Developments in Volatility Derivatives Pricing

Jim Gatheral

Merrill Lynch

Global Derivatives 2007
Paris, May 23, 2007
Motivation

- We would like to be able to price consistently at least
  1. options on SPX
  2. options on VIX
- Over the last year, we have seen increased liquidity in both options on variance and options on VIX
  - so now we have decent market price data
- Which dynamics are consistent with market prices?
Outline

1. Historical development
   - Problems with one-factor stochastic volatility models.
   - Motivations for adding another factor.
   - Historical attempts to add factors.

2. Variance curve models
   - Dupire’s single-factor example.
   - Bergomi’s variance curve model.
   - Buehler’s consistent variance curve functionals.

3. Fitting the market
   - Comparison of Double Heston and Double Lognormal fits

4. Conclusion
Problems with one-factor stochastic volatility models

- All volatilities depend only on the instantaneous variance $\nu$
  - Any option can be hedged perfectly with a combination of any other option plus stock
  - Skew, appropriately defined, is constant
- We know from PCA of volatility surface time series that there are at least three important modes of fluctuation:
  - level, term structure, and skew
- It makes sense to add at least one more factor.
Other motivations for adding another factor

- Adding another factor with a different time-scale has the following benefits:
  - One-factor stochastic volatility models generate an implied volatility skew that decays as $1/T$ for large $T$. Adding another factor generates a term structure of the volatility skew that looks more like the observed $1/\sqrt{T}$.
  - The decay of autocorrelations of squared returns is exponential in a one-factor stochastic volatility model. Adding another factor makes the decay look more like the power law that we observe in return data.
  - Variance curves are more realistic in the two-factor case. For example, they can have humps.
Historical attempts to add factors

- Dupire’s unified theory of volatility (1996)
  - Local variances are driftless in the butterfly measure.
  - We can impose dynamics on local variances.

- Stochastic implied volatility (1998)
  - The implied volatility surface is allowed to move.
  - Under diffusion, complex no-arbitrage condition, impossible to work with in practice.

- Variance curve models (1993-2005)
  - Variances are tradable!
  - Simple no-arbitrage condition.
**Dupire’s unified theory of volatility**

- The price of the calendar spread $\partial_T C(K, T)$ expressed in terms of the butterfly $\partial_{K,K} C(K, T)$ is a martingale under the measure $Q_{K,T}$ associated with the butterfly.

- Local variance $\nu_L(K, T)$ is given by (twice) the current ratio of the calendar spread to the butterfly.

- We may impose any dynamics such that the above holds and local variance stays non-negative.

- For example, with one-factor lognormal dynamics, we may write:

\[
\nu(S, t) = \nu_L(S, t) \frac{\exp \left\{ - \frac{b^2}{2} t - b W_t \right\}}{\mathbb{E} \left[ \exp \left\{ - \frac{b^2}{2} t - b W_t \right\} \mid S_t = S \right]} 
\]

where it is understood that $\nu_L(\cdot)$ is computed at time $t = 0$. Note that the denominator is hard to compute!
Stochastic implied volatility

• The evolution of implied volatilities is modeled directly as in
  \[ \sigma_{BS}(k, T, t) = G(z; k, T - t) \] with \( z = \{ z_1, z_2, ..., z_n \} \) for some factors \( z_i \).
  
  • For example, the stochastic factors \( z_i \) could represent level, term structure and skew.

• The form of \( G(\cdot) \) is highly constrained by no-arbitrage conditions
  
  • An option is valued as the risk-neutral expectation of future cashflows – it must therefore be a martingale.
  
  • Even under diffusion assumptions, the resulting no-arbitrage condition is very complicated.

• Nobody has yet written down an arbitrage-free solution to a stochastic implied volatility model that wasn’t generated from a conventional stochastic volatility model.
  
  • SABR is a stochastic implied volatility model, albeit without mean reversion, but it’s not arbitrage-free.

• Stochastic implied volatility is a dead end!
Why model variance swaps?

- Dupire’s UTV is hard to implement because local variances are not tradable.
- Stochastic implied volatility isn’t practical because implied volatilities are not tradable.
- Variance swaps are tradable.
  - Variance swap prices are martingales under the risk-neutral measure.
  - Moreover variance swaps are now relatively liquid and forward variance swaps are natural hedges for cliquets and other exotics.
- Thus, as originally suggested by Dupire in 1993, and then latterly by Duanmu, Bergomi, Buehler and others, we should impose dynamics on forward variance swaps.
Modelsing forward variance

Denote the variance curve as of time $t$ by

$$\hat{W}_t(T) = \mathbb{E} \left[ \int_0^T v_s \, ds \mid \mathcal{F}_t \right].$$

The forward variance $\zeta_t(T) := \mathbb{E} [v_T \mid \mathcal{F}_t]$ is given by

$$\zeta_t(T) = \partial_T \hat{W}_t(T)$$

A natural way of satisfying the martingale constraint whilst ensuring positivity is to impose lognormal dynamics as in Dupire’s (1993) example:

$$\frac{d\zeta_t(T)}{\zeta_t(T)} = \sigma(T - t) \, dW_t$$

for some volatility function $\sigma(\cdot)$. Lorenzo Bergomi does this and extends the idea to n-factors.
Bergomi’s model

In the 2-factor version of his model, we have

\[
\frac{d\zeta_t(T)}{\zeta_t(T)} = \xi_1 e^{-\kappa(T-t)} \, dW_t + \xi_2 e^{-c(T-t)} \, dZ_t
\]

This has the solution

\[
\zeta_t(T) = \zeta_0(T) \exp \left\{ \xi_1 e^{-\kappa(T-t)} X_t + \xi_2 e^{-c(T-t)} Y_t + \text{drift terms} \right\}
\]

with

\[
X_t = \int_0^t e^{-\kappa(t-s)} \, dW_s; \quad Y_t = \int_0^t e^{-c(t-s)} \, dZ_s;
\]

Thus, both \(X_t\) and \(Y_t\) are Ornstein-\(\ddot{\text{U}}\)hlenbeck processes. In particular, they are easy to simulate. The Bergomi model is a market model: \(\mathbb{E} [\zeta_t(T)] = \zeta_0(T)\) for any given initial forward variance curve \(\zeta_0(T)\).
Variance curve models

- The idea (similar to the stochastic implied volatility idea) is to obtain a factor model for forward variance swaps. That is,

\[ \zeta_t(T) = G(z; T - t) \]

with \( z = \{z_1, z_2, ..., z_n\} \) for some factors \( z_j \) and some variance curve functional \( G(\cdot) \).

- Specifically, we want \( z \) to be a diffusion so that

\[ d\zeta_t = \mu(\zeta_t) \, dt + \sum_{j}^{d} \sigma^j(\zeta_t) \, dW^j_t \quad (1) \]

- Note that both \( \mu \) and \( \sigma \) are \( n \)-dimensional vectors.
Buehler’s consistency condition

**Theorem**

The variance curve functional $G(z_t, \tau)$ is consistent with the dynamics (1) if and only if

$$
\partial \tau G(z; \tau) = \sum_{i=1}^{n} \mu_i(z) \partial z_i G(z; \tau)
$$

$$
+ \frac{1}{2} \sum_{i,k=1}^{n} \left( \sum_{j=1}^{d} \sigma_i^j(z) \sigma_k^j(z) \right) \partial z_i, z_k G(z; \tau)
$$

To get the idea, apply Itô’s Lemma to $\zeta_t(T) = G(z, T-t)$ with $dz = \mu \, dt + \sigma \, dW$ to obtain

$$
\mathbb{E} \left[ d\zeta_t(T) \right] = 0 = \left\{-\partial \tau G(z, \tau) + \mu \partial z G(z, \tau) + \frac{1}{2} \sigma^2 \partial z, z G(z, \tau)\right\} dt
$$
Example: The Heston model

- In the Heston model, \( G(\nu, \tau) = \nu + (\nu - \bar{\nu}) e^{-\kappa \tau}. \)
  - This variance curve functional is obviously consistent with Heston dynamics with time-independent parameters \( \kappa, \rho \) and \( \eta. \)

- Imposing the consistency condition, Buehler shows that the mean reversion rate \( \kappa \) cannot be time-dependent.

- By imposing a similar martingale condition on forward entropy swaps, Buehler further shows that the product \( \rho \eta \) of correlation and volatility of volatility cannot be time-dependent.
Consider the following variance curve functional:

\[ G(z; \tau) = z_3 + (z_1 - z_3) e^{-\kappa \tau} + (z_2 - z_3) \frac{\kappa}{\kappa - c} \left( e^{-c \tau} - e^{-\kappa \tau} \right) \]

- This looks like the Svensson parametrization of the yield curve.
- The short end of the curve is given by \( z_1 \) and the long end by \( z_3 \).
- The middle level \( z_2 \) adds flexibility permitting for example a hump in the curve.
Buehler’s affine variance curve functional is consistent with double mean reverting dynamics of the form:

\[
\frac{dS}{S} = \sqrt{v} \, dW
\]

\[
dv = -\kappa (v - v') \, dt + \eta_1 v^\alpha \, dZ_1
\]

\[
dv' = -c (v' - z_3) \, dt + \eta_2 v'^\beta \, dZ_2
\]

for any choice of \( \alpha, \beta \in [1/2, 1] \).

- We will call the case \( \alpha = \beta = 1/2 \) Double Heston,
- the case \( \alpha = \beta = 1 \) Double Lognormal,
- and the general case Double CEV.

All such models involve a short term variance level \( v \) that reverts to a moving level \( v' \) at rate \( \kappa \). \( v' \) reverts to the long-term level \( z_3 \) at the slower rate \( c < \kappa \).
Check of consistency condition

- Because $G(\cdot)$ is affine in $z_1$ and $z_2$, we have that

$$\partial z_i, z_j G (\{z_1, z_2\}; \tau) = 0 \quad i, j \in \{1, 2\}.$$ 

- Then the consistency condition reduces to

$$\partial \tau G (\{z_1, z_2\}; \tau) = \sum_{i=1}^{2} \mu_i (\{z_1, z_2\}) \partial z_i G (\{z_1, z_2\}; \tau)$$

$$= -\kappa (z_1 - z_2) \partial z_1 G - c (z_2 - z_3) \partial z_2 G$$

- It is easy to verify that this holds for our affine functional.
- In fact, the consistency condition looks this simple for affine variance curve functionals with any number of factors!
Dufresne’s trick for computing moments

Dufresne (2001) shows how to compute any desired moment of the state variables in the Heston model through repeated application of Itô’s Lemma.

For example, suppose we want to compute the second moment of integrated variance $W_T := \int_0^T \nu_t \, dt$. We first note that

$$dW_t = \nu_t \, dt$$

Then,

$$d(W_t)^2 = 2 \, W_t \, \nu_t \, dt$$

so

$$\mathbb{E} \left[ (W_t)^2 \right] = 2 \int_0^t \mathbb{E} \left[ W_s \, \nu_s \right] \, ds$$
We may repeat this procedure to compute \( \mathbb{E} [ W_t v_t ] \). Specifically, applying Itô’s Lemma,

\[
d(W_t v_t) = W_t dv_t + v_t dW_t = W_t \left[ -\kappa (v_t - \bar{v}) dt + \eta \sqrt{v_t} dZ \right] + v_t^2 dt
\]

Thus

\[
\mathbb{E} [ W_t v_t ] = \kappa \bar{v} \int_0^t e^{-\kappa (t-s)} \mathbb{E} [ W_s ] ds + \int_0^t e^{-\kappa (t-s)} \mathbb{E} [ v_s^2 ] ds
\]

We can apply Itô’s Lemma once more to find \( \mathbb{E} [ v_t^2 ] \) and integrate to get our result.

- This trick will also work for
  - the Double Heston model (so long as \( \mathbb{E} [dZ_1 dZ_2] = 0 \))
  - for the Double Lognormal model (even if \( \mathbb{E}[dZ_1 dZ_2] \neq 0 \)).
A digression: formulations of lognormal stochastic volatility

There are at least two obvious ways of writing down a lognormal stochastic volatility model:

\[ dv = -\kappa (v - \bar{v}) \, dt + \xi v \, dZ \]  \hspace{1cm} (2)

and

\[ d(\log v) = -\kappa (\log v - \theta) \, dt + \xi \, dZ \]  \hspace{1cm} (3)

(2) allows for easy computation of moments, including moments of integrated variance, using Dufresne’s trick. On the other hand, with the Ornstein-Ühlenbeck formulation (3), \( \log v \) is normally distributed with easy expressions for the mean and variance, so exact big-step Monte Carlo becomes possible.
Double Lognormal vs Bergomi

- Recall that the Bergomi model has dynamics (with $\tau = T - t$)
  \[
  \frac{d\zeta_t(T)}{\zeta_t(T)} = \xi_1 e^{-\kappa \tau} dZ_1 + \xi_2 e^{-c \tau} dZ_2
  \]

- Now in the Double Lognormal model
  \[
  d\zeta_t(T) = dG(v, v'; \tau)
  = \xi_1 v e^{-\kappa \tau} dZ_1 + \xi_2 v' \frac{\kappa}{\kappa - c} \left( e^{-c \tau} - e^{-\kappa \tau} \right) dZ_2
  \]

  - We see that the two sets of dynamics are very similar.
  - Bergomi’s model is a market model and Buehler’s affine model is a factor model.
  - However any variance curve model may be made to fit the initial variance curve by writing
    \[
    \zeta_t(T) = \frac{\zeta_0(T)}{G(z_0, T)} G(z_t, T)
    \]
SPX option implied volatilities as of 03-Apr-2007

Expiry: 20070405
Expiry: 20070421
Expiry: 20070519
Expiry: 20070616
Expiry: 20070629
Expiry: 20070922
Expiry: 20070928
Expiry: 20071222
Expiry: 20071231
Expiry: 20080322
Expiry: 20080331
Expiry: 20080621
Expiry: 20081220
Expiry: 20091219
We note that skews are steeply positive and that implied volatilities decline with time to expiry.
How to price options on VIX

A VIX option expiring at time $T$ with strike $K_{\text{VIX}}$ is valued at time $t$ as

\[
\mathbb{E}_t \left[ \left( \sqrt{\mathbb{E}_T \left[ \int_T^{T+\Delta} \nu_s \, ds \right]} - K_{\text{VIX}} \right)^+ \right]
\]

where $\Delta$ is around one month (we take $\Delta = 1/12$). In the affine models under consideration, the inner expectation is linear in $\nu_T$, $\nu'_T$ and $z_3$ so that

\[
VIX_T^2 = \mathbb{E}_T \left[ \int_T^{T+\Delta} \nu_s \, ds \right] = a_1 \nu_T + a_2 \nu'_T + a_3 z_3
\]

with some coefficients $a_1, a_2$ and $a_3$ that depend only on $\Delta$. 
A simple lognormal model

As in Friz-Gatheral, assume (wrongly of course) that \( VIX \) is lognormally distributed: \( \log VIX \sim N(\mu, s^2) \). Then \( VIX^2 \) is also lognormal with \( \log VIX^2 \sim N(2\mu, 4s^2) \). Then

\[
\begin{align*}
\mathbb{E}_t [VIX^2_T] &= \mathbb{E}_t [a_1 v_T + a_2 v'_T + a_3 z_3] = \exp \{2\mu + 2s^2\} \\
\mathbb{E}_t [VIX^4_T] &= \mathbb{E}_t \left[ (a_1 v_T + a_2 v'_T + a_3 z_3)^2 \right] = \exp \{4\mu + 8s^2\}
\end{align*}
\]

- \( \mathbb{E}_t [a_1 v_T + a_2 v'_T + a_3 z_3] \) is easy to evaluate; the result does not depend on whether we choose Heston or lognormal dynamics.
- \( \mathbb{E}_t \left[ (a_1 v_T + a_2 v'_T + a_3 z_3)^2 \right] \) may be computed using the Dufresne trick; in this case, the result does depend on our choice of dynamics.
Calibration to VIX options

- We now have explicit expressions for $\mu$ and $s$ under both Heston and lognormal dynamics.
- Moreover, with our lognormal assumption, the volatility smile of VIX options will be flat at the level $s/\sqrt{t}$.
- We proceed by fitting the model parameters jointly to the term structure of VIX forwards and ATM VIX implied volatilities.
Calibration to VIX options

- Of course, because VIX is not lognormally distributed, this calibration doesn’t work very well.
  - It gets us close enough to be able to fit with manual tweaking of parameters
- In the next slide, we see the result of tweaking.
Fit of Double Lognormal model to VIX options

From Monte Carlo simulation with parameters

\[ z_1 = 0.0137; \quad z_2 = 0.0208; \quad z_3 = 0.0421; \quad \kappa = 12; \quad \xi_1 = 7; \quad c = 0.34; \quad \xi_2 = 0.94; \]

we get the following fits (orange lines):
Fit of Double Heston model to VIX options

From Monte Carlo simulation with parameters

\[ z_1 = 0.0137; \quad z_2 = 0.0208; \quad z_3 = 0.0421; \quad \kappa = 12; \quad \eta_1 = 0.7; \quad c = 0.34; \quad \eta_2 = 0.14; \]

we get the following fits (orange lines):
In terms of densities of VIX

- When we draw the densities of VIX for the last expiration ($T = 1.13$) under each of the two modeling assumptions, we see what’s happening:

  ![Double log simulation](image1)
  ![Double Heston simulation](image2)

- In the (double) Heston model, $\nu_t$ spends too much time in the neighborhood of $\nu = 0$ and too little time at high volatilities.
In terms of densities of $\sqrt{v_T}$

- We see this really clearly when we plot the densities of $\sqrt{v_T}$ (again with $T = 1.13$):

![Double log simulation](image1)

![Double Heston simulation](image2)

- The distribution of $v_T$ in the Heston model is completely unrealistic. What do you think is the probability of instantaneous volatility being less than 2%?
Fit to SPX variance swaps

Variance swap fits are independent of the specific dynamics. Then as before with
\[ z_1 = 0.0137; \quad z_2 = 0.0208; \quad z_3 = 0.0421; \quad \kappa = 12; \quad c = 0.34, \]
we obtain the following fit (green points are market prices):
Fit of double lognormal model to SPX options

From Monte Carlo simulation with the same parameters as before plus $\rho_1 = -0.66$, $\rho_2 = -0.60$, we get the following fits (blue lines):
Observations

- The double lognormal model clearly fits the market better than double Heston.
- We can fit both short and long expirations with the same parameters in contrast to single-factor stochastic volatility models.
- The fitted lognormal parameters satisfy
  \[ \frac{\xi_1^2}{\kappa} = \frac{7^2}{12} \approx 4.1; \quad \kappa \gg 1; \quad \frac{\kappa}{c} = \frac{12}{0.34} \approx 35.3 \gg 1 \]
- We therefore have time-scale separation and can apply the methods presented by Philippe Balland in 2006.
Implied vs Historical

- Just as option traders like to compare implied volatility with historical volatility, we would like to compare the risk-neutral parameters that we got by fitting the Double Lognormal model to the VIX and SPX options markets with the historical behavior of the variance curve.

- First, we check to see (in the time series data) how many factors are required to model the variance curve.
PCA on historical variance swap data

- We proxy variance swaps by the log-strip for each expiration.
- Spline-interpolate to get standardized variance curves.
- Perform PCA on first differences to obtain the following two factors:

![First factor loading](image1)

![Second factor loading](image2)

- The blue line is conventional PCA and the red line is robust PCA.
Extracting time series for $z_1$ and $z_2$

- In our affine model, given estimates of $\kappa$, $c$ and $z_3$, we may estimate $z_1$ and $z_2$ using linear regression.
- From two years of SPX option data with parameters $\kappa = 12$, $c = 0.34$ and $z_3 = 0.0421$, we obtain the following time series for $\sqrt{z_1}$ (orange) and $\sqrt{z_2}$ (green):
Statistics of \( z_1 \) and \( z_2 \)

- Let’s naïvely compute the standard deviations of log-differences of \( z_1 \) and \( z_2 \). We obtain

<table>
<thead>
<tr>
<th>Factor</th>
<th>Historical vol.</th>
<th>Implied vol. (from VIX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( z_1 )</td>
<td>8.6</td>
<td>7.0</td>
</tr>
<tr>
<td>( z_2 )</td>
<td>0.84</td>
<td>0.94</td>
</tr>
</tbody>
</table>

- The two factors have the following autocorrelation plots
Observations

- Historical and implied volatilities are similar
  - in contrast to single-factor stochastic volatility models.
- Historical decay rates are greater than implied
  - price of risk effect just as in single-factor stochastic volatility models.
Summary I

- It makes sense to model tradables such as variance swaps rather than non-tradables such as implied volatilities.

- The affine variance curve functional introduced by Buehler is particularly attractive.
  - It is consistent with many dynamics of interest.

- Dufresne’s trick of recursively applying Itô’s Lemma allows us to compute moments for both Heston and lognormal dynamics.

- Although Double Heston is more analytically tractable, Double Lognormal agrees much better with the market.
  - Whilst the rough levels of VIX option implied volatilities are determined by SPX option prices, VIX option skews are seen to be very sensitive to dynamical assumptions.
By adding a second volatility factor, we have achieved the following:

- The term structure of SPX skew seems right even for short expirations with no need for jumps.
- We are able to fit VIX options with time-homogeneous parameters.
- Historical and risk-neutral estimates of the volatilities of the factors are similar
  - Recall that implied and historical vol. of vol. are very different in single-factor volatility models.
Current and future research

- Develop efficient algorithms for pricing and calibration.
- Investigate alternative dynamics:
  - More general CEV models with $\alpha, \beta \neq 1/2$ or 1
  - Add jumps in volatility and stock price.
- Add more tradable factors.
References

Philippe Balland.
Forward smile.

Lorenzo Bergomi.
Smile dynamics II.

Hans Buehler.
Consistent variance curve models.

Daniel Dufresne.
The integrated square-root process.

Bruno Dupire.
Model art.

Bruno Dupire.
A unified theory of volatility.

Jim Gatheral.
*The Volatility Surface: A Practitioner’s Guide*.