A New Simple Approach for Constructing Implied Volatility Surfaces

Peter Carr and Liuren Wu

NYU & Baruch College

February 28, 2011
at *Columbia University*

The Literature

1. **Stochastic instantaneous volatility models:**
   (Hull-White, Heston, Hagan et. al.,...)
   - *Starting point:* Known initial stock price level and financing.
   - *Assumptions:* Stock price and instantaneous return volatility dynamics
   - *Implications:* The level and shape of the implied volatility surface (across strike and maturity); risk exposures...
   - *Calibration:* Parameters governing the price/volatility dynamics and the initial volatility level can be calibrated to a finite number of option observations. The calibrated model can be used to construct the whole implied volatility surface.
   - *Drawbacks:*
     - Initial instantaneous volatility level is not observable.
     - Slow and/or difficult to calibrate.

The Literature (continued)

2. **Market models of implied volatilities:**
(Avellaneda & Zhu, Schonbucher, Ledoit & Santa-Clara, ...)
   - *Starting point:* Known initial option implied volatility level (on a single option, a curve, or over the whole surface)
   - *Assumptions:* The martingale component of the implied volatility dynamics.
   - *Implications:* The drift of the implied volatility dynamics; prices on exotic contracts; risk exposures...
   - *Calibration:* ?
   - *Drawbacks:*
     - Given an entire initial implied volatility surface, one is not free to choose *any* martingale component of dynamics.
     - For example, if initial smile slopes down in strike, correlation cannot be positive.

A new approach in constructing implied volatility surfaces

somewhere in between the two existing approaches:

- *Starting point:* Initial stock price level and financing.
- *Assumptions:* Stock price and *option implied volatility* dynamics (both drift and diffusion), instead of instantaneous return volatility dynamics.
- *Implications:* The level and shape of the initial implied volatility surface (across strike and maturity) at a given date.
- *Calibration:*
  - Parameters governing the implied volatility dynamics and the initial instantaneous volatility level can be calibrated to a finite number of vanilla option implied volatility observations.
  - The calibrated model can be used to construct the whole implied volatility surface.
  - Calibration does not go through option price calculation. It is directly from implied volatility dynamics to implied volatility surface.
  - 100 times faster than calibrating standard option pricing models of similar complexities.

Why so entrenched in implied volatility?

- Implied volatility is calculated from the Black-Merton-Scholes (BMS) model.
- The fact that practitioners use the BMS model to quote options does *not* mean they agree with the BMS assumptions.
- *Why so entrenched in implied volatility?*
  1. **Information:** It is much easier to gauge/express views in terms of implied volatility than in terms of option prices.
     - IV is invariant to a change in units in spot/strike/option premium.
     - IV does not depend on intrinsic value; option prices do — intrinsic has no information value.
     - IV has the normal return distribution (BMS model) as a benchmark.
       $\Rightarrow$ Deviation from a flat line (across strike) reveals return deviation from normality.
       $\Rightarrow$ A higher IV for OTM puts (low strikes) than for OTM calls (high strikes) says that the left tail is heavier than the right tail.
       $\Rightarrow$ Higher IVs for OTM options than for ATM options suggests fatter tails (leptokurtosis).

Why so entrenched in implied volatility?

2. **No arbitrage constraints:**
   - Merton (1973): model-free bounds based on no-arb. arguments:

     Type I: No-arbitrage between European options of a fixed strike and maturity vs. the underlying and cash:
     call/put prices $\geq$ intrinsic;
     call prices $\leq$ (dividend discounted) stock price;
     put prices $\leq$ (present value of the) strike price;
     put-call parity.

     Type II: No-arbitrage between options of different strikes and maturities:
     bull, bear, calendar, and butterfly spreads ≥ 0.
   - Hodges (1996): These bounds can be expressed in implied volatilities.

     Type I: Implied volatility is positive.

   $\Rightarrow$*If market makers quote options in terms of a positive implied volatility surface, all Type I no-arbitrage conditions are automatically guaranteed.*
3. **Technological:** In the absence of options order flow, IV surface does not need to be updated as frequently as option prices.

This paper: Through assumptions on IV dynamics, we obtain tighter no-arbitrage constraints on the shape of the implied volatility surface.

Implied volatility dynamics and no-arbitrage conditions

- Zero rates for notational clarity.
- Diffusion stock price dynamics: $dS_t/S_t = s_t dW_t$.
- The dynamics of the instantaneous return volatility (s_t) is left unspecified.
- For each option struck at K and expiring at T, its implied volatility $I_t(K,T)$ follows a continuous process,

$$dI_t(K,T) = \mu_t dt + \omega_t dZ_t, \text{ for all } K > 0 \text{ and } T > t.$$

  - μ_t (drift) and ω_t (volvol) can depend on K, T, and $I(K,T)$.
  - One Brownian motion Z_t drives the whole implied volatility surface.
  - Correlation between implied volatility and return $\rho_t dt = \mathbb{E}[dW_t dZ_t]$.
- $I_t(K,T) > 0$ guarantees no dynamic arbitrage between any option (K,T) and the underlying stock (and cash).
- We further require that no dynamic arbitrage (NDA) be allowed between any option at (K,T) and a basis option at (K_0,T_0) and the stock.

From NDA to the fundamental PDE

NDA: No dynamic arbitrage is allowed between any option at (K, T) and a basis option at (K_0, T_0) and the stock.

- Let $P_t(K, T)$ denote the option value, which we can represent in the Black-Merton-Scholes formula $B(\cdot)$: $P_t(K, T) = B(S_t, I_t(K, T), t)$.
- NDA implies that we can hedge away the risk in $P_t(K, T)$ by using the stock and the basis option, such that

$$\mathbb{E}\left[dP_t(K, T) - B_S s_t S_t dW_t - B_\sigma \omega_t dZ_t\right] = 0, \text{ for } t \in [0, T_0 \wedge T)$$

- The fundamental PDE:

$$-B_t = \mu_t B_\sigma + \frac{s_t^2}{2} S_t^2 B_{SS} + \rho_t \omega_t s_t S_t B_{S\sigma} + \frac{\omega_t^2}{2} B_{\sigma\sigma}.$$

- The PDE defines a linear relation between the theta (B_t) of the option and its vega (B_σ), dollar gamma ($S_t^2 B_{SS}$), dollar vanna ($S_t B_{S\sigma}$), and volga ($B_{\sigma\sigma}$).
- We christen the class of implied volatility surfaces defined by the fundamental PDE as the **Vega-Gamma-Vanna-Volga (VGVV)** model.

From the PDE to an algebraic equation

- From the PDE,

$$-B_t = \mu_t B_\sigma + \frac{s_t^2}{2} S_t^2 B_{SS} + \rho_t \omega_t s_t S_t B_{S\sigma} + \frac{\omega_t^2}{2} B_{\sigma\sigma}.$$

- Plug in the partial derivatives of the BMS formula:

$$\begin{array}{rclrcl} B_t & = & -\frac{\sigma^2}{2} S^2 B_{SS}, & B_\sigma & = & \sigma\tau S^2 B_{SS}, \\ SB_{\sigma S} & = & -d_2\sqrt{\tau} S^2 B_{SS}, & B_{\sigma\sigma} & = & d_1 d_2 \tau S^2 B_{SS}. \end{array}$$

- The PDE reduces to an algebraic equation for $I_t(K, T)$,

$$\frac{I_t^2}{2} - \mu_t I_t \tau - \left[\frac{s_t^2}{2} - \rho_t \omega_t s_t \sqrt{\tau} d_2 + \frac{\omega_t^2}{2} d_1 d_2 \tau \right] = 0.$$

- If (μ_t, ω_t) do not depend on $I_t(K, T)$, we can solve the whole implied volatility surface as the solution to a quadratic equation.

SRV: Square root implied variance dynamics

- Represent the implied volatility surface in terms of standardized moneyness z and term $\tau = T - t$, $v_t(z, \tau) \equiv I_t(K, T)$.
  - The standardized moneyness $z_t = \frac{\ln(K/S_t) + \frac{1}{2}I^2\tau}{I\sqrt{\tau}}$, represents the number of std. dev's of future log spot that the log strike is above its mean
- Square-root implied variance dynamics (SRV):
  $dI_t^2 = \kappa\left[\theta - I_t^2\right]dt + 2we^{-\eta(T-t)}I_t dZ_t,$
- The implied volatility surface $v(z, \tau)$ solves a *quadratic equation*:

$$(1 + \kappa\tau)\, v_t^2(z, \tau) + \left(w^2 e^{-2\eta\tau}\tau^{3/2} z\right) v_t(z, \tau) - \left[\left(\kappa\theta - w^2 e^{-2\eta\tau}\right)\tau + s_t^2 + 2\rho w s_t e^{-\eta\tau}\sqrt{\tau}z + w^2 e^{-2\eta\tau}\tau z^2\right] = 0.$$

  - In the limit $\tau = 0$: $v_t^2(z, 0) = s_t^2$ (continuous price dynamics),
    In the limit $\tau = \infty$, $v_t^2(z, \infty) = \theta$ (central limit theorem).
  - ATM implied variance ($z = 0$) term structure:
    $a_t^2(\tau) = \frac{\left(\kappa\theta - w^2 e^{-2\eta\tau}\right)\tau + s_t^2}{(1 + \kappa\tau)}$,
    only a function of $\mu_t = \frac{1}{2}\left(\frac{\left(\kappa\theta - w^2 e^{-2\eta\tau}\right)}{I_t(K, T)} - \kappa_t I_t(K, T)\right)$.

LNV: Log-normal implied variance dynamics

- Represent the implied volatility surface in terms of log relative strike and term, $\hat{I}_t(k,\tau) \equiv I_t(K,T)$
  - OTC Equity index option implied volatilities are quoted in terms of log relative strike $k_t = \ln(K/S_t)$ and term.
- Log-normal implied variance dynamics (LNV):
  $dI_t^2(K,T) = \kappa[\theta - I_t^2(K,T)]dt + 2we^{-\eta(T-t)} I_t^2(K,T)dZ_t.$
- Implied variance surface $(\hat{I}_t^2(k,\tau))$ solves a *quadratic equation*:

$$\frac{w^2}{4} e^{-2\eta\tau}\tau^2 \hat{I}_t^4(k,\tau) + \left[1 + \kappa\tau + w^2 e^{-2\eta\tau}\tau - \rho s_t w e^{-\eta\tau}\tau\right] \hat{I}_t^2(k,\tau) - \left[s_t^2 + \kappa\theta\tau + 2\rho s_t w e^{-\eta\tau} k + w^2 e^{-2\eta\tau} k^2\right] = 0.$$

  - In the limit of $\tau = 0$, $\hat{I}_t^2(k,0) = w^2k^2 + 2\rho s_t wk + s_t^2$.
    In the limit of $\tau = \infty$, $\hat{I}_t^2(k,\infty) = \theta$.
  - ATM implied variance ($z = 0$) term structure: $a_t^2(\tau) = \frac{\kappa\theta\tau + s_t^2}{1+(\kappa + w^2 e^{-2\eta\tau})\tau}$, only a function of $\mu_t = \frac{1}{2}\left(\frac{\kappa\theta}{I_t(K,T)} - \left(\kappa + w^2 e^{-2\eta\tau}\right) I_t(K,T)\right)$.

Recap: Two tractable implied volatility dynamics

- Mean-reverting square root or log-normal implied variance dynamics (SRV and LNV).
  - Six potentially time-varying coefficients $(\kappa_t, \theta_t, w_t, \eta_t, \rho_t, s_t)$.
  - Given time-t values on the six coefficients, the whole implied volatility surface at time t can be solved as the solution to quadratic equations.
- Benchmark: Heston (1993) assumes mean-reverting square-root dynamics on the instantaneous variance rate (s_t^2).
  - Five coefficients $(\kappa_t, \theta_t, w_t, \rho_t, s_t)$.
  - Given values on the five coefficients, the implied volatility surface can be computed as follows:
    - Derive analytical solution for the return characteristic function.
    - Perform numerical integration to obtain option values (quadrature or FFT).
    - Solve the implied volatility from the option value.
  - About 100 times slower, and not as accurate.

A fast and robust approach for dynamic calibration

- Treat the six or five coefficients as the state vector X_t.
- Assume that the state vector propagates like a random walk:
  $X_t = X_{t-1} + \sqrt{\Sigma_x}\varepsilon_t$
  - Transform the coefficients so that the state X_t can take values on the whole real line.
  - Assume diagonal matrix for Σ_x.
- Assume that all implied volatilities are observed with errors,
  $y_t = h(X_t) + \sqrt{\Sigma_y}e_t$.
  - $h(\cdot)$ denote the model value (quadratic solution for SRV and LNV, complicated numerical calculation for Heston).
  - For SRV and LNV, take implied volatilities for y_t. For Heston, define y_t as vega weighted out-of-the-money option value.
  - Assume IID error, $\Sigma_y = \sigma_e^2 I_n$.
- The set-up introduces 6-7 auxiliary parameters (Σ_x, σ_e^2) controlling the relative update speed of the coefficients.

Unscented Kalman filter

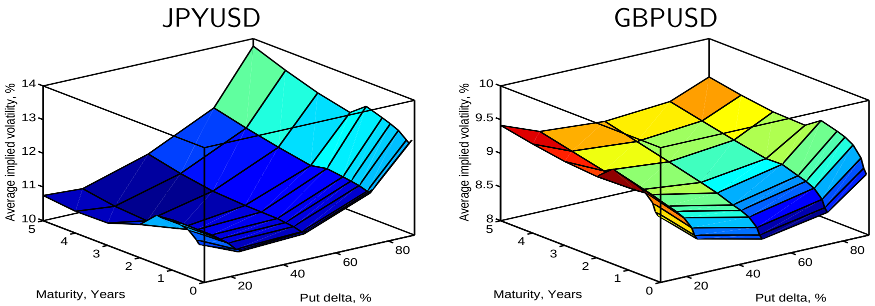
- Given the auxiliary parameters, the implied volatility surface can be fitted quickly via unscented Kalman filter:

$$
\begin{aligned}
&\overline{X}_t = \widehat{X}_{t-1}, \quad \overline{V}_{x,t} = \widehat{V}_{x,t-1} + \Sigma_x, \\
&\chi_{t,0} = \overline{X}_t, \quad \chi_{t,i} = \overline{X}_t \pm \sqrt{(k+\delta)(\overline{V}_{x,t})_j}, \\
&\overline{y}_t = \sum_{i=0}^{2k} w_i \zeta_{t,i}, \quad \overline{V}_{y,t} = \sum_{i=0}^{2k} w_i \left[\zeta_{t,i} - \overline{y}_t\right]\left[\zeta_{t,i} - \overline{y}_t\right]^\top + \Sigma_y, \\
&\overline{V}_{xy,t} = \sum_{i=0}^{2k} w_i \left[\chi_{t,i} - \overline{X}_t\right]\left[\zeta_{t,i} - \overline{y}_t\right]^\top, K_t = \overline{V}_{xy,t}\left(\overline{V}_{y,t}\right)^{-1}, \\
&\widehat{X}_t = \overline{X}_t + K_t\left(y_t - \overline{y}_t\right), \quad \widehat{V}_{x,t} = \overline{V}_{x,t} - K_t \overline{V}_{y,t} K_t^\top.
\end{aligned}
$$

- The whole sample (573 weeks) of implied volatility surfaces can be fitted in about half a second (versus about 1 minute for Heston).

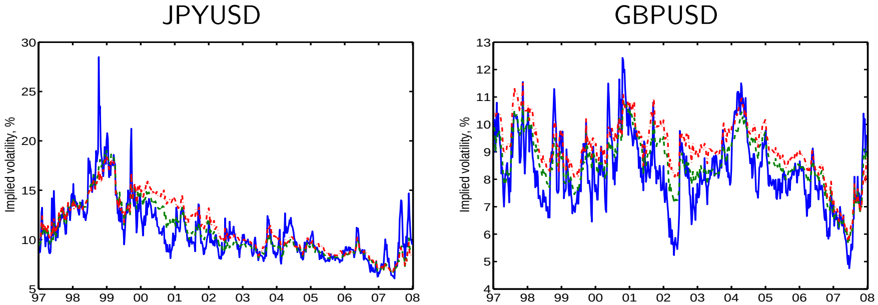
- Choose the auxiliary parameters to minimize the sum of squared pricing errors: $\sum_{t=1}^{N}(y_t - \widehat{y}_t)^\top (y_t - \widehat{y}_t)$.

Application to OTC currency option implied volatilities



- OTC currency options are quoted in
  - Delta-neutral straddle (ATMV): (call + put) with zero delta $\Rightarrow d_1 = 0$.
  - 25-delta Risk reversal (RR): $IV^{25c} - IV^{25p}$
  - 25-delta butterfly spread (BF): $(IV^{25c} + IV^{25p})/2 - ATMV$
  - 10-delta risk reversals and butterfly spreads.
- ATMV, RR, and BF measure the level, slope (skew), and curvature (kurtosis) of the IV smile (return distribution).

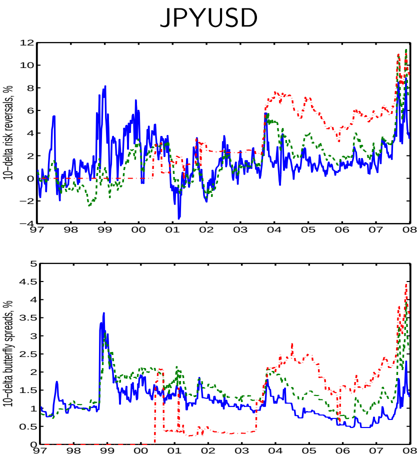
Time variation in currency option volatility levels



The three lines are at one month (solid lines), three months (dashed lines), and five years (dashdotted lines).

- Implied volatilities across different maturities (from one month to 5 years) vary together and at similar levels.

Time variation in currency return skewness and kurtosis



- Before 2001, long-term implied volatilities do not smile.
- Now, they smile, smirk, and are constantly switching into different faces. Long-term smile more than short term.

Pricing performance comparison on currency options

- Weekly from January 8, 1997 to December 26, 2007, 573 weeks.
- 5 delta $\times$ 11 maturities from 1 month to 5 years, 31,515 options.
- Average performance:

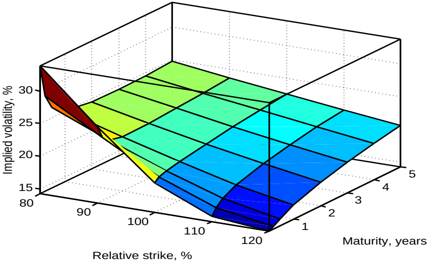
	JPYUSD			GBPUSD		
	SRV	LNV	Heston	SRV	LNV	Heston
RMSE	0.40	0.36	0.37	0.13	0.12	0.14
R^2	0.98	0.99	0.98	0.98	0.99	0.98
Auto	0.77	0.78	0.85	0.71	0.74	0.78

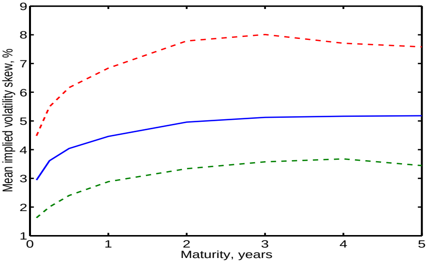
RMSE root mean squared pricing error in IV volatility points.
Auto autocorrelation of pricing errors in IV.

- All three models perform reasonably well.
- LNV is the best of the three for both currency pairs.

Application to OTC SPX option implied volatilities

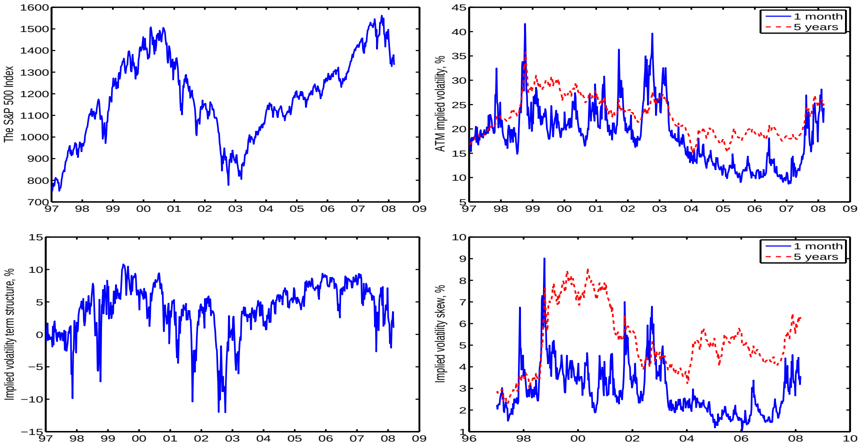
- SPX option implied volatilities over the same sample period.
- 5 moneyness levels at 80, 90, 100, 110, 120 percent of spot.
- 8 maturities from 1 month to 5 years, 30,120 options.





- When measured against a standardized moneyness measure $d = \ln(\mathcal{K}/100)/(IV\sqrt{\tau})$, the skew defined as, $SK_{t,T} = \frac{IV_{t,T}(80\%) - IV_{t,T}(120\%)}{|d_{t,T}(80\%) - d_{t,T}(120\%)|}$, does not flatten as maturity increases.

Time variation in SPX volatilities and skews



- Upward sloping term structure most of the time, except during crisis.
- Heavily negatively skewed all the time; more so at longer term.

Pricing performance comparison on SPX options

	SRV	LNV	Heston
RMSE	0.87	0.67	1.12
R^2	0.99	0.99	0.95
Auto	0.84	0.77	0.85

RMSE root mean squared pricing error in IV volatility points.

Auto autocorrelation of pricing errors in IV.

Compared to Heston, the LNV model

- generates half the root mean squared error,
- explains 4% more variation,
- generates errors with lower serial correlation,
- can be calibrated 100 times faster.

Concluding remarks

- Options traders are *deeply* entrenched in BMS implied volatilities, and for good reasons.
- Directly modeling implied volatility dynamics and generating direct implications on the implied volatility surface shape are both attractive ideas.
- "Market models of implied volatilities" try to do the former while taking the latter as given.
  - The latter (the shape of the implied volatility surface) can put severe (but many times unknown) constraints on what the former (implied volatility dynamics) can be, or vice versa.
- We directly model the implied volatility dynamics, and we *derive* the dynamic-no-arbitrage implication on the shape of the implied volatility surface.
  - The two (dynamics and surface shapes) are guaranteed to be consistent.
  - Market deviations from model implications can serve as relative trading opportunities.

Promise and future research

- Our new approach generates very promising results.
  - Two models with extreme simplicity: The whole implied volatility surface becomes solutions to quadratic equations — 6th grade math.
  - Great performance on both currency options and equity index options.
  - 100 times faster than standard option pricing models, ideal for automated options market making.
- Many open questions remain, for future research.
  - The PDE guarantees dynamic no-arbitrage between any option and a basis option under a single-factor continuous implied volatility dynamics. It remains open on how to guarantee (static) no-arbitrage among many options across different strikes and maturities.
  - How to link the implied volatility dynamics to the dynamics of the instantaneous return variance rate.
  - How to accommodate multiple factors and discontinuous dynamics in both prices and implied volatilities.