

Implied volatility surface

The widespread practice of quoting option prices in terms of their Black-Scholes implied volatilities (IVs) in no way implies that market participants believe underlying returns to be lognormal. On the contrary, the variation of IVs across option strike and term to maturity, which is widely referred to as the *volatility surface*, can be substantial. In this brief review, we highlight some empirical observations that are most relevant for the construction and validation of realistic models of the volatility surface for equity indices.

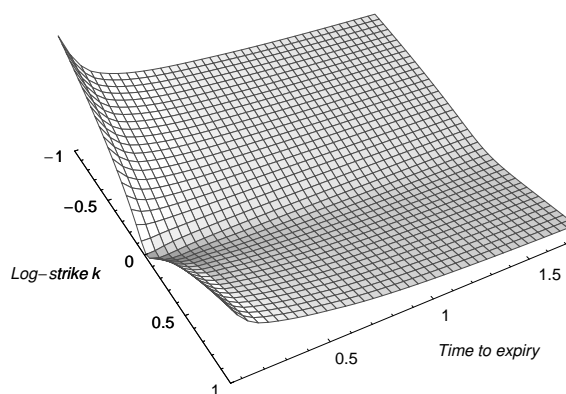
The Shape of the Volatility Surface

Ever since the 1987 stock market crash, volatility surfaces for global indices have been characterized by the *volatility skew*: For a given expiration date, implied volatilities increase as strike price decreases for strikes below the current stock price (spot) or current forward price. This tendency can be seen clearly in the S&P500 volatility surface shown in Figure 1. For short-dated expirations, the cross-section of IVs as a function of strike is roughly V-shaped, but has a rounded vertex and is slightly tilted. Generally this V-shape softens and becomes flatter for longer dated expirations, but the vertex itself may rise or fall depending on whether the term structure of ATM volatility is upward or downward sloping.

Conventional explanations for the volatility skew include:

- *The leverage effect*: Stocks tend to be more volatile at lower prices than at higher prices.
- Volatility moves and spot moves are anti-correlated.
- Big jumps in spot tend to be downwards rather than upwards.
- *The risk of default*: There is a nonzero probability for the price of a stock to collapse if the issuer defaults.
- *Supply and demand*: Investors are net long of stock and so tend to be net buyers of downside puts and sellers of upside calls.

Figure 1: Graph of the SPX implied volatility surface as of the close on September 15, 2005, the day before triple witching.

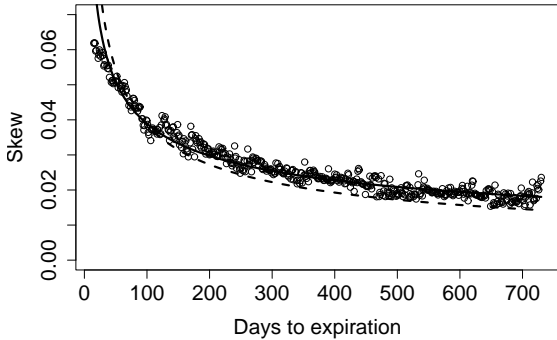


The volatility skew probably reflects all of these factors.

Conventional stochastic volatility (SV) models imply a relationship between the assumed dynamics of the instantaneous volatility and the volatility skew (see Chapter 8 of [8]). Empirically, volatility is well-known to be roughly lognormally distributed [1],[4] and in this case, the derivative of implied volatility with respect to log-strike in a SV model is approximately independent of volatility [8]. This motivates a simple measure of skew: For a given term to expiration, the “95-105” skew is simply the difference between the implied volatilities at strikes of 95% and 105% of the forward price. Figure 2 shows the historical variation of this measure as a function of term to expiration as calculated from end-of-day SPX volatility surfaces generated from listed options prices between Jan 2, 2001 to Feb 6, 2009. In order to fairly compare across different dates and overall volatility levels, all volatilities for a given date are scaled uniformly to ensure that the one-year ATM volatility equals its historical median value over this period (18.80%). The skews for all listed expirations are binned by their term to expiration; the median value for each 5-day bin is plotted along with fits to both $1/\sqrt{T}$ and the best fitting power-law dependence on T .

The important conclusion to draw here is that

Figure 2: Decay of SPX 95%-105% skew with respect to term to expiration. Dots show the median value for each 5-day bin; lines represent best fit to $1/\sqrt{T}$ (dashed) and to power-law behavior: $T^{-0.39}$ (solid).



the term structure of skew is approximately consistent with square-root (or at least power-law) decay. Moreover, this rough relationship continues to hold for longer expirations that are typically traded OTC.

Significantly, this empirically observed term structure of the volatility skew is inconsistent with the $1/T$ dependence for longer expirations typical of popular one-factor stochastic volatility models (see Chapter 7 of [8] for example): Jumps affect only short-term volatility skews so adding jumps doesn't resolve this disagreement between theory and observation. Introducing more volatility factors with different timescales [3] does help but doesn't entirely eliminate the problem. Market models of implied volatility (obviously fit the term structure of skew by construction, but such models are in general time inhomogeneous and in any case, have not so far proven to be tractable. In summary, fitting the term structure of skew remains an important and elusive benchmark by which to gauge models of the volatility surface.

Volatility Surface Dynamics

Volatility surfaces cannot have arbitrary shape; they are constrained by no-arbitrage conditions (such as the convexity of price with respect to

strike). In practice, these restrictions are not onerous and generally are met provided there are no large gradients anywhere on the surface. This observation, together with the fact that index options in most markets trade actively over a wide range of strikes and expiration dates might lead one to expect the dynamics of the volatility surface to be quite complicated. On the contrary, several principal component analysis (PCA) studies have found that an overwhelming fraction of the total daily variation in volatility surfaces is explained by just a few factors, typically three.

Table 1 makes clear that a level mode, one where the entire volatility surface shifts upwards or downwards in tandem, accounts for the vast majority of variation; this result holds across different markets, historical periods, sampling frequencies and statistical methodologies. Although the details vary, generally this mode is not quite flat; short-term volatilities tend to move more than longer ones, as evidenced by the slightly upward tilt in Figure 3(a).

In most of the studies, a term structure mode (TS) is the next most important mode: Here short term volatilities move in the opposite direction from longer term ones, with little variation across strikes. In the Merrill Lynch data, which is sampled at 30, 91, 182, 273, 365 and 547 days to expiration, the pivot point is close to the 91-day term (see Figure 3(b)). In all studies where TS is the second most important mode, the third mode is always a skew mode: one where strikes below the spot (or forward) move in the opposite direction from those above and where the overall magnitude is attenuated as term increases (Figure 3(c)).

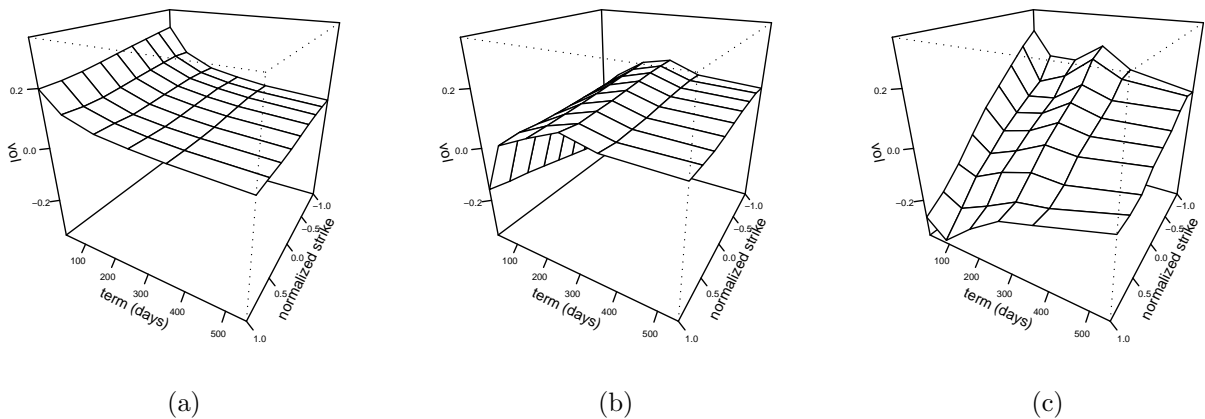
It is also worth noting that the two studies [5],[9] that looked at two different markets during comparable periods found very similar patterns of variation; the modes and their relative importance were very similar, suggesting strong global correlation across index volatility markets.

In the study by Cont and da Fonseca [5], a term structure mode does not figure in the top three modes; instead a skew mode and another strike-related mode related to the curvature emerge as number two and three. This likely reflects the atypically low variation in term structure over the historical sample period and is not due to any methodological differences with the other studies. As in the other studies, the patterns of variation were very similar across markets (S&P500 and FTSE100).

Table 1: PCA studies of the volatility surface. GS= Goldman Sachs study [9]; ML= Merrill Lynch proprietary data

Source	Market	Top 3 modes	Var. explained by		Correlation of 3 modes with spot
			First mode	Top 3	
GS	S&P500, weekly, '94-'97	Level, TS, Skew	81.6%	90.7%	-0.61, -0.07, 0.07
GS	Nikkei, daily, '94-'97	Level, TS, Skew	85.6%	95.9%	-0.67, -0.05, 0.04
Cont, et. al.	S&P500, daily, '00-'01	Level, Skew, Curvature	94%	97.8%	-0.66, ~ 0 , 0.27
Cont, et. al.	FTSE100, daily, '99-'01	Level, Skew, Curvature	96%	98.8%	-0.70, 0.08, 0.7
Daglish et. al,	S&P500, monthly, '98-'02	Level, TS, Skew	92.6%	99.3%	n.a
ML	S&P500, daily, '01-'09	Level, TS, Skew	95.3%	98.2%	-0.87, -0.11, ~ 0

Figure 3: PCA Modes for Merrill Lynch S&P500 Volatility Surfaces: (a) Level; (b) Term-structure; (c) Skew



Changes in Spot and Volatility are Negatively Correlated Perhaps the sturdiest empirical observation of all is simply that changes in spot and changes in volatility (by pretty much any measure) are negatively and strongly correlated. Within the results we have surveyed here, this can be inferred from the high R^2 obtained in the regressions of daily ATMF volatility changes shown in Table 1, as well as directly from the correlations between spot return and PCA modes shown in Table 2. It is striking that the correlation between the level mode and spot return is consistently high across studies, ranging from -0.66 to -0.87 . Correlation between the spot return and the other modes is significantly weaker and less stable.

This high correlation between spot returns and changes in volatility persists even in the most extreme market conditions. For example, during the turbulent period following the collapse of Lehman

Brothers in September 2008 which was characterized by both high volatility and high volatility of volatility, spot-volatility correlation remained at historically high levels: -0.92 for daily changes between September 15, 2008 and end December 31, 2008). On the other hand, the skew mode, which is essentially uncorrelated with spot return in the full historical period (see Table 1), did exhibit stronger correlation in this period (-0.55), while the term structure mode did not. These observations underscore the robustness of the level-spot correlation as well as the time-varying nature of correlations between spot returns and the other modes of fluctuation of the volatility surface.

Other studies have also commented on the robustness of the spot-volatility correlation. For example, using a maximum-likelihood technique, Aït-Sahalia and Kimmel [1] carefully estimated the parameters of Heston, CEV and GARCH models from

S&P500 and VIX data between January 2, 1990 and September 30, 2003; the correlation between spot and volatility changes varied little between these models and various estimation techniques, all estimates were around -0.76 for the period studied.

A related question that was studied by Bouchaud et al. [4] is whether spot changes drive *realized* volatility changes or vice versa. By computing the correlations of leading and lagging returns and squared-returns, they find that for both stocks and stock indices, price changes lead volatility changes. In particular, there is no volatility feedback effect whereby changes in volatility affect future stock prices. Moreover, unlike the decay of the implied volatility correlation function itself, which is power-law with an exponent of around 0.3 for SPX, the decay of the spot-volatility correlation function is exponential with a short half-life of a few days. Supposing the general level of implied volatility (the variation of which accounts for most of the variation of the volatility surface) to be highly correlated with realized volatility, these results apply also to the dynamics of the implied volatility surface. Under diffusion assumptions, the relationship between implied and realized volatility is even more direct: Instantaneous volatility is given by the implied volatility of the ATM option with zero time to expiration.

Skew Relates Statics to Dynamics

Volatility changes are related to changes in spot: as mentioned earlier, volatility and spot tend to move in opposite directions, and large moves in volatility tend to follow large moves in the spot.

It is reasonable to expect skew to play a role in relating the magnitudes of these changes. For example, if all the variation in at-the-money-forward (ATMF) volatility were explained simply by movement along a surface that is unchanged as a function of strike when spot changes, then we would expect

$$\Delta\sigma_{ATMF}(T) = \beta_T \frac{d\sigma}{d(\log K)} \frac{\Delta S}{S} \quad (1)$$

with $\beta_T = 1$ for all terms to expiration (T).

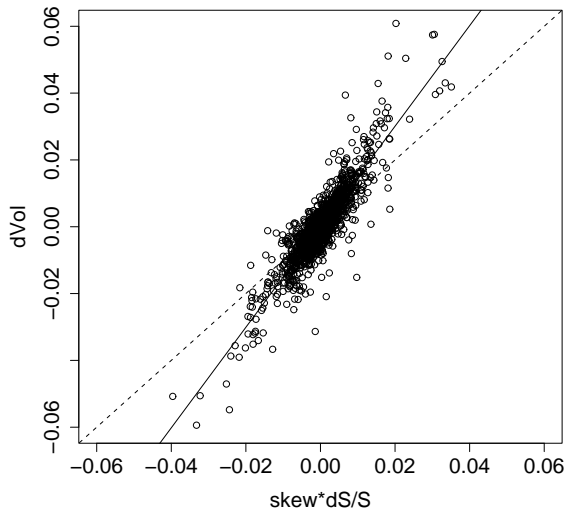
The empirical estimates of β_T shown in Table 2 are based on the daily changes in SP500 ATMF volatilities from Jan 2, 2001 to Feb 6, 2009 (volatilities tied to fixed expiration dates are interpolated

to arrive at volatilities for a fixed number of days to expiration.) Two important conclusions may be drawn: (1) β is not 1.0, rather it is closer to 1.5; (2) remarkably β_T does not change appreciably by expiration. Put differently, although the volatility skew systematically underestimates the daily change in volatility, it does so by roughly the same factor for all maturities.

Table 2: Historical estimates of β_T

T	β_T (std. err.)	R^2
30	1.55 (0.02)	0.774
91	1.50 (0.02)	0.825
182	1.48 (0.02)	0.818
365	1.49 (0.02)	0.791

Figure 4: Regression of 91-day volatility changes vs spot returns. A zero-intercept least squares fit to model (1) leads to $\beta_{91} = 1.50$ (solid lines). The $\beta = 1$ (“sticky-strike”) prediction (dashed line) clearly doesn’t fit.



Although empirical relationships between changes in ATMF volatility and changes in spot are clearly relevant to volatility trading and risk management, the magnitude of β_T itself has direct implications for volatility modeling as well. In both local and stochastic volatility models, $\beta_T \rightarrow 2$

in the short-expiration limit. Under stochastic volatility, β_T is typically a decreasing function of T whereas under the local volatility assumption where the local volatility surface is fixed with respect to a given level of the underlying, β_T is typically an increasing function of T .

Market participants often adopt a phenomenological approach and characterize surface dynamics as following one of these rules: “sticky strike”, “sticky delta” or “local volatility”; each rule has an associated value of β_T . Under the sticky strike assumption, $\beta_T = 1$ and the volatility surface is fixed by strike, under the sticky delta assumption, $\beta_T = 0$ and the volatility surface is a fixed function of K/S , and under the local volatility assumption, as mentioned earlier, $\beta_T = 2$ for short expirations.

Neither “sticky strike” nor “sticky delta” rules imply reasonable dynamics[2]: In a sticky-delta model, the log of the spot has independent increments and the only arbitrage-free sticky-strike model is Black-Scholes (where there is no smile).

Although the estimates of β_T in Table 2 are all around 1.5, consistent with stochastic volatility, that does not exclude the possibility that there may be periods where the β_T may substantially depart from these average values. Derman [7] identified seven distinct regimes for S&P500 daily volatility changes between September 1997 and November 1998 finding evidence for all three of the alternatives listed above. A subsequent study [6] looked at S&P500 monthly data between June 1998 and April 2002 (47 points) and found that for that period, the data was much more consistent with the sticky-delta rule than the sticky-strike rule.

Michael Kamal and Jim Gatheral.

References

- [1] Ait-Sahalia, Yacine, and Robert Kimmel (2007) Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics* **83**, 413–452.
- [2] Balland, Philippe (2002) Deterministic implied volatility models, *Quantitative Finance* **2**, 31–44.
- [3] Bergomi, Lorenzo (2008) Smile dynamics III, *Risk* **21**, 90–96.
- [4] Bouchaud, Jean-Philippe, and Marc Potters

(2003), *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*, Cambridge University Press: Cambridge.

[5] Cont, Rama, and José da Fonseca (2002) Dynamics of implied volatility surfaces. *Quantitative Finance* **2**, 45–60.

[6] Daglish, Toby, John Hull, and Wulin Suo (2007). Volatility surfaces: Theory, rules of thumb, and empirical evidence. *Quantitative Finance* **7**, 507–524.

[7] Derman, Emanuel (1999) Regimes of Volatility, *Risk* **12**, 55–59.

[8] Gatheral, Jim (2006) *The Volatility Surface*, John Wiley & Sons, Hoboken.

[9] Kamal, Michael, and Emanuel Derman (1997). The patterns of change in implied index volatilities, in *Goldman Sachs Quantitative Research Notes*, Goldman Sachs, New York.