an option?



Black-Scholes robustness (El Karoui-Jeanblanc-Shreve 1998)

- For simplicity, assume $r = q = 0$. Consider a general diffusive model for S:

$$dS_t = \sigma_t S_t dW_t$$

- **No assumption is made on σ_t**: it is a general stochastic process.
- Let B denote the Black price:

$$\begin{aligned} B(t,S;\sigma) &= SN\left(d_+\left(S,\sigma^2(T-t),K\right)\right) - KN\left(d_-\left(S,\sigma^2(T-t),K\right)\right) \\ d_\pm(S,v,K) &= \frac{\ln S/K}{\sqrt{v}} \pm \frac{1}{2}\sqrt{v} \end{aligned}$$

Black-Scholes robustness (El Karoui-Jeanblanc-Shreve 1998)

- $(S_T - K)_+ = B(T, S_T; \sigma)$ for any σ, so by Itô's formula,

$$(S_T - K)_+ = B(0, S_0; \sigma) + \int_0^T \left(\partial_t B + \frac{1}{2}\sigma_t^2 S_t^2 \partial_S^2 B \right)(t, S_t; \sigma)dt + \int_0^T \partial_S B(t, S_t; \sigma) dS_t$$

- Since $\partial_t B + \frac{1}{2}\sigma^2 S^2 \partial_S^2 B = 0$, we get

$$(S_T - K)_+ = B(0, S_0; \sigma) + \frac{1}{2}\int_0^T \left(\sigma_t^2 - \sigma^2\right) S_t^2 \partial_S^2 B(t, S_t; \sigma)dt + \int_0^T \partial_S B(t, S_t; \sigma) dS_t \tag{2}$$

- $\Rightarrow$ The price of the call in the general diffusive model is the Black price with arbitrary constant volatility parameter σ plus a volatility adjustment which is half the time-space average of $(\sigma_t^2 - \sigma^2)\times$ Black Dollar Gamma:

$$\mathbb{E}[(S_T - K)_+] = B(0, S_0; \sigma) + \frac{1}{2}\mathbb{E}\left[\int_0^T \left(\sigma_t^2 - \sigma^2\right) S_t^2 \partial_S^2 B(t, S_t; \sigma)dt\right]$$

- For a call option the Gamma is positive, so if the model volatility is always above (below) σ then the model price of the call $\mathbb{E}[(S_T - K)_+]$ is larger (smaller) than the Black price $B(0, S_0; \sigma)$.

A quizz

- You sell a call option, maturity 1 year, strike K, for an implied volatility of 20%. At maturity, you observe that the realized volatility over the year was 19%. Did you make money? Did you lose money?
- If you are a risk-taking (directional) trader (no delta-hedging), the P&L depends on S_T, not on the volatility.
- Now, assume you are a risk-averse trader who delta-hedges the call.
- From Equation (2), it is tempting to say that, if you delta-hedged using the Black delta with implied volatility $\sigma = 0.2$, you made money, since you sold the call for $B(0, S_0; 0.2)$, and $0.19^2 < 0.2^2$.
- However, observing a 1-year realized volatility of 19% does not mean that for all t, $\sigma_t = 0.19$. Rather, it means that $\frac{1}{T}\int_0^T \sigma_t^2 dt = 0.19^2$.
- From Equation (2), (unlikely) scenarios where, for t close to maturity, S_t has large volatility while S_t stays close to K (so that the Gamma weight is very large) can lead to a negative P&L.
- This reasoning assumes that you delta-hedge using the Black-Scholes Delta. Using other deltas can be beneficial, e.g., use the business-time hedge (Bick 1995).

Business-time hedging (Bick 1995)

- Assume $r = q = 0$ so calendar time does not enter the picture.
- Let $Q_t := \int_0^t \sigma_u^2 du$, $\bar{Q} > 0$ a variance budget, and $\tau_{\bar{Q}} := \inf\{t \geq 0 : Q_t \geq \bar{Q}\}$.
- Consider the payoff $g(S_{\tau_{\bar{Q}}})$, paid at $\tau_{\bar{Q}}$ ("timer option").
- Calendar time t has been replaced by the random business time Q_t, and the maturity T by $\bar{Q}$.
- Assume the price of the timer option at time $t \leq T$ is a smooth function V of (Q_t, S_t), $V_t = V(Q_t, S_t)$. We have $V(\bar{Q}, S) = g(S)$.
- By Itô's lemma,

$$dV_t = \left(\mu S_t \frac{\partial V}{\partial S} + \sigma_t^2 \frac{\partial V}{\partial Q} + \frac{1}{2} \sigma_t^2 S_t^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \sigma_t S_t \frac{\partial V}{\partial S} dW_t$$

- We repeat the crucial delta-hedging argument of Black and Scholes (and Merton) but we now use the new value of $\Delta := \partial V / \partial S$ for the new function $V(Q, S)$.

Business-time hedging (Bick 1995)

- The total profit or loss of the hedged portfolio Π is

$$
\begin{aligned}
d\Pi_t &= -dV_t + \frac{\partial V}{\partial S} dS_t \\
&= -\left(\left(\mu S \frac{\partial V}{\partial S} + \sigma_t^2 \frac{\partial V}{\partial Q} + \frac{1}{2}\sigma_t^2 S^2 \frac{\partial^2 V}{\partial S^2}\right) dt + \sigma_t S \frac{\partial V}{\partial S}\, dW_t\right) \\
&\quad + \frac{\partial V}{\partial S} dS_t \\
&= \left(-\sigma_t^2 \frac{\partial V}{\partial Q} - \frac{1}{2}\sigma_t^2 S^2 \frac{\partial^2 V}{\partial S^2}\right) dt
\end{aligned}
$$

- No more dW_t, no more uncertaintly! $\Longrightarrow$ we must have $d\Pi_t = 0$, i.e.,

$$
\frac{\partial V}{\partial Q} + \frac{1}{2} S^2 \frac{\partial^2 V}{\partial S^2} = 0, \quad Q < \bar{Q}, \qquad V(\bar{Q}, S) = g(S).
$$

- No more σ in the equation, as it is absorbed by the Q variable.

Business-time hedging (Bick 1995)

- Back to the quizz: now assume that you delta-hedge the call option in business time using $\bar{Q} = \sigma^2 T$, where $\sigma = 0.2$ and $T = 1$ year.
- Similarly as (2), since $V(\bar{Q}, S) = (S - K)_+$, we have

$$(S_{\tau_{\bar{Q}}} - K)_+ = V(Q_{\tau_{\bar{Q}}}, S_{\tau_{\bar{Q}}}) = V(0, S_0) + \int_0^{\tau_{\bar{Q}}} \partial_S V(Q_t, S_t) dS_t$$

  Perfect delta-hedging of the payoff $g(S_{\tau_{\bar{Q}}})$. But payoff maturity is random!
- Since the volatility realized at $0.19 < 0.2$, we have that $Q_T = 0.19^2 T < \bar{Q}$ and $\tau_{\bar{Q}} > T$.
- At maturity $T < \tau_{\bar{Q}}$, since $Q_T < \bar{Q}$, you are left with the wealth

$$V(0, S_0) + \int_0^T \partial_S V(Q_t, S_t) dS_t = V(Q_T, S_T) > V(\bar{Q}, S_T) = g(S_T) = (S_T - K)_+$$

  You surely make money!
- Symmetrically, if the volatility realizes above the implied volatility used for pricing, selling the derivative and hedging it in business time leads to a sure loss.

Contents

Local Volatility model (LV)

- Local Volatility model (deterministic rate and div yield):

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \sigma(t, S_t)dW_t$$

where r_t and q_t are deterministic, and $\sigma(t, S)$ is a function of t and S.

- Model local volatility function from model call option prices (**Dupire's Formula**)

$$\sigma^2(T, K) = \frac{\frac{\partial C}{\partial T} + (r_T - q_T)\,K\,\frac{\partial C}{\partial K} + q_T C}{\frac{1}{2}K^2\,\frac{\partial^2 C}{\partial K^2}}$$

where

$$C(T, K) = e^{-\int_0^T r_t\,dt}\mathbb{E}^{\mathbb{Q}}\left[(S_T - K)_+\right]$$

is the model price of the call option on S with maturity T and strike K.

- Market local volatility function from market call option prices

$$\sigma^2_{\mathsf{Dup}}(T, K) := \sigma^2_{\mathsf{mkt}}(T, K) = \frac{\frac{\partial C_{\mathsf{mkt}}}{\partial T} + (r_T - q_T)\,K\,\frac{\partial C_{\mathsf{mkt}}}{\partial K} + q_T C_{\mathsf{mkt}}}{\frac{1}{2}K^2\,\frac{\partial^2 C_{\mathsf{mkt}}}{\partial K^2}}$$

where $C_{\mathsf{mkt}}(T, K)$ is the market price of the call option.

Local Volatility model (LV)

$$\sigma^2_{\mathsf{Dup}}(T,K) := \sigma^2_{\mathsf{mkt}}(T,K) = \frac{\frac{\partial C_{\mathsf{mkt}}}{\partial T} + (r_T - q_T)\, K \frac{\partial C_{\mathsf{mkt}}}{\partial K} + q_T C_{\mathsf{mkt}}}{\frac{1}{2} K^2 \frac{\partial^2 C_{\mathsf{mkt}}}{\partial K^2}}$$

- Absence of calendar arbitrage $\iff$ numerator is ≥ 0.
- Absence of butterfly arbitrage $\iff$ denominator is ≥ 0.
- In the absence of arbitrage, if denominator is positive, the r.h.s. is indeed ≥ 0.

Local Volatility model (LV)

- From Implied Volatility to Local Volatility:

$$\sigma_{\mathsf{Dup}}^2(t,x) = \frac{\frac{\partial w}{\partial T}}{1 - \frac{x}{w}\frac{\partial w}{\partial x} + \frac{1}{4}\left(-\frac{1}{4} - \frac{1}{w} + \frac{x^2}{w^2}\right)\left(\frac{\partial w}{\partial x}\right)^2 + \frac{1}{2}\frac{\partial^2 w}{\partial x^2}}$$

  where log-moneyness $x = \log(K/F_T)$, $F_T = S_0 e^{\int_0^T (r_t - q_t)dt}$ is the forward of maturity T, and $w(T,x) = \sigma_{\mathsf{BS}}^2(K,T)T$.

- Gyöngy + Dupire $\Longrightarrow$ Model

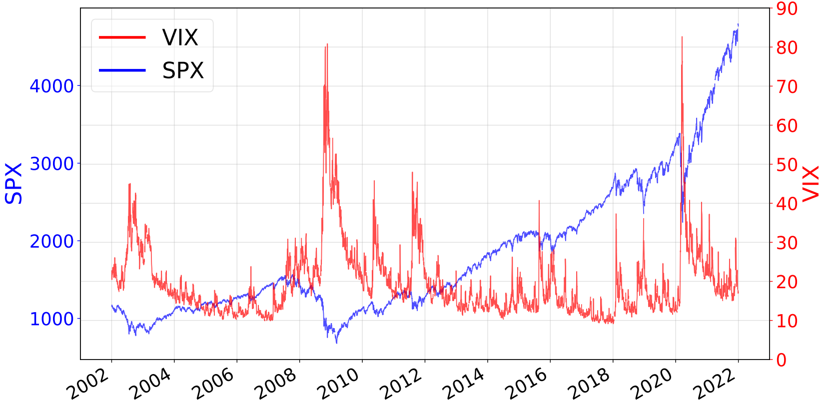
$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \sigma_t\,dW_t$$

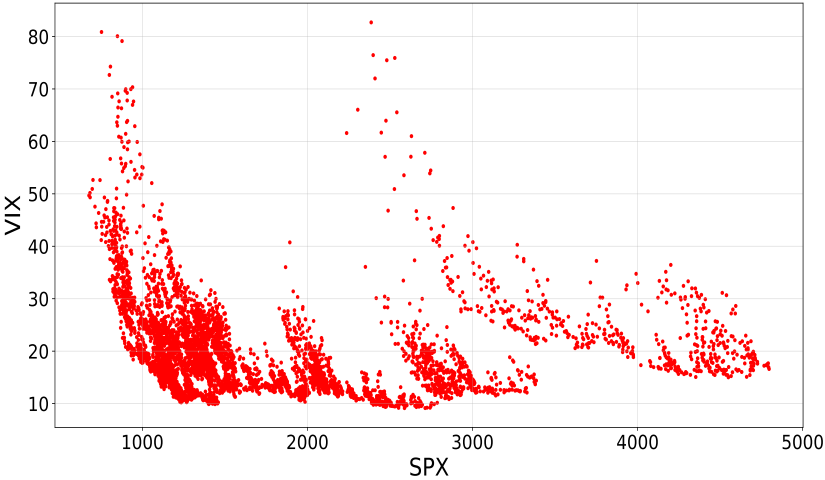
  with deterministic rates and repos is calibrated to the market smile if and only if

$$\forall t, K, \qquad \mathbb{E}[\sigma_t^2 | S_t = K] = \sigma_{\mathsf{Dup}}(t,K)^2$$

Local Volatility model (LV)

- Local volatility model is the simplest model that is consistent with the market smile.
- However, it has poor dynamics.




VIX
SPX
80
70
60
50
40
30
20
10
1000
2000
3000
4000
5000

Contents

Stochastic Volatility (SV) models

- Exotic options bear exotic risks such as volatility-of-volatility risk, forward smile risk, or spot/volatility correlation risk.
- To price and hedge these risks, one should not use the Black-Scholes model, or the local volatility model, because such models give no control on them.
- One would rather use a Stochastic Volatility (SV) model

$$dS_t = a_t S_t \, dW_t \tag{3}$$

with a_t the stochastic volatility, a random process (e.g., Heston, SABR, double-lognormal) which allows us to handle exotic risks through parameters such as volatility-of-volatility, spot/volatility correlation, or mean reversion.

Heston Stochastic Volatility Model - An example of SV Model

SDE

$$\frac{dS_t}{S_t} = (r_t - q_t)\, dt + \sqrt{V_t}\, dW_t^S$$
$$dV_t = k\left(V^0 - V_t\right) dt + \omega \sqrt{V_t}\, dW_t^V$$

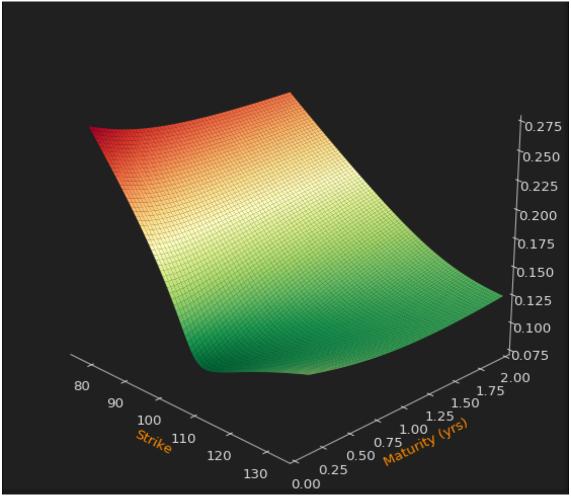
with

$$d\langle W^S, W^V \rangle_t = \rho\, dt.$$

Model Parameters

- k: mean reversion rate
- V_0: initial variance
- V^0: Long term variance
- ρ: spot-vol correlation
- ω: volatility of variance

Heston Stochastic Volatility Model - Implied Volatility Surface



$$V_0 = 0.0165, k = 1.3093, V^0 = 0.0541, \omega = 0.6344, \rho = -0.744, S_0 = 100$$

Examples of SV models

$$\begin{aligned}\frac{dS_t}{S_t} &= \sigma(t,S_t)a_t(\rho dZ_t + \sqrt{1-\rho^2}dB_t), \qquad S_0 = x \\ da_t &= b(a_t)dt + \sigma(a_t)dZ_t, \qquad a_0 = \alpha\end{aligned}$$

Examples of 1-factor SV models:

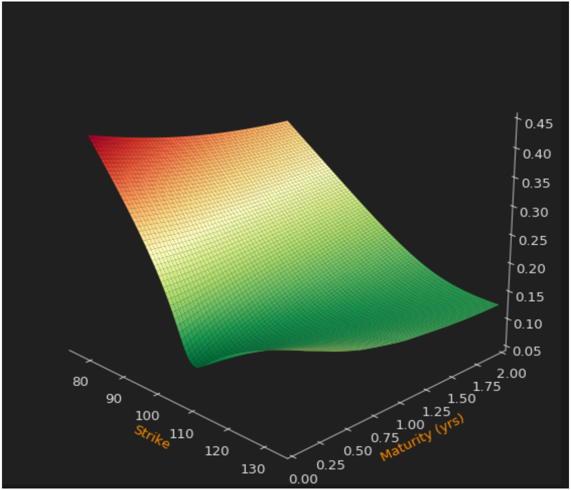
Name	SDE
Stein-Stein	$\frac{dS_t}{S_t} = a_t dW_t$ $da_t = \lambda(a_t - \bar{a})dt + \zeta dZ_t \ , \ d\langle W,Z\rangle_t = 0$
Geometric	$\frac{dS_t}{S_t} = a_t dW_t$ $da_t = \lambda a_t dt + \zeta a_t dZ_t \ , \ d\langle W,Z\rangle_t = \rho\, dt$
3/2-model	$\frac{dS_t}{S_t} = a_t dW_t$ $da_t^2 = \lambda(a_t^2 - \bar{v}a_t^4)dt + \zeta a_t^3 dZ_t \ , \ d\langle W,Z\rangle_t = \rho\, dt$
SABR	$\frac{dS_t}{S_t} = a_t S_t^{\beta-1} dW_t$ $da_t = \nu a_t dZ_t \ , \ d\langle W,Z\rangle_t = \rho\, dt$
Scott-Chesney	$\frac{dS_t}{S_t} = \sigma_0 e^{y_t} dW_t$ $dy_t = \lambda\,(\bar{y} - y_t)\,dt + \zeta dZ_t \ , \ d\langle W,Z\rangle_t = 0$
Heston	$\frac{dS_t}{S_t} = a_t dW_t$ $da_t^2 = \lambda(\bar{v} - a_t^2)dt + \zeta a_t dZ_t \ , \ d\langle W,Z\rangle_t = \rho\, dt$

Equivalent local volatility of an SV model

$$\sigma(t,S)^2 = \mathbb{E}[a_t^2 | S_t = S]$$

- Small or large $S \Longrightarrow$ the asset price has experienced large volatility between 0 and $t \Longrightarrow a_t^2$ tends to be large.
- Values of S around S_0 are compatible with both small and large a_t^2.
- Zero spot-vol correlation: $S \mapsto \sigma(t,S)$ is U-shaped.
- Negative spot-vol correlation: $S \mapsto \sigma(t,S)$ decreases around the money.
- Even with negative spot-vol correlation, $\sigma(t,S)$ increases for large S, because of the first item.
- More volatility of volatility ("vol-of-vol") $\Longrightarrow$ more convexity.

Heston Stochastic Volatility Model - Equivalent local Volatility Surface



$$V_0 = 0.0165, k = 1.3093, V^0 = 0.0541, \omega = 0.6344, \rho = -0.744, S_0 = 100$$

Heston Stochastic Volatility Model: drawbacks

- Variance swap variances are given by

$$\hat{\sigma}_T^2(t) = V^0 + \frac{1 - e^{-k(T-t)}}{k(T-t)}(V_t - V^0)$$

- Dynamics of variance swap volatilities:

$$d\hat{\sigma}_T(t) = \cdots dt + \frac{\omega}{2} \frac{1 - e^{-k(T-t)}}{k(T-t)} \frac{\sqrt{V_t}}{\hat{\sigma}_T(t)} dW_t^V$$

- For long maturities T, the instantaneous volatility of $\hat{\sigma}_T(t)$ decays like $\frac{1}{T-t}$. Indeed, for a flat term-structure of variance swaps ($V_t = V^0$),

$$\text{vol}(\hat{\sigma}_T(t)) \sim \frac{1 - e^{-k(T-t)}}{k(T-t)}$$

- Not in line with what we observe in equity markets.
- This problem comes from the fact that V has a single time scale of mean reversion.

Heston Stochastic Volatility Model: drawbacks

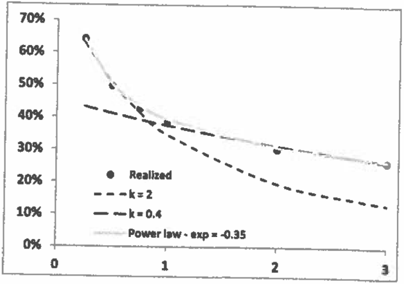


Figure 6.1: Volatility of VS volatilities of the Euro Stoxx 50 index as a function of maturity (years), evaluated on the period [2005, 2010] (dots), along with (a) volatilities of VS volatilities in the Heston model given by (6.9) for two different values of k (dotted lines), (b) a power-law fit $\propto T^{-0.35}$.

From L. Bergomi, Stochastic Volatility Modeling (Chapman & Hall, 2016)

An example of one-factor dynamics: the Heston model

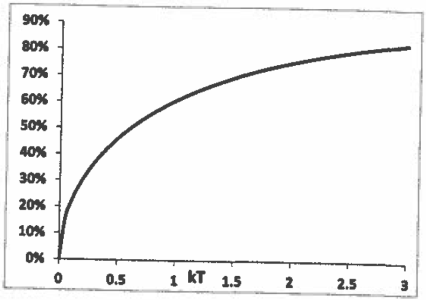


Figure 6.3: $\frac{\widehat{\sigma}_T^{\min}(t)}{\sqrt{V^0}}$ as a function of $k(T-t)$.

From L. Bergomi, Stochastic Volatility Modeling (Chapman & Hall, 2016)

Heston Stochastic Volatility Model: drawbacks

- For a flat term-structure of VS volatilities, at order one in vol-of-vol (see later), the ATMF skew of the Heston model is given by

$$\mathcal{S}_T = \frac{\rho\omega}{2\sqrt{V}} \frac{e^{-kT} - (1 - kT)}{(kT)^2}$$

- For $T \gg \frac{1}{k}$, this decreases (in absolute value) as $\frac{1}{T}$. Not in line with what we observe in equity markets.
- This problem comes from the fact that V has a single time scale of mean reversion.
- Market data suggests using at least two time scales of mean reversion, or using more complex, non-Markovian volatility dynamics that use non-exponential kernels (power law, time-shifted power law, etc.); see later.

Heston Stochastic Volatility Model: drawbacks

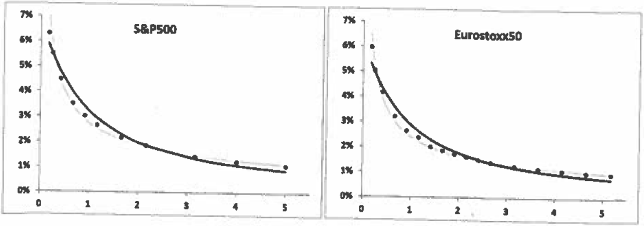


Figure 6.5: The ATMF skew as a function of maturity for the Euro Stoxx 50 and S&P 500 indexes (dots), observed on October 22, 2010, expressed as the difference in volatility points of the implied volatilities of strikes $0.95F_T$ and $1.05F_T$. A best fit using a power law with exponent 0.55 (lighter line) as well as a best fit using formula (6.20) ($k = 2.9$) (darker line) are shown as well.

From L. Bergomi, Stochastic Volatility Modeling (Chapman & Hall, 2016)

Heston Stochastic Volatility Model: drawbacks

- In particular, for short maturities,

$$\mathcal{S}_T = \frac{\rho\omega}{4\sqrt{V}}$$

- The smaller V, the larger the short-term skew (in absolute value). Not in line with what we observe in equity markets.
- This problem comes from the $\sqrt{V}$ in the volatility of V. Replacing $\sqrt{V}$ by V^α leads to

$$\mathcal{S}_0 = \frac{\rho\omega}{4} V^{\alpha-1}$$

Equity market data suggests using $\alpha = 1$.

Heston Stochastic Volatility Model: drawbacks

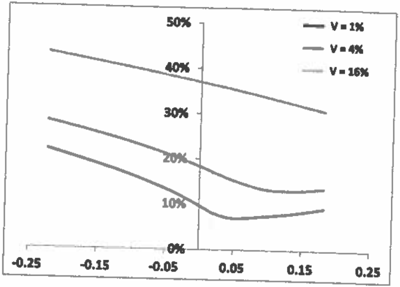


Figure 6.4: Implied volatilities $\widehat{\sigma}_{KT}$ as a function of $\ln\left(\frac{K}{F_T}\right)$ for a maturity $T = 3$ months, in the Heston model, generated with the following parameters: $V^0 = 0.04$, $k = 1$, $\sigma = 0.6$, $\rho = -80\%$, for three different values of V.

From L. Bergomi, Stochastic Volatility Modeling (Chapman & Hall, 2016)

Contents

One-factor lognormal forward instantaneous variance models

- $\mathbb{E}[\sigma_u^2|\mathcal{F}_t]$ is a **forward instantaneous variance**. We denote it by
$$\xi_t^u := \mathbb{E}[\sigma_u^2|\mathcal{F}_t].$$
- It is well known (Dupire, Bergomi, Buehler) 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** under risk-neutral measures.
- $(\xi_0^u, u \geq 0)$ is read from the market at time 0.
- For simplicity, we assume that forward instantaneous variances are lognormal and all driven by a single standard one-dimensional $(\mathcal{F}_t)$-Brownian motion Z, correlated with W:
$$\frac{d\xi_t^u}{\xi_t^u} = K(t,u)\, dZ_t. \tag{4}$$
- Here the kernel K is deterministic.
$$\frac{dS_t}{S_t} = \sigma_t\, dW_t, \qquad \sigma_t^2 := \xi_t^t, \qquad d\langle W, Z\rangle_t = \rho\, dt.$$

One-factor lognormal forward instantaneous variance models

- The solution to (4) is simply

$$\xi_t^u = \xi_0^u \exp\left(\int_0^t K(s,u)\,dZ_s - \frac{1}{2}\int_0^t K(s,u)^2\,ds\right) \tag{5}$$

which yields

$$\sigma_u^2 = \xi_0^u \exp\left(\int_0^u K(s,u)\,dZ_s - \frac{1}{2}\int_0^u K(s,u)^2\,ds\right). \tag{6}$$

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

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

- Financially, we expect the kernel $K : \mathbb{R}_+ \to \mathbb{R}_+$ to be decreasing: The further away the instantaneous forward variance maturity u, the less volatile the instantaneous forward variance.

Exponential kernel

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

- One-factor Bergomi model (Dupire 1993, Bergomi 2005).
- ξ_t^u admits a one-dimensional Markov representation: $\xi_t^u = \xi_0^u f^u(t, X_t)$

$$f^u(t,x) = \exp\left(\omega e^{-k(u-t)}x - \frac{\omega^2}{2} e^{-2k(u-t)}\mathrm{Var}(X_t)\right), \quad \mathrm{Var}(X_t) = \frac{1-e^{-2kt}}{2k}$$

Ornstein-Uhlenbeck process $X_t = \int_0^t e^{-k(t-s)} dZ_s$:

$$dX_t = -kX_t\, dt + dZ_t, \quad X_0 = 0. \tag{7}$$

- (S_t, X_t) Markov process.
- $k =$ strength of mean reversion.
- $\omega =$ volatility of variance: it is the instantaneous (lognormal) volatility of the instantaneous variance $\sigma_t^2 := \xi_t^t$.
- Exercise: $d(\sigma_t^2) = ?$

Power-law kernel

$$K(\theta) = \nu \theta^{H-\frac{1}{2}}, \qquad \nu > 0, \qquad H \in \left(0, \frac{1}{2}\right)$$

- Rough Bergomi model (Bayer, Friz, Gatheral, 2014).
- $\lim_{\theta \to 0^+} K(\theta) = +\infty$.
- $H =$ Hurst exponent.

$$\begin{aligned} \xi_t^u &= \xi_0^u \exp\left(\nu X_t^u - \frac{\nu^2}{2}\mathrm{Var}(X_t^u)\right) \\ X_t^u &= \int_0^t (u-s)^{H-\frac{1}{2}}\, dZ_s, \qquad \mathrm{Var}(X_t^u) = \frac{u^{2H} - (u-t)^{2H}}{2H}. \end{aligned}$$

- $H > 0$ ensures that $\mathrm{Var}(X_t^t)$ is finite.
- Exercise: $d(\sigma_t^2) = ?$
- Non-Markovian model. And the instantaneous volatility is not a semimartingale!

Heston model as a variance curve model

$$\xi_t^u = \mathbb{E}[V_u|\mathcal{F}_t] = V^0 + e^{-k(u-t)}(V_t - V^0) = V^0 + e^{-k(u-t)}(\xi_t^t - V^0)$$
$$d\xi_t^u = \omega e^{-k(u-t)}\sqrt{\xi_t^t}dW_t^V$$

- One-factor model for forward variances.
- The instantaneous volatility of all forward variances ξ_t^u is proportional to the instantaneous volatility $\sqrt{\xi_t^t}$.
- All forward variances are 100% correlated.
- But the Heston model should *not* be considered a variance curve model: it is not able to accommodate general term structures of VS volatilities. Can only accommodate

$$\xi_0^u = V^0 + e^{-ku}(\xi_0^0 - V^0)$$

Two-factor Bergomi model

- Idea: Combine two exponentials to mimick power-law decay
- Variance curve stays Markovian in low dimension (two)
- Variance curve driven by two factors (long end/short end)

$$
\begin{aligned}
\frac{d\xi_t^u}{\xi_t^u} &= \omega \xi_t^u \alpha_\theta \left((1-\theta) e^{-k_1(u-t)} dW_t^1 + \theta e^{-k_2(u-t)} dW_t^2 \right) \\
\alpha_\theta &= \left((1-\theta)^2 + \theta^2 + 2\rho_{12}\theta(1-\theta) \right)^{-1/2}
\end{aligned}
$$

$$
\begin{aligned}
dX_t^i &= -k_i X_t^i \, dt + dW_t^i, \quad X_0^i = 0, \quad i \in \{1,2\} \\
\mathrm{Var}(X_t^i) &= \frac{1 - e^{-2k_i t}}{2k_i} \\
\mathrm{Cov}(X_t^1, X_t^2) &= \frac{1 - e^{-(k_1+k_2)t}}{k_1 + k_2}
\end{aligned}
$$

Two-factor Bergomi model

$$
\begin{aligned}
x_t^u &= \alpha_\theta\left((1-\theta)e^{-k_1(u-t)}X_t^1+\theta e^{-k_2(u-t)}X_t^2\right)\\
d\xi_t^u &= \omega\xi_t^u\,dx_t^u\\
\xi_t^u &= \xi_0^u f^u(t,x_t^u)\\
f^u(t,x) &= \exp\left(\omega x-\frac{\omega^2}{2}\chi(t,u)\right)\\
\chi(t,u) &= \alpha_\theta^2\Big((1-\theta)^2e^{-2k_1(u-t)}\mathsf{Var}(X_t^1)+\theta^2e^{-2k_2(u-t)}\mathsf{Var}(X_t^2)\\
&\quad +2\theta(1-\theta)e^{-(k_1+k_2)(u-t)}\mathsf{Cov}(X_t^1,X_t^2)\Big)
\end{aligned}
$$

Calibration procedure suggested by Bergomi (2005):

- Calibrate ω, θ, k_1, k_2, ρ_{12} to mimick the power-law-like term-structure of volatilities of VS volatilities on a wide range of maturities (say, 1 month to 5 years)
- Several sets are possible. Pick one that gives the desired correlation structure between forward variances
- Finally choose ρ_{S1} and ρ_{S2} to best fit the vanilla smile (see closed form formula of the smile at order 2 in vol-of-vol in the next chapter)

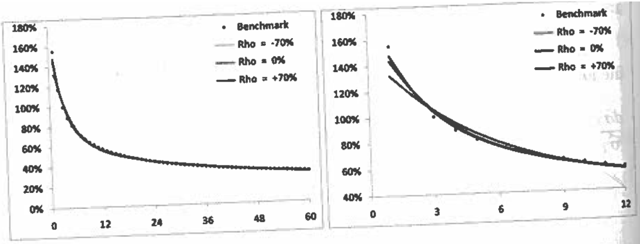
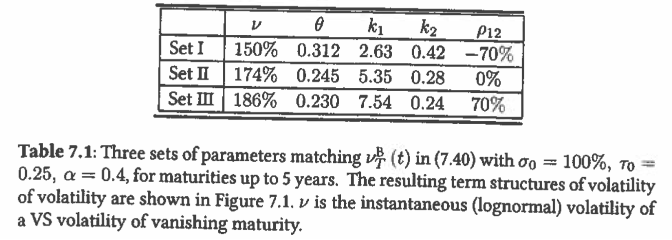


Figure 7.1: The left-hand graph displays the term structure of instantaneous volatilities at $t = 0$ of VS volatilities $\nu_T(t)$ (y axis) as a function of T (x axis, in months) generated by the benchmark form (7.40) as well as the two-factor model, with the different sets of parameters listed in Table 7.1. The right-hand graph focuses on maturities less than 1 year.

From L. Bergomi, Stochastic Volatility Modeling (Chapman & Hall, 2016)

Two-factor Bergomi model

$\omega = 2\nu$



	ν	θ	k_1	k_2	ρ_{12}
Set I	150%	0.312	2.63	0.42	−70%
Set II	174%	0.245	5.35	0.28	0%
Set III	186%	0.230	7.54	0.24	70%

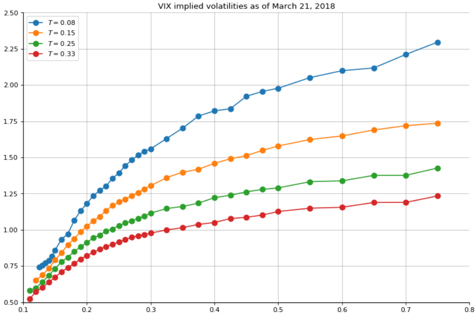
Table 7.1: Three sets of parameters matching $\nu_T^B(t)$ in (7.40) with $\sigma_0 = 100\%$, $\tau_0 = 0.25$, $\alpha = 0.4$, for maturities up to 5 years. The resulting term structures of volatility of volatility are shown in Figure 7.1. ν is the instantaneous (lognormal) volatility of a VS volatility of vanishing maturity.

From L. Bergomi, Stochastic Volatility Modeling (Chapman & Hall, 2016)

Two-factor Bergomi model

- In the classical version of the 2-factor Bergomi model, the VIX smile is almost flat, as forward instantaneous variances are lognormal
- The VIX smile can simply be modeled by changing the function

$$f^u(t,x) \quad = \quad (1-\gamma_u)\exp\left(\omega_u x - \frac{\omega_u^2}{2}\chi(t,u)\right) + \gamma_u \exp\left(\beta_u\omega_u x - \frac{\beta_u^2\omega_u^2}{2}\chi(t,u)\right)$$



Contents

Outline

- Motivation
- Expansion of the smile at order 2 in vol-of-vol
- Short maturities and long maturities
- First example: a family of Heston-like models
- Second example: the Bergomi model with 2 factors on the variance curve
- Numerical experiments
- Rederiving the link between skew and skewness of log-returns
- Conclusion

Motivation

- Consider the following general dynamics for a diffusive stochastic volatility model:

$$\begin{aligned} dX_t &= -\frac{1}{2}\xi_t^t dt + \sqrt{\xi_t^t} dW_t^1, \qquad X_0 = x \\ d\xi_t^u &= \lambda(t, u, \xi_t^\cdot) \cdot dW_t, \qquad \xi_0^u = y^u \end{aligned} \tag{8}$$

- $X_t = \ln S_t$
- $\xi_t^\cdot \equiv (\xi_t^u, t \leq u)$: instantaneous forward variance curve from t onwards. ξ^u = driftless process; initial value y^u read on market prices of variance swap contracts: $\xi_0^u = \frac{d}{du}\left(\hat{\sigma}_u^2 u\right)$, where $\hat{\sigma}_u$ is the implied variance swap volatility for maturity u.
- $\lambda = (\lambda_1, \ldots, \lambda_d)$: volatility of forward instantaneous variances.
- $W = \left(W^1, \ldots, W^d\right)$ = a d-dimensional Brownian motion. W^1 drives the spot dynamics.
- No dividend. Zero rates and repos (for the sake of simplicity)

- **No closed-form formula** available **for the price of vanilla options** in Model (8).
- Approximations available in a few particular cases of "first generation" stochastic volatility models (e.g., Heston)
- Our goal: find a general approximation of the smile **which does not depend on a particular specification of the model**, i.e., on a particular choice of λ.

- $\Rightarrow$ We will derive **general asymptotic expansion of the smile, for small volatility of volatility, at second order**.
- Scaling factor ε: $\lambda \to \varepsilon\lambda$. X and ξ then depend on ε: $X \to X^\varepsilon$ and $\xi \to \xi^\varepsilon$.
- Two important assumptions: **no local volatility component**, and λ **does not depend on the asset value**.

- Smile produced by stochastic volatility models is generated by the covariance of forward variances with themselves and spot.
- Our goal: to **pinpoint exactly** which functionals of these covariances determine the vanilla smile

- Important to ensure, while varying ε, that implied volatilities of some specific payoffs are unchanged, so that the overall volatility level is not altered in the model.
- In our framework, **VS volatilities are unchanged** as ε is varied.

Expansion of the price of a vanilla option

- Consider the vanilla option delivering $g(X_T^\varepsilon)$ at time T.
- Price $P^\varepsilon\left(t, X_t^\varepsilon, \xi_t^{\cdot,\varepsilon}\right)$. We write $P^\varepsilon(t,x,y)$: **the variable $y \equiv (y^u, t \le u \le T)$ is a curve**.
- P^ε solves the PDE $(\partial_t + L^\varepsilon) P^\varepsilon = 0$ with terminal condition $P^\varepsilon(T,x,y) = g(x)$, where $L^\varepsilon = L_0 + \varepsilon L_1 + \varepsilon^2 L_2$ with

$$
\begin{aligned}
L_0 &= -\frac{1}{2} y^t \partial_x + \frac{1}{2} y^t \partial_x^2 \\
L_1 &= \int_t^T du \; \mu(t,u,y) \; \partial^2_{xy^u} \\
L_2 &= \frac{1}{2} \int_t^T du \int_t^T du' \; \nu(t,u,u',y) \; \partial^2_{y^u y^{u'}} \\
\mu(t,u,y) &= \sqrt{y^t} \lambda_1(t,u,y) = \frac{\mathbb{E}\left[dX_t d\xi_t^u \middle| \xi_t = y\right]}{dt} = \frac{\mathbb{E}\left[\frac{dS_t}{S_t} d\xi_t^u \middle| \xi_t = y\right]}{dt} \\
\nu(t,u,u',y) &= \sum_{i=1}^d \lambda_i(t,u,y)\lambda_i(t,u',y) = \frac{\mathbb{E}\left[d\xi_t^u d\xi_t^{u'} \middle| \xi_t = y\right]}{dt}
\end{aligned}
$$

The perturbation equations

- Assume that $P^\varepsilon = P_0 + \varepsilon P_1 + \varepsilon^2 P_2 + \varepsilon^3 P_3 + \cdots$

$$\begin{aligned}
0 &= \left(\partial_t + L_0 + \varepsilon L_1 + \varepsilon^2 L_2\right)\left(P_0 + \varepsilon P_1 + \varepsilon^2 P_2 + \varepsilon^3 P_3 + \cdots\right) \\
&= \left(\partial_t + L_0\right) P_0 + \varepsilon\left(\left(\partial_t + L_0\right) P_1 + L_1 P_0\right) \\
&\quad + \varepsilon^2\left(\left(\partial_t + L_0\right) P_2 + L_1 P_1 + L_2 P_0\right) \\
&\quad + \varepsilon^3\left(\left(\partial_t + L_0\right) P_3 + L_1 P_2 + L_2 P_1\right) + \cdots
\end{aligned}$$

- $\Rightarrow$ We need to solve the following equations:

$$\begin{aligned}
\left(\partial_t + L_0\right) P_0 &= 0, \quad P_0(T,x,y) = g(x) \\
\left(\partial_t + L_0\right) P_1 + L_1 P_0 &= 0, \quad P_1(T,x,y) = 0 \\
\left(\partial_t + L_0\right) P_n + L_1 P_{n-1} + L_2 P_{n-2} &= 0, \quad P_n(T,x,y) = 0, \qquad \forall n \geq 2
\end{aligned}$$

- L_0 = infinitesimal generator associated to X^0, the unperturbed diffusion for which $\varepsilon = 0$. L_0 = standard **one-dimensional Black-Scholes operator** with deterministic volatility $\sqrt{y^t}$ at time t.
- Each P_n = solution to the traditional **one-dimensional** diffusion equation **with a source term** $H_n = L_1 P_{n-1} + L_2 P_{n-2}$:

$$(\partial_t + L_0) P_n + H_n = 0$$

- Feynmann-Kac theorem $\Rightarrow$

$$\begin{aligned} P_0(t,x,y) &= \mathbb{E}\left[g\left(X_T^{0,t,x}\right)\right], \\ P_n(t,x,y) &= \mathbb{E}\left[\int_t^T H_n(s, X_s^{0,t,x}, y) ds\right], \qquad \forall n \geq 1 \end{aligned}$$

where $X^{0,t,x}$ is the unperturbed process where $\varepsilon = 0$, starting at log-spot x at time t:

$$dX_s^{0,t,x} = -\frac{1}{2} y^s ds + \sqrt{y^s} dW_s^1, \qquad X_t^{0,t,x} = x$$

The price at order 0

- P_0 is just the Black-Scholes price with time-dependent volatility $\sqrt{y^t}$:

$$P_0(t,x,y) = \mathbb{E}\left[g\left(x + \int_t^T \sqrt{y^s}dW_s^1 - \frac{1}{2}\int_t^T y^s ds\right)\right] = P_{BS}\left(x, \int_t^T y^s ds\right)$$

where

$$P_{BS}(x,v) = \mathbb{E}\left[g\left(x + \sqrt{v}G - \frac{1}{2}v\right)\right], \qquad G \sim \mathcal{N}(0,1) \tag{9}$$

- $v = \int_t^T y^s ds$ is the total variance of X^0 integrated from t to T.
- $P_0(t,x,y)$ depends on the curve $y \equiv (y^s, t \leq s \leq T)$ only through v.
- P_{BS} is solution to the PDE

$$\partial_v P_{BS} = \frac{1}{2}\left(\partial_x^2 - \partial_x\right) P_{BS}, \qquad P_{BS}(x,0) = g(x) \tag{10}$$

Links the vega and gamma of a vanilla option in the unperturbed state.

The price at order 0

An important observation:

- Because L_0 incorporates no local volatility, L_0 and ∂_x commute so $(\partial_t + L_0)\,\partial_x^p P_0 = \partial_x^p\,(\partial_t + L_0)\,P_0 = 0$.
- $\Rightarrow \partial_x^p P_{BS}\left(X_t^0, \int_t^T y^s ds\right) \equiv \partial_x^p P_0(t, X_t^0, y)$ is a martingale for all integer p.
- Equation (10) then shows that **for all integers m, n**, $\partial_v^m \partial_x^n P_{BS}\left(X_t^0, \int_t^T y^s ds\right)$ **is a martingale**.
- This is crucial in the computations of P_1 and P_2.

The price at order 1

- Let us define the **integrated spot-variance covariance** function $C_t^{X\xi}(y)$:

$$C_t^{X\xi}(y) = \int_t^T ds \int_s^T du\ \mu(s,u,y) = \int_t^T ds \int_s^T du\ \frac{\mathbb{E}\left[\frac{dS_s}{S_s} d\xi_s^u | \xi_s = y\right]}{ds}$$

- We then have

$$\begin{aligned}
P_1(t,x,y) &= \mathbb{E}\left[\int_t^T L_1 P_0(s, X_s^{0,t,x}, y) ds\right] \\
&= \mathbb{E}\left[\int_t^T ds \int_s^T du\ \mu(s,u,y)\ \partial_{y^u}\left(\partial_x P_{BS}\left(X_s^{0,t,x}, \int_s^T y^r dr\right)\right)\right] \\
&= \mathbb{E}\left[\int_t^T ds \int_s^T du\ \mu(s,u,y)\ \partial^2_{xv} P_{BS}\left(X_s^{0,t,x}, \int_s^T y^r dr\right)\right] \\
&= \int_t^T ds \int_s^T du\ \mu(s,u,y) \mathbb{E}\left[\ \partial^2_{xv} P_{BS}\left(X_s^{0,t,x}, \int_s^T y^r dr\right)\right] \\
&= C_t^{X\xi}(y)\ \partial^2_{xv} P_{BS}\left(x, \int_t^T y^r dr\right)
\end{aligned}$$

The price at order 2

A similar result holds for the second order correction:

$$
\begin{aligned}
P_2 &= P_2^{L_2 P_0} + P_2^{L_1 P_1} \\
P_2^{L_2 P_0}(t,x,y) &= \frac{1}{2} C_t^{\xi\xi}(y)\ \partial_v^2 P_{BS}\left(x, \int_t^T y^r\, dr\right) \\
P_2^{L_1 P_1} &= P_{2,0}^{L_1 P_1} + P_{2,1}^{L_1 P_1} \\
P_{2,0}^{L_1 P_1}(t,x,y) &= \frac{1}{2} C_t^{X\xi}(y)^2\ \partial_x^2 \partial_v^2 P_{BS}\left(x, \int_t^T y^r\, dr\right) \\
P_{2,0}^{L_1 P_1}(t,x,y) &= C_t^{\mu}(y)\ \partial_x^2 \partial_v P_{BS}\left(x, \int_t^T y^r\, dr\right)
\end{aligned}
$$

$$
\begin{aligned}
C_t^{\xi\xi}(y) &= \int_t^T ds \int_s^T du \int_s^T du'\ \nu(s,u,u',y) = \int_t^T ds \int_s^T du \int_s^T du'\ \frac{\mathbb{E}\left[d\xi_s^u d\xi_s^{u'} | \xi_s = y\right]}{ds} \\
C_t^{\mu}(y) &= \int_t^T ds \int_s^T du\ \mu(s,u,y)\ \partial_{y^u}\left(C_s^{X\xi}(y)\right)
\end{aligned}
$$

$C_t^{\xi\xi}(y)$: **integrated variance-variance covariance** function

Expansion of the implied volatility

- We write $C^{X\xi} = C_0^{X\xi}(y)$, $C^{\xi\xi} = C_0^{\xi\xi}(y)$ and $C^{\mu} = C_0^{\mu}(y)$.
- **In the general diffusive stochastic volatility model (8), at second order in the vol-of-vol ε, the implied volatility for maturity T and strike K is quadratic in $L = \ln\left(\frac{K}{S_0}\right)$:**

$$\widehat{\sigma}^{\varepsilon}(T,K) \;=\; \widehat{\sigma}_T^{\mathsf{ATM}} + \mathcal{S}_T \ln\left(\frac{K}{S_0}\right) + \mathcal{C}_T \ln^2\left(\frac{K}{S_0}\right) + O(\varepsilon^3) \qquad (11)$$

- Coefficients are

$$\begin{aligned}
\widehat{\sigma}_T^{\mathsf{ATM}} &= \hat{\sigma}_T^{\mathsf{VS}}\left[1 + \frac{\varepsilon}{4v}C^{X\xi} + \frac{\varepsilon^2}{32v^3}\left(12\left(C^{X\xi}\right)^2 - v\,(v+4)\,C^{\xi\xi} + 4v\,(v-4)\,C^{\mu}\right)\right] \\
\mathcal{S}_T &= \hat{\sigma}_T^{\mathsf{VS}}\left[\frac{\varepsilon}{2v^2}C^{X\xi} + \frac{\varepsilon^2}{8v^3}\left(4C^{\mu}v - 3\left(C^{X\xi}\right)^2\right)\right] \\
\mathcal{C}_T &= \hat{\sigma}_T^{\mathsf{VS}}\frac{\varepsilon^2}{8v^4}\left(4C^{\mu}v + C^{\xi\xi}v - 6\left(C^{X\xi}\right)^2\right)
\end{aligned}$$

- $v = \int_0^T \xi_0^s ds$ and $\hat{\sigma}_T^{\mathsf{VS}} = \sqrt{\frac{v}{T}}$, the VS implied volatility for maturity T.

Comments

ATM implied volatility:

$$\widehat{\sigma}_T^{\mathrm{ATM}} = \hat{\sigma}_T^{\mathrm{VS}} \left[1 + \frac{\varepsilon}{4v} C^{X\xi} + \frac{\varepsilon^2}{32v^3}\left(12\left(C^{X\xi}\right)^2 - v(v+4)C^{\xi\xi} + 4v(v-4)C^{\mu}\right)\right]$$

- ATM implied volatility = variance swap volatility + spread. At first order, spread $= \frac{C^{X\xi}}{4\sqrt{vT}}\varepsilon$.
- Typically, on the equity market, $C^{X\xi} < 0$: ATM implied volatility lies below the variance swap volatility.
- When spot returns and forward variances are uncorrelated, $C^{X\xi} = C^{\mu} = 0$ so that

$$\widehat{\sigma}_T^{\mathrm{ATM}} = \hat{\sigma}_T^{\mathrm{VS}} \left(1 - \frac{\varepsilon^2}{32v^3} v(v+4) C^{\xi\xi}\right)$$

  Because $C^{\xi\xi} \geq 0$, ATM implied volatility lies again below variance swap volatility. The higher the volatility of variances, the smaller the ATM implied volatility.

Comments (continued)

ATM skew: $\mathcal{S}_T = \hat{\sigma}_T^{\mathsf{VS}} \left[\frac{\varepsilon}{2v^2} C^{X\xi} + \frac{\varepsilon^2}{8v^3} \left(4C^\mu v - 3 \left(C^{X\xi} \right)^2 \right) \right]$

- ATM skew $\mathcal{S}_T$ is of order ε. It has the sign of $C^{X\xi}$. $\mathcal{S}_T$ vanishes when spot returns and forward variances are uncorrelated, even at second order. ATM skew is produced only by the spot-variance correlation.
- Link ATM vol-VS vol-ATM skew:

$$\widehat{\sigma}_T^{\mathsf{ATM}} = \hat{\sigma}_T^{\mathsf{VS}} + \frac{\left(\hat{\sigma}_T^{\mathsf{VS}}\right)^2 T}{2} \mathcal{S}_T$$

- At first order in ε, ATM skew has same sign as the difference between ATM implied volatility and variance swap volatility.

ATM convexity: $\mathcal{C}_T = \hat{\sigma}_T^{\mathsf{VS}} \frac{\varepsilon^2}{8v^4} \left(4C^\mu v + C^{\xi\xi} v - 6 \left(C^{X\xi} \right)^2 \right)$

- Curvature $\mathcal{C}_T$ is of order ε^2.
- Not only does it involve variance/variance covariance: spot/variance covariance (squared) contributes as well.
- If spot and variances are uncorrelated, $\mathcal{C}_T = \frac{C^{\xi\xi}}{8v^{5/2}\sqrt{T}} \varepsilon^2 \geq 0$.

Another derivation which stays at the level of operators

- Recall that the price P^ε of the vanilla option is solution to $(\partial_t + L_t^\varepsilon) P^\varepsilon = 0$ with $L_t^\varepsilon = L_{0,t} + \varepsilon L_{1,t} + \varepsilon^2 L_{2,t}$, and terminal condition $P^\varepsilon(T, x, y) = g(x)$.
- Price can be expressed in terms of the semigroup $(U_{st}^\varepsilon, 0 \le s \le t \le T)$ attached to the family of differential operators L_t^ε: $P^\varepsilon(t, \cdot) = U_{tT}^\varepsilon g$.
- The semigroup is defined by

$$U_{st}^\varepsilon = \lim_{n\to\infty} (1 - \delta t L_{t_0}^\varepsilon)(1 - \delta t L_{t_1}^\varepsilon) \cdots \left(1 - \delta t L_{t_{n-1}}^\varepsilon\right), \quad \delta t = \frac{t-s}{n}, \quad t_i = s + i\delta t$$

- It satisfies $U_{rt}^\varepsilon = U_{rs}^\varepsilon U_{st}^\varepsilon$ for $0 \le r \le s \le t \le T$, hence the notation $:\exp\left(\int_s^t L_\tau^\varepsilon d\tau\right):$, where :: denotes time ordering.
- **We can directly expand U_{st}^ε in powers of ε.** Usual time-dependent perturbation technique in quantum mechanics. U_{st}^0 is called the free propagator.

- Consider the general situation where a differential operator L_t is perturbed by another operator H_t: $L_t^\varepsilon = L_t + \varepsilon H_t$
- From the definition of the semigroup, $U_{st}^\varepsilon = U_{st}^{(0)} + \varepsilon U_{st}^{(1)} + \varepsilon^2 U_{st}^{(2)} + \cdots$ with

$$\begin{aligned} U_{st}^{(1)} &= \int_s^t d\tau\ U_{s\tau}^0 H_\tau U_{\tau t}^0 \\ U_{st}^{(2)} &= \int_s^t d\tau_1 \int_{\tau_1}^t d\tau_2\ U_{s\tau_1}^0 H_{\tau_1} U_{\tau_1 \tau_2}^0 H_{t_2} U_{\tau_2 t}^0 \end{aligned}$$

- $\Rightarrow P^\varepsilon = P_0 + \varepsilon P_1 + \varepsilon^2 P_2 + \cdots$, with

$$\begin{aligned} P_1 &= \int_t^T d\tau\ U_{t\tau}^0 L_{1,\tau} U_{\tau T}^0 g \\ P_2 &= \int_t^T d\tau\ U_{t\tau}^0 L_{2,\tau} U_{\tau T}^0 g + \int_t^T d\tau_1 \int_{\tau_1}^T d\tau_2\ U_{t\tau_1}^0 L_{1,\tau_1} U_{\tau_1 \tau_2}^0 L_{1,\tau_2} U_{\tau_2 T}^0 g \end{aligned}$$

- We recover the expressions of P_1 and P_2.

Short maturity

- Assume $d\xi_t^t = \cdots dt + \varepsilon(\xi_t^t)^{\varphi} dB_t$
- Let ρ_{SV} be the correlation between S_t and instantaneous variance $V_t = \xi_t^t$
- Heston: $\varphi = \frac{1}{2}$, $\rho_{SV} = \rho$;
  Bergomi: $\varphi = 1$, $\rho_{SV} = \alpha_\theta \left((1-\theta)\rho_{SX} + \theta\rho_{SY}\right)$
- Then for short maturities

$$\mathcal{S}_0 \simeq \frac{\varepsilon}{4}\rho\left(\widehat{\sigma}^{\mathrm{ATM}}\right)^{2\varphi-2} \quad (12)$$

$$\mathcal{C}_0 \simeq \varepsilon^2\left(\left(\frac{1}{12}\varphi - \frac{7}{48}\right)\rho^2 + \frac{1}{24}\right)\left(\widehat{\sigma}^{\mathrm{ATM}}\right)^{4\varphi-5} \quad (13)$$

- $\Rightarrow$ Short-term ATM skew does not depend on short-term ATM volatility iff $\varphi = 1$ (observed in equity markets)
- $\Rightarrow$ Short-term ATM convexity does not depend on short-term ATM volatility iff $\varphi = \frac{5}{4}$. And $(\forall \rho_{SV},\, \mathcal{C}_0 \geq 0) \iff \varphi \geq \frac{5}{4}$

Long-term asymptotics of implied volatility

- Assume the term-structure of variance swaps volatilities is flat: $\xi_0^t \equiv \xi$.
- Assume that for large $u-t$, $\mu(t,u,y) \propto (u-t)^{-\alpha}$, $\alpha > 0$.
  Then at higher order in ε, for long maturities,

$$\begin{aligned} \mathcal{S}_T &\propto T^{-\alpha} \qquad \text{if } \alpha < 1 \\ \mathcal{S}_T &\propto T^{-1} \qquad \text{if } \alpha > 1 \end{aligned}$$

  α is exactly a signature of the long-time decay of the spot/variance covariance function.
- Assume that for large $u-t$ and $u'-t$,
  $\nu(t,u,u',y) \propto (u-t)^{-\beta}(u'-t)^{-\beta}$, $\beta > 0$.
  Also assume that spots and volatilities are uncorrelated ($\mu \equiv 0$). Then at higher order in ε, for long maturities,

$$\begin{aligned} \mathcal{C}_T &\propto T^{-2\beta} \qquad \text{if } \beta < 1 \\ \mathcal{C}_T &\propto T^{-2} \qquad \text{if } \beta > 1 \end{aligned}$$

- Exponential decay $\leftrightarrow \beta > 1$.

First example: a Heston-like model

$$\begin{aligned} dX_t &= -\frac{1}{2}V_t dt + \sqrt{V_t}dW_t^1, \qquad X_0 = x \qquad (14) \\ dV_t &= -k\left(V_t - V_\infty\right)dt + \lambda\left(V_t\right)^\varphi\left(\rho dW_t^1 + \sqrt{1-\rho^2}dW_t^2\right), \qquad V_0 = V \end{aligned}$$

- The instantaneous forward variance reads

$$\xi_t^u = \mathbb{E}\left[V_u | V_t\right] = V_\infty + \left(V_t - V_\infty\right)e^{-k(u-t)}$$

and its dynamics is:

$$d\xi_t^u = \lambda e^{-k(u-t)}\left(\xi_t^t\right)^\varphi\left(\rho dW_t^1 + \sqrt{1-\rho^2}dW_t^2\right)$$

- The initial term-structure of instantaneous forward variances is

$$y^u \equiv \xi_0^u = v_\infty + \left(v - v_\infty\right)e^{-ku}$$

- Like **in all classic "first generation" stochastic volatility models, this term-structure is determined by the model parameters**, and the current value of the instantaneous volatility.

- The volatility $\lambda(t,u,y)$ of instantaneous forward variances depends on the instantaneous forward variance curve $y = (y^s, t \le s \le T)$ only through the instantaneous spot variance y^t:

$$\begin{aligned}
\lambda_1(t,u,y) &= \rho \left(y^t\right)^{\varphi} e^{-k(u-t)} \\
\lambda_2(t,u,y) &= \sqrt{1-\rho^2} \left(y^t\right)^{\varphi} e^{-k(u-t)}
\end{aligned}$$

- As a consequence,

$$\begin{aligned}
C^{X\xi} &= \frac{\rho}{k} \int_0^T ds \ \left(y^s\right)^{\varphi+\frac{1}{2}} \left(1 - e^{-k(T-s)}\right) \\
C^{\xi\xi} &= \sum_{i=1}^{2} \int_0^T ds \left(\int_s^T du \, \lambda_i(s,u,y) \right)^2 = \frac{1}{k^2} \int_0^T ds \ \left(y^s\right)^{2\varphi} \left(1 - e^{-k(T-s)}\right)^2 \\
C^{\mu} &= \left(\varphi + \frac{1}{2}\right) \frac{\rho^2}{k} \int_0^T ds \ \left(y^s\right)^{\varphi+\frac{1}{2}} \int_s^T du \ \left(y^u\right)^{\varphi-\frac{1}{2}} e^{-k(u-s)} \left(1 - e^{-k(T-u)}\right)
\end{aligned}$$

- This coincides with Equations (3.7) to (3.10) in Lewis [5], where $J^{(1)} = C^{X\xi}$, $J^{(3)} = \frac{1}{2} C^{\xi\xi}$, and $J^{(4)} = C^{\mu}$

Second example: the Bergomi model

$$\begin{aligned} dx_t &= -\frac{1}{2}\xi_t^t dt + \sqrt{\xi_t^t} dW_t^S \\ d\xi_t^u &= \xi_t^u \alpha_\theta \omega \left((1-\theta)\, e^{-k_X(u-t)} dW_t^X + \theta e^{-k_Y(u-t)} dW_t^Y \right) \\ &= \lambda(t,u,\xi_t^\cdot) \cdot dW_t \end{aligned}$$

$d\langle W^S, W^X\rangle_t = \rho_{SX} dt$, $d\langle W^S, W^Y\rangle_t = \rho_{SY} dt$, $d\langle W^X, W^Y\rangle_t = \rho_{XY} dt$.

- The normalizing factor

$$\alpha_\theta = \left((1-\theta)^2 + 2\rho_{XY}\theta\,(1-\theta) + \theta^2 \right)^{-1/2}$$

  is such that the very-short term variance $\xi_t^{t,\omega}$ has log-normal volatility ω.

- We pick $k_X > k_Y$, θ is a parameter which mixes the short-term factor W^X and the long-term factor W^Y.

- After a Cholesky transform, this can be restated using independent Brownian motions W^1, W^2 and W^3 as follows:

$$
\begin{aligned}
W^S &= W^1 \\
W^X &= \rho_{SX} W^1 + \sqrt{1-\rho_{SX}^2}\, W^2 \\
W^Y &= \rho_{SY} W^1 + \chi_{XY}\sqrt{1-\rho_{SY}^2}\, W^2 + \sqrt{\left(1-\chi_{XY}^2\right)\left(1-\rho_{SY}^2\right)}\, W^3
\end{aligned}
$$

where $\chi_{XY} = \frac{\rho_{XY}-\rho_{SX}\rho_{SY}}{\sqrt{1-\rho_{SX}^2}\sqrt{1-\rho_{SY}^2}}$

- ρ_{SX}, ρ_{SY} and ρ_{XY} define a correlation matrix $\iff \chi_{XY} \in [-1, 1]$.
- The volatility of variance $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ reads

$$
\begin{aligned}
\lambda_1(t,u,y) &= y^u \alpha_\theta \left((1-\theta)\,\rho_{SX} e^{-k_X(u-t)} + \theta \rho_{SY} e^{-k_Y(u-t)} \right) \\
\lambda_2(t,u,y) &= y^u \alpha_\theta \left((1-\theta)\sqrt{1-\rho_{SX}^2}\, e^{-k_X(u-t)} + \theta \chi_{XY} \sqrt{1-\rho_{SY}^2}\, e^{-k_Y(u-t)} \right) \\
\lambda_3(t,u,y) &= y^u \alpha_\theta \theta \sqrt{\left(1-\chi_{XY}^2\right)\left(1-\rho_{SY}^2\right)}\, e^{-k_Y(u-t)}
\end{aligned}
$$

- We write $\lambda_i(t,u,y) = y^u \alpha_\theta \left(w_{iX} e^{-k_X(u-t)} + w_{iY} e^{-k_Y(u-t)} \right)$

The covariance functions read

$$
\begin{aligned}
C^{X\xi} &= \int_0^T du \int_0^u dt \sqrt{y^t}\lambda_1(t,u,y) \\
&= \alpha_\theta\,(1-\theta)\,\rho_{SX}\int_0^T du\ y^u \int_0^u dt\sqrt{y^t}e^{-k_X(u-t)} \\
&\quad +\alpha_\theta\theta\rho_{SY}\int_0^T du\ y^u\int_0^u dt\sqrt{y^t}e^{-k_Y(u-t)} \\
C^{\xi\xi} &= \sum_{i=1}^{3}\int_0^T ds\left(\int_s^T du\,\lambda_i(s,u,y)\right)^2 \\
&= \alpha_\theta^2\sum_{i=1}^{3}\int_0^T ds\left(w_{iX}\int_s^T du\ y^u e^{-k_X(u-s)} + w_{iY}\int_s^T du\ y^u e^{-k_Y(u-s)}\right)^2 \\
C^{\mu} &= \int_0^T ds\int_s^T du\,\sqrt{y^s}\lambda_1(s,u,y^u)\left(\frac{1}{2\sqrt{y^u}}\int_u^T dt\lambda_1(u,t,y^t)\right. \\
&\quad \left. +\int_s^u dr\sqrt{y^r}\frac{\partial\lambda_1}{\partial z}(r,u,z)|_{z=y^u}\right)
\end{aligned}
$$

In the case of a flat initial term structure of variance swaps ($y_0^t \equiv \xi$), this reads

$$\begin{aligned}
C^{x\xi} &= \alpha_\theta \omega \xi^{3/2} T^2 \left(w_{1X} \mathcal{J}(k_X T) + w_{1Y} \mathcal{J}(k_Y T) \right) \\
C^{\xi\xi} &= \alpha_\theta^2 \omega^2 \xi^2 T^3 \left(w_0 + w_X \mathcal{I}(k_X T) + w_Y \mathcal{I}(k_Y T) \right. \\
&\quad \left. + w_{XX} \mathcal{I}(2k_X T) + w_{YY} \mathcal{I}(2k_Y T) + w_{XY} \mathcal{I}((k_X + k_Y) T) \right) \\
C^{\mu} &= \alpha_\theta^2 \omega^2 \xi^2 T^3 \left(C_1^\mu + C_2^\mu \right)
\end{aligned}$$

with

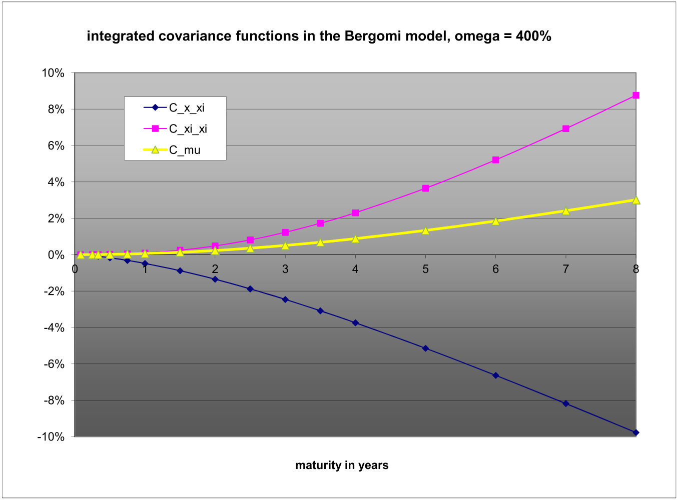
$$\begin{aligned}
\mathcal{I}(\alpha) &= \frac{1 - e^{-\alpha}}{\alpha}, \quad \mathcal{J}(\alpha) = \frac{\alpha - 1 + e^{-\alpha}}{\alpha^2} \\
\mathcal{K}(\alpha) &= \frac{1 - e^{-\alpha} - \alpha e^{-\alpha}}{\alpha^2}, \quad \mathcal{H}(\alpha) = \frac{\mathcal{J}(\alpha) - \mathcal{K}(\alpha)}{\alpha} \\
w_0 &= \sum_{i=1}^{3} \left(\frac{w_{iX}}{k_X T} + \frac{w_{iY}}{k_Y T} \right)^2, \quad w_X = -2 \sum_{i=1}^{3} \frac{w_{iX}}{k_X T} \left(\frac{w_{iX}}{k_X T} + \frac{w_{iY}}{k_Y T} \right) \\
w_Y &= -2 \sum_{i=1}^{3} \frac{w_{iY}}{k_Y T} \left(\frac{w_{iX}}{k_X T} + \frac{w_{iY}}{k_Y T} \right), \\
w_{XX} &= \sum_{i=1}^{3} \frac{w_{iX}^2}{k_X^2 T^2}, \quad w_{YY} = \sum_{i=1}^{3} \frac{w_{iY}^2}{k_Y^2 T^2}, \quad w_{XY} = 2 \sum_{i=1}^{3} \frac{w_{iX} w_{iY}}{k_X k_Y T^2}
\end{aligned}$$

and

$$\begin{aligned}
C_1^\mu &= \frac{1}{2}w_{1X}^2\mathcal{H}(k_X T) + \frac{1}{2}w_{1Y}^2\mathcal{H}(k_Y T) - w_{1X}w_{1Y}\frac{\mathcal{J}(k_Y T) - \mathcal{J}(k_X T)}{(k_Y - k_X)T} \\
C_2^\mu &= w''_X\mathcal{J}(k_X T) + w''_Y\mathcal{J}(k_Y T) + w''_{XX}\mathcal{J}(2k_X T) + w''_{YY}\mathcal{J}(2k_Y T) \\
&\quad + w''_{XY}\mathcal{J}((k_X + k_Y)T)
\end{aligned}$$

with

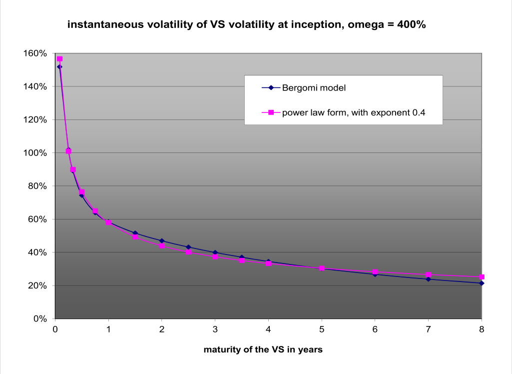
$$\begin{aligned}
w''_X &= \frac{w_{1X}^2}{k_X T} + \frac{w_{1X}w_{1Y}}{k_Y T}, \qquad w''_Y = \frac{w_{1Y}^2}{k_Y T} + \frac{w_{1X}w_{1Y}}{k_X T} \\
w''_{XX} &= -\frac{w_{1X}^2}{k_X T}, \qquad w''_{YY} = -\frac{w_{1Y}^2}{k_Y T}, \qquad w''_{XY} = -\frac{w_{1X}w_{1Y}}{k_X T} - \frac{w_{1X}w_{1Y}}{k_Y T}
\end{aligned}$$


integrated covariance functions in the Bergomi model, omega = 400%
10%
8%
6%
4%
2%
0%
-2%
-4%
-6%
-8%
-10%
0
1
2
3
4
5
6
7
8
C_x_xi
C_xi_xi
C_mu
maturity in years

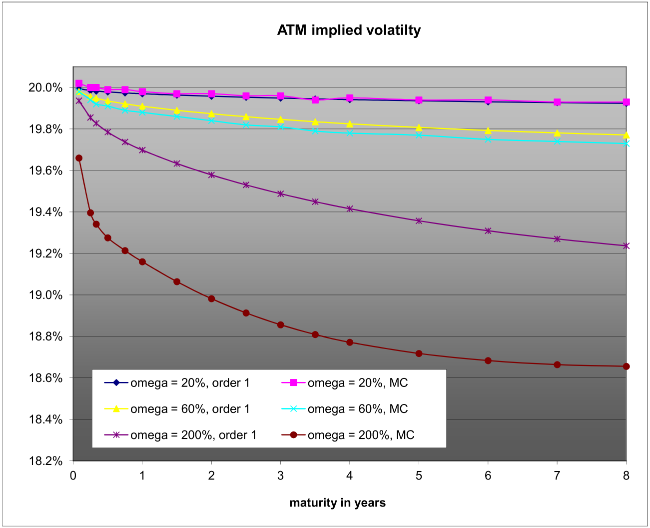
Numerical experiments

We pick the Bergomi model with a flat initial term structure of variance swap prices and

θ	k_X	k_Y	ρ_{SX}	ρ_{SY}	ρ_{XY}	χ_{XY}	ξ
0.25	8	0.35	-0.8	-0.48	0	-0.73	$(0.2)^2$

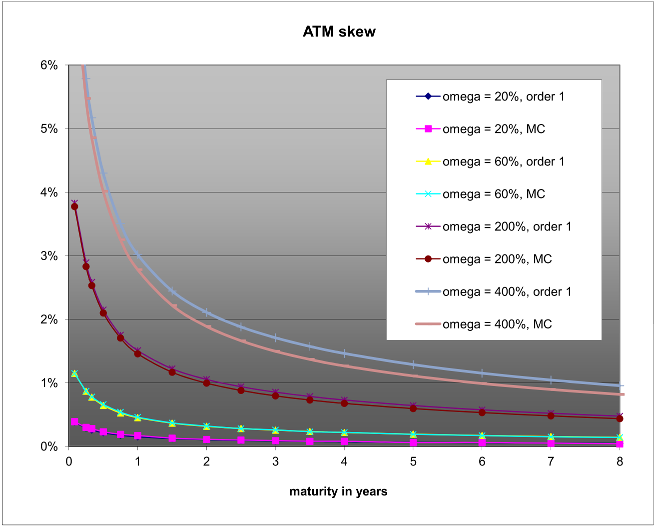


First order

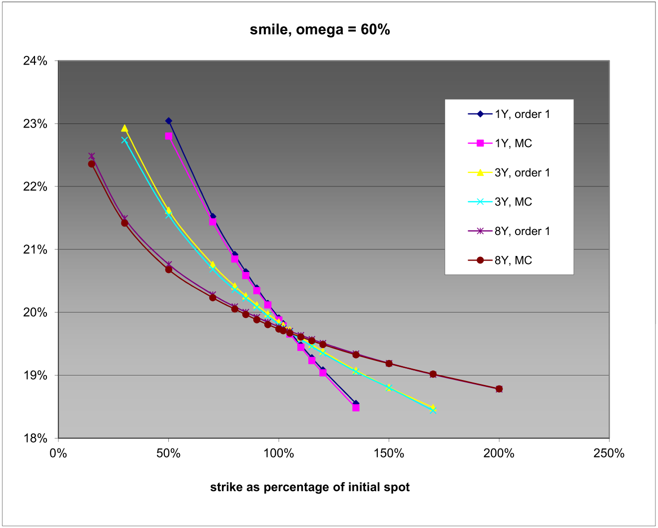


ATM implied volatilty
20.0%
19.8%
19.6%
19.4%
19.2%
19.0%
18.8%
18.6%
18.4%
18.2%
0
1
2
3
4
5
6
7
8
maturity in years
omega = 20%, order 1
omega = 20%, MC
omega = 60%, order 1
omega = 60%, MC
omega = 200%, order 1
omega = 200%, MC

First order

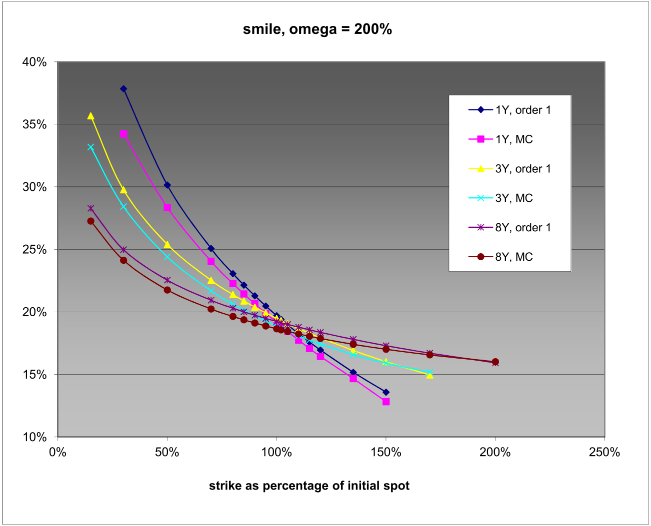

ATM skew
6%
5%
4%
3%
2%
1%
0%
0
1
2
3
4
5
6
7
8
maturity in years
omega = 20%, order 1
omega = 20%, MC
omega = 60%, order 1
omega = 60%, MC
omega = 200%, order 1
omega = 200%, MC
omega = 400%, order 1
omega = 400%, MC

First order



smile, omega = 60%
24%
23%
22%
21%
20%
19%
18%
0%
50%
100%
150%
200%
250%
strike as percentage of initial spot
1Y, order 1
1Y, MC
3Y, order 1
3Y, MC
8Y, order 1
8Y, MC

First order


smile, omega = 200%
40%
35%
30%
25%
20%
15%
10%
0%
50%
100%
150%
200%
250%
strike as percentage of initial spot
1Y, order 1
1Y, MC
3Y, order 1
3Y, MC
8Y, order 1
8Y, MC

First order

- **ATM skew very sharply estimated by the first order expansion**, even for large values of the volatility of variance ω.
- **ATM volatility well captured by the expansion at first order in ω only for small values of ω** (say, up to 60%).
- True ATM implied volatilities are below their first order approximates $\Rightarrow$ ATM volatility is a very concave function of ω, around $\omega = 0$. In view of the expression for $\widehat{\sigma}_T^{\text{ATM}}$, this means that, for the set of parameters picked,

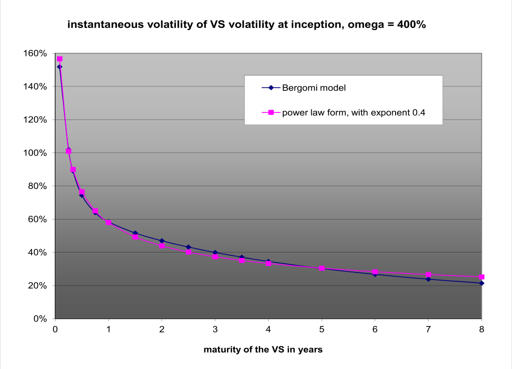
$$12 C^{X\xi 2} - C^{\xi\xi} v (v+4) + 4 C^{\mu} v (v-4) \leq 0$$

- **Global shape of smile well captured by first order expansion**: the true implied volatility for strike K is indeed approximately affine in $\ln(K/S_0)$.
- But level of smile well captured only for small values of ω.

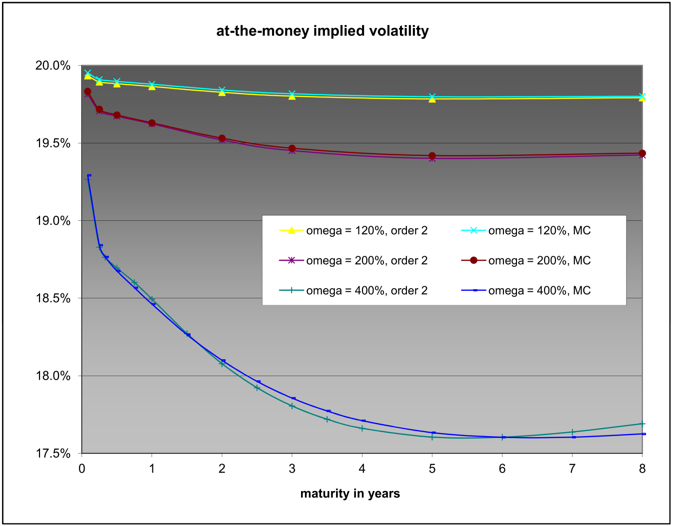
Second order

We first consider the situation when spot returns and forward variances are uncorrelated. In this case, the ATM skew vanishes, and so does its expansion at second order in ω. We pick

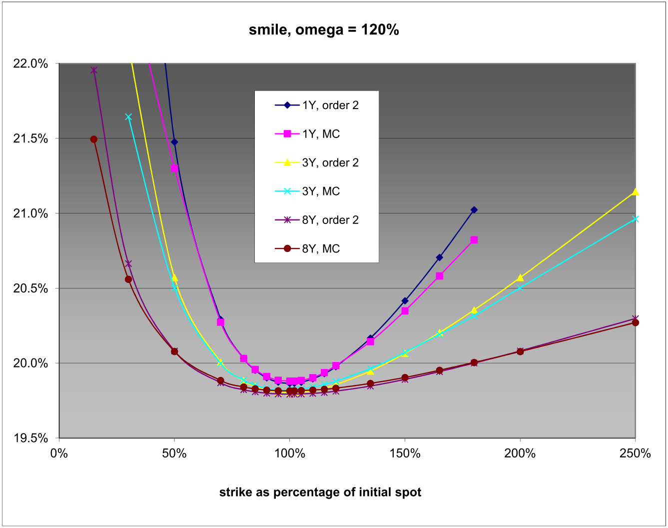
θ	k_X	k_Y	ρ_{SX}	ρ_{SY}	ρ_{XY}	ξ
0.25	8	0.35	0	0	0	$(0.2)^2$



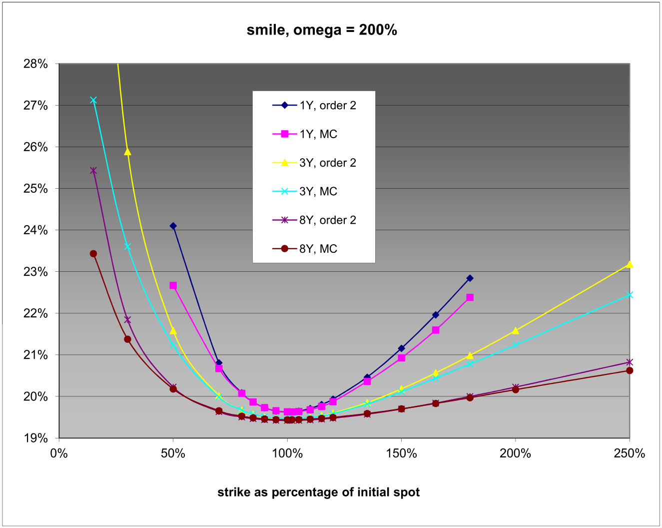
Second order



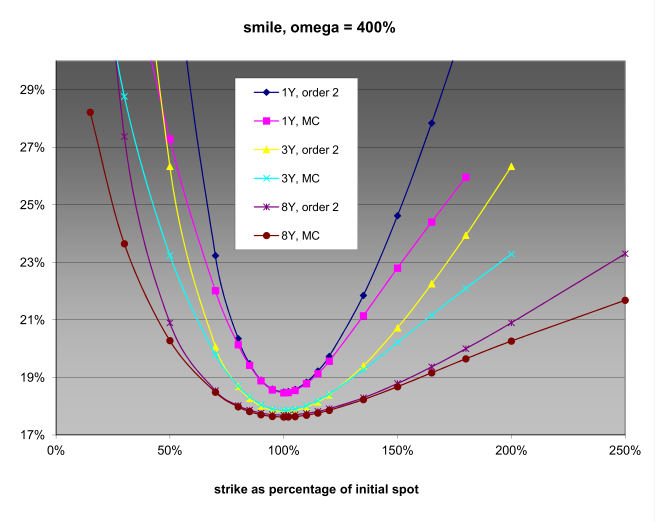
Second order



Second order



Second order

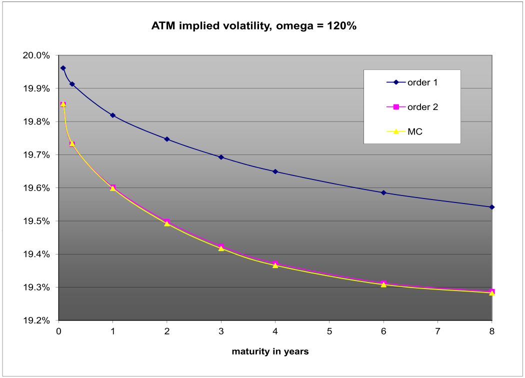


Second order

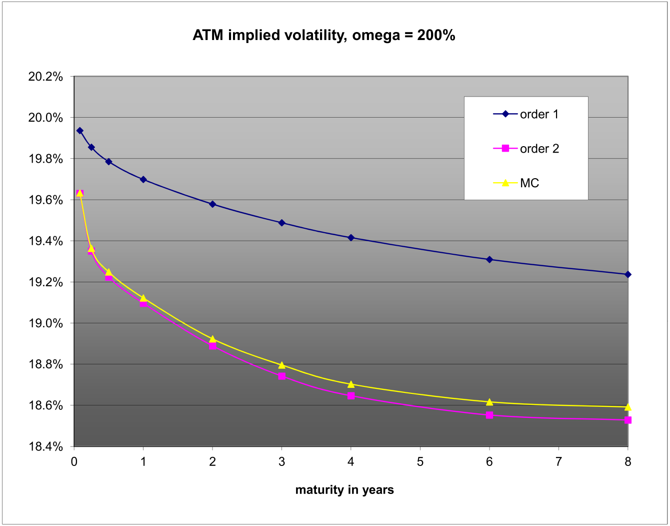
- **ATM implied volatility very sharply estimated by the second order expansion**, even up to $\omega = 400\%$ and to long maturities. For $T = 15$ years, estimate is less than 15 bps above true ATM volatility.
- Looking at the whole smile: second order expansion of the implied volatility is excellent around the money, but becomes too large for strikes far from the money.
- Not surprising: No arbitrage $\Rightarrow$ for very small and very large strikes, $\widehat{\sigma}(T,K)^2$ grows at most linearly with $\ln(K/S_0)$ (see Lee [4]), whereas second order estimate for $\widehat{\sigma}(T,K)^2$ grows like $\ln^4(K/S_0)$, see (11). Remainder $O(\omega^3) = R(\omega, T, K)$ is large for large K, for finite ω.
- Nevertheless, even for $\omega = 400\%$, a maturity of 8 years and an out-the-money strike of 250%, the error is only 1.5 point of volatility.

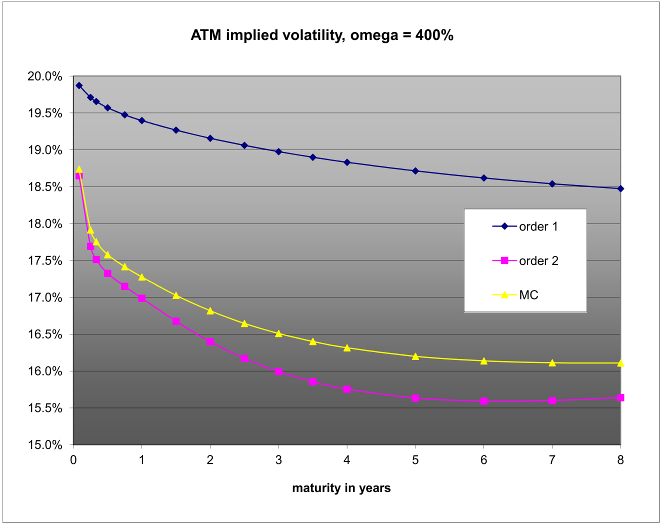
Second order

We now check numerically the accuracy of the second order expansion of the smile in the general case of correlated spot returns and variances.

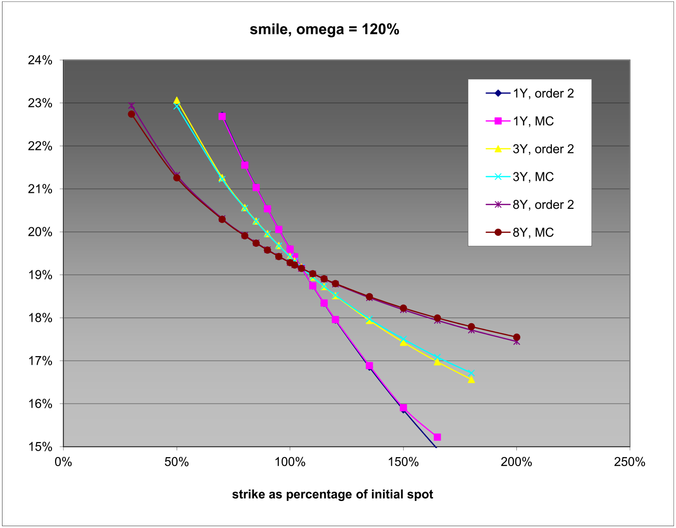


Second order

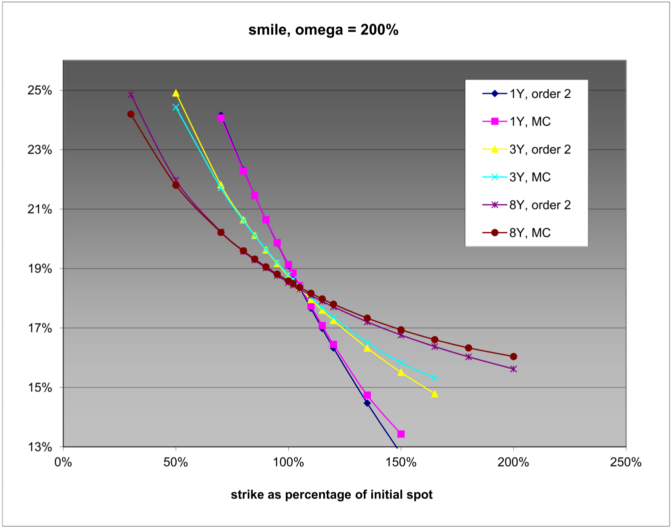



ATM implied volatility, omega = 400%
20.0%
19.5%
19.0%
18.5%
18.0%
17.5%
17.0%
16.5%
16.0%
15.5%
15.0%
0
1
2
3
4
5
6
7
8
maturity in years
order 1
order 2
MC

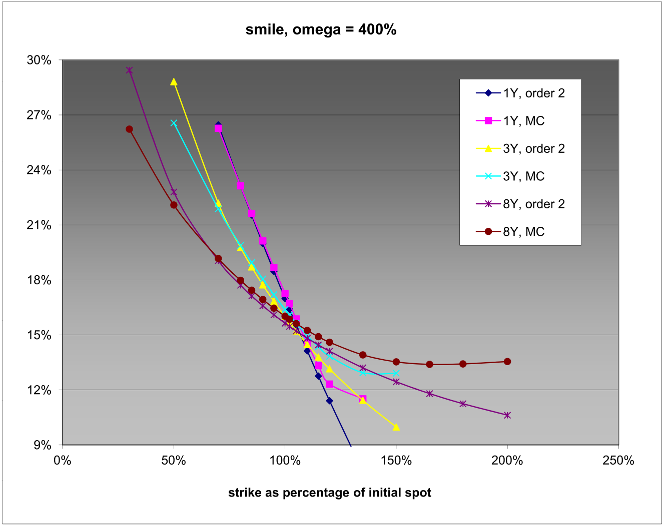
Second order



Second order



Second order



- Remember $\mathcal{S}_T = \frac{C^{X\xi}}{2v^{3/2}\sqrt{T}}\varepsilon + O\left(\varepsilon^2\right)$
- Let us now compute the skewness s_T of log-returns:

$$s_T = \frac{\mathbb{E}\left[\mathcal{X}_T^3\right]}{\mathbb{E}\left[\mathcal{X}_T^2\right]^{3/2}}, \qquad \mathcal{X}_T = X_T - \mathbb{E}\left[X_T\right] = \int_0^T \sqrt{\xi_t^{t,\varepsilon}} dW_t^1$$

- We have $\mathbb{E}\left[\mathcal{X}_T^2\right] = \int_0^T \xi_0^t dt + O(\varepsilon)$ and

$$\mathbb{E}\left[\mathcal{X}_T^3\right] = 3\varepsilon C^{X\xi} + O\left(\varepsilon^2\right)$$

- At first order in the vol-of-vol, the skewness of (the distribution of) $\ln\left(S_T/S_0\right)$ is thus

$$s_T = \frac{3\varepsilon C^{X\xi}}{\left(\int_0^T \xi_0^t dt\right)^{3/2}}$$

- The ATM skew $\mathcal{S}_T$ simply reads

$$\mathcal{S}_T = \frac{s_T}{6\sqrt{T}} + O(\varepsilon^2)$$

Conclusion

- We provide an expansion at order two in volatility-of-volatility for **general stochastic volatility models** based on a forward variance formulation.
- VS volatilities for all maturities are unchanged as ε is varied.
- At order two in ε, **the smile is exactly quadratic in log-moneyness** and **depends on only three model-dependent dimensionless quantities**:
  - $C^{x\xi}$, the integrated spot/variance covariance function,
  - $C^{\xi\xi}$, the integrated variance/variance covariance function,
  - C^{μ}, which, like $C^{x\xi}$, depends only on instantaneous spot/variance covariances.
- We shed light on the significance of $C^{x\xi}$ by establishing a simple link between the ATM skew and the skewness of $\ln S_T$.

- From our general expression we derive the short-maturity limits of ATM volatility, skew, curvature: we give structural dependencies of the ATM skew and curvature on ATM volatility.
- We also link the long-term decay of the ATM skew and curvature to the decay of spot/variance and variance/variance covariance functions.
- Numerical experiments in the case of a two-factor version of the Bergomi model show good agreement of the order one expression for the ATM skew, and of the order two expression for the ATM volatility, for values of the volatility of short-dated variance (around 400%) that are typical of implied levels of equity indices.

- Bergomi L., *Smile Dynamics 2*, Risk Magazine, pages 67-73, October 2005.
- Bergomi L., *Smile Dynamics 4*, Risk Magazine, December 2009.
- Backus D., Foresi S., Li K. and Wu L., *Accounting for Biases in Black-Scholes*, unpublished.
- Lee R., *The moment formula for implied volatility at extreme strikes*, Stanford University and Courant Institute, 2002.
- Lewis A., *Option valuation under stochastic volatility*, Finance Press, 2000.

Contents

Stochastic Local Volatility (SLV) models

- SV models produce a smile of implied volatilities (see the seminal papers by Hull-White, Renault-Touzi; see also Bergomi-Guyon for an analysis of the smile in general multi-factor second-generation SV models).
- Since they have a finite number of parameters (typically, 3 to 10), SV models cannot be perfectly calibrated to full (inter- and extrapolated) market smiles, indexed by all strikes and maturities (until some final maturity T).
- Now, vanilla options often provide a good hedge of exotic options. Dynamic or static trading of vanilla options often results in reducing the variance of the final P&L. In those cases, it is important that the SV model incorporates the correct initial prices of the hedging instruments—the vanilla options.

Stochastic Local Volatility (SLV) models

- How do we build an SV model that calibrates exactly to a full surface of implied volatilities?
- Due to the double infinity of constraints, indexed by strikes and maturities, one needs to introduce a double infinity of parameters.
- A natural way to do so, and a common practice in the foreign exchange market, is to embed a local volatility $\sigma(t, S)$ (also called "leverage function") into the SV model:

$$dS_t = a_t S_t \sigma(t, S_t)\, dW_t \tag{15}$$

  We speak of stochastic local volatility (SLV) models.
- The main issue we address in these lecture notes is how to build the local volatility $\sigma(t, S)$ to ensure that the market smile is exactly calibrated.
- Note that this local volatility function differs from Dupire's local volatility. For instance, if the smile produced by the "naked" SV model (3) is close to the market smile, one expects the calibrated local volatility $\sigma(t, S)$ to be uniformly close to 1.

Calibration of SLV models to market smiles

- SLV model (deterministic rate and div yield):

$$\frac{dS_t}{S_t} = (r_t - q_t)dt + \sigma(t, S_t)a_t dW_t$$

- Model exactly calibrated to market smile iff for all t, K

$$\sigma(t,K)^2 \mathbb{E}[a_t^2 | S_t = K] = \sigma_{\mathsf{Dup}}(t,K)^2$$

- $\Rightarrow$ **Nonlinear** SDE:

$$\frac{dS_t}{S_t} = (r_t - q_t)dt + \frac{\sigma_{\mathsf{Dup}}(t, S_t)}{\sqrt{\mathbb{E}[a_t^2 | S_t]}} a_t dW_t \tag{16}$$

- The local volatility function depends on the joint pdf $p\,(t, S, a)$ of (S_t, a_t):

$$\sigma(t, S, p) = \sigma_{\mathsf{Dup}}(t, S) \sqrt{\frac{\int p(t, S, a')da'}{\int a'^2 p(t, S, a')da'}}$$

- Equation (16) is an example of **McKean SDE**

Contents

The calibration condition

Proposition. Let us consider the following dynamics for an asset S, where the volatility α_t, the interest rate r_t, and the repo q_t, inclusive of the dividend yield, are all stochastic processes:

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \alpha_t\,dW_t \tag{17}$$

Model (17) is exactly calibrated to the market smile of S if and only if

$$\begin{aligned}\frac{\mathbb{E}[D_{0t}\alpha_t^2|S_t = K]}{\mathbb{E}[D_{0t}|S_t = K]} &= \sigma_{\mathsf{Dup}}(t,K)^2 \\ &- \frac{\mathbb{E}\left[D_{0t}\left(r_t - q_t - (r_t^0 - q_t^0)\right)\mathbf{1}_{S_t > K}\right]}{\frac{1}{2}K\partial_K^2 C_{\mathsf{mkt}}(t,K)} + \frac{\mathbb{E}\left[D_{0t}\left(q_t - q_t^0\right)(S_t - K)^+\right]}{\frac{1}{2}K^2\partial_K^2 C_{\mathsf{mkt}}(t,K)}\end{aligned} \tag{18}$$

for all (t,K), where $D_{0t} = \exp\left(-\int_0^t r_s ds\right)$ is the discount factor, r_t^0 and q_t^0 are deterministic rates and repos, and

$$\sigma_{\mathsf{Dup}}(t,K)^2 = \frac{\partial_t C_{\mathsf{mkt}}(t,K) + (r_t^0 - q_t^0)K\partial_K C_{\mathsf{mkt}}(t,K) + q_t^0 C_{\mathsf{mkt}}(t,K)}{\frac{1}{2}K^2\partial_K^2 C_{\mathsf{mkt}}(t,K)} \tag{19}$$

with $C_{\mathsf{mkt}}(t,K)$ the market price of the call option on S with strike K and maturity t

Case where rates and repo are deterministic

Proposition. In the case where rates r_t and repo/dividend yield q_t are deterministic, Model (17) is exactly calibrated to the market smile of S if and only if

$$\mathbb{E}[\alpha_t^2|S_t = K] = \sigma_{\mathsf{Dup}}(t,K)^2$$

for all (t, K), where

$$\sigma_{\mathsf{Dup}}(t,K)^2 = \frac{\partial_t C_{\mathsf{mkt}}(t,K) + (r_t - q_t)K\partial_K C_{\mathsf{mkt}}(t,K) + q_t C_{\mathsf{mkt}}(t,K)}{\frac{1}{2}K^2\partial_K^2 C_{\mathsf{mkt}}(t,K)}. \quad (20)$$

When $\alpha_t = \sigma(t, S_t)a_t$ the calibration condition reads

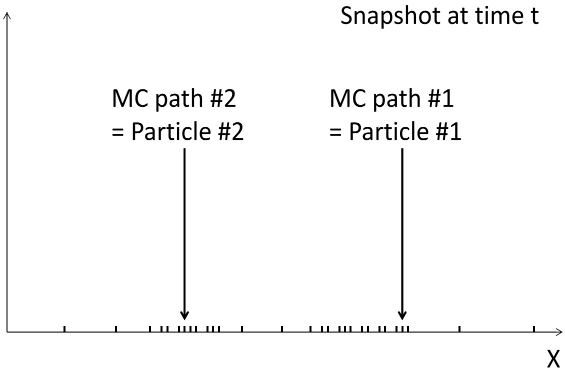
$$\sigma_{\mathrm{loc}}(t,K)^2 = \sigma(t,K)^2\mathbb{E}[a_t^2|S_t = K]$$

How to simulate a McKean SDE?

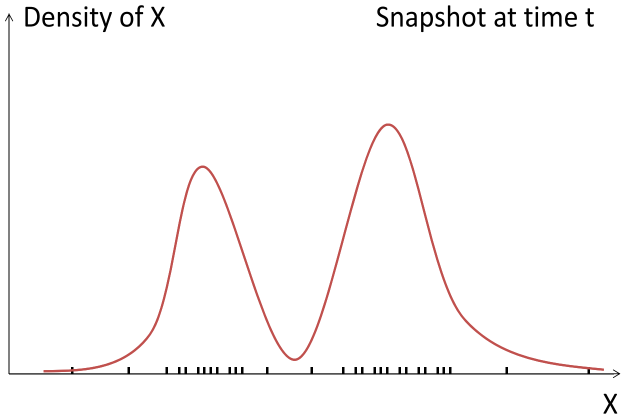
$$dX_t = b(t, X_t, \mathbb{P}_t)\ dt + \sigma(t, X_t, \mathbb{P}_t) \cdot dW_t, \qquad \mathbb{P}_t = \text{Law}\,(X_t)$$

How to simulate a McKean SDE?

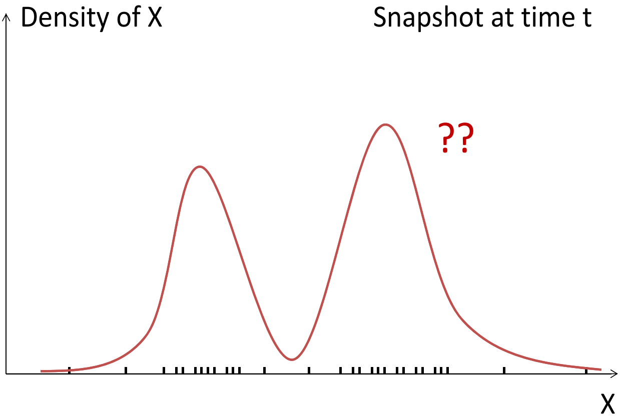
$$dX_t = b(t, X_t, \mathbb{P}_t)\ dt + \sigma(t, X_t, \mathbb{P}_t) \cdot dW_t, \qquad \mathbb{P}_t = \text{Law}\,(X_t)$$



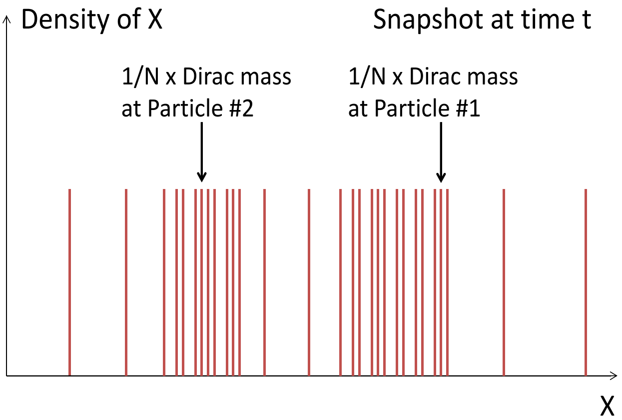
How to simulate a McKean SDE?



How to simulate a McKean SDE?



How to simulate a McKean SDE?



Particle algorithm

- Principle: replace the law $\mathbb{P}_t$ by the empirical distribution

$$\mathbb{P}_t^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^{i,N}}$$

where the $X_t^{i,N}$ are solution to the $(\mathbb{R}^n)^N$-dimensional **classical** SDE

$$dX_t^{i,N} = b\left(t, X_t^{i,N}, \mathbb{P}_t^N\right)\, dt + \sigma\left(t, X_t^{i,N}, \mathbb{P}_t^N\right)\cdot dW_t^i, \qquad \mathrm{Law}\left(X_0^{i,N}\right) = \mathbb{P}_0$$

- $\{X_t^i\}_{1\le i\le N}$ = system of N **interacting particles**. The interaction with the other $N-1$ particles comes from $\mathbb{P}_t^N$
- Then (see Sznitman, Méléard, Jourdain, Bossy, etc.) propagation of chaos $\Longrightarrow$

$$\frac{1}{N}\sum_{i=1}^{N} f(X_t^{i,N}) \xrightarrow[N\to\infty]{\mathrm{L}^1} \int_{\mathbb{R}^d} f(x)p(t,x)dx$$

- In the large N limit, the $(\mathbb{R}^n)^N$-dimensional **linear** Fokker-Planck PDE approximates the **nonlinear** low-dimensional (n-dimensional) Fokker-Planck PDE

Particle algorithm

$$D_n = \sum_{i \leq n} D(i) + \binom{n}{1}^{-1} \sum_{i<j\leq n} D(ij) + \binom{n}{2}^{-1} \sum_{i<j<k\leq n} D(ijk) + \ldots + D(12\ldots n).$$

[*A backward* operator D is the dual of a *forward* infinitesimal operator D^* as described in §1.] $D(1)$ is an (additive) 1-molecule backward infinitesimal operator acting on the first coordinate of Q^n. $D(i)$ is a copy of $D(1)$ acting on the ith coordinate ($i \leq n$). $D(12) = D(21)$ is a 2-molecule operator acting on the pair of coordinates 12. $D(ij)$ is a copy of $D(12)$ acting on the pair ij ($i < j \leq n$), etc. $D(1)$ governs the motion of molecule number 1 in isolation, $D(12)$ governs the interaction (double collisions) between molecules 1 and 2, $D(123)$ the (triple) collisions between molecules 1, 2, 3, etc. $D_n{}^*$ commutes with permutations of the coordinates so that if the initial n-molecule distribution f^n is symmetrical, then so is $v = e(tD_n{}^*)f^n$ at any later time.

Given a 1-molecule distribution function f, let f^n be the outer product $f\otimes\ldots\otimes f$ so that the molecules are independent at time $t = 0$ with common distribution f. A formal computation indicates that as $n \uparrow \infty$, $v = e(tD_n{}^*)f^n$ tends to the infinite outer product $u^\infty = u\otimes u\otimes$, etc., of the 1-molecule distribution function $u = \lim_{n\uparrow\infty} P_{f^n}[x_1(t) \in db]$. Kac,[4] describing the first known instance of this phenomenon, called it *the propagation of chaos*. u is the (formal) solution of $\partial u/\partial t =$

An archetypal example: the McKean-Vlasov SDE

- For $1 \leq k \leq n$ and $1 \leq l \leq d$

$$\begin{aligned} b_k(t, x, \mathbb{P}_t) &= \int b_k(t, x, y)\mathbb{P}_t(dy) &= \mathbb{E}[b_k(t, x, X_t)] \\ \sigma_{kl}(t, x, \mathbb{P}_t) &= \int \sigma_{kl}(t, x, y)\mathbb{P}_t(dy) &= \mathbb{E}[\sigma_{kl}(t, x, X_t)] \end{aligned}$$

- Particle method:

$$dX_t^{i,N} = \left(\int b(t, X_t^{i,N}, y) d\mathbb{P}_t^N(y) \right) dt + \left(\int \sigma(t, X_t^{i,N}, y) d\mathbb{P}_t^N(y) \right) \cdot dW_t^i$$

- This is equivalent to

$$dX_t^{i,N} = \frac{1}{N} \sum_{j=1}^{N} b\left(t, X_t^{i,N}, X_t^{j,N}\right) dt + \frac{1}{N} \sum_{j=1}^{N} \sigma\left(t, X_t^{i,N}, X_t^{j,N}\right) \cdot dW_t^i$$

Back to the calibration of SLV models

- In the SLV model, the approximated conditional expectation is

$$\mathbb{E}^{\mathbb{P}_t^N}[a_t^2|S_t = S] = \frac{\int a'^2 p_N(t,S,a')da'}{\int p_N(t,S,a')da'} = \frac{\sum_{i=1}^N (a_t^{i,N})^2 \delta(S_t^{i,N} - S)}{\sum_{i=1}^N \delta(S_t^{i,N} - S)}$$

- Instead of the Dirac delta function $\delta(\cdot)$, we use a kernel $\delta_{t,N}(\cdot)$ and approximate $\mathbb{E}^{\mathbb{P}_t}[a_t^2|S_t = S]$ by (simple nonparametric regression!)

$$\bar{a}(t,S)^2 = \frac{\sum_{i=1}^N \left(a_t^{i,N}\right)^2 \delta_{t,N}\left(S_t^{i,N} - S\right)}{\sum_{i=1}^N \delta_{t,N}\left(S_t^{i,N} - S\right)} \quad (21)$$

- Then we define

$$\sigma_N(t,S) = \frac{\sigma_{\mathrm{loc}}(t,S)}{\bar{a}(t,S)}$$

  and simulate

$$\frac{dS_t^{i,N}}{S_t^{i,N}} = (r_t - q_t)dt + \sigma_N(t, S_t^{i,N}) a_t^{i,N} dW_t^i \quad (22)$$

- $O(N^2)$ operations at each discretization date $\Rightarrow$ it is crucial to design acceleration techniques

How to make the particle method efficient in practice

- Computing $\sigma_N(t, S_t^{i,N})$ for all i is useless $\Rightarrow$ Compute $\sigma_N(t, S)$ for only a grid of values $G_{S,t}$ of S, of a size much smaller than N, and then inter- and extrapolate
- In the sums in (21), a large number of terms make a negligible contribution $\Rightarrow$ Disregard $S_t^{i,N}$ when it is far from S, say, when $\delta_{t,N}(S_t^{i,N} - S) < \eta$ $\Rightarrow$ Sort particles according to spot value. Cost of sorting, $O(N \ln N)$, is more than compensated by the acceleration in the evaluations of (21)
- Alternative methods for estimating such conditional expectations include B-spline techniques (see recent work by Corlay)
- Calibration and pricing can be achieved in the course of the same Monte-Carlo simulation
- If we need more paths for pricing, we can reuse the calibration paths, and simulate new paths using the already calibrated local volatility function

Particle algorithm (Guyon and Henry-Labordère, 2011)

1. Initialize $k = 1$ and set $\sigma_N(t,S) = \frac{\sigma_{\text{loc}}(0,S)}{\alpha}$ for all $t \in [t_0 = 0, t_1]$
2. Simulate the N processes $\{S_t^{i,N}, a_t^{i,N}\}_{1 \le i \le N}$ from t_{k-1} to t_k using a discretization scheme for (22) - say a log-Euler scheme
3. Sort particles according to spot value. For $S \in G_{S,t_k}$, find the smallest index $\underline{i}(S)$ and the largest index $\overline{i}(S)$ for which $\delta_{t_k,N}\left(S_{t_k}^{i,N} - S\right) > \eta$, and compute the local volatility

$$\sigma_N(t_k,S) = \sigma_{\text{loc}}(t_k,S)\sqrt{\frac{\sum_{i=\underline{i}(S)}^{\overline{i}(S)} \delta_{t_k,N}\left(S_{t_k}^{i,N} - S\right)}{\sum_{i=\underline{i}(S)}^{\overline{i}(S)} \left(a_{t_k}^{i,N}\right)^2 \delta_{t_k,N}\left(S_{t_k}^{i,N} - S\right)}}$$

   Interpolate the square root using cubic splines, and extrapolate (e.g., flat). Set $\sigma_N(t,S) \equiv \sigma_N(t_k,S)$ for all $t \in [t_k, t_{k+1}]$
4. Set $k := k+1$. Iterate steps 2 and 3 up to the maturity date T

Adding stochastic rates

- SIR-SLV model:

$$\frac{dS_t}{S_t} = r_t dt + \sigma(t, S_t) a_t dW_t$$

- Model exactly calibrated to market smile iff for all t, K

$$\sigma(t,K)^2 \frac{\mathbb{E}[D_{0t} a_t^2 | S_t = K]}{\mathbb{E}[D_{0t} | S_t = K]} = \sigma_{\text{loc}}(t,K)^2 - \frac{\mathbb{E}[D_{0t} \left(r_t - r_t^0\right) 1_{S_t > K}]}{\frac{1}{2} K \partial_K^2 \mathcal{C}(t,K)} \quad (23)$$

where r_t^0 is a deterministic rate and

$$\sigma_{\mathsf{Dup}}(t,K)^2 = \frac{\partial_t C_{\mathsf{mkt}}(t,K) + r_t^0 K \partial_K C_{\mathsf{mkt}}(t,K)}{\frac{1}{2} K^2 \partial_K^2 C_{\mathsf{mkt}}(t,K)}$$

- Equivalently, using the t-forward measure $\mathbb{P}^t$,

$$\sigma(t,K)^2 \mathbb{E}^{\mathbb{P}^t}[a_t^2 | S_t = K] = \sigma_{\text{loc}}(t,K)^2 - P_{0t} \frac{\mathbb{E}^{\mathbb{P}^t}[\left(r_t - r_t^0\right) 1_{S_t > K}]}{\frac{1}{2} K \partial_K^2 \mathcal{C}(t,K)}$$

- Usually $r_t^0 = -\partial_t \ln P_{0t} = \mathbb{E}^{\mathbb{P}^t}[r_t]$
- Does the r.h.s. of (23) stay nonnegative?

Adding stochastic rates

- $\Rightarrow$ The calibrated model follows the nonlinear SDE

$$\frac{dS_t}{S_t} = r_t\, dt + \sigma\left(t, S_t, \mathbb{P}_t\right)\, a_t\, dW_t$$

where

$$\sigma\left(t, K, \mathbb{P}_t\right)^2 = \left(\sigma_{\mathsf{Dup}}(t,K)^2 - \frac{\mathbb{E}[D_{0t}\left(r_t - r_t^0\right) 1_{S_t > K}]}{\frac{1}{2} K \partial_K^2 C_{\mathsf{mkt}}(t,K)}\right) \frac{\mathbb{E}^{\mathbb{P}}[D_{0t} | S_t = K]}{\mathbb{E}^{\mathbb{P}}[D_{0t} a_t^2 | S_t = K]} \tag{24}$$

- Equivalently:

$$\frac{df_t}{f_t} = \sigma\left(t, P_{tT} f_t, \mathbb{P}_t^T\right)\, a_t\, dW_t^T - \sigma_{\mathrm{P}}^T(t).dB_t^T$$

where

$$\sigma\left(t, K, \mathbb{P}_t^T\right)^2 = \left(\sigma_{\mathrm{loc}}(t,K)^2 - P_{0T} \frac{\mathbb{E}^{\mathbb{P}^T}[P_{tT}^{-1}\left(r_t - r_t^0\right) 1_{S_t > K}]}{\frac{1}{2} K \partial_K^2 C_{\mathsf{mkt}}(t,K)}\right) \frac{\mathbb{E}^{\mathbb{P}^T}[P_{tT}^{-1} | S_t = K]}{\mathbb{E}^{\mathbb{P}^T}[P_{tT}^{-1} a_t^2 | S_t = K]} \tag{25}$$

Particle algorithm (Guyon and Henry-Labordère) - Case I

If we use (24) we define

$$\sigma_N(t,S)^2 = \left(\sigma_{\mathsf{Dup}}(t,S)^2 - \frac{\frac{1}{N}\sum_{i=1}^{N} D_{0t}^{i,N}\left(r_t^{i,N} - r_t^0\right) 1_{S_t^{i,N} > S}}{\frac{1}{2} S \partial_K^2 C_{\mathsf{mkt}}(t,S)}\right) \times \frac{\sum_{i=1}^{N} D_{0t}^{i,N} \delta_{t,N}\left(S_t^{i,N} - S\right)}{\sum_{i=1}^{N} D_{0t}^{i,N}\left(a_t^{i,N}\right)^2 \delta_{t,N}\left(S_t^{i,N} - S\right)}$$

and simulate

$$\frac{dS_t^{i,N}}{S_t^{i,N}} = r_t^{i,N} dt + \sigma_N\left(t, S_t^{i,N}\right) a_t^{i,N} dW_t^i$$

Particle algorithm (Guyon and Henry-Labordère) - Case II

In many commonly used short rate models, P_{tT} bond has a closed form formula, we can use (25), define

$$\sigma_N(t,S)^2 = \left(\sigma_{\text{Dup}}(t,S)^2 - P_{0T} \frac{\frac{1}{N}\sum_{i=1}^N \left(P_{tT}^{i,N}\right)^{-1} \left(r_t^{i,N} - r_t^0\right) 1_{S_t^{i,N} > S}}{\frac{1}{2} S \partial_K^2 C_{\text{mkt}}(t,S)} \right)$$

$$\times \frac{\sum_{i=1}^N \left(P_{tT}^{i,N}\right)^{-1} \delta_{t,N}\left(S_t^{i,N} - S\right)}{\sum_{i=1}^N \left(P_{tT}^{i,N}\right)^{-1} \left(a_t^{i,N}\right)^2 \delta_{t,N}\left(S_t^{i,N} - S\right)}$$

and simulate

$$df_t^{i,N} = f_t^{i,N} \sigma_N\left(t, f_t^{i,N} P_{tT}^{i,N}\right) a_t^{i,N} dW_t^i - f_t^{i,N} \sigma_{\mathrm{P}}^{T,i,N}(t).dB_t^i$$

where W^i and B^i are $\mathbb{P}^T$-Brownian motions.

Malliavin representation of the local volatility (Guyon and Henry-Labordère, 2011)

- With Malliavin calculus we transform $P_{0t}\frac{\mathbb{E}^{\mathbb{P}^t}[(r_t-r_t^0)1_{S_t>K}]}{\frac{1}{2}K\partial_K^2\mathcal{C}(t,K)}$ into a conditional expectation (more accurate on the wings)
- For $r_t^0 = -\partial_t \ln P_{0t}$, Clark-Ocone's formula + Itô's isometry give ("**disintegration by parts**")

$$P_{0t}\frac{\mathbb{E}^{\mathbb{P}^t}\left[\left(r_t - r_t^0\right) 1_{S_t>K}\right]}{\frac{1}{2}K\partial_K^2\mathcal{C}(t,K)} = \frac{2}{K}\int_0^t \mathbb{E}^{\mathbb{P}^t}\left[\sigma_{\mathrm{r}}^t(s).D_s^{B^t} S_t | S_t = K\right] ds \qquad (26)$$

where

$$r_t - r_t^0 = \int_0^t \sigma_{\mathrm{r}}^t(s) dB_s^t$$

- Example: in the Ho-Lee and Hull-White models, the r.h.s. of (26) can be computed by simulating 5 extra processes
- Extension to Libor Market Models possible

Adding stochastic repo/dividend yield (Guyon, 2013)

- SIR-SDY-SLV model:

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \sigma(t, S_t)a_t\,dW_t$$

- Model exactly calibrated to market smile iff for all (t, K)

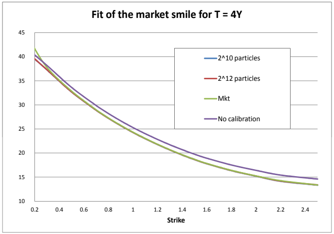
$$\begin{aligned}\sigma(t,K)^2 \frac{\mathbb{E}[D_{0t}a_t^2|S_t = K]}{\mathbb{E}[D_{0t}|S_t = K]} \quad &= \quad \sigma_{\mathsf{Dup}}(t,K)^2 - \frac{\mathbb{E}\left[D_{0t}\left(r_t - q_t - (r_t^0 - q_t^0)\right)1_{S_t > K}\right]}{\frac{1}{2}K\partial_K^2 C_{\mathsf{mkt}}(t,K)} \\ &\quad + \frac{\mathbb{E}\left[D_{0t}\left(q_t - q_t^0\right)(S_t - K)^+\right]}{\frac{1}{2}K^2\partial_K^2 C_{\mathsf{mkt}}(t,K)}\end{aligned}$$

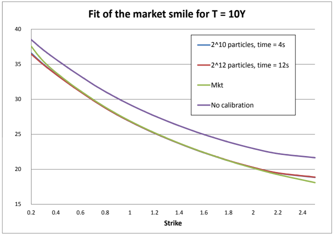
where r_t^0 and q_t^0 are deterministic rates and repos and

$$\sigma_{\mathsf{Dup}}(t,K)^2 = \frac{\partial_t C_{\mathsf{mkt}}(t,K) + (r_t^0 - q_t^0)K\partial_K C_{\mathsf{mkt}}(t,K) + q_t^0 C_{\mathsf{mkt}}(t,K)}{\frac{1}{2}K^2\partial_K^2 C_{\mathsf{mkt}}(t,K)}$$

- The calibrated model follows a McKean SDE
- Does $\sigma(t,K)^2$ stay nonnegative?

- DAX market smiles (30-May-11)
- Ho-Lee parameters: $\sigma_{\mathrm{r}} = 1\%$, $\rho = 40\%$
- Particle algorithm: $\Delta = 1/100$, $N = 2^{10}$ particles. MC pricing: $N = 2^{15}$ paths





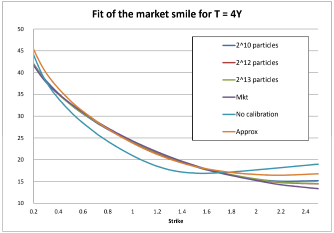
LV + Ho-Lee

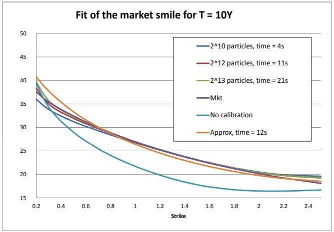
Strike	0.5	0.7	0.8	0.9	1	1.1	1.2	1.3	1.5	1.8
with Malliavin	14	10	10	9	8	7	6	5	3	1
without Malliavin	16	8	7	4	1	1	1	3	3	5

Table: DAX (30-May-11) Implied volatilities $T = 10$Y. Errors in bps using the particle method with $N = 2^{10}$ particles

LV + Bergomi

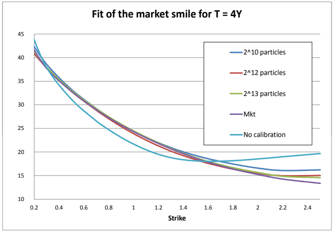
- Parameters of Bergomi model: $\sigma = 200\%$, $\theta = 22.65\%$, $k_X = 4$, $k_Y = 12.5\%$, $\rho = 30\%$, $\rho_{\mathrm{SX}} = -50\%$, $\rho_{\mathrm{SY}} = -50\%$

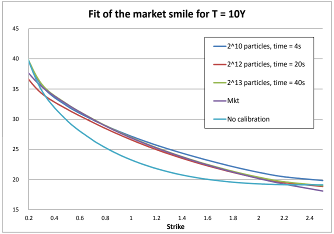




LV + Bergomi + Ho-Lee

- Ho-Lee parameters: $\sigma_{\mathrm{r}} = 1\%$, $\rho_{\mathrm{Sr}} = 40\%$
- Parameters of Bergomi model: $\sigma = 200\%$, $\theta = 22.65\%$, $k_X = 4$, $k_Y = 12.5\%$, $\rho = 30\%$, $\rho_{\mathrm{SX}} = -50\%$, $\rho_{\mathrm{SY}} = -50\%$. Correlation Vol/Rate $= 0$
- Requires solving a 4d-PDE!



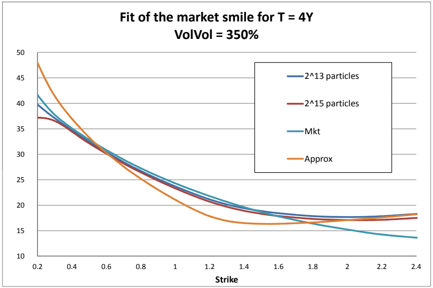


No optimization at all!

- Classical calibration procedures involve optimization over parameters: minimize the mean squared error between prices in the model and market prices.
- Time consuming, there could be local minima...
- And calibration to whole smile is imperfect (due to the finite number of parameters)
- It is remarkable that the particle method involves **no optimization at all**! And it gives perfect calibration.

Existence?

- Parameters of Bergomi model: $\sigma = 350\%$, $\theta = 22.65\%$, $k_X = 4$, $k_Y = 12.5\%$, $\rho = 30\%$, $\rho_{\mathrm{SX}} = -50\%$, $\rho_{\mathrm{SY}} = -50\%$.

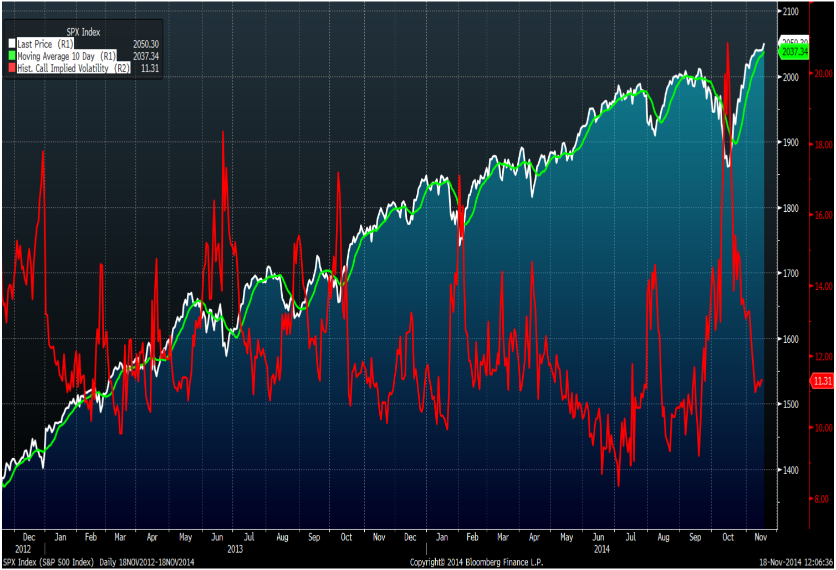


Existence?

$$\sigma(t,S,p) = \sigma_{\text{loc}}(t,S)\sqrt{\frac{\int p(t,S,a)da}{\int a^2 p(t,S,a)da}}$$

- Lipschitz condition w.r.t. $\mathbb{P}$ not satisfied $\Longrightarrow$ **uniqueness and existence results for calibrated SLV models are not at all obvious**
- Partial result by Abergel and Tachet (2010): the calibration problem for a SLV model is proved to be well posed (a) until some maturity T^*, (b) if the vol-of-vol is small enough, and (c) for suitably regularized initial conditions - not for a Dirac mass!
- Partial result by Jourdain and Zhou (2016): Regime-switching LV model = when the SV (a_t) is a pure jump process taking finitely many values, with jump intensities depending on the spot level. Range of (a_t^2) not too large $\Longrightarrow$ existence of a weak solution to the McKean SDE.
- Our numerical experiments with SLV show that the calibration fails for large vol-of-vol, whatever the algorithm used: PDE, particle method. Numerical issue? Non-existence?

Contents


SPX Index
Last Price (R1) 2050.30
Moving Average 10 Day (R1) 2037.34
Hist. Call Implied Volatility (R2) 11.31
2100
2000
1900
1800
1700
1600
1500
1400
20.00
18.00
16.00
14.00
12.00
11.31
10.00
8.00
Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov
2012
2013
2014
SPX Index (S&P 500 Index) Daily 18NOV2012-18NOV2014
Copyright© 2014 Bloomberg Finance L.P.
18-Nov-2014 12:06:36

Why study and use path-dependent volatility?

- Path-dependent volatility (PDV) models have drawn little attention compared with local volatility (LV) and stochastic volatility (SV) models
- This is unfair: PDV models combine benefits from both LV and SV, and even go beyond
- Like LV: complete and can fit exactly the market smile
- Like SV: produce a wide variety of joint spot-vol dynamics
- Not only that:
  1. Can generate spot-vol dynamics that are not attainable using SV models
  2. Can also capture prominent historical patterns of volatility

Outline

- A short recap on volatility modeling
- PDV models: a (necessarily) brief history, and what we want from them
- Smile calibration of PDV models
- Choose a particular PDV to generate desired spot-vol dynamics
- Choose a particular PDV to capture historical patterns of volatility
- Concluding remarks
- Discussion

A short recap on volatility modeling

- Constant volatility (Bachelier, 1905; Black and Scholes, 1973) and LV (Dupire, 1994) are complete models: every payoff admits a unique self-financing replicating portfolio consisting of cash and the underlying asset $\Longrightarrow$ unique price
- LV flexible enough to fit exactly any arbitrage-free smile—but no more flexibility is left
- SV models are incomplete $\Longrightarrow$ no unique price. But they give control on key risk factors such as vol-of-vol, forward skew, and spot-vol correlation. Unlike LV, they generate rich joint dynamics of the asset and its implied volatilities
- To allow SV models to perfectly calibrate to the market smile, one can use SLV models + particle method. Modifies spot-vol dynamics, but only slightly (except maybe for small times t): usually the LV component (leverage function) flattens as t grows

Can we combine benefits from both LV and SV?

- Can we build complete models that have all the nice properties of SLV models, namely, rich spot-vol dynamics, and calibration to market smile?
- For instance, can we build a complete model that is calibrated to a flat smile, and yet produces very negative short term forward skews?
- Tempting but wrong to quickly answer 'no', by arguing that the only complete model calibrated to the smile is the LV model.
- This is not true: we will show that PDV models, which are complete, can produce rich spot-vol dynamics and, on top of that, can be perfectly calibrated to the market smile
- Benefits of model completeness: price uniqueness and parsimony. All properties of SLV models can be captured using a single Brownian motion. Although perfect delta-hedging is unrealistic, incorporating the path-dependency of volatility into the delta is likely to improve the delta-hedge.
- PDV models actually go beyond SLV models: they can generate spot-vol dynamics that are not attainable using SLV models.

PDV models

- PDV models are those models where the instantaneous volatility σ_t depends on the path followed by the asset price so far:

$$\frac{dS_t}{S_t} = \sigma(t, (S_u, u \leq t))\, dW_t$$

- In practice, $\sigma_t \equiv \sigma(t, S_t, X_t)$ where $X_t =$ finite set of path-dependent variables: running or moving averages, maximums or minimums, realized variances, etc.
- Most famous examples: ARCH/GARCH models. Discrete-time and hardly used in the derivatives industry
- Discrete setting version of Bergomi's SV model = a mixed SV-PDV model: given a realization of the (random) variance swap volatility at time $T_i = i\Delta$ for maturity T_{i+1}, $\sqrt{\xi^i_{T_i}}$, the (continuous time) volatility of the underlying on $[T_i, T_{i+1}]$ is path-dependent: $\sigma(S_t/S_{T_i})$, where σ is calibrated to both $\xi^i_{T_i}$ and a desired value of the forward ATM skew for maturity Δ.

The Hobson-Rogers model

- Main contribution on continuous-time pure PDV models so far
- $\sigma_t = \sigma(X_t)$; $X_t = (X_t^1, \dots, X_t^n)$ where the X_t^m are exponentially weighted moments of all the past log increments of the asset price:

$$X_t^m = \int_{-\infty}^t \lambda e^{-\lambda(t-u)} \left(\ln \frac{S_t}{S_u} \right)^m du$$

- $n = 1$: σ_t depends only on $X_t^1 = \ln S_t - \int_{-\infty}^t \lambda e^{-\lambda(t-u)} \ln S_u \, du =$ the difference between current log price and a weighted average of past log prices $\Longrightarrow$ volatility determined by local trend of the asset price over a period of order $1/\lambda$ years (e.g., 1 month if $\lambda = 12$)
- Supported by empirical studies (see later)
- Choice of an infinite time window and exponential weights only guided by computational convenience: ensures that (S_t, X_t) is a Markovian process $\Longrightarrow$ price of a vanilla option reads $u(t, S_t, X_t)$ where u is the solution to a second order parabolic PDE
- Implied volatilities at time 0 in the model depend not only on the strike, maturity, and S_0, but also on all the past asset prices through X_0

Four natural and important questions arise

1. Can we specify $\sigma(\cdot)$ and λ so that the model fits exactly the market smile? Platania & Rogers, Figà-Talamanca & Guerra only gave approximate calibration results
2. Does the calibrated model have desired dynamics of implied volatility, such as large negative short term forward skew for instance?
3. In the definition of X_t, can we use general weights and a finite time window $[t - \Delta, t]$ instead of $(-\infty, t]$, so that the volatility truly depends only a limited portion of the past? The generalization in Foschi & Pascucci is partial as it requires positive weights on $[0, t]$.
4. Much more importantly: how do we generalize to other choices of X_t? The generalization in Hubalek *et al.*, where the volatility depends on a particular modified version of the offset X_t^1, is also very partial

Our approach will solve these four questions all at once

- First we choose any set of path-dependent variables X_t and any function $\sigma(t, S, X)$ so that the PDV model with $\sigma_t = \sigma(t, S_t, X_t)$ has desired spot-vol dynamics and/or captures historical patterns of volatility
- Then we define a new model by multiplying $\sigma(t, S_t, X_t)$ by a leverage function $l(t, S_t)$ and we perfectly calibrate l to the market smile of S using the particle method
- Usually, multiplying $\sigma(t, S_t, X_t)$ by the calibrated leverage function distorts only slightly the spot-vol dynamics
- This way we mimic SLV models, with the 'pure' PDV $\sigma(t, S_t, X_t)$ playing the role of SV, but we stay in the world of complete models
- Not only that: thanks to their huge flexibility, PDV models can generate spot-vol dynamics that are not attainable using SLV models
- Same program can be run by choosing two functions $a(t, S, X)$ and $b(t, S, X)$ instead of only one function $\sigma(t, S, X)$, and then defining $\sigma_t^2 = a(t, S_t, X_t) + b(t, S_t, X_t)l(t, S_t)$
  $b \equiv 1$: complete analogue of incomplete *additive* SLV models

Smile calibration of PDV models: Particle method

- Given a PDV $\sigma(t,S,X)$, we can uniquely build the leverage function function $l(t,S)$ such that the PDV model

$$\frac{dS_t}{S_t} = \sigma(t,S_t,X_t)l(t,S_t)\,dW_t \tag{27}$$

fits exactly the market smile of S

- From Itô-Tanaka's formula, Model (27) is exactly calibrated to the market smile of S if and only if

$$\mathbb{E}^{\mathbb{Q}}\left[\sigma(t,S_t,X_t)^2\middle|S_t\right]l(t,S_t)^2 = \sigma^2_{\mathsf{Dup}}(t,S_t)$$

where $\mathbb{Q}$ denotes the unique risk-neutral measure and σ_{Dup} the Dupire LV $\Longrightarrow$ calibrated model satisfies the nonlinear McKean stochastic differential equation

$$\frac{dS_t}{S_t} = \frac{\sigma(t,S_t,X_t)}{\sqrt{\mathbb{E}^{\mathbb{Q}}[\sigma(t,S_t,X_t)^2|S_t]}}\sigma_{\mathsf{Dup}}(t,S_t)\,dW_t$$

- The particle method (Guyon & Henry-Labordère, 2011) computes the above conditional expectation, hence the leverage function $l(t,S) = \sigma_{\mathsf{Dup}}(t,S)\ /\sqrt{\mathbb{E}^{\mathbb{Q}}[\sigma(t,S_t,X_t)^2|S_t=S]}$, on the go while simulating the paths

Smile calibration of PDV models

- Brunick and Shreve (2013): Given a general Itô process $dS_t = \sigma_t S_t \, dW_t$ and a special type of path-dependent variable X, there exists a PDV $\sigma(t, S_t, X_t)$ such that, for each t, the joint distribution of (S_t, X_t) is the same in both models:

$$\sigma(t, S_t, X_t)^2 = \mathbb{E}^{\mathbb{Q}}[\sigma_t^2 | S_t, X_t]$$

- Only X's satisfying a type of Markov property are admissible though: running averages are admissible, but moving averages are not; instead, one must pick $X_t = (S_u, t - \Delta \leq u \leq t)$
- Take $X_t = (S_u, 0 \leq u \leq t)$: Brunick-Shreve $\Longrightarrow$ the price process produced by any SV/SLV model has the same distribution, as a process, as a PDV model (not only the marginal distributions) $\Longrightarrow$ There always exists a PDV model that produces exactly the same prices of, not only vanilla options, but all options, including path-dependent, exotic options $\Longrightarrow$ No surprise that PDV models can reproduce popular SLV spot-vol dynamics (see below)

Now the crucial question is:

How to choose a particular PDV?

Two main possible goals:

1. Generate desired spot-vol dynamics
2. Capture historical features of volatility

These two goals are not mutually exclusive: it might very well happen, and it is desirable, that a given choice of a PDV fulfills both objectives at a time

Choose a particular PDV
to generate desired spot-vol dynamics

Choose a particular PDV to generate desired spot-vol dynamics

- Can we choose a PDV $\sigma(t,S,X)$ that, for instance, generates large negative short term forward skews, even when it is calibrated to a flat smile?
- SLV analogy $\Longrightarrow$ We need $\sigma(t,S_t,X_t)$ to be negatively correlated with S_t
- May be achieved by picking a decreasing function σ of S alone, but smile calibration would bring us back to pure LV model:

$$\frac{dS_t}{S_t} = \frac{\sigma(t,S_t,X_t)}{\sqrt{\mathbb{E}^{\mathbb{Q}}[\sigma(t,S_t,X_t)^2|S_t]}}\sigma_{\text{Dup}}(t,S_t)\,dW_t \tag{28}$$

- What we actually need is $\sqrt{\eta(t,S_t,X_t)}$ to be negatively correlated with S_t, where

$$\eta(t,S,X) \equiv \frac{\sigma(t,S,X)^2}{\mathbb{E}^{\mathbb{Q}}[\sigma(t,S_t,X_t)^2|S_t=S]}$$

- $\eta(t,S,X)$ = PDLVR = 'path-dependent to local variance ratio'
- The PDLVR or alternatively $D(t,S) = \mathrm{Var}(\eta(t,S_t,X_t)|S_t=S) = \mathbb{E}[(\eta(t,S_t,X_t)-1)^2|S_t=S]$ measures deviation from LV: LV $\Longleftrightarrow \eta \equiv 1 \Longleftrightarrow D \equiv 0$

Choose a particular PDV to generate desired spot-vol dynamics

- Recall that we want

$$\sqrt{\eta(t,S,X)} \equiv \frac{\sigma(t,S,X)}{\sqrt{\mathbb{E}^{\mathbb{Q}}[\sigma(t,S_t,X_t)^2|S_t=S]}}$$

  to tend to be large when S is small, and conversely
- $\Longrightarrow \sigma(t,S,X)$ must be negatively linked to S, but not perfectly: target correl of the levels of spot and volatility is more around, say, -50% than around -1% or -99%. Moderate correlation property
- Usual SLV models: Mean reversion in the SV $\Longrightarrow$ Moderate correlation, even if increments of spot and volatility are extremely negatively correlated

Example 1

Ex.	X_t	$\sigma(S, X)$ producing large forward skew
1	$S_{t-\Delta}$	$\overline{\sigma} 1_{\{\frac{S}{X} \leq 1\}} + \underline{\sigma} 1_{\{\frac{S}{X} > 1\}}$

Comparing with SV models:

- $\overline{\sigma} - \underline{\sigma} \longleftrightarrow$ vol-of-vol: we need it to be large enough to generate large negative short term forward skew
- $\Delta \longleftrightarrow$ spot-vol correlation:
  - S_t small $\Longrightarrow$ more likely that S_t be smaller than $S_{t-\Delta} \Longrightarrow$ more likely that σ_t be large
  - The larger Δ, the larger the correlation
- $\Delta \longleftrightarrow$ mean reversion too: the smaller Δ, the more ergodic the volatility, hence the flatter the forward smile (cf. Fouque-Papanicolaou-Sircar, 2000)

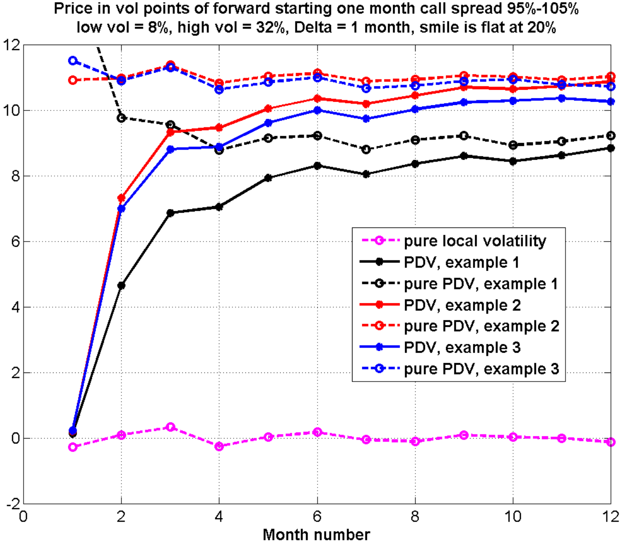
Examples 2 and 3

Ex.	X_t	$\sigma(S,X)$ producing large forward skew
1	$S_{t-\Delta}$	$\overline{\sigma}1_{\left\{\frac{S}{X}\leq 1\right\}}+\underline{\sigma}1_{\left\{\frac{S}{X}>1\right\}}$
2	$\overline{S}_t^{\Delta}$	as above
3	$(m_t^{\Delta}, M_t^{\Delta})$	$\overline{\sigma}1_{\left\{\frac{S-m}{M-m}\leq\frac{1}{2}\right\}}+\underline{\sigma}1_{\left\{\frac{S-m}{M-m}>\frac{1}{2}\right\}}$

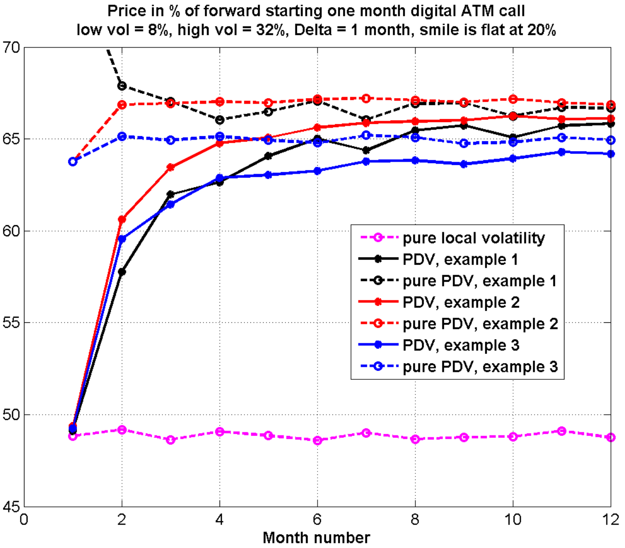
$$\overline{S}_t^{\Delta}=\frac{\int_0^{\Delta}w_\tau S_{t-\tau}\,d\tau}{\int_0^{\Delta}w_\tau\,d\tau},\qquad m_t^{\Delta}=\inf_{t-\Delta\leq u\leq t}S_u,\qquad \text{and}\qquad M_t^{\Delta}=\sup_{t-\Delta\leq u\leq t}S_u$$

- Ex. 2: $S_{t-\Delta}$ replaced by moving average $\overline{S}_t^{\Delta}$. Makes more financial sense: why put all the weight w_τ on $\tau=\Delta$?
- Ex. 3 uses that $\frac{S_t-m_t^{\Delta}}{M_t^{\Delta}-m_t^{\Delta}}$ is positively correlated with S_t. The larger Δ, the larger the correlation

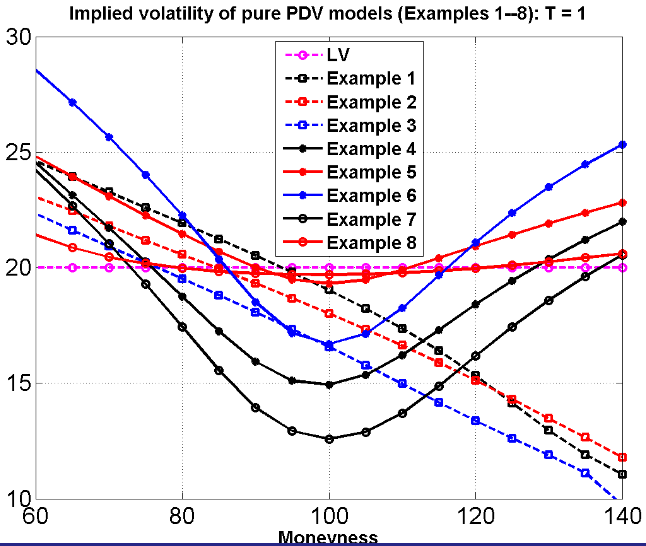
Forward starting 1M call spread 95%-105%



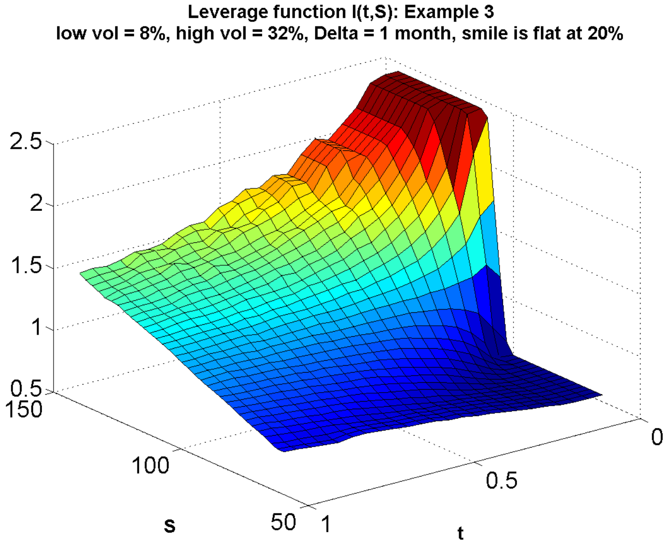
Forward starting 1M ATM digital call



Smiles of pure PDV models



Example 3: Leverage function $l(t, S)$



Short term forward skew: comparison with SLV models

- Consider the SLV model where the SV $=$ exponential O-U process
- To reproduce Example 2 prices, a huge positive return in volatility is needed when asset returns are negative $\Longrightarrow$ one must use a very large vol-of-vol, beyond 550%, together with an extreme spot-vol correlation (say, -99%)
- $\Longrightarrow$ A very large mean reversion above 20 is then somehow artificially needed to keep volatility within a reasonable range
- By comparison PDV models, which can directly relate the asset returns to the volatility levels, look much more handy and can more naturally generate large forward skews

- Large positive short term forward skew: exchange $\overline{\sigma}$ and $\underline{\sigma}$
- One may use smoothed versions of the PDV by replacing the Heaviside function by $\frac{1}{2}(1+\tanh(\lambda x))$ for instance

U-shaped short term forward smile

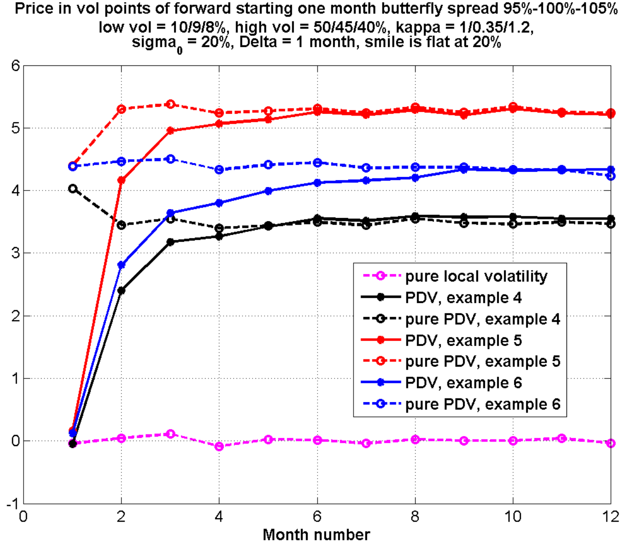
- What if we want a PDV model calibrated to a flat smile and yet that generates a pronounced U-shaped short term ($\tau = 1\mathrm{M}$) forward smile?
- $\Longrightarrow$ We need

$$\sqrt{\eta(t,S,X)} \equiv \frac{\sigma(t,S,X)}{\sqrt{\mathbb{E}^{\mathbb{Q}}[\sigma(t,S_t,X_t)^2|S_t = S]}}$$

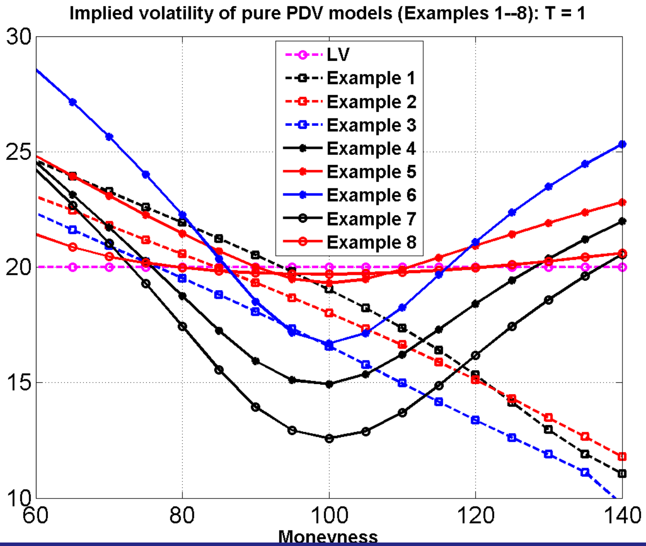
  to be highly volatile and uncorrelated with S
- Examples 1–3 cannot capture this:
  - $\Delta \ll \tau \Longrightarrow$ ergodic volatility $\Longrightarrow$ flat forward smile
  - $\Delta \approx \tau \Longrightarrow \sqrt{\eta(t,S,X)}$ correlated with S
  - $\Delta \gg \tau \Longrightarrow \sqrt{\eta(t,S,X)}$ almost constant
- Examples 4–6 are natural candidates: volatility is large if and only if recent asset returns (up or down) are as well. Produce vanishing ATM forward skew

Ex.	X_t	$\sigma(S,X)$ producing U-shaped forward smile
4	$S_{t-\Delta}$	$\overline{\sigma}1_{\left\{\left\lvert\frac{S}{X}-1\right\rvert>\kappa\sigma_0\sqrt{\Delta}\right\}} + \underline{\sigma}1_{\left\{\left\lvert\frac{S}{X}-1\right\rvert\leq\kappa\sigma_0\sqrt{\Delta}\right\}}$
5	$\overline{S}_t^{\Delta}$	as above
6	$(m_t^{\Delta}, M_t^{\Delta})$	$\overline{\sigma}1_{\left\{\frac{M}{m}-1>\kappa\sigma_0\sqrt{\Delta}\right\}} + \underline{\sigma}1_{\left\{\frac{M}{m}-1\leq\kappa\sigma_0\sqrt{\Delta}\right\}}$

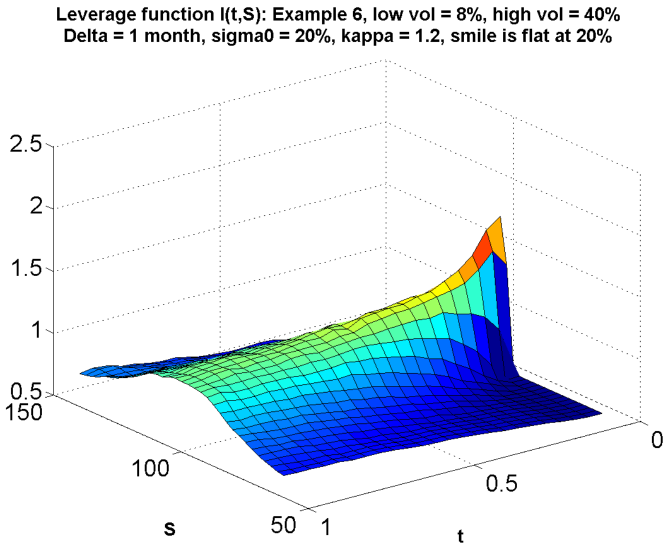
Forward starting 1M butterfly spread 95%-100%-105%



Smiles of pure PDV models



Example 6: Leverage function $l(t, S)$



Spot-vol dynamics beyond what SLV models can attain

- PDV models have so many degrees of freedom—the path-dependent variables X, and the function $\sigma(t,S,X)$—that they can generate spot-vol dynamics that are not attainable using SLV models
- Example: imagine that a sophisticated client asks a quote on the conditional variance swap with payoff

$$H_T = \sum_{i=1}^{n-1} r_{i+1}^2 1_{\{r_i \le 0\}} \approx \int_0^T \sigma_t^2 1_{\left\{\frac{S_t}{S_{t-\Delta}} \le 1\right\}} \, dt, \quad r_i = \frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}}, \quad \Delta = t_i - t_{i-1} = 1\,\text{day}$$

- SLV model: for a given risk-neutral probability $\mathbb{Q}$,

$$\mathbb{E}^{\mathbb{Q}}\left[\sigma_t^2 1_{\left\{\frac{S_t}{S_{t-\Delta}} \le 1\right\}} \middle| S_t\right] \approx \mathbb{E}^{\mathbb{Q}}\left[\sigma_t^2 \middle| S_t\right] \mathbb{E}^{\mathbb{Q}}\left[1_{\left\{\frac{S_t}{S_{t-\Delta}} \le 1\right\}} \middle| S_t\right] \approx \frac{1}{2}\sigma_{\mathsf{Dup}}^2(t,S_t)$$

$\Longrightarrow$ Both the SLV price and the LV price are very close to the variance swap price halved:

$$\text{SLV price} \approx \text{LV price} \quad \approx \quad \frac{1}{2}\int_0^T \mathbb{E}^{\mathbb{Q}}\left[\sigma_{\mathsf{Dup}}^2(t,S_t)\right] dt = \frac{1}{2}\text{var swap price}$$

whatever the choice of $\mathbb{Q}$

Spot-vol dynamics beyond what SLV models can attain

- However, if client requests such quote, it is probably because they observed that for this asset r_{i+1}^2 is large when $r_i \leq 0$, and small otherwise, expect this to continue, and try to statistically arbitrage a counterparty
- All the models commonly used in the industry today would fail to capture this risk but the PDV model of Ex. 1, with $\Delta = t_i - t_{i-1} = 1$ day, grasps it very well
- Δ small $\Longrightarrow \mathbb{E}^{\mathbb{Q}}\left[\sigma(t_i, S_{t_i}, S_{t_{i-1}})^2 \,|S_{t_i}\right] \approx \frac{\overline{\sigma}^2+\underline{\sigma}^2}{2}$ is almost constant $\Longrightarrow$

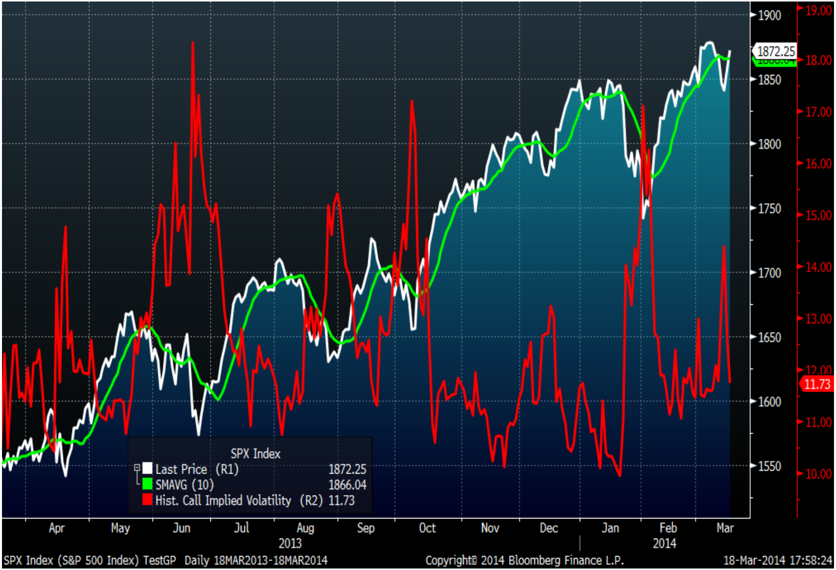
$$\text{PDV price} \approx \int_0^T \mathbb{E}^{\mathbb{Q}}\left[\frac{\overline{\sigma}^2}{\frac{\overline{\sigma}^2+\underline{\sigma}^2}{2}}\sigma_{\mathsf{Dup}}^2(t,S_t)1_{\left\{\frac{S_t}{S_{t-\Delta}}\leq 1\right\}}\right] dt \approx \frac{\overline{\sigma}^2}{\overline{\sigma}^2+\underline{\sigma}^2}\text{var swap price}$$

- For reasonable values of $(\underline{\sigma}, \overline{\sigma})$, e.g., $(10\%, 40\%)$, this is close to the (unconditional) var swap price $=$ twice the SLV price and twice the LV price. In such a case, an investment bank equipped with PDV may avoid a large mispricing

Choose a particular PDV
to capture historical patterns of volatility

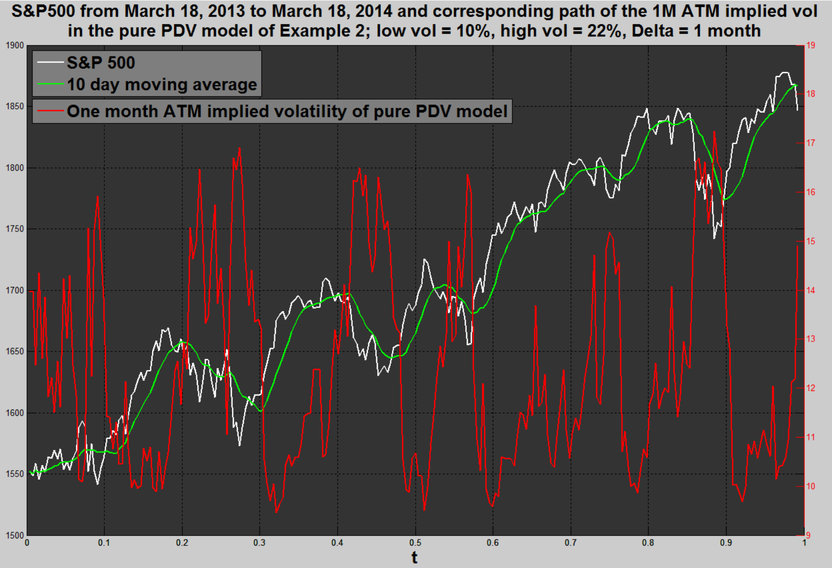
Choose a particular PDV to capture historical patterns of volatility

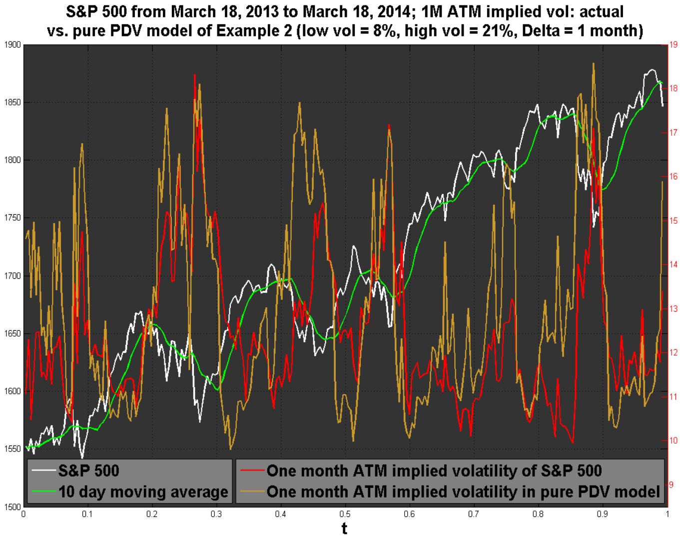
- Like the local correlation models presented in Guyon (2013), PDV models are flexible enough to reconcile implied calibration (e.g., calibration to the market smile) with historical calibration (calibration from historical time series of asset prices):
  1. one chooses a PDV $\sigma(t, S, X)$ from the observation of the time series, e.g., the short term ATM volatility is a certain function of $S_t/\overline{S}_t^{\Delta}$
  2. one then multiplies it by a leverage function l and eventually calibrates l to the market smile using the particle method
- By construction, PDV models are flexible enough to capture any path-dependency of the volatility. For a given choice of PDV, what remains to be numerically checked is
  1. how much and how long the smile calibration distorts the link between past prices and current instantaneous volatility
  2. whether the model produces suitable dynamics of implied volatility

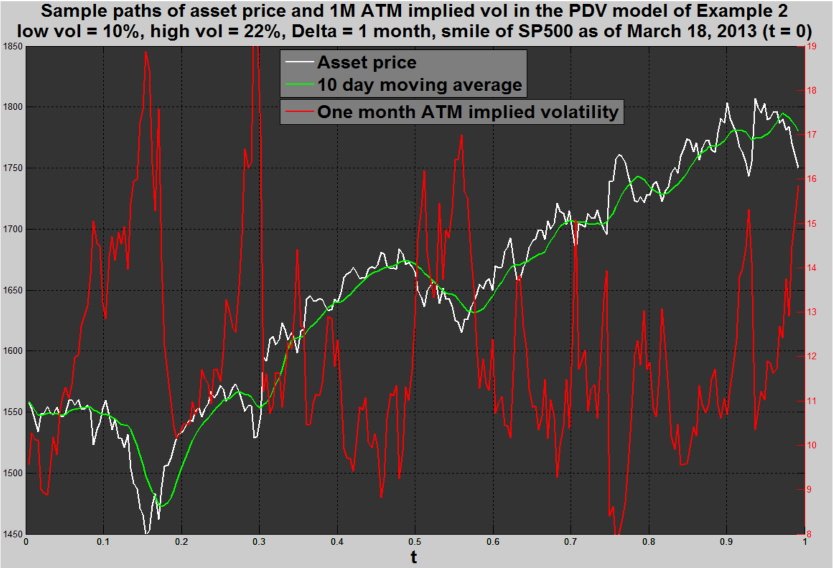

1900
1850
1800
1750
1700
1650
1600
1550
19.00
18.00
17.00
16.00
15.00
14.00
13.00
12.00
11.00
10.00
1872.25
11.73
SPX Index
Last Price (R1) 1872.25
SMAVG (10) 1866.04
Hist. Call Implied Volatility (R2) 11.73
Apr
May
Jun
Jul
Aug
2013
Sep
Oct
Nov
Dec
Jan
Feb
2014
Mar
SPX Index (S&P 500 Index) TestGP Daily 18MAR2013-18MAR2014
Copyright© 2014 Bloomberg Finance L.P.
18-Mar-2014 17:58:24

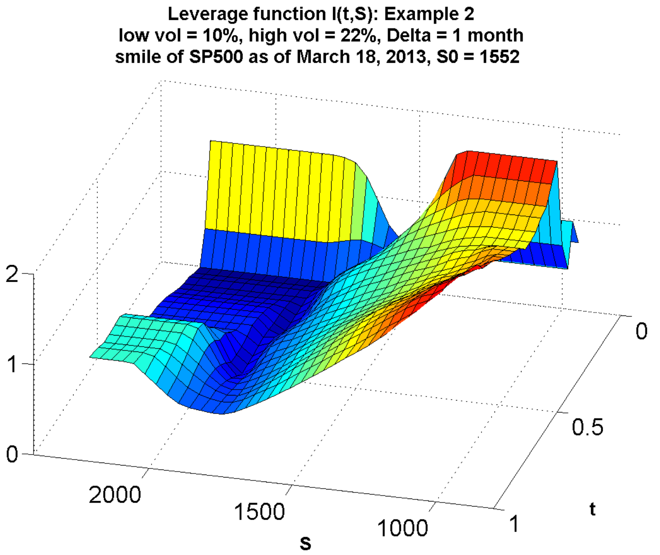
Choose a particular PDV to capture historical patterns of volatility

- For the S&P 500, the volatility level is not determined by the asset price level, but by the recent changes in the asset price
- Examples 1–3, which relate volatility levels to recent asset price returns, easily capture this
- Actually, the two basic quantities that possess a natural scale are the volatility levels and the asset returns so we believe that a good model should relate these two quantities
- LV model links the volatility level to the asset level, does not make much financial sense: well designed PDV models need not be recalibrated as often as the LV model
- SV models connect the change in volatility to the relative change in the asset price. Has limitations:
  - Only unreasonable levels of vol-of-vol allow large movements (e.g., a 70% return in 2 weeks) of instantaneous volatility
  - Therefore a large mean reversion needs to be artificially added to keep volatility within its natural range
- By contrast PDV models can easily capture such large changes in volatility

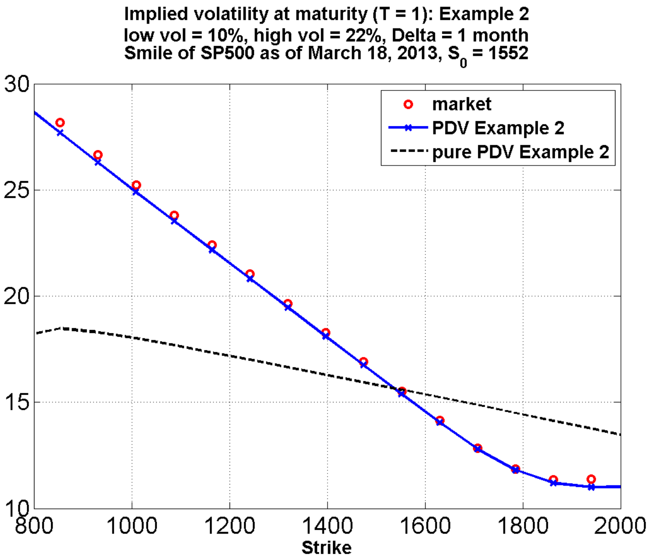

S&P500 from March 18, 2013 to March 18, 2014 and corresponding path of the 1M ATM implied vol in the pure PDV model of Example 2; low vol = 10%, high vol = 22%, Delta = 1 month
S&P 500
10 day moving average
One month ATM implied volatility of pure PDV model
t
1900
1850
1800
1750
1700
1650
1600
1550
1500
19
18
17
16
15
14
13
12
11
10
9
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


S&P 500 from March 18, 2013 to March 18, 2014; 1M ATM implied vol: actual vs. pure PDV model of Example 2 (low vol = 8%, high vol = 21%, Delta = 1 month)
S&P 500
10 day moving average
One month ATM implied volatility of S&P 500
One month ATM implied volatility in pure PDV model
t
1500
1550
1600
1650
1700
1750
1800
1850
1900
9
10
11
12
13
14
15
16
17
18
19
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Sample paths of asset price and 1M ATM implied vol in the PDV model of Example 2
low vol = 10%, high vol = 22%, Delta = 1 month, smile of SP500 as of March 18, 2013 (t = 0)
Asset price
10 day moving average
One month ATM implied volatility
t
1850
1800
1750
1700
1650
1600
1550
1500
1450
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
19
18
17
16
15
14
13
12
11
10
9
8

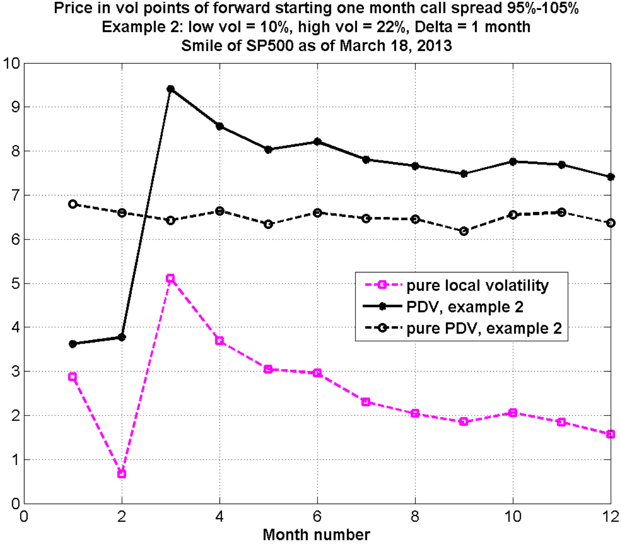


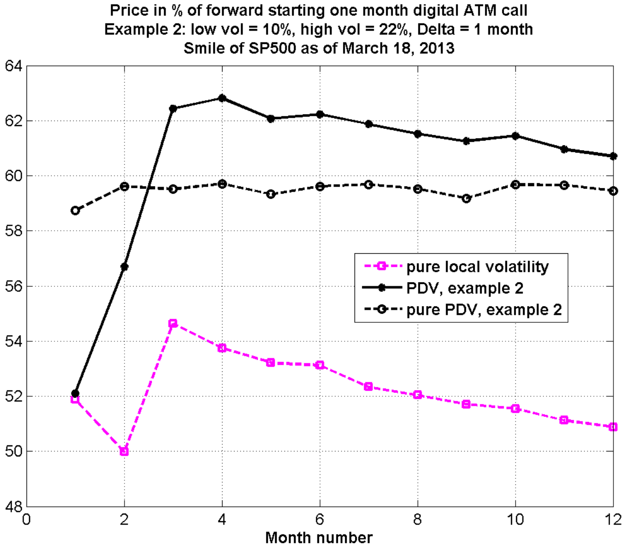
Leverage function l(t,S): Example 2
low vol = 10%, high vol = 22%, Delta = 1 month
smile of SP500 as of March 18, 2013, S0 = 1552
0
1
2
2000
1500
1000
S
0
0.5
1
t


Implied volatility at maturity (T = 1): Example 2
low vol = 10%, high vol = 22%, Delta = 1 month
Smile of SP500 as of March 18, 2013, S0 = 1552
market
PDV Example 2
pure PDV Example 2
30
25
20
15
10
800
1000
1200
1400
1600
1800
2000
Strike

Example 2: $\underline{\sigma} = 10\%$, $\overline{\sigma} = 22\%$, $\Delta = 1$ month

- The implied volatility varies continuously (with spikes when the market is locally bearish) $\Longrightarrow$ Modeling instantaneous volatility as a pure jump process is not problematic: no one has ever seen such quantity—it may actually not exist
- With such parameter values, the PDV model of Example 2 captures what we believe is a major pattern of the historical joint behaviour of the S&P 500 and its short term implied volatilities
- What about pricing? As the volatility interval $[\underline{\sigma}, \overline{\sigma}]$ is not as wide as $[8\%, 32\%]$, the forward skew is not as expensive. However, still sizeably larger than the LV forward skew


Price in vol points of forward starting one month call spread 95%-105%
Example 2: low vol = 10%, high vol = 22%, Delta = 1 month
Smile of SP500 as of March 18, 2013
pure local volatility
PDV, example 2
pure PDV, example 2
Month number


Price in % of forward starting one month digital ATM call
Example 2: low vol = 10%, high vol = 22%, Delta = 1 month
Smile of SP500 as of March 18, 2013
64
62
60
58
56
54
52
50
48
0
2
4
6
8
10
12
Month number
pure local volatility
PDV, example 2
pure PDV, example 2

Generalized local ARCH/GARCH models

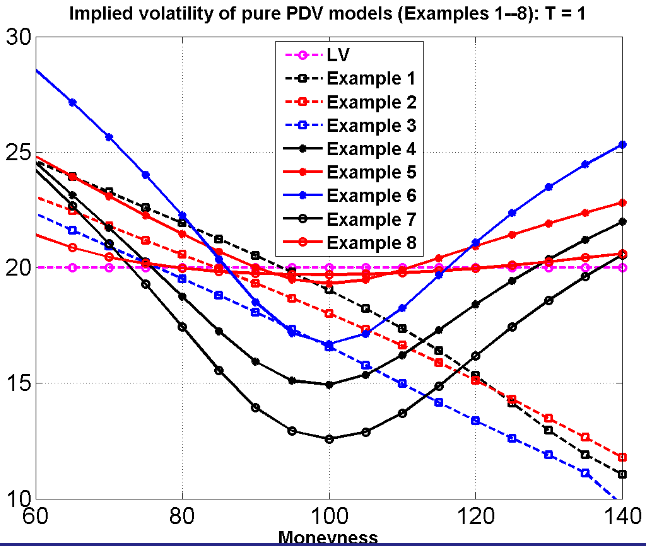
- That volatility depends on recent asset returns was also supported by other statistical analyses (Platania-Rogers, 2003; Foschi-Pascucci, 2007)
- Some empirical studies show that volatility may depend on recent realized volatility. So far, only the ARCH (Engle, 1982) and GARCH (Bollersev, 1986) models, and their descendants, could capture this
- ARCH/GARCH capture tail heaviness, volatility clustering and dependence without correlation, like Examples 1–6 above
- Our approach generalizes them by defining *local ARCH models*, in which the ARCH volatility is multiplied by a leverage function in order to fit a smile and the function $\sigma(X)$ is arbitrary:

$$\frac{dS_t}{S_t} = \sigma(X_t) l(t, S_t)\, dW_t, \qquad X_t = \sum_{t-\Delta < t_i \leq t} r_i^2, \qquad r_i = \frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}}$$

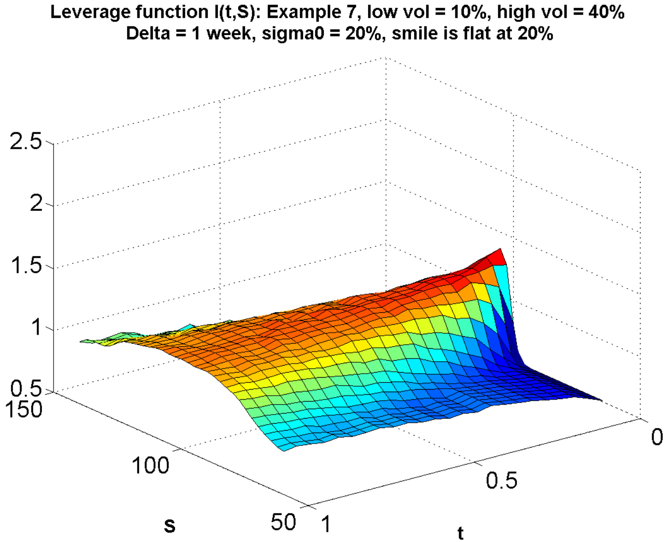
$$\frac{dS_t}{S_t} = \sigma(X_t) l(t, S_t)\, dW_t, \qquad X_t = \sum_{t-\Delta < t_i \le t} r_i^2, \qquad r_i = \frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}}$$

- Ex. 7: $\sigma(t, X) = \underline{\sigma}$ if $X \le \sigma_0$ and $\overline{\sigma}$ otherwise. Vanishing ATM forward skew. Forward starting butterfly spreads cost around 2.4 points of volatility
- Ex. 8 (to mimic ARCH models): $\sigma(X)^2 = \alpha + \beta X$ with $\alpha > 0$, $\beta < 1$. Much flatter pure PDV smile and a much flatter leverage function l. Vanishing ATM forward skew. Forward starting butterfly spreads around 0.7 point of volatility

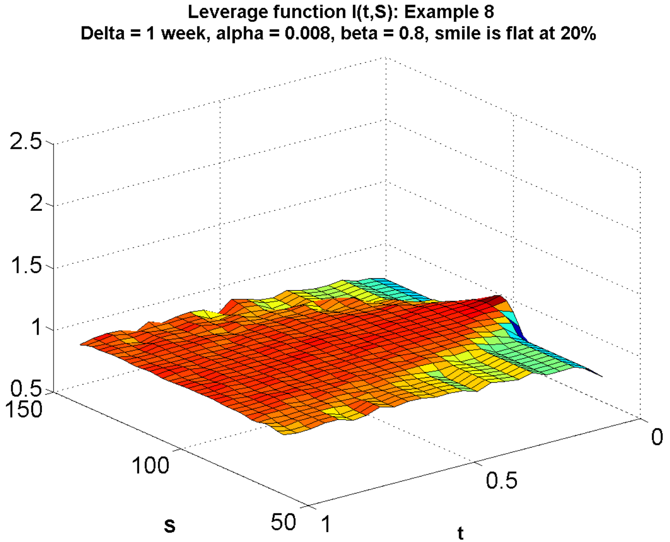
Smiles of pure PDV models



Example 7: Leverage function $l(t, S)$



Example 8: Leverage function $l(t, S)$



Conclusion

- PDV models are excellent candidates to challenge the duopoly of LV and SV which has dominated option pricing for 20 years
- Like the LV model: complete and can be calibrated to the market smile $\Longrightarrow$ all derivatives have a unique price which is consistent with today's prices of vanilla options
- Like SV models: can produce rich spot-vol dynamics, such as large negative short term forward skews or large forward smile curvatures
- Huge flexibility: one can choose any set of path-dependent variables X and any PDV $\sigma(t, S, X) \Longrightarrow$ PDV models actually span a much broader range of spot-vol dynamics than SV models, and can also capture important historical features of asset returns, such as volatility *levels* depending on recent asset *returns*, tail heaviness, volatility clustering, and dependence without correlation

Conclusion

- In practice, the particle method is so simple and efficient that the smile calibration is not a problem $\Longrightarrow$ Efforts can be concentrated on the choice of a convenient PDV, depending on the market and derivative under consideration
- Beyond the ability to produce desired spot-vol dynamics and capture spot-vol historical patterns, an important criterion to assess the quality of a PDV model should be its hedging performance on backtests

Risk magazine, October 2014

Path-dependent volatility

So far, path-dependent volatility models have drawn little attention compared with local volatility and stochastic volatility models. In this article, Julien Guyon shows they combine benefits from both and can also capture prominent historical patterns of volatility

Three main volatility models have been used so far in the finance industry: constant volatility, local volatility (LV) and stochastic volatility (SV). The first two models are complete: since the asset price is driven by a single Brownian motion, every payoff admits a unique self-financing replicating portfolio consisting of cash and the underlying asset. Therefore, its price is uniquely defined as the initial value of the replicating portfolio, independent of utilities or preferences. Unlike the constant volatility models, the LV model is flexible enough to fit any arbitrage-free surface of implied volatilities (henceforth, 'smile'), but then no more flexibility is left. Calibrating to the market smile is useful when one sells an exotic option whose risk is well mitigated by trading vanilla options – then the model correctly prices the hedging instruments at inception.

For their part, SV models are incomplete: the volatility is driven by one of several extra Brownian motions, and as a result perfect replication and price uniqueness are lost. Modifying the drift of the SV leaves the model arbitrage-free, but changes option prices.

Using SV models allows us to gain control of key risk factors such as volatility of volatility (vol-of-vol), forward skew and spot-vol correlation. SV models generate joint dynamics of the asset and its implied

price uniqueness and parsimony: it is remarkable that so many popular properties of SLV models can be captured using a single Brownian motion. Although perfect delta-hedging is unrealistic, incorporating the path-dependency of volatility into the delta is likely to improve the delta-hedge. Not only that, we will see that, thanks to their huge flexibility, PDV models can generate spot-vol dynamics that are not attainable using SLV models.

Below, we first introduce the class of PDV models and then explain how we calibrate them to the market smile. Subsequently, we investigate how to pick a particular PDV.

Path-dependent volatility models

PDV models are those models where the instantaneous volatility σ_t depends on the path followed by the asset price so far:

$$\frac{\mathrm{d}S_t}{S_t} = \sigma(t, (S_u, u \leqslant t))\,\mathrm{d}W_t$$

where, for simplicity, we have taken zero interest rates, repo and dividends. In practice, the volatility $\sigma_t \equiv \sigma(t, S_t, X_t)$ will often be

A few selected references

- Bergomi, L.: *Smile dynamics II*, Risk, October, 2005.
- Bollerslev, T.: *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, 31:307-327, 1986.
- Brunick, G. and Shreve, S.: *Mimicking an Itô process by a solution of a stochastic differential equation*, Ann. Appl. Prob., 23(4):1584–1628, 2013.
- Dupire, B.: *Pricing with a smile*, Risk, January, 1994.
- Engle, R.: *Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation*, Econometrica 50:987-1008, 1982.
- Figà-Talamanca, G. and Guerra, M.L.: *Fitting prices with a complete model*, J. Bank. Finance 30(1), 247?258, 2006.
- Foschi, P, and Pascucci, A.: *Path Dependent Volatility*, Decisions Econ. Finan., 2007.
- Fouque, J.-P., Papanicolaou, G. and Sircar, R.: *Derivatives in Financial Markets with Stochastic Volatility*, Cambridge University Press, 2000.

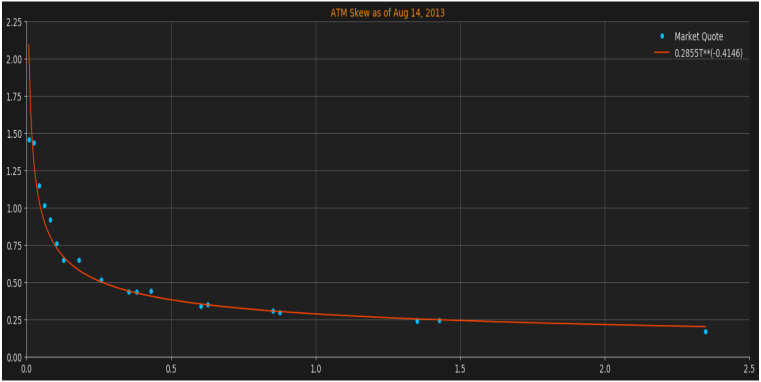
A few selected references

- Guyon, J. and Henry-Labordère, P.: *Being Particular About Calibration*, Risk, January, 2012.
- Guyon, J. and Henry-Labordère, P.: *Nonlinear Option Pricing*, Chapman & Hall/CRC Financial Mathematics Series, 2013.
- Guyon, J.: *Local correlation families*, Risk, February, 2014.
- Hobson, D. G. and Rogers, L. C. G.: *Complete models with stochastic volatility*, Mathematical Finance 8 (1), 27?48, 1998.
- Hubalek, F., Teichmann, J. and Tompkins, R.: *Flexible complete models with stochastic volatility generalising Hobson-Rogers*, working paper, 2004.
- Klüppelberg, C., Lindner, A. and Maller, R.: *A continuous time GARCH process driven by a Lévy process: stationarity and second order behaviour*, J. Appl. Probab., 41(3):601–622, 2004.
- Platania, A. and Rogers, L.C.G.: *Putting the Hobson-Rogers model to the test*, working paper, 2003

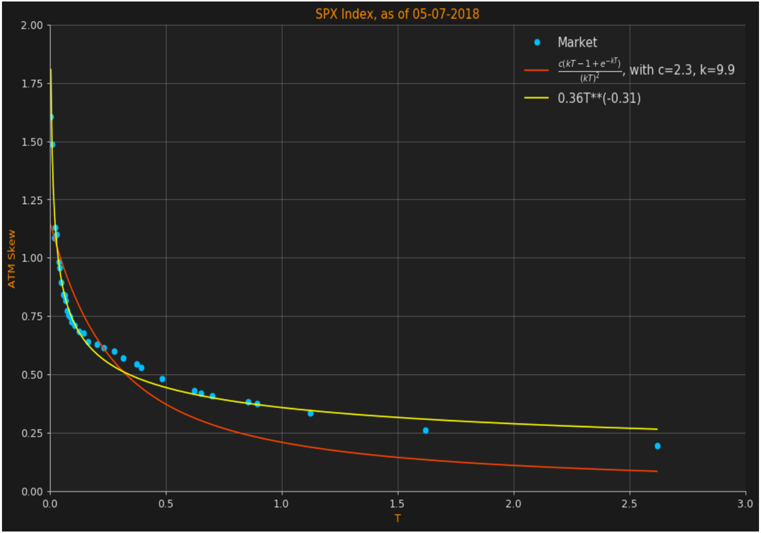
Contents

Term-structure of SPX ATM skew

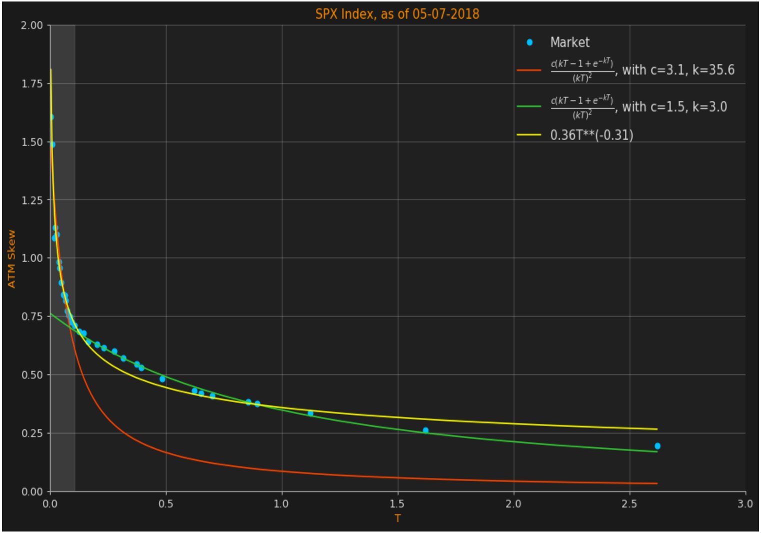
One-factor Bergomi model with large mean reversion and vol-of-vol: $\mathcal{S}_T \sim \frac{1}{T}$.
To mimic or produce a power-law decay $\mathcal{S}_T \sim \frac{1}{T^\alpha}$: 2-factor Bergomi model and rough volatility model.



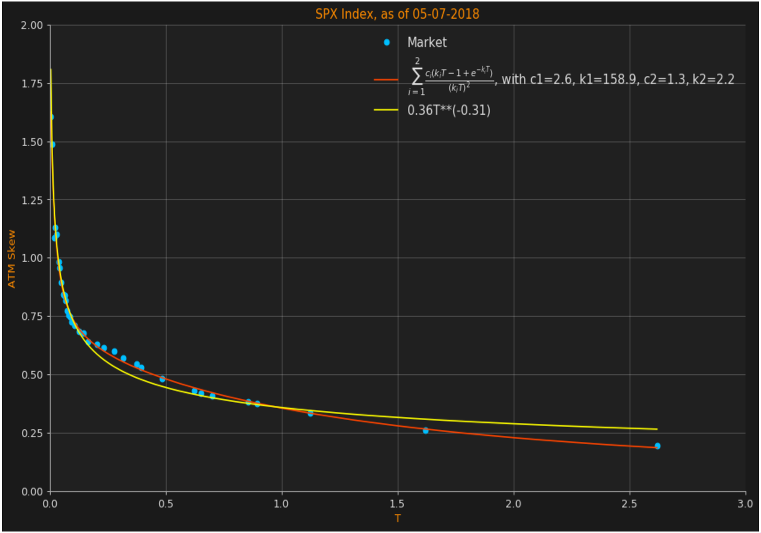
SPX ATM skew, May 7, 2018



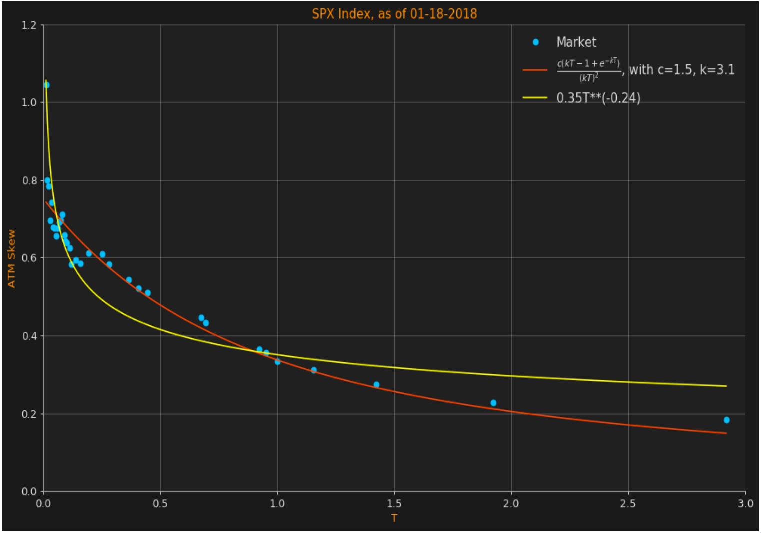
SPX ATM skew, May 7, 2018



SPX ATM skew, May 7, 2018



However... SPX ATM skew, Jan 18, 2018



Power-law kernel

$$K(\theta) = \nu \theta^{H-\frac{1}{2}}, \qquad \nu \geq 0, \qquad H \in \left(0, \frac{1}{2}\right) \tag{29}$$

- Rough Bergomi model (Bayer, Friz, Gatheral, 2014).
- $\lim_{\theta \to 0^+} K(\theta) = +\infty$.
- $H =$ Hurst exponent.

$$\begin{aligned} \xi_t^u &= \xi_0^u \exp\left(\nu X_t^u - \frac{\nu^2}{2}\mathrm{Var}(X_t^u)\right) \\ X_t^u &= \int_0^t (u-s)^{H-\frac{1}{2}}\, dZ_s, \qquad \mathrm{Var}(X_t^u) = \frac{u^{2H} - (u-t)^{2H}}{2H}. \end{aligned}$$

- $H > 0$ ensures that $\mathrm{Var}(X_u^u)$ is finite.

Power-law kernel: $K(\theta) = \nu\theta^{H-\frac{1}{2}}$

- **No Markov representation** for ξ_t^u.
- **Instantaneous variance $\sigma_t^2 := \xi_t^t$ is not a semimartingale**. One cannot write Itô dynamics $d\xi_t^t = \cdots dt + \cdots dZ_t$ for the instantaneous variance. **No notion of a dynamic volatility of instantaneous spot variance**.
- However we can compare the values of $\mathrm{Var}\left(\ln \frac{\xi_t^u}{\xi_0^u}\right)$ in the power-law and exponential kernel models:

$$\nu^2 \frac{u^{2H} - (u-t)^{2H}}{2H} \quad \longleftrightarrow \quad \omega^2 e^{-2k(u-t)} \frac{1-e^{-2kt}}{2k} \tag{30}$$

$$u = t \to 0: \qquad \nu^2 \frac{t^{2H}}{2H} \quad \longleftrightarrow \quad \omega^2 \frac{1-e^{-2kt}}{2k} \approx \omega^2 t \tag{31}$$

$$\nu \frac{t^{H-\frac{1}{2}}}{\sqrt{2H}} \quad \longleftrightarrow \quad \omega \tag{32}$$

- **$\nu \frac{t^{H-\frac{1}{2}}}{\sqrt{2H}}$ can be interpreted as a short term volatility of instantaneous spot variance.**
- $[\nu] = \text{time}^{-H}$; $\left[\nu\theta^{H-\frac{1}{2}}\right] = \left[\nu \frac{t^{H-\frac{1}{2}}}{\sqrt{2H}}\right] =$ vol.

Power-law kernel: $K(\theta) = \nu\theta^{H-\frac{1}{2}}$

- Short-term ATM skew in SV models $\sim \rho\omega$. **Explains why the ATM skew in such rough volatility models behaves like $T^{H-\frac{1}{2}}$ for short maturities T** (Alós, Fukasawa...), which is one of the reasons why this model has been introduced (Gatheral, Jaisson, Rosenbaum, Friz, Bayer).
- In the limit $H \to 0$, for fixed ν, $\nu^2 \frac{t^{2H}}{2H} \to +\infty$ for any $t > 0$.
- In order for $\mathrm{Var}(\sigma_t^2)$ to tend to a finite limit, we must impose that $\frac{\nu^2}{2H}$ tend to a finite limit $\Longrightarrow$ A natural limiting regime, analogous to the **ergodic regime** described above for the exponential kernel, is **$H, \nu \to 0$, with $\frac{\nu^2}{2H}$ kept constant**.
- However in this ergodic limit the SPX skew is $\sim \sqrt{H} T^{H-\frac{1}{2}}$...

Contents

The uncertain volatility model

- We model an asset S_t by a positive local Itô $(\mathcal{F}_t, \mathbb{Q})$-martingale (zero interest rates, repos, and dividends for simplicity):

$$dS_t = \sigma_t S_t \, dW_t$$

- The volatility σ_t is unspecified for the moment.
- In the Uncertain Volatility Model (UVM) introduced independently by Avellaneda-Levy-Paras and Lyons (1995), the volatility is uncertain. As a minimal modeling hypothesis, we only assume that:
  - the volatility is valued in a compact interval $[\underline{\sigma}, \overline{\sigma}]$, and
  - it cannot look into the future (i.e., it is an $(\mathcal{F}_t)$-adapted process).

The uncertain volatility model

- Consider an option delivering a payoff F_T at maturity T which is a function of the asset path $(S_t, 0 \leq t \leq T)$.
- We define the time-t value u_t of the option in the UVM as (worst case pricing):

$$u_t = \sup_{(\sigma_s, s \in [t,T]) \text{ adapted}, \ \sigma_s \in [\underline{\sigma}, \overline{\sigma}]} \mathbb{E}^{\mathbb{Q}}[F_T | \mathcal{F}_t] \tag{33}$$

- This model is useful when we price and hedge options on underlyings that have no liquid options market, e.g., funds or funds of funds. Indeed, in this case, we cannot hedge against volatility movements.
- We *cannot* control the volatility of the underlying. However, the pricing formulation is equivalent to that of the value function in a stochastic control problem.

Optimal control problems and the Hamilton-Jacobi-Bellman PDE

Controlled Diffusion:

$$dX_t^{\alpha,i} = b_i(t, X_t^\alpha, \alpha_t)\, dt + \sum_{j=1}^{d} \sigma_{i,j}(t, X_t^{\alpha,i}, \alpha_t)\, dW_t^j, \qquad i \in \{1, \ldots, n\}$$

$$u(t,x) = \sup_{(\alpha_s, s\in[t,T]) \text{ adapted},\ \alpha_s \in \mathcal{A}} \mathbb{E}^{\mathbb{Q}} \left[\int_t^T f(s, X_s^\alpha, \alpha_s) ds + g(X_T^\alpha) \middle| X_t^\alpha = x \right]$$

- For a fixed control α:

$$\begin{aligned} \partial_t u(t,x) + \mathcal{L}^\alpha u(t,x) + f(t,x,\alpha) &= 0 \\ u(T,x) &= g(x) \end{aligned}$$

- The Hamilton-Jacobi-Bellman PDE is a consequence of Bellman's dynamic programming principle:

$$\begin{aligned} \partial_t u(t,x) + \sup_{\alpha \in \mathcal{A}} \{ f(t,x,\alpha) + \mathcal{L}^\alpha u(t,x) \} &= 0 \\ u(T,x) &= g(x) \end{aligned}$$

Stochastic optimal control problems

- $(\Omega, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{Q})$ denotes a probability space equipped with a d-dimensional Brownian motion W; $(\mathcal{F}_t)$ denotes the natural filtration of W.
- We introduce the following SDE

$$dX_t^{\alpha,i} = b_i(t, X_t^\alpha, \alpha_t)\,dt + \sum_{j=1}^{d} \sigma_{i,j}(t, X_t^\alpha, \alpha_t)\,dW_t^j, \qquad i \in \{1,\dots,n\} \quad (34)$$

where the drift b and the volatility σ depend on t, X_t^α and a control parameter α_t, which is an $\mathcal{F}_t$-adapted process valued in a domain $\mathcal{A}$, not necessarily compact.

- We denote

$$\mathcal{A}_{t,T} = \{(\alpha_s)_{t\le s\le T} \text{ adapted} \mid \forall s \in [t,T],\ \alpha_s \in \mathcal{A}\}$$

Stochastic optimal control problems

Controlled diffusion process

$$dX_t^{\alpha,i} = b_i(t, X_t^\alpha, \alpha_t)\, dt + \sum_{j=1}^d \sigma_{i,j}(t, X_t^\alpha, \alpha_t)\, dW_t^j, \qquad i \in \{1, \dots, n\}$$

with Itô generator

$$\mathcal{L}^a = \sum_{i=1}^n b_i(t,x,a)\partial_i + \frac{1}{2}\sum_{i,j=1}^n \sum_{k=1}^d \sigma_{i,k}(t,x,a)\sigma_{j,k}(t,x,a)\partial_{ij}$$

- In order to ensure that SDE (35) admits a strong solution, we assume that b, σ satisfy:

  Assum(SDE_α**)**: b and σ are Lipschitz-continuous functions in x uniformly in t and α, and satisfy a linear growth condition:

$$\begin{aligned} |b(t,x,\alpha) - b(t,y,\alpha)| + |\sigma(t,x,\alpha) - \sigma(t,y,\alpha)| &\leq C|x-y| \\ |b(t,x,\alpha)| + |\sigma(t,x,\alpha)| &\leq C\,(1+|x|) \end{aligned}$$

  for every $t \geq 0$, $x, y \in \mathbb{R}^n$ and $\alpha \in \mathcal{A}$ where C is a positive constant.
- This condition ensures that

$$\mathbb{E}^{\mathbb{Q}}\left[\sup_{0 \leq s \leq T} |X_s^\alpha|^2\right] < \infty \tag{35}$$

Stochastic optimal control problems

- A standard stochastic optimal control problem consists of maximizing a reward function (or minimizing a cost function) J given by

$$J(t,x,\alpha) := \mathbb{E}^{\mathbb{Q}}\left[\int_t^T f(s,X_s^\alpha,\alpha_s)\,ds + g(X_T^\alpha)\middle| X_t^\alpha = x\right]$$

  with respect to the control $\alpha \in \mathcal{A}_{t,T}$:

$$u(t,x) := \sup_{\alpha\in\mathcal{A}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[\int_t^T f(s,X_s^\alpha,\alpha_s)\,ds + g(X_T^\alpha)\middle| X_t^\alpha = x\right] \qquad (36)$$

  where the initial condition is $X_t^\alpha = x$.

- In order to ensure that the reward function is finite, we assume that the supremum is taken over the subset $\mathcal{A}^*_{t,T}$ of $\mathcal{A}_{t,T}$ satisfying

$$\mathbb{E}^{\mathbb{Q}}\left[\int_t^T |f(s,X_s^\alpha,\alpha_s)|\,ds + |g(X_T^\alpha)|\right] < \infty \qquad (37)$$

Stochastic optimal control problems: standard form ($f = 0$)

- By augmenting the state variables X_t^α with the path-dependent continuous variable

$$dZ_t = f(t, X_t^\alpha, \alpha_t)\, dt, \qquad Z_0 = 0$$

our stochastic control problem can be written without loss of generality as

$$u(t, \tilde{x}) = \sup_{\alpha \in \mathcal{A}^*_{t,T}} \mathbb{E}^{\mathbb{Q}}[\tilde{g}(\tilde{X}_T^\alpha) | \tilde{X}_t^\alpha = (x, z)] - z$$

with $\tilde{X}^\alpha = (X^\alpha, Z)$ and $\tilde{g}(\tilde{x}) = g(x) + z$.

- We shall appeal to this standard form (i.e., $f = 0$)

$$u(t, x) := \sup_{\alpha \in \mathcal{A}^*_{t,T}} \mathbb{E}^{\mathbb{Q}}[g(X_T^\alpha) | X_t^\alpha = x] \tag{38}$$

to review results on stochastic control.

- **Assum(g)**: g has quadratic growth, i.e., there exists $C \geq 0$ such that for all $x \in \mathbb{R}^n$, $|g(x)| \leq C \left(1 + |x|^2\right)$.
- From the integrability result (35), this assumption on g implies that $\mathcal{A}^*_{t,T} = \mathcal{A}_{t,T}$.

Bellman's principle

Bellman's principle, also known as dynamic programming principle (DPP):

$$u(t,x) = \sup_{\alpha \in \mathcal{A}_{t,t+h}} \mathbb{E}^{\mathbb{Q}}[u(t+h, X^{\alpha}_{t+h}) | X^{\alpha}_t = x] \tag{39}$$

Bellman's principle also holds for any stopping time τ with $t \leq \tau \leq T$, $\mathbb{Q}-$a.s.

$$u(t,x) = \sup_{\alpha \in \mathcal{A}_{t,\tau}} \mathbb{E}^{\mathbb{Q}}[u(\tau, X^{\alpha}_{\tau}) | X^{\alpha}_t = x]$$

Proof.

(i). Using iterated conditional expectation, one gets

$$u(t,x) = \sup_{\alpha \in \mathcal{A}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}\left[g\left(X^{\alpha}_T\right) \middle| \mathcal{F}_{t+h} \right] \middle| X^{\alpha}_t = x \right]$$

By the Markov property of X^{α} and the definition of u, one has

$$\mathbb{E}^{\mathbb{Q}}\left[g\left(X^{\alpha}_T\right) \middle| \mathcal{F}_{t+h} \right] = \mathbb{E}^{\mathbb{Q}}\left[g\left(X^{\alpha}_T\right) \middle| X^{\alpha}_{t+h} \right] \leq u\left(t+h, X^{\alpha}_{t+h}\right), \quad \forall \alpha \in \mathcal{A}_{t+h,T}.$$

Hence

$$u(t,x) \leq \sup_{\alpha \in \mathcal{A}_{t,t+h}} \mathbb{E}^{\mathbb{Q}}\left[u\left(t+h, X^{\alpha}_{t+h}\right) \middle| X^{\alpha}_t = x \right].$$

Bellman's principle

(ii). For any $\alpha \in \mathcal{A}_{t,t+h}$, there is an $\alpha^\epsilon \in \mathcal{A}_{t+h,T}$ such that

$$u\left(t+h, X^\alpha_{t+h}\right) \leq \mathbb{E}^{\mathbb{Q}}\left[g\left(X_T^{\alpha^\epsilon}\right) \middle| X^\alpha_{t+h}\right] + \epsilon.$$

We glue together α and α^ϵ:

$$\tilde{\alpha}^\epsilon_s = \begin{cases} \alpha_s & \text{if } s \in [t, t+h) \\ \alpha^\epsilon_s & \text{if } s \in [t+h, T] \end{cases}.$$

Then $\tilde{\alpha}^\epsilon_s \in \mathcal{A}_{t,T}$ and

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[u\left(t+h, X^\alpha_{t+h}\right) \middle| X^\alpha_t = x\right] &\leq \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}\left[g\left(X_T^{\alpha^\epsilon}\right) \middle| X^\alpha_{t+h}\right] \middle| X^\alpha_t = x\right] + \epsilon \\
&= \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}\left[g\left(X_T^{\alpha^\epsilon}\right) \middle| \mathcal{F}_{t+h}\right] \middle| X^\alpha_t = x\right] + \epsilon \\
&= \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}\left[g\left(X_T^{\tilde{\alpha}^\epsilon}\right) \middle| \mathcal{F}_{t+h}\right] \middle| X^\alpha_t = x\right] + \epsilon \\
&= \mathbb{E}^{\mathbb{Q}}\left[g\left(X_T^{\tilde{\alpha}^\epsilon}\right) \middle| X^\alpha_t = x\right] + \epsilon \leq u(t,x) + \epsilon.
\end{aligned}$$

Letting $\epsilon \to 0$, we obtain the reverse inequality.

Formal derivation of the HJB PDE

- Let us take an arbitrary constant control $\alpha_s = a$ with $a \in \mathcal{A}$ during the interval $[t, t+h]$. From the Bellman principle, we get

$$u(t,x) \geq \mathbb{E}^{\mathbb{Q}}[u(t+h, X^a_{t+h}) | X^a_t = x]$$

- By applying Itô's lemma to $u(t+h, X^\alpha_{t+h})$ (u should be smooth enough, $u \in C^{1,2}([0,T) \times \mathbb{R}^n)$), we get

$$u(t,x) \geq \mathbb{E}^{\mathbb{Q}}\left[u(t,x) + \int_t^{t+h} \left(\partial_s + \mathcal{L}^a\right) u(s, X^a_s)\, ds + M \middle| X^a_t = x\right]$$

where $M := \int_t^{t+h} Du(s, X^a_s)\sigma(s, X^a_s, a)\, dW_s$ and $\mathcal{L}^a$ is the Itô generator associated to X_t (35) controlled by the constant $a \in \mathcal{A}$.

- By assuming that M is not only a local martingale but a true martingale, we obtain

$$\mathbb{E}^{\mathbb{Q}}\left[\int_t^{t+h} \left(\partial_s + \mathcal{L}^a\right) u(s, X_s)\, ds \middle| X_t = x\right] \leq 0$$

Dividing by $h > 0$, letting $h \to 0$, and permuting the limit and the expectation using the dominated convergence theorem, we finally get

$$\partial_t u(t,x) + \mathcal{L}^a u(t,x) \leq 0, \qquad \forall a \in \mathcal{A} \tag{40}$$

Formal derivation of the HJB PDE

- Then we take the optimal control α^* that realizes the supremum in Equation (39). By following a similar route, we get

$$\partial_t u(t,x) + \mathcal{L}^{\alpha_t^*} u(t,x) = 0 \qquad (41)$$

- By combining Equations (40) and (41), we get that u, as given by Equation (38), is a solution to the following parabolic second order fully nonlinear PDE, the so-called HJB equation:

$$\partial_t u(t,x) + \sup_{a \in \mathcal{A}} \mathcal{L}^a u(t,x) = 0, \qquad u(T,x) = g(x) \qquad (42)$$

- The interpretation is clear: For a deterministic constant "control" a, the linear PDE reads

$$-\partial_t u(t,x) = \mathcal{L}^a u(t,x), \qquad u(T,x) = g(x)$$

  At each time, we actually have the possibility to choose a in $\mathcal{A}$ so as to maximize $u(0,\cdot)$. Since u is known at the final date, $a \in \mathcal{A}$ must be chosen so that $-\partial_t u$ is maximized. This yields

$$-\partial_t u(t,x) = \sup_{a \in \mathcal{A}} \mathcal{L}^a u(t,x), \qquad u(T,x) = g(x)$$

  which is Equation (42).

HJB PDE with source term f and discount factor

- For our initial stochastic control problem with a source term (36), the HJB PDE reads

$$\partial_t u(t,x) + \sup_{a\in\mathcal{A}}\{\mathcal{L}^a u(t,x) + f(t,x,a)\} = 0, \qquad u(T,x) = g(x) \tag{43}$$

- When we include a discount factor:

$$\begin{aligned} u(t,x) = \sup_{\alpha\in\mathcal{A}_{t,T}} \mathbb{E}^{\mathbb{Q}}\Big[g(X_T^\alpha)e^{-\int_t^T r(s,X_s^\alpha,\alpha_s)\,ds} + \\ \int_t^T e^{-\int_t^s r(u,X_u^\alpha,\alpha_u)du} f(s,X_s^\alpha,\alpha_s)\,ds \Big| X_t = x\Big] \end{aligned} \tag{44}$$

  the HJB PDE reads

$$\begin{aligned} \partial_t u(t,x) + \sup_{a\in\mathcal{A}}\{\mathcal{L}^a u(t,x) + f(t,x,a) - r(t,x,a)u(t,x)\} &= 0 \quad (45) \\ u(T,x) &= g(x) \end{aligned}$$

- Note that if the control can only be chosen at discrete dates, the HJB equation is not applicable because it assumes that the control can be chosen continuously in time. In these cases, we must rely on the (discrete) Bellman's dynamic programming equation.

Derivation of HJB PDE directly with source term f

- Bellman's dynamic programming principle:

$$u(t,x) = \sup_{\alpha_{s\in[t,t+h]}\in\mathcal{A}} \mathbb{E}^{\mathbb{P}}\left[\int_t^{t+h} f(s,X_s^\alpha,\alpha_s)\,ds + u(t+h,X_{t+h}^\alpha)\middle| X_t^\alpha = x\right]$$

- For any constant control a on $[t,t+h]$,

$$u(t,X_t^a) \geq \mathbb{E}^{\mathbb{Q}}\left[\int_t^{t+h} f(s,X_s^a,a)\,ds + u(t+h,X_{t+h}^a)\middle| X_t^a = x\right]$$

$$= \mathbb{E}^{\mathbb{Q}}\left[\int_t^{t+h} f(s,X_s^a,a)\,ds + u(t,X_t^a) + \int_t^{t+h} (\partial_s + \mathcal{L}^a)\,u(s,X_s^a)\,ds + M\middle| X_t^a = x\right]$$

$$0 \geq \mathbb{E}^{\mathbb{Q}}\left[\int_t^{t+h} f(s,X_s^a,a)\,ds + \int_t^{t+h} (\partial_s + \mathcal{L}^a)\,u(s,X_s^a)\,ds\middle| X_t^a = x\right]$$

Letting $h \to 0$, one gets

$$f(t,x,a) + (\partial_t + \mathcal{L}^a)\,u(t,x) \leq 0.$$

- For the optimal control α^*, the inequality becomes equality:

$$f(t,x,\alpha_t^*) + \left(\partial_t + \mathcal{L}^{\alpha_t^*}\right) u(t,x) = 0$$

$$\partial_t u(t,x) + \sup_{\alpha\in\mathcal{A}} \{f(t,x,\alpha) + \mathcal{L}^\alpha u(t,x)\} = 0$$

Verification

Theorem (Verification Theorem)

Let $U \in C^{1,2}([0,T) \times \mathbb{R}^n) \cap C^0([0,T] \times \mathbb{R}^n)$ be a function with quadratic growth in x, uniformly in t, and for all $(t,x) \in [0,T) \times \mathbb{R}^n$, there exists $\alpha^*(t,x) \in \mathcal{A}$ such that

$$-\partial_t U(t,x) - \sup_{\alpha \in \mathcal{A}} \{\mathcal{L}^\alpha U(t,x)\} = -\partial_t U(t,x) - \mathcal{L}^{\alpha^*} U(t,x) = 0$$

and that the SDE with the control α^* admits a strong solution $X_t^{\alpha^*}$. We also assume that $U(T,\cdot) = g$. Then $U = u$ in $[0,T] \times \mathbb{R}^n$.

- For any admissible control α, Itô's lemma gives
$$U(T, X_T^\alpha) = U(t,x) + \int_t^T (\partial_s U + \mathcal{L}^\alpha U(s, X_s^\alpha))\, ds + \text{martingale}.$$
Taking expectation $\mathbb{E}[\cdot | X_t^\alpha = x]$ yields $U(t,x) \geq \sup_\alpha \mathbb{E}\left[g(X_T^\alpha) \middle| X_t^\alpha = x\right]$. Setting $\alpha = \alpha^*$ one has $U(t,x) = \mathbb{E}\left[g(X_T^{\alpha^*}) \middle| X_t^{\alpha^*} = x\right]$. Thus $U = u$.
- For non-smooth solutions, consider viscosity solutions.

The uncertain volatility model: vanilla options

- For vanilla payoffs $F_T = g(S_T)$, the price $u(t,x)$ is solution to a nonlinear PDE:

$$\partial_t u(t,x) + \frac{1}{2} \sup_{\sigma \in [\underline{\sigma}, \overline{\sigma}]} \sigma^2 x^2 \partial_x^2 u(t,x) = 0, \qquad u(T,x) = g(x)$$

- Taking the supremum over σ, we get a fully nonlinear PDE called the Black-Scholes-Barenblatt equation (in short BSB)

$$\begin{aligned} \partial_t u(t,x) + \frac{1}{2} x^2 \Sigma\left(\partial_x^2 u(t,x)\right)^2 \partial_x^2 u(t,x) &= 0, \quad (t,x) \in [0,T) \times \mathbb{R}_+^* \\ u(T,x) &= g(x), \quad x \in \mathbb{R}_+^* \end{aligned}$$

  with $\Sigma(\Gamma) = \overline{\sigma} 1_{\Gamma \geq 0} + \underline{\sigma} 1_{\Gamma < 0}$.

- For a convex payoff, u coincides with the Black-Scholes price with the upper volatility $\overline{\sigma}$.

Solving the pricing equation

1D Black-Scholes-Barenblatt (BSB) PDE:

$$\partial_t u(t,x) + \frac{1}{2} x^2 \Sigma(\partial_x^2 u)^2 \partial_x^2 u(t,x) = 0, \quad (t,x) \in [0,T) \times \mathbb{R}_+^*$$

with $u(T,x) = g(x)$ and $\Sigma(\Gamma) = \overline{\sigma} 1_{\Gamma \geq 0} + \underline{\sigma} 1_{\Gamma < 0}$

- No analytical solution
- Finite-difference schemes
- We must turn to Monte Carlo methods when the number of variables (underlyings and path-dep variables) is 3 or more. We will present some MC methods in this course, based on Backward Stochastic Differential Equations (BSDEs). Need to estimate conditional expectations $\mathbb{E}[Y|X=x] \longrightarrow$ ML techniques

The uncertain volatility model: A finite difference scheme

In practice, this nonlinear PDE is not analytically solvable. In low dimension, we can implement a finite difference scheme. By setting $z = \ln \frac{x}{S_0}$, we consider the following θ-scheme:

1. Set $u_i^N = g(S_0 e^{z_i})$.
2. Predictor:
$$\frac{u_i^{k+1} - u_i^k}{\Delta t} + \frac{1}{2}\Sigma(\Gamma_i^{k+1})^2 \left(\theta \Gamma_i^{k+1} + (1-\theta)\Gamma_i^{k,(1)}\right) = 0$$
with $\Gamma_i^{k+1} = \frac{u_{i+1}^{k+1} + u_{i-1}^{k+1} - 2u_i^{k+1}}{\Delta z^2} - \frac{u_{i+1}^{k+1} - u_{i-1}^{k+1}}{2\Delta z}$ and $\Gamma_i^{k,(1)} = \frac{u_{i+1}^{k} + u_{i-1}^{k} - 2u_i^{k}}{\Delta z^2} - \frac{u_{i+1}^{k} - u_{i-1}^{k}}{2\Delta z}$. This scheme is explicit in $\Sigma(\Gamma_i^{k+1})$.
3. Corrector: Set $\Gamma_i^k = \frac{1}{2}\left(\Gamma_i^{k,(1)} + \Gamma_i^{k+1}\right)$ and solve
$$\frac{u_i^{k+1} - u_i^k}{\Delta t} + \frac{1}{2}\Sigma(\Gamma_i^{k})^2 \left(\theta \Gamma_i^{k+1} + (1-\theta)\Gamma_i^{k,(1)}\right) = 0$$

Robust super-replication

- We consider the set $\mathcal{P}$ of all measures $\mathbb{Q}$ under which S_t is a local martingale such that

$$\underline{\sigma}^2 \leq \frac{d\langle \ln S \rangle_t}{dt} \equiv \sigma_t^2 \leq \overline{\sigma}^2$$

- The seller's super-replication price in the UVM is defined by

$$\mathcal{S}_t(F_T) = \inf \Big\{ z \in \mathcal{F}_t \;\Big|\; \text{there exists an admissible portfolio } \Delta \text{ such that } \pi_T^S \equiv z + \int_t^T \Delta_s \, dS_s - F_T \geq 0 \quad \mathbb{Q}\text{−a.s} \quad \forall \mathbb{Q} \in \mathcal{P} \Big\}$$

Theorem

$$\mathcal{S}_t(F_T) \quad = \quad \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}^{\mathbb{Q}}[F_T|\mathcal{F}_t] \quad = \quad \sup_{[t,T]} \mathbb{E}[F_T|\mathcal{F}_t]$$

This theorem justifies our definition of the UVM price.

Robust super-replication: proof

Proof. We assume that $F_T = g(S_T)$ with $g \in C_b^3(\mathbb{R}_+)$ and $\overline{\sigma} < \infty$.

(i) We set $z = u(t, S_t) = \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}^{\mathbb{Q}}[g(S_T)|\mathcal{F}_t]$ and $\Delta_s = \partial_x u(s, S_s)$ with u the solution of the BSB PDE. As $u \in C^{1,2}([0,T) \times \mathbb{R}_+)$, Itô's lemma gives

$$z + \int_t^T \Delta_s \, dS_s - g(S_T) = \int_t^T \Big(- \partial_s u(s, S_s) - \frac{1}{2} S_s^2 \sigma_s^2 \partial_x^2 u(s, S_s) \Big) \, ds$$

Using the BSB PDE, we get

$$z + \int_t^T \Delta_s \, dS_s - g(S_T) = \frac{1}{2} \int_t^T S_s^2 \partial_x^2 u(s, S_s) (\Sigma(\partial_x^2 u(s, S_s))^2 - \sigma_s^2) \, ds \geq 0$$

$\mathbb{Q}$-a.s. for all $\mathbb{Q} \in \mathcal{P}$ as $\left(\Sigma \left(\partial_x^2 u(s, S_s) \right)^2 - \sigma_s^2 \right) \partial_x^2 u(s, S_s) \geq 0$ for all adapted σ_s such that for all $s \in [t, T]$, $\sigma_s \in [\underline{\sigma}, \overline{\sigma}]$. This implies that $\mathcal{S}_t(g(S_T)) \leq u(t, S_t) := \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}^{\mathbb{Q}}[g(S_T)|\mathcal{F}_t]$ as we have obtained that the delta-hedging strategy $(z = u(t, S_t), \Delta_s = \partial_x u(s, S_s))$ super-replicates the payoff.

Robust super-replication: proof

(ii) (Weak duality) Let z be such that there exists an admissible portfolio Δ such that

$$z + \int_t^T \Delta_s \, dS_s - g(S_T) \geq 0 \quad \mathbb{Q}\text{−a.s} \quad \forall \mathbb{Q} \in \mathcal{P}$$

By taking the conditional expectation we obtain $z \geq \mathbb{E}^{\mathbb{Q}}[g(S_T)|\mathcal{F}_t]$ for all $\mathbb{Q} \in \mathcal{P}$. We have used that the local martingale $\int_t^T \Delta_s \, dS_s$ bounded from below is a supermartingale. This implies that

$$\mathcal{S}_t(g(S_T)) \geq \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}^{\mathbb{Q}}[g(S_T)|\mathcal{F}_t]$$

Selected references

- Andersen, L. (1999): *A Simple Approach to the pricing of Bermudan Swaptions in the Multi-Factor Libor Market Model* : Journal of Computational Finance, 3:5–32.
- Avellaneda, M., Levy, A., Paras, A. (1995): *Pricing and hedging derivative securities in markets with uncertain volatilities* : Journal of Applied Finance, Vol 1.
- Bouchard, B., Touzi, N. (2004): *Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations*, Stochastic Processes and their Applications, 111, 175–206.
- Cheridito, P., Soner, M., Touzi, N., Victoir, N. (2007): *Second Order Backward Stochastic Differential Equations and Fully Non-Linear Parabolic PDEs*, Communications on Pure and Applied Mathematics, Volume 60, Issue 7, 1081–1110.

Selected references

- El Karoui, N., Peng, S., Quenez, M.C. (1997): *Backward Stochastic Differential Equations in Finance*, Mathematical Finance, Vol. 47, No. 1, 1–77.
- Guyon, J., Henry-Labordère, P. (2011) : *Uncertain Volatility Model: A Monte Carlo approach*, Journal of Computational Finance, 14(3):37–71.
- Lyons, T.J. (1995): *Uncertain volatility and the risk-free synthesis of derivatives*, Applied Mathematical Finance, 2(2):117–133.
- Pardoux, E., Peng, S. (1992): *Backward Stochastic Differential Equations and Quasilinear Parabolic Partial Differential Equations*, Lecture Notes in CIS, vol. 176. Springer-Verlag, 200–217.

Contents