Implied Volatility Surface

Liuren Wu

Zicklin School of Business, Baruch College

Options Markets

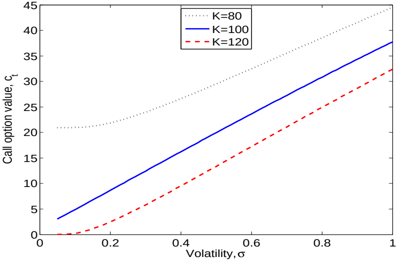
Implied volatility

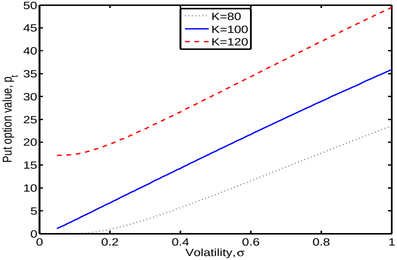
Recall the BMS formula:

$$c(S,t,K,T) = e^{-r(T-t)}\left[F_{t,T}N(d_1) - KN(d_2)\right], d_{1,2} = \frac{\ln\frac{F_{t,T}}{K} \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}$$

- The BMS model has only one free parameter, the asset return volatility σ.
- Call and put option values increase monotonically with increasing σ under BMS.
- Given the contract specifications (K, T) and the current market observations (S_t, F_t, r), the mapping between the option price and σ is a unique one-to-one mapping.
- The σ input into the BMS formula that generates the *market observed option price* (or sometimes a model-generated price) is referred to as the implied volatility (IV).
- Practitioners often quote/monitor implied volatility for each option contract instead of the option invoice price.

The relation between option price and σ under BMS





- An option value has two components:
  - Intrinsic value: the value of the option if the underlying price does not move (or if the future price = the current forward).
  - Time value: the value generated from the underlying price movement.
- Since options give the holder only rights but no obligation, larger moves generate bigger opportunities but no extra risk — Higher volatility increases the option's time value.
- At-the-money option price is approximately linear in implied volatility.

Implied volatility versus σ

- If the real world behaved just like BMS, σ would be a constant.
  - In this BMS world, we could use one σ input to match market quotes on options at all days, all strikes, and all maturities.
  - Implied volatility is the same as the security's return volatility (standard deviation).
- In reality, the BMS assumptions are violated. With one σ input, the BMS model can only match one market quote at a specific date, strike, and maturity.
  - The IVs at different (t, K, T) are usually different — direct evidence that the BMS assumptions do not match reality.
  - IV no longer has the meaning of return volatility.
  - IV still reflects the time value of the option.
  - The intrinsic value of the option is model independent (e.g., $e^{-r(T-t)}(F-K)^+$ for call), modelers should only pay attention to time value.

Implied volatility at (t, K, T)

- At each date t, strike K, and expiry date T, there can be two European options: one is a call and the other is a put.
- The two options should generate the same implied volatility value to exclude arbitrage.
  - Recall put-call parity: $c - p = e^{r(T-t)}(F - K)$.
  - The difference between the call and the put at the same (t, K, T) is the forward value.
  - The forward value does not depend on (i) model assumptions, (ii) time value, or (iii) implied volatility.
- At each (t, K, T), we can write the in-the-money option as the sum of the intrinsic value and the value of the out-of-the-money option:
  - If $F > K$, call is ITM with intrinsic value $e^{r(T-t)}(F - K)$, put is OTM. Hence, $c = e^{r(T-t)}(F - K) + p$.
  - If $F < K$, put is ITM with intrinsic value $e^{r(T-t)}(K - F)$, call is OTM. Hence, $p = c + e^{r(T-t)}(K - F)$.
  - If $F = K$, both are ATM (forward), intrinsic value is zero for both options. Hence, $c = p$.

Implied volatility quotes on OTC currency options

- At each date t and each fixed time-to-maturity $\tau = (T - t)$, OTC currency options are quoted in terms of
  - Delta-neutral straddle implied volatility (ATMV):
    A straddle is a portfolio of a call & a put at the same strike. The strike here is set to make the portfolio delta-neutral:
    $e^{q(T-t)}N(d_1) - e^{q(T-t)}N(-d_1) = 0 \Rightarrow N(d_1) = \frac{1}{2} \Rightarrow d_1 = 0.$
  - 25-delta risk reversal: $RR_{25} = IV(\Delta_c = 25) - IV(\Delta_p = 25)$.
  - 25-delta butterfly spreads:
    $BF_{25} = (IV(\Delta_c = 25) + IV(\Delta_p = 25))/2 - ATMV$.
- When trading currency options OTC, options and the corresponding BMS delta of the underlying are exchanged at the same time.

From delta to strikes

- Given these quotes, we can compute the IV at the three moneyness (strike) levels:

$$\begin{array}{rclll} IV & = & ATMV & at & d_1 = 0 \\ IV & = & BF_{25} + ATMV + RR_{25}/2 & at & \Delta_c = 25\% \\ IV & = & BF_{25} + ATMV - RR_{25}/2 & at & \Delta_p = 25\% \end{array}$$

- The three strikes at the three deltas can be inverted as follows:

$$\begin{array}{rclll} K & = & F \exp\left(\frac{1}{2} IV^2 \tau\right) & at & d_1 = 0 \\ K & = & F \exp\left(\frac{1}{2} IV^2 \tau - N^{-1}(\Delta_c e^{q\tau}) IV \sqrt{\tau}\right) & at & \Delta_c = 25\% \\ K & = & F \exp\left(\frac{1}{2} IV^2 \tau + N^{-1}(|\Delta_p| e^{q\tau}) IV \sqrt{\tau}\right) & at & \Delta_p = 25\% \end{array}$$

- Put-call parity is guaranteed in the OTC quotes: one implied volatility at each delta/strike.
- *These are not exactly right... They are simplified versions of the much messier real industry convention.*

Example

- On 9/29/2004, I obtain the following quotes on USDJPY: $S_t = 110.87$, $ATMV = 8.73, 8.5, 8.66$, $RR_{25} = -0.53, -0.7, -0.98$, and $BF_{25} = 0.24, 0.26, 0.31$ at 3 fixed maturities of 1, 3, and 2 months. (The quotes are in percentages).
- The USD interest rates at 3 maturities are: $1.82688, 1.9, 2.31$. The JPY rates are $0.03625, 0.04688, 0.08125$. (Assume that they are continuously compounding). All rates are in percentages.
- Compute the implied volatility at three moneyness levels at each of the three maturities.
- Compute the corresponding strike prices.
- Compute the invoice prices for call and put options at these strikes and maturities.

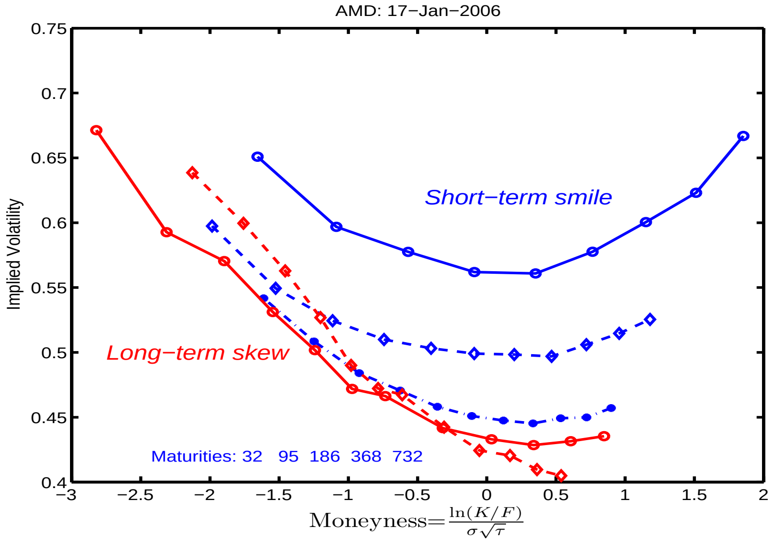
Violations of put-call parities

- On the exchange, calls and puts at the same maturity and strike are quoted/traded separately. It is possible to observe put-call parity being violated at some times.
  - Violations can happen when there are market frictions such as short-sale constraints.
  - For American options (on single name stocks), there only exists a put-call inequality.
  - The "effective" maturities of the put and call American options with same strike and expiry dates can be different.
    - The option with the higher chance of being exercised early has effectively a shorter maturity.
- Put-call violations can predict future spot price movements.
  - One can use the call and put option prices at the same strike to compute an "option implied spot price:" $S_t^o = \left(c_t - p_t + e^{-r\tau} K\right) e^{q\tau}$.
  - The difference between the implied spot price and the price from the stock market can contain predictive information: $S_t - S_t^o$.
- Compute implied forward for option pricing model estimation.

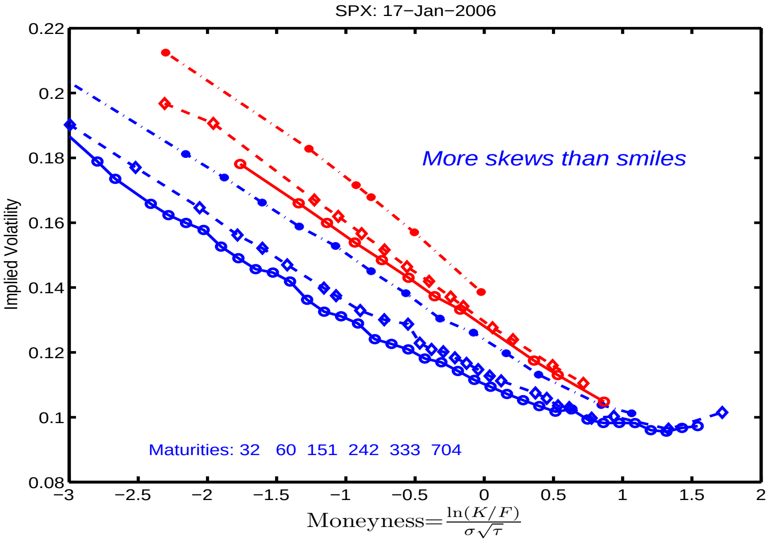
The information content of the implied volatility surface

- At each time t, we observe options across many strikes K and maturities $\tau = T - t$.
- When we plot the implied volatility against strike and maturity, we obtain an implied volatility surface.
- If the BMS model assumptions hold in reality, the BMS model should be able to match all options with one σ input.
  - The implied volatilities are the same across all K and τ.
  - The surface is flat.
- We can use the shape of the implied volatility surface to determine what BMS assumptions are violated and how to build new models to account for these violations.
  - For the plots, do not use K, $K - F$, K/F or even $\ln K/F$ as the moneyness measure. Instead, use a standardized measure, such as $\frac{\ln K/F}{ATMV\sqrt{\tau}}$, d_2, d_1, or delta.
  - Using standardized measure makes it easy to compare the figures across maturities and assets.

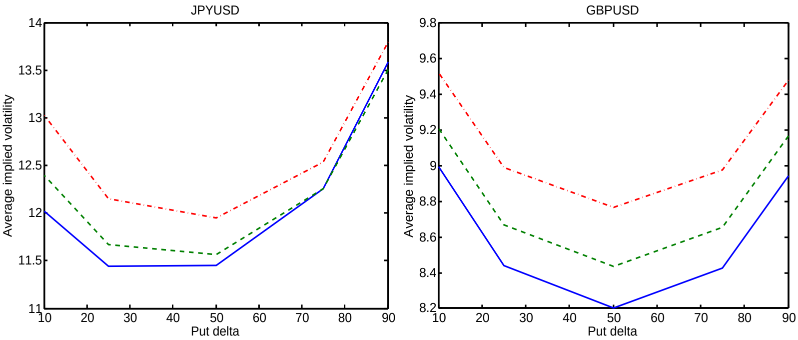
Implied volatility smiles & skews on a stock



Implied volatility skews on a stock index (SPX)



Average implied volatility smiles on currencies



Maturities: 1m (solid), 3m (dashed), 1y (dash-dotted)

Return non-normalities and implied volatility smiles/skews

- BMS assumes that the security returns (continuously compounding) are normally distributed. $\ln S_T/S_t \sim N\left((\mu - \frac{1}{2}\sigma^2)\tau, \sigma^2\tau\right)$.
  - $\mu = r - q$ under risk-neutral probabilities.
- A smile implies that actual OTM option prices are more expensive than BMS model values.
  - $\Rightarrow$ The probability of reaching the tails of the distribution is higher than that from a normal distribution.
  - $\Rightarrow$ Fat tails, or (formally) leptokurtosis.
- A negative skew implies that option values at low strikes are more expensive than BMS model values.
  - $\Rightarrow$ The probability of downward movements is higher than that from a normal distribution.
  - $\Rightarrow$ Negative skewness in the distribution.
- Implied volatility smiles and skews indicate that the underlying security return distribution is not normally distributed (under the risk-neutral measure —We are talking about cross-sectional behaviors, not time series).

Quantifying the linkage

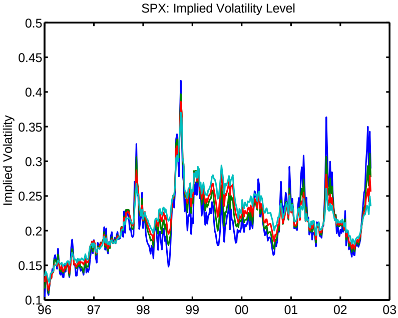
$$IV(d) \approx ATMV \left(1 + \frac{\text{Skew.}}{6} d + \frac{\text{Kurt.}}{24} d^2 \right), \; d = \frac{\ln K/F}{\sigma\sqrt{\tau}}$$

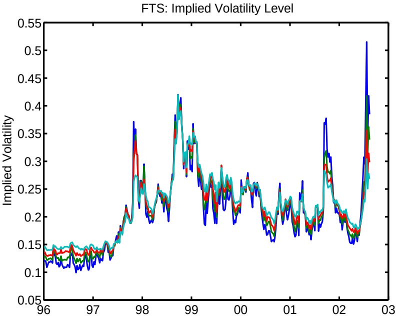
- If we fit a quadratic function to the smile, the slope reflects the skewness of the underlying return distribution.
- The curvature measures the excess kurtosis of the distribution.
- A normal distribution has zero skewness (it is symmetric) and zero excess kurtosis.
- This equation is just an approximation, based on expansions of the normal density (Read "Accounting for Biases in Black-Scholes.")
- The currency option quotes: Risk reversals measure slope/skewness, butterfly spreads measure curvature/kurtosis.
  Check the VOLC function on Bloomberg.

Revisit the implied volatility smile graphics

- For single name stocks (AMD), the short-term return distribution is highly fat-tailed. The long-term distribution is highly negatively skewed.
- For stock indexes (SPX), the distributions are negatively skewed at both short and long horizons.
- For currency options, the average distribution has positive butterfly spreads (fat tails).
- Normal distribution assumption does not work well.
- Another assumption of BMS is that the return volatility (σ) is constant — Check the evidence in the next few pages.

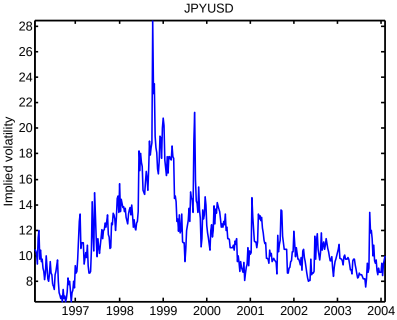
Stochastic volatility on stock indexes





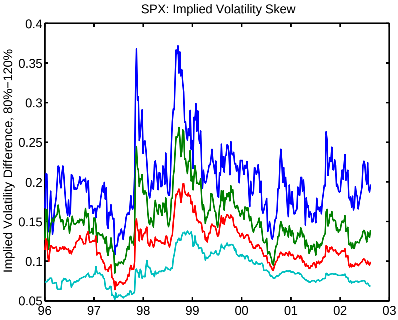
At-the-money implied volatilities at fixed time-to-maturities from 1 month to 5 years.

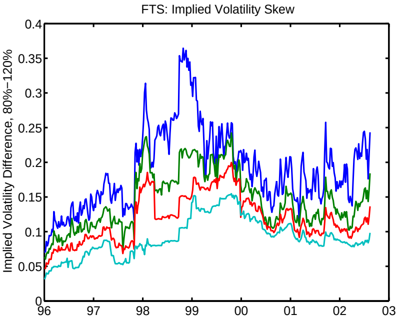
Stochastic volatility on currencies



Three-month delta-neutral straddle implied volatility.

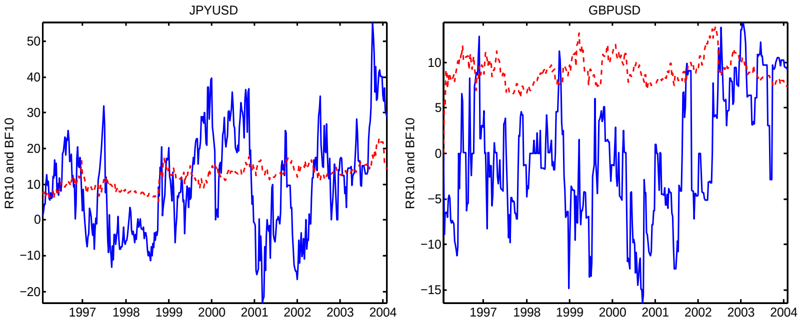
Stochastic skewness on stock indexes





Implied volatility spread between 80% and 120% strikes at fixed time-to-maturities from 1 month to 5 years.

Stochastic skewness on currencies



Three-month 10-delta risk reversal (blue lines) and butterfly spread (red lines).

What do the implied volatility plots tell us?

- Returns on financial securities (stocks, indexes, currencies) are not normally distributed.
  - They all have fatter tails than normal (most of the time).
  - The distribution is also skewed, mostly negative for stock indexes (and sometimes single name stocks), but can be of either direction (positive or negative) for currencies.
- The return distribution is not constant over time, but varies strongly.
  - The volatility of the distribution is not constant.
  - Even higher moments (skewness, kurtosis) of the distribution are not constant, either.
- A good option pricing model should account for return non-normality and its stochastic (time-varying) feature.

Institutional applications of options:

$\Rightarrow$ gain volatility exposures

- Buy/sell an out-of-the-money option and *delta hedge*.
  - Daily delta hedge is a requirement for most institutional volatility investors and options market makers.
  - Call or put is irrelevant. What matter is whether one is long/short vega.
  - Delta hedged P&L is equal to dollar-gamma weighted variance difference: $PL_T = \int_0^T e^{r(T-t)}(\sigma_t^2 - IV_0^2)\frac{S_t^2}{2}\frac{n(d_1(S_t,t,IV_0))}{S_t IV_0\sqrt{T-t}}dt$.
- Views/quotes are expressed not in terms of dollar option prices, but rather in terms of *implied volatilities* (IV).
  - Implied volatilities are 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.
  - Rather, they use the BMS model as a way to transform/standardize the option price, for several practical benefits.

Why BMS implied volatility?

There are several practical benefits in transforming option prices into BMS implied volatilities.

1. **Information:** It is much easier to gauge/express views in terms of implied volatilities than in terms of option prices.
   - Option price behaviors all look alike under different dynamics: Option prices are monotone and convex in strike...
   - By contrast, how implied volatilities behavior against strikes reveals the shape of the underlying risk-neutral return distribution.
     - ★ A flat implied volatility plot against strike serves as a benchmark for a normal return distribution.
     - ★ Deviation from a flat line reveals deviation from return normality.
       $\Rightarrow$ Implied volatility smile — leptokurtotic return distribution
       $\Rightarrow$ Implied volatility smirk — asymmetric return distribution

Why BMS 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 must be 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. **Delta hedge:** The standard industry practice is to use the BMS model to calculate delta with the implied volatility as the input.
   No delta modification consistently outperforms this simple practice in all practical situations.

When to get away from implied volatility?

At very short maturities for out-of-the-money options.

- The implied volatility is computed from a pure diffusion model, under which out-of-money option value approaches zero with an exponential rate (very fast) as the time to maturity nears zero.
  - The chance of getting far out there declines quickly in a diffusive world.
- In the presence of jumps, out-of-money option value approaches zero at a much lower rate (linear).
  - The chance of getting far out there can still be significant if the process can jump over.
- The two difference convergence rates imply that implied volatilities for out-of-money options will blow up (become infinity) as the option maturity shrinks to zero, if the underlying process can jump.
- One can use the convergence rate difference to test whether the underlying process is purely continuous or contains jumps (Carr & Wu (2003, JF).)
- Bid-ask spread can also make the implied volatility calculation difficult/meaningless at very short maturities.