LEHMAN BROTHERS | Fixed-Income Research
Interest Rate Derivatives

Volatility Skews

September 30, 2003

Fei Zhou
1-212-526-8886
feiz@lehman.com

The variation of the Black volatilities with respect to option strikes, known as the volatility skew, represents an intriguing and problematic aspect of the volatility market. However, due to its inherent complexity, the matter has not received sufficient coverage. This article expounds on various aspects of skewed volatilities, including market conventions, historical observations, supply and demand mechanisms, the necessity for integrated approaches, mathematical models, and effects on complex products.

INTRODUCTION

Interest rate volatilities, which measure the extent of rate moves, have evolved into an indispensable element of the interest rate market. Along with levels of rates, they constitute an integral part of market dynamics.

Participants in the interest rate market are all concerned with volatilities. However, the extent to which they are exposed to and affected by them varies. While volatilities may help those who primarily contemplate rate levels to visualize the variability of rate changes, the quantification of volatility is normally beyond their concern. In contrast, option users cannot afford such a casual observation. The valuation of option products depends crucially on the accurate measurement of volatility. To them, the precision of volatilities is as important as that of rates.

In fact, there are benchmark swaptions on which the market relies to quantify volatility levels, which are analogous to the benchmark swaps on which the market relies to gauge interest rate levels. A benchmark swaption is an option contract written on an interest rate swap. It is an European option either to receive or to pay the par rate on a swap that commences at the option's expiry. Consistent with this structure, the market convention prescribes that interest rate volatilities be quoted for various option expirations and various underlying swap tenors. Thus, unlike swap rates that are quoted for a list of benchmark maturities, swaption volatilities are benchmarked against a two-dimensional maturity lattice of option expiry and swap tenor. This lattice is nicknamed the vol-grid by option users.

Despite the establishment of standardized grid-based quotations, the description is far from complete. Implicit in the grid volatilities is a quoting mechanism via the industry standard Black formula. It is a closed-form expression of an option price derived from the assumption that the size of fluctuation of an interest rate is proportional to its level. The Black assumption is known to be at odds with market observations; the Black formula itself has, nonetheless, been retained by the industry for the purpose of quoting volatilities. Given an option price, the equivalent percentage volatility in the Black formula is found to match that price. This equivalent volatility is known as the Black volatility and is the market convention for volatility quotes on the vol-grid.

Also implicit in the grid quotation is the convenient choice of at-the-money option strikes; that is, the grid volatilities are quoted exclusively for strikes that are the same as the underlying swap rates. A direct extrapolation of an at-the-money Black volatility to other strikes would not be appropriate since the actual underlying rate does not move in exact proportion to its level as assumed in the Black formula. Despite its benchmark status, the vol-grid provides only one layer of volatility information, applicable to just at-the-money swaptions. In practice, though, just as a freshly traded benchmark swap quickly becomes off market because of ever-changing rates, any at-the-money swaption inevitably falls in or out of the money for the same reason. The subsequent off-the-money[1] volatilities also have to be quoted. Knowledge of off-the-money volatilities is not only complementary to that of the at-the-money vol-grid, but also essential to pricing and managing a majority of options, including those originally at the money.

Hence, any discussion of the volatility market would be incomplete without an analysis of off-the-money volatilities. For a comprehensive grasp of the volatility market, it is necessary that the vol-grid be expanded to include an additional dimension in strike and to establish multiple layers of off-the-money volatility quotations. In the volatility market, the pattern of volatility along the strike dimension is referred to as the volatility skew.

Unfortunately, there has generally been a lack of attention paid to understanding the volatility skew of interest rates, due partially to a greater degree of complexity in its behavior and perhaps also due to insufficient discussion of the subject. Lay persons are often content with just a conceptual appreciation of the volatility skew and passing acknowledgement of its numerical significance. Serious practitioners, on the other hand, are compelled to master a broader and deeper comprehension of the subject. In the absence of coherent information on skews, dedicated practitioners keep track of skew formation and evolution themselves and develop proprietary methods accordingly to handle skew-related pricing and hedging.

This article addresses the scarcity of coverage by providing an elaborate account of the issues related to this important market phenomenon.

REFERENCE SKEWS

As already pointed out, the vol-grid quotes are calculated with the standard Black formula. Even though the Black model nowadays serves only as a vehicle for quoting volatility levels, its hypothetical rate process is still of interest and is considered as a useful reference in analyzing other processes. The Black model assumes a constant percentage volatility for a rate, or proportional basis point volatility. The resultant Black formula uses this percentage volatility as an input in pricing options of all strikes. In this case, the Black volatility coincides with the constant percentage volatility, and there is no volatility skew for off-the-money strikes, or, equivalently, a flat skew.

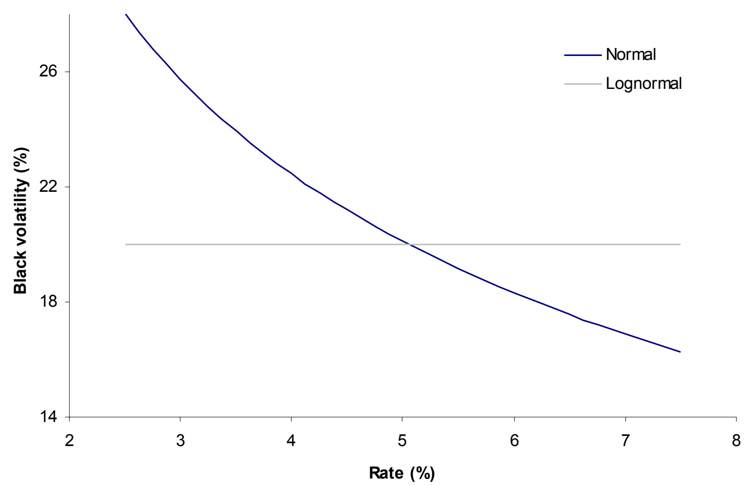
Another interesting case is where the basis point volatility of an underlying rate is independent of its level. Its nominal percentage volatility is, thus, higher for a lower rate and lower for a higher rate. It is then reasonable to argue that an option struck at a higher rate would have a lower Black volatility and the same option struck at a lower rate would have a higher Black volatility. As a result, its Black volatility curve is expected to decrease from low to high strikes, forming a downward skew shape. In order precisely to quantify the volatility skew in this case, the exact price of an option should be first computed in accordance with this level-independent rate process; then, with this price as the target, the equivalent Black volatility is found by matching the Black formula's price to this target.

The process of the constant basis point volatility considered here is known as the normal process; the volatility skew of this process is known accordingly as the normal skew. Similar

[1] The term off-the-money refers to strikes away from the money, including both in- and out-of-the-money ones.

to the flat skew for the lognormal rate process, the normal skew is also a useful reference. Understandably, neither assumption of the two processes conforms to market observations. However, they help view observed skew patterns in perspective. The two special rate processes are uncomplicated and familiar to practitioners, and the comparison between observed skews and the two special ones often sheds lights on actual rate processes. In practice, a skew curve is indeed viewed in relation to a flat skew levelled at the at-the-money strike. The magnitude of skew is defined as the difference between the off-the-money volatility and the at-the-money one. In light of this measurement, the normal skew curve, for its downward sloping shape, displays positive skew for low strikes and negative skew for high strikes. Figure 1 shows an example of the Black volatility patterns under these two processes.

Figure 1. Respective Reference Skews under the Normal and Lognormal Processes with the At-the-Money Rate of 5% and the At-the-Money Black volatility of 20%



The lognormal and the normal interest rate processes are relatively transparent and are well studied. Their volatility skew patterns are viewed as yardsticks in comparing and judging other skew shapes. These two reference skews will be referred to repeatedly in the discussion below.

HISTORICAL OBSERVATION

Although the volatility market has been in existence for decades, the documented history of volatility skews has only been for the last few years. In the early Nineties and before, market participants paid little attention to volatility variation over option strikes. Instead, they largely focused on applying various interest rate methodologies to replicate the at-the-money volatilities, regardless of the models' skew implications. As a result, this pre-skew period was marked by the adoption of a plethora of interest rate models that assumed all types of the underlying process, including the aforementioned constant and proportional basis point volatilities. Those models were put forth mostly for tractability and for their ability to explain the observed at-the-money vol-grid. Knowingly or not, users were generally blind to the skew incompatibility among those models, which was excusable due to a dearth of skew information at the time.

Following this long stretch of ambivalence about volatility skews, the market started to notice their emergence around the mid Nineties. By then, the volumes in the options market had picked up visibly. Off-the-money volatilities were seen being marked away from at-the-money levels. Compared with an at-the-money volatility at the time, the volatility traded somewhat higher for a low-strike option and lower for a high-strike one. It is true that bid-offer spreads for off-the-money volatilities then were considerably wide and actual skew levels were disputable, but the onset of the volatility skew was unambiguous. Figure 2 sketches such an incipient skew pattern for this early period.

Figures 2-5. Various Patterns of Volatility Skews Observed in the Past; Shown Relative to the Normal Skew Represented by the Solid Line

Figure 2. Nascency in the mid-Nineties

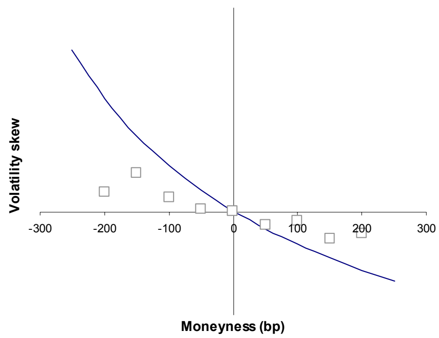


Figure 3. Smirk of Late 1998

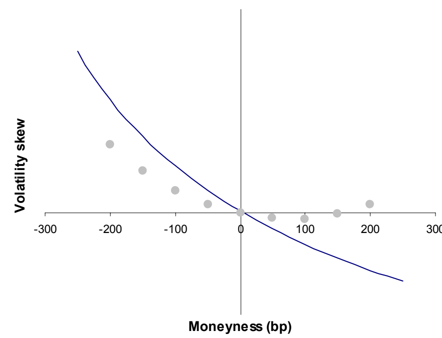


Figure 4. Supernormal of Early 2001

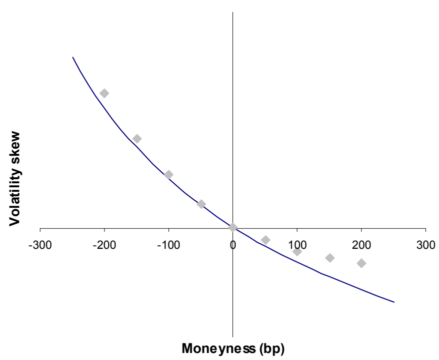
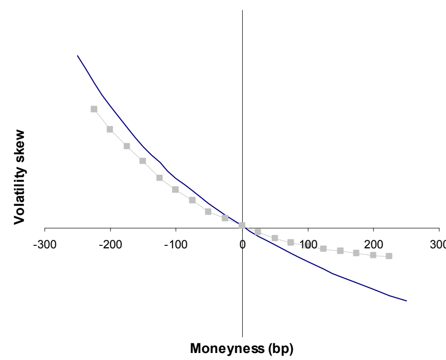


Figure 5. Sandwiched Skew of Late



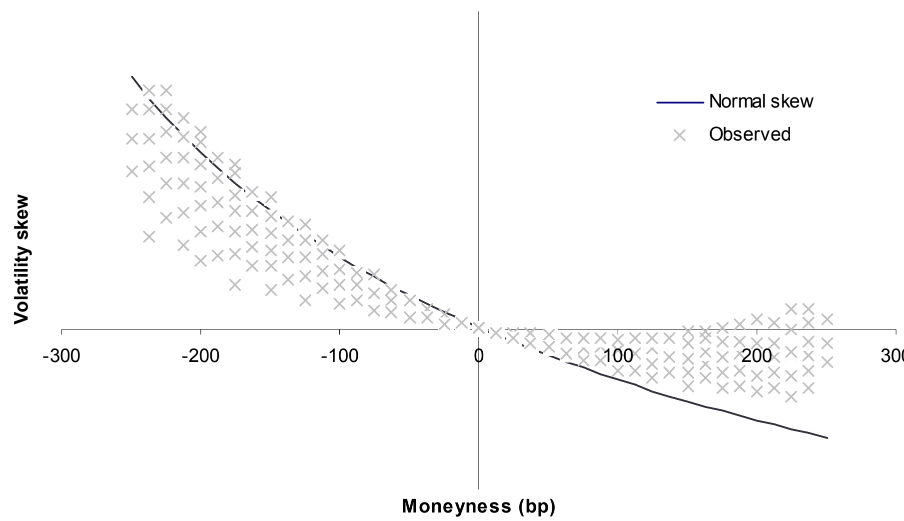
This downward skew pattern mostly persisted for the next several years. It was not until late 1998 that the skew deviated markedly from its original form. It was a time when the global financial crisis was running rampant, triggered by the Russian debt default. Interest rate volatilities then reached inordinate levels. For fear of out-sized swings in interest rates, options with both low and high strikes were in heavy demand, which lifted the volatility levels of all off-the-money strikes. As a result, the skew steepened for low strikes and flattened for high ones. Not only did the skew curve become more convex, but, more noticeably, volatilities for very high strikes were above the at-the-money level for the first time. The new appearance resembled a smile. Actually, it was a wry smile due to an overlaying downward slope and was dubbed a smirk by observers. Shown in Figure 3 is this crisis induced volatility smirk.

As the crisis subsided, off-the-money volatilities lowered their levels gradually. High strike volatilities reverted back below the at-the-money level and the smirk subsequently receded. The volatility curve resumed its downward looking shape. Then, the economic downturn began, which forced interest rates considerably lower. This downward move in rates further softened volatilities for high strike options, but brought protection against potentially worrisome low rates into high demand. Such a market imbalance gradually pushed up volatilities for low-strike options. Eventually, the low-strike skew reached its steepest formation in early 2001. What was remarkable this time was that the low-strike skew broke through the reference level of the normal skew and the actual skew was entirely above the normal level. Although the low-strike skew crossed this special level for a brief time and only by a small margin, the occurrence had its psychological significance and inspired skew watchers to elicit a new description—the supernormal skew—for this unprecedented event. Figure 4 recounts this super shape recorded in early 2001.

Since the supernormal incident, the volatility skew has largely become steady. Up until now, it has mainly moved within the reference levels of the normal and flat skews. If yet another distinctive depiction is in order for this stable period, it can be characterized as a sandwiched skew. It is important to point out that while confined within a range, the sandwiched skew level tends to be closer to the normal skew for low strikes and closer to the flat skew for high strikes. Figure 5 displays such a skew sandwich.

The above anecdotal recounting not only provides some eventful snapshots of historic skew formations, but also proffers an account of the skew evolution over time. In order to appreciate what has taken place in its entirety, Figure 6 is assembled from the skew data available from the past. The shaded area marks the range over which volatility skews have traversed thus far. Again, the normal skew is superimposed for perspective. The depicted range may be only roughly accurate, yet this compendium is instructive for its historical revelation.

Figure 6. Historical Manifestation of Volatility Skews; Confined Between the Normal and Lognormal Levels, with Occasional Breakouts to the Upside



SUPPLY AND DEMAND

The interest rate volatility market in the U.S. has largely reflected the structure and evolution of the housing market. It has been a long-held U.S. government policy to encourage affordable home ownership. To fulfil this course, several government sponsored enterprises (GSEs) were formed to provide liquidity in the residential mortgage market. These housing GSEs borrow money in capital markets and use the proceeds either to purchase mortgage loans from or to provide mortgage funding to lenders. The lenders, some of whom are also known as mortgage servicers, make loans directly to homebuyers. Facilitated by these government agencies, homebuyers have gained easy access to investment capital, and a gigantic housing market has gradually been created.

Also created in the process has been a vast demand for interest rate options. It results from the fact that mortgage loans invariably carry a provision that allows borrowers to repay principal anytime before stated maturities, the so-called the prepayment option. Borrowers take this right for granted and automatically acquire this optionality along with their loans. Thus, every time a mortgage loan is originated, passive demand for volatility is generated. Normally, GSEs handle this demand by securitizing the mortgages acquired from lenders and passing them directly to mortgages investors. The investors receive higher mortgage coupon payments in exchange for bearing the prepayment risk. However, in recent years, an increasing amount of mortgage loans have been retained by GSEs as their own assets. In doing so, they face the task of actively managing their volatility exposure. They usually buy interest rate options from dealers to hedge out this risk, effectively passing the original demand to the volatility market.

On the other hand, GSEs tap public capital to finance their retained mortgage assets or to finance mortgages indirectly with stable funds provided to qualified lenders. Usually, GSEs issue debt in the agency market. To attract a broad range of investors, the agencies often structure bonds with yields higher than prevailing rates. This is achieved by embedding a call provision in these bonds that allows the GSEs to redeem bonds earlier than their nominal maturities. Typically, the embedded options in callable bonds are toward high strikes. Contrary to homeowners, who are long the prepayment option and demand volatility from markets, investors of agency callable bonds are effectively short the call option and supply a large amount of volatility, usually in high strikes. Although the GSEs do absorb some of the supply themselves to offset their retained demand of prepayment options, they very often pass on the supply to the volatility market by swapping these fixed-rate callable bonds to floating rates and selling the optionality to volatility dealers. This supply in volatility has become steadier and more dependable in recent years owing to more predictable schedules of agency issuance.

Another important effect of the mortgage market on the supply-demand balance of volatility skews stems from the convexity of mortgages portfolios. The price movement of a mortgage does not vary linearly with its underlying rate; rather, it changes in a curved fashion. Contrary to Treasury bonds or swaps that have moderate positive convexity, mortgages exhibit negative convexity, and their curvature is considerably bent. This is evinced in that a market rally shortens mortgage duration due to increasing prepayments, and a sell-off extends its duration due to diminishing prepayment activities. In other words, a duration-hedged mortgage appreciates much slower in rallies than the hedge decreases, and the mortgage decreases much faster in sell offs than the hedge appreciates. Managing this unfavorable bias for mortgage assets, also known as the negative convexity risk,[2] has always posed a challenge to active hedgers such as GSEs and servicers. Often, these mortgage hedgers enter the volatility market and buy convexity in options to offset their negative convexity exposure. Usually, low strike options are bought to hedge the prepayment risk and higher strike options for the extension risk. These activities frequently generate waves of demand for off-the-money options and have a large impact on the volatility skew.

There are also other sources that influence volatility skews. Though less dominant or predictable than those described above, they do play a visible role in the supply and demand balance of the volatility market. The following are particularly noticeable.

The Federal Home Loan Banks, which are GSEs, often provide fixed-rate term financing to their member banks though a cancellable arrangement, which entitles the lender to terminate the funding earlier than the specified maturity. It is so structured to reduce the nominal coupon costs for the members, yet at the risk of member's financing being called away if rates go higher. The member banks have effectively sold volatility at low strikes to their Federal Home Loan Banks, which usually swap the trade with option dealers, resulting in a supply of low strike volatility to the market.

[2] For a recent discussion on managing this risk, see *Mortgage Convexity Risk,* S. Modukuri et al, Lehman Brothers, 2003.

As alluded to earlier, investors holding securitized mortgages are generally passive volatility hedgers. They sell prepayment options for potential higher yields in return. But under certain market conditions, they do decide to mitigate their volatility and convexity risks with purchases of both at- and off-the-money options. In doing so, they turn their implicit need of volatility into active market demand.

There are also non-mortgage-related volatility sources. For example, corporate debt with embedded call options is very popular with retail investors, as they are ready to sell optionality in return for higher coupons. Like GSEs, corporations usually swap these deals and pass the embedded options to dealers. Depending on the structure, off-the-money options are usually supplied to the volatility market.

The interest rate volatility market in the U.S. is well structured, primarily reflecting the hedging needs of mortgage portfolios against the supply from callable issuance. Seated in the center of all the activities are option dealers. They make markets to absorb or redistribute the supply and demand in volatility, regardless of whether it originates from homeowners, callable bond buyers, or others. The net of all the supply and demand eventually falls into dealer's books. Dealers have both long and short volatility positions, and some general offsetting is expected, but a certain concentration of volatility risks in their books is almost unavoidable. As described above, the demand often concentrates in product structures different from that of the supply. Not only is it true for option maturities across the vol-grid, it is also true along the strike dimension. Depending on past activities, as well as on current conditions and future prospects, dealers may find themselves accumulating substantial amounts of long and short volatility exposure around certain low or high strikes. Since such concentrations in strikes are susceptible to unexpected market moves and must be actively managed, they undoubtedly affect skew levels traded in the volatility market.

COPING WITH SKEWS

A Practical Viewpoint

As noted before, the volatility market was initially unconcerned with skews. Players might have been aware of the skew implications of their model assumptions, but they were hesitant to address the problem due to a lack of market observation. As the market evolved, they gradually realized the significance of skewed volatilities and paid close attention to their levels and dynamics. Naturally, handling the volatility skew became an inescapable issue.

Since the advent of skews in the volatility market, a considerable number of participants have, for quite a period of time, been content with just making manual skew adjustment to volatilities. Many of them believe that manifestations of skews are simply a result of a supply and demand imbalance, a reasonable assessment considering that the volatility market largely serves the needs of the mortgage market with distinctive volatility requirements. For example, investors would attribute occasional high levels of high-strike volatility to mortgage portfolios' demand in high-strike options for extension protection. Dealers are obliged to offer them; but they are willing to do so only at levels with which they feel comfortable. They normally hedge the resulting high-strike exposure with at-the-money options, knowing full well that if rates do move up, their hedged position would become net short in volatility. Since the market is one-sided, subsequent rebalancing could drive up volatility and be costly. Thus, dealers often push up high-strike volatilities in anticipation of demand for extension protection trades.

This pragmatic viewpoint about skews can be effective in dealing with a particular type of skew-related trades. There is no mystery about skews with this view; they are simply an artificial effect from an imbalanced volatility market. This understanding has led quite a number of practitioners to believe that it is unnecessary to resort to complicated approaches toward skews so long as their behaviour could be explained and understood in plain supply-demand arguments.

However, reality is generally more complex than the isolated portrayal given above. A broader perspective on skew-related issues may lead participants to argue otherwise. Considerations of the increased skew information available in the market, the close interplay between European and Bermudan volatilities, the skew effects on complex structures, and the desire for a unified risk management all argue for a systematic treatment of volatility skews.

Integrated Approaches

It is true that the main focus of the volatility market is on the needs of the mortgage market and any heightened activity from a particular need can move some volatility skew to unusual levels. However, as the market has matured, artificial skew levels are often unsustainable. Increased sophistication in arbitrage has enabled savvy dealers to benefit from abnormal skew levels. Meanwhile, misplaced skew levels have also attracted a growing number of speculators who pay close attention to and take quick action against them. It has been increasingly evident in the current market that large skew moves are often met and explored quickly by vigilant dealers and speculators and that dislocated skew levels are usually short lived.

Not only have these skew imbalances been frequently corrected by observant participants, mortgage hedgers have also recently begun bridging the gap themselves. For example, those hedgers normally demand European options to manage their volatility and convexity exposures and supply Bermudan options by swapping callable debt. As a result, European volatilities have usually been offered at higher levels than comparable Bermudans. Some hedgers already recognized this value differential and started substituting pricey Europeans with inexpensive Bermudans. While still not prevalent, the increased hedge with Bermudans in place of Europeans has already blurred the traditional supply and demand distinction. Indeed, the gap between European and Bermudan volatilities has recently been narrowed noticeably. Though traditionally viewed as exotic swaptions, Bermudans are commonplace in the current market and are now almost considered a vanilla product. As yet another sign of convergence of these two sources of volatility, European-Bermudan switch trades have been frequently quoted in the inter-dealer broker market.

The expanded utilization of Bermudan swaptions has added another level of complexity to the analysis of the volatility skew. Most notably, the strike of a Bermudan option applies to all allowable exercise dates. Depending on curve shapes, the at-the-money rates for individual exercise dates of a Bermudan cover a wide range. For instance, in a upward sloping curve environment, the at-the-money rates for near exercise dates of the Bermudan are lower than those of distant dates. As will be explained later, a nominally defined at-the-money Bermudan strike is actually of high strike for closer exercise dates and low strike for more remote dates. A Bermudan swaption may be of both high- and low-strike volatilities at the same time, which renders the simple European skew characterization inadequate in Bermudans. A more systematic treatment is thus required to understand and manage Bermudan skews.

Beside standard Bermudans, truly exotic swaptions include, for example, those with step-up coupons or coupons even carrying optionality themselves, such as in callable range notes. They may also carry spreads, variable upfront costs, flexible redemption or surrender values, etc. For consistency, those structures are normally priced and hedged in close reference to European volatility skews. The ever-increasing sophistication in exotic products has increased the need for a modernization of skew management. For the purpose of capturing skew effects on those products, a model-based skew approach is undoubtedly more reliable than the old-fashioned ad hoc adjustments.

As conveyed earlier, the volatility business is not yet an efficient two-way market. There is no doubt that inefficiency still exists across different segments of the volatility market. This shall serve as yet another justification for modelling the volatility skew systematically. The identification of inefficiencies in the volatility market is important from a relative value perspective. Model-based skew approaches provide a more objective framework to uncover anomalies than the naked eye and limited experience.

Lastly, the risk management of volatilities among different products also calls for unified skew approaches. Volatility exposures in Bermudans and exotic swaptions are routinely hedged with a portfolio of Europeans. Even for Europeans themselves, off-the-money volatilities are often hedged with at-the-money ones. The risk management of skew dynamics among various products and across different strikes has always been a priority for practitioners. A systematic modelling approach aggregates skew risks hidden in different instruments and makes disparate volatilities manageable at portfolio levels. Prudent volatility participants have long been dependent on sound skew models to safeguard their positions, as well as exploit opportunities.

Volatility products have become increasingly integrated. Model-based skew approaches have the advantages of revealing the underlying cause of skew formations, identifying relative skew values from aberrant market activities, facilitating analyses of skew effects on complex products, and allowing for aggregation of volatility risks.

SKEW MODELLING

The Normal Skew

As mentioned earlier, the volatility skew of the normal interest rate process serves as a useful reference in measuring other skew patterns because the underlying assumption is intuitive and easy to analyze. In fact, under the normal model, European options can be valued directly with closed-form formulas. For example, a payer swaption, with initial rate F, strike K and expiry T, can be priced exactly with

$$V = \sigma_N \sqrt{T}\left[N(d)d + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d^2} \right], \qquad d = \frac{F-K}{\sigma_N \sqrt{T}}, \tag{1}$$

where σ_N is the constant basis point volatility and $N(d)$ is the cumulative normal distribution function.[3]

It was argued earlier that the volatility skew of the normal model is a downward sloping curve, which is deduced from the fact that its nominal percentage volatility is higher for low rates and lower for high rates. It may also be tempting to further deduce that the equivalent Black volatility for strike K is simply σ_N/K based on the inverse relationship between the percentage volatility and the rate level. This extension proves to be too simplistic. To an acceptable degree of accuracy, the Black volatility for the normal model, denoted as B_N, takes on the form

$$B_N(K) \approx \frac{\sigma_N}{\sqrt{FK}}. \tag{2}$$

$B_N(K)$ is actually an inverse function of the geometric average of F and K, or an inverse square root function with respect to the strike K.

[3]The full swaption price is actually the product of V and a sum of the discounted accrual factors of the underlying swap. The function $N(d)$ is defined as the definite integral of the normal distribution function from negative infinity to d,

$$N(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d} e^{-\frac{1}{2}x^2} dx.$$

In general, the Black volatility B depends on K for a given rate process and the function $B(K)$ quantifies its volatility skew. The exact form of $B(K)$ is usually obtained with a numerical search procedure that matches a given option price V to that of the standard Black formula

$$V = FN\left(\frac{X}{B\sqrt{T}} + \frac{1}{2}B\sqrt{T}\right) - KN\left(\frac{X}{B\sqrt{T}} - \frac{1}{2}B\sqrt{T}\right), \tag{3}$$

with

$$X = \ln\left(\frac{F}{K}\right). \tag{4}$$

This formula is the solution to the lognormal model in which the basis point volatility is strictly proportional to rate. With v denoting in general the basis point volatility of a process, the basis point volatility for the lognormal model is

$$v_{LN} = \sigma_{LN}\, f, \tag{5}$$

where σ_{LN} is the constant percentage volatility and f is the random underlying rate with the initial value F. By construction, the Black volatility for the lognormal model does not very with strikes; that is, a flat skew function $B_{LN}(K) = \sigma_{LN}$.

The compact expression (2) offers a quick way to examine the dependency of the normal skew on strikes. But its accuracy sometimes is inadequate for practical purposes. A preferred approximation of the normal skew reads[4]

$$B_N(K) \cong \frac{\sigma_N}{F}\left(1 + \frac{\sigma_N^2 T}{24F^2} + \frac{1}{2}X + \frac{1}{12}X^2\right). \tag{6}$$

Although the expression (6) appears more involved, it is still an analytic form. Not only is it more accurate, it is also more explicit in revealing the skew pattern. Note that the expression comprises three distinctive types of term. The first one is independent of strike, the second one depends linearly on X, and the third one is a quadratic function of X. Defined in (4), X is the lognormal difference of F and K and is viewed as a lognormal measure of the option's moneyness. Recognizing the functional form of (6), it is natural to identify the constant term as the at-the-money volatility, the linear term as the clean skew, and the quadratic term as the added smile.

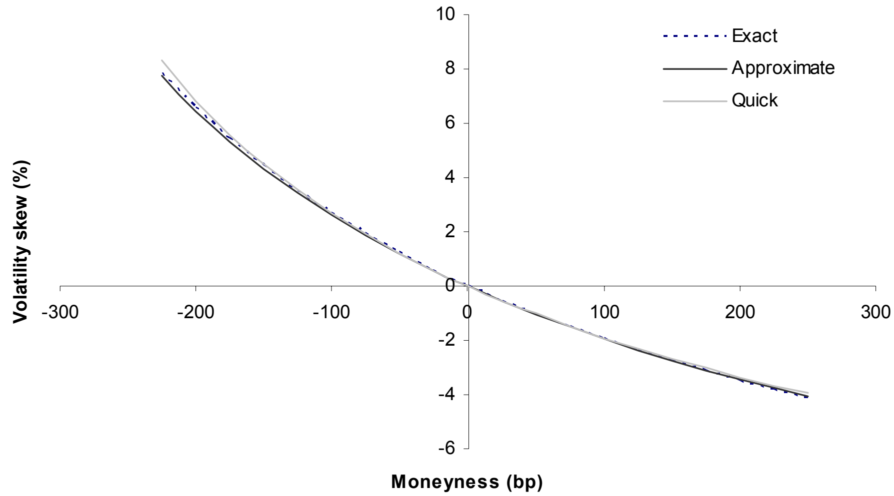
The quadratic contribution indicates that the normal model, generally believed to give rise to a skewed volatility curve, embeds a certain amount of smile. Moreover, it is noted that for the at-the-money strike ($X = 0$), the Black volatility is not simply σ_N/F as estimated routinely in practice by market participants. It may seem surprising that according to (6), there is a correction to this popular conversion. The magnitude of this correction is not always negligible, especially for long-dated expiries or for high levels of volatility.

The expression (6), though still approximate, is attractive for its explicit display of the skew and the smile under the normal model. Its accuracy can be readily checked against the exact Black volatility by substituting the normal option price (1) into the Black formula (3) and solving for the equivalent volatility B at various strikes. As a comparison, Figure 7 displays the exact normal skew curve alongside the favored approximation (6), as well as the nimble substitute (2).

---

[4] For a derivation of this approximation as well as other skew expressions referred to later, see F. Zhou, 2003. "Black Smirks." *Risk,* May, pages 87-91.

Figure 7. Comparison of Three Alternative Results of the Volatility Skew Curve under the Normal Model with σ_N = 0.01 and for a Three-Year Option with F = 4.5%



As seen in the graph, the difference between the approximation (6) and the exact result is nearly indistinguishable over a wide range of strikes. Even the quick skew expression (2) is sufficiently close to the exact result.

The Street Favorites

Like the lognormal model, the normal model is too idealistic to represent reality. Street modellers have long been searching for an interest rate process that is not overly complicated but flexible enough to embody the observed strike dependency of the Black volatilities.

Traditionally, the Street settled on a process known as the constant elasticity of variance (CEV). In this model, the basis point volatility of a rate process is assumed to be proportional to a power of the rate, namely,

$$v_{CEV} = \sigma_\beta f^\beta,$$

where the power β is called the CEV coefficient and σ_β the CEV volatility. The process employs an extra parameter and is, indeed, more flexible than the standard normal and lognormal processes. What is attractive about the CEV model is that it includes the two reference processes as special cases. For β = 0, the rate dependency of the basis point volatility disappears and the CEV model reverts to the normal version; for β = 1, the proportional volatility resumes and the CEV model recovers the lognormal version. This flexible construction allows users to study the skew variation continuously over the range spun by the normal and lognormal processes. As shown earlier in Figure 6, the historic evolvement of skews has mostly been confined within this range.

Unfortunately, this hybrid process does not have exact analytical solutions, except for some special β. In order to examine the skew implication of the CEV model, it is still preferable to have an analytic expression, albeit approximate. The equivalent Black volatility of a standard option under the CEV model, for example, can be written as

$$B_{CEV}(K) \cong \frac{\sigma_\beta}{F^{1-\beta}}\left[1 + \frac{\sigma_\beta^2 T(1-\beta)^2}{24F^{2(1-\beta)}} + \frac{1}{2}(1-\beta)X + \frac{1}{12}(1-\beta)^2 X^2\right]. \tag{7}$$

Similar to the form of the normal skew (6), the CEV result also possesses three modulated terms. Not surprisingly, the CEV skew becomes a constant in the case of $\beta = 1$ and reduces to expression (6) for $\beta = 0$, correctly recovering the two special results. For $\beta < 1$, the CEV skew displays various degrees of downward sloping curves. Although some curvature is present due to the quadratic contribution, it is insignificant compared with the linear term.

In practice, modellers adjust the coefficient β to fit to the observed market skews. Though imperfect, the CEV model is able to take a bulk of market skews into account. It has been used widely on the Street, especially in the earlier years of the skew history. Figure 8 illustrates the skews generated by the CEV model with various β-coefficients.

Figures 8-9. Volatility Skews for a Three-Year Option with F = 4.5% under the CEV and the Shifted Lognormal Models with Various Beta and Shift Parameters, Respectively

Figure 8. CEV Skews

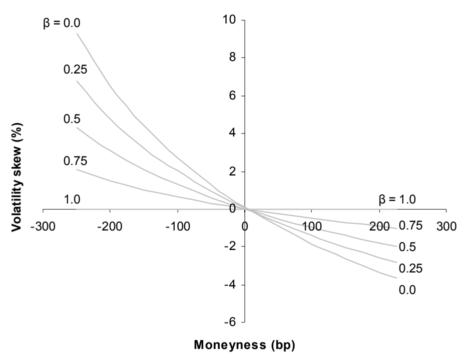
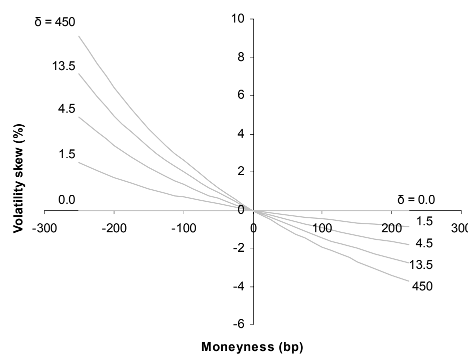


Figure 9. Shifted Lognormal Skews



More recently, though, modellers started looking into an appealing alternative. It has been found that a modified lognormal process, known as the shifted lognormal model, demonstrates very similar skew behaviors, as does the traditional CEV process. Instead of a pure proportional volatility assumption, this modified lognormal process assumes that the basis point volatility of an interest rate is proportional to its rate plus a shift. Explicitly, it is

$$v_{SLN} = \sigma_{SLN} (f + \delta), \tag{8}$$

where δ is the shift rate introduced in the modification. Similar to the CEV coefficient β, the shift parameter δ incorporates more flexibility than the standard normal and lognormal models. Like the CEV model, the shifted lognormal model also includes these two processes as its special cases. For $\delta = 0$, the basis point volatility is strictly proportional to the rate level and the model recaptures the pure lognormal version; for a shift rate much larger than the rate itself ($\delta >> f$), the dependency on f of the basis point volatility is negligible and the model approaches the normal version.

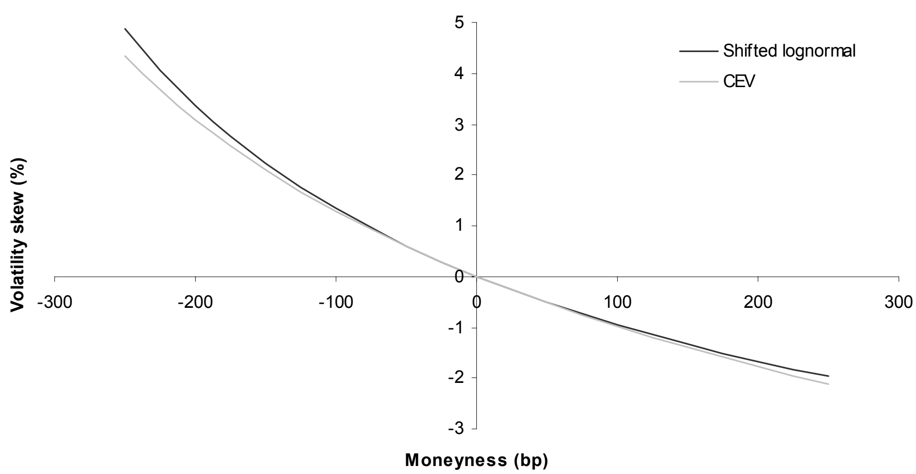
Apart from its similarity to the CEV model, the shifted lognormal model is more attractive to users because it is analytically tractable. The price of a standard option under the shifted lognormal model can be obtained with the Black formula (3) provided that the input replacements of f by $f + \delta$, K by $K + \delta$, and σ_{LN} by σ_{SLN} are made. With its price exactly known, its equivalent Black volatility can then be computed with the usual inversion of the original Black formula (3). However, sometimes it is still desirable to examine the skew curve of the shifted lognormal model through an explicit, though approximate, analytic expression, such as the one below

$$B_{SLN}(K) \cong \frac{\sigma_{SLN}(F+\delta)}{F}\left[1+\frac{\sigma_{SLN}^2 T(\delta^2+2\delta F)}{24F^2}+\frac{\delta}{2(F+\delta)}X+\frac{(\delta^2+2\delta F)}{12(F+\delta)^2}X^2\right]. \tag{9}$$

As expected, the result for the shifted lognormal model (9) reduces to a constant in the lognormal limit ($\delta = 0$) and approaches to the normal expression (6) in the normal limit ($\delta >> f$). Figure 9 illustrates the shifted lognormal model skews generated with various shift parameters.

Despite the apparent difference between (7) and (9), the two are comparable in terms of functional behaviors. For instance, if the model parameters are chosen such that $\beta = F/(F + \delta)$, their skew terms match perfectly. Although this results in a mismatch in the smile terms (with the shifted lognormal smile slightly more curved), the effect is negligible because of rather small quadratic contributions in these two models. In light of this compatibility, these two models can be used almost interchangeably to represent a Black volatility curve that is overwhelmingly skewed. For illustration, Figure 10 compares the skew shapes of these two models under the matching condition specified above.

Figure 10. Compatibility of the Volatility Skews Between the CEV and the Shifted Lognormal Models for a Three-Year Option with F = 4.5% under the Model Parameters β = 0.5, δ = 4.5%, σ_β = 0.0042 and σ_δ = 0.10.



The modest complexity of these two models as well as their flexibility to produce reasonable skew profiles has made the CEV and the shifted lognormal models two favorite choices by skew modellers on the Street.

Advanced Skew Models

It is understood that the two popular models just discussed have limited capability in producing smiles. This limitation has led modellers to extend the dependency of the basis point volatility on the underlying rate.

The basis point volatility of a rate process can be assumed to depend on a flexible function of the rate level. Specifically, the general level dependency, denoted with the subscript *lvl*, can be written as

$$v_{lvl} = \sigma f \alpha_{lvl}(f), \qquad \alpha_{lvl}(F) = 1, \tag{10}$$

where σ is an overall percentage volatility and $\alpha_{lvl}(f)$ is an arbitrary function of f with unit initial value. The level volatility v_{lvl} is so defined such that it generalizes the standard lognormal process (5) and it simply becomes (5) in case of a level-independent α_{lvl} function.

Because of the rather general form of the level dependency in (10), the model is not expected to have analytic solutions. Numerical effort is needed to compute the equivalent Black volatility curve. Fortunately, a reliable closed form approximation is still possible for this generalized model. It is found that its equivalent Black volatility can be elegantly expressed as

$$B_{lvl}(K) \cong \sigma\left[1 + \frac{\sigma^2 T}{24}\left(2\lambda - \lambda^2 + 2\zeta\right) - \frac{1}{2}\lambda X + \frac{1}{12}\left(2\lambda - \lambda^2 + 2\zeta\right)X^2\right], \qquad (11)$$

with

$$\lambda = F\alpha_{lvl}^{'}(F), \qquad \zeta = F^2\alpha_{lvl}^{''}(F).$$

The primes on $\alpha_{lvl}(F)$ denote either the first or second derivative of the function $\alpha_{lvl}(f)$ with respect to f at its initial value F.

Similar to the results seen so far, the general Black volatility curve (11) still possesses three distinctive terms: constant, linear, and quadratic functions of the moneyness X. The skew of this general level model is regulated by the first derivative of the level function α_{lvl} while the smile component is linked to both the first and second derivatives of α_{lvl}. In order for a substantial smile to emerge, the function $\alpha_{lvl}(f)$ has to be considerably curved.

Like the simple normal skew expression (2), a nifty yet less accurate skew function is available for this generalized level model. If the smile and the constant terms are insignificant in (11), the remaining skew term can be rearranged, without further loss of accuracy, as

$$B_{lvl}(K) \approx \sigma\left(\frac{K}{F}\right)^{-\frac{\lambda}{2}}.$$

In this reduced form of the Black volatility, the strike dependency is simply through the half power λ whose magnitude is directly tied to the first derivative of the level function. As should be expected, the above expression contains the normal version (2) as a special case.

Despite the generality of the level model, its shortcomings cannot be overlooked. For instance, awkward level dependency sometimes has to be assumed to produce an isolated volatility smile. This handicap has led modellers to infer that observed volatility skews cannot be entirely explained by just the level dependency of the basis point volatility. This inference is corroborated by empirical evidence. Close observations indicate that volatilities are fluctuating themselves and are driven by random sources other than just that of underlying rates. This has motivated modellers to consider a volatility of volatility approach in which volatilities themselves are assumed to be volatile.

Under this more advanced framework, the basis point volatility of a rate process can be written as

$$v_{vol} = \sigma f \alpha_{vol}, \qquad (12)$$

with the subscript *vol* denoting a volatile volatility. The process (12) is of the same form as the level function (10). But instead of a deterministic α-function, the volatility-of-volatility model assumes that the α_{vol} is itself random, with its own "basis point" volatility defined as

$$v_\alpha = b\alpha_{vol}, \qquad \alpha_{vol}(0) = 1,$$

where the constant b is the volatility-of-volatility parameter and the random function α_{vol} is scaled to take a unit value initially. Here, the volatility of α_{vol} is assumed to be proportional to its level, and this dependency makes the parameter b resemble a percentage value, similar to the overall percentage volatility σ. Furthermore, there is no reason to believe that the underlying rate and the volatility processes are independent of each other. Volatility variation is often coupled with changes in the level of rates. It is, thus, further assumed that the two processes are correlated with a non-zero correlation ρ.

The volatility-of-volatility process described above does not allow for an exact analytic solution. Instead, one has to rely on either numerical technique or analytic approximation. In the latter case, the equivalent Black volatility can be written in the following manageable form

$$B_{vol}(K) \cong \sigma\left[1+\frac{\sigma^2 T}{24}\left(6\tau\rho+2\tau^2-3\tau^2\rho^2\right)-\frac{1}{2}\tau\rho X+\frac{1}{12}\tau^2\left(2-3\rho^2\right)X^2\right], \tag{13}$$

with $\tau = b/\sigma$, the ratio of the volatility of volatility and the overall volatility. Although the volatility-of-volatility process is a two-factor process, its equivalent Black volatility function still conforms to the modulated form. Compared to the level function model, a noticeable advantage of the volatility-of-volatility approach is that by varying the correlation parameter, one can manipulate the relative size of the skew and smile effects much more easily. For example, with a correlation equal to the square root of 2/3, the smile term vanishes, while a zero correlation suppresses the skew term instead. With careful choice of the correlation parameter ρ, a desirable mix of the skew and the smile can be achieved. This allows for richer shape of the Black volatility curves, especially for smile related formations. This adaptability is not easily seen in the generalized level-dependent model discussed above.

A Practical Combination

The volatility-of-volatility model appears more suitable for manifesting smiles, but sometimes the fitted values of model parameters cause concerns. They may be too large, to the extent that they are irreconcilable with empirical estimates. For a realistic description of the skew and the smile effects, a combination of the level-dependent and the volatility-of-volatility processes is more appropriate. A reasonable form of a level function is chosen to capture the bulk of skew dynamics, and a volatility-of-volatility process is superimposed for fine-tuning desired smiles. Such a mixed process can be composed as

$$v_{mix} = \sigma f \alpha_{lvl}(f)\alpha_{vol},$$

where α_{lvl} is defined in the level process (10) and α_{vol} is the volatility process described by (12). With certain added algebraic complexity, an analytical approximation to the Black volatility curve is still available for this coupled process. Its skew function can be written, for example, as

$$\Delta B_{mix}(K) \cong \Delta B_{lvl}(K)+\Delta B_{vol}(K)+\frac{1}{4}\rho\sigma^2 bT\lambda, \tag{14}$$

where $\Delta B(K) = B(K) - \sigma$, and the corresponding B_{lvl} and B_{vol} are given by (11) and (13), respectively. Not too surprisingly, the skew of the combined model is essentially the sum of the two individual ones. There is a cross effect between the level function and the volatility of volatility as shown by the last term of (14); but it only introduces a flat correction to the overall Black volatility, which vanishes in case of vanishing correlation. The result (14) is rather general, valid for an arbitrary level-function coupled with a volatility-of-volatility process.

For practical use, though, a particular level function $\alpha_{lvl}(f)$ has to be specified, and a CEV function is preferable. Discussed before and shown here in the scaled form, it is

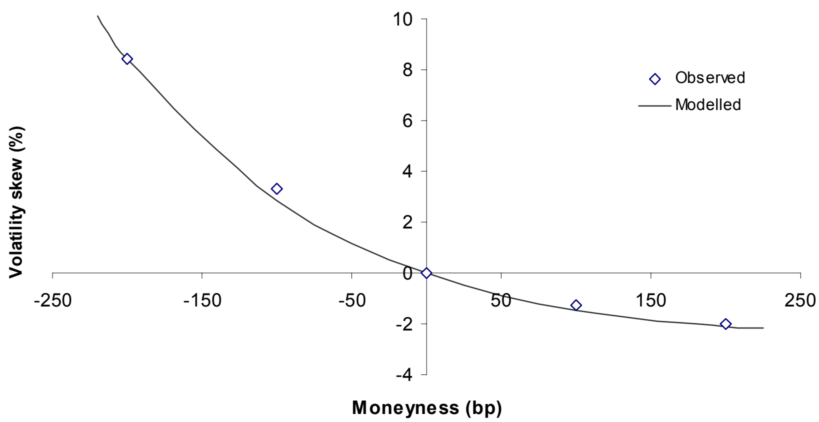
$$\alpha_{CEV}(f) = \left(\frac{f}{F}\right)^{\beta-1}.$$

The Black volatility for this particular mixed model becomes

$$B_{mix,\beta}(K) \cong \sigma\left\{ 1 + \frac{\sigma^2 T}{24}\left[(1-\beta)^2 + \tau^2(2-3\rho^2) + 6\tau\rho\beta\right] + \frac{1}{2}(1-\beta-\tau\rho)X + \frac{1}{12}\left[(1-\beta)^2 + \tau^2(2-3\rho^2)\right]X^2 \right\}. \tag{15}$$

This result takes advantage of both the level-dependent and the volatility-of-volatility processes and provides enough flexibility for the calibration of observed Black volatility curves. Specifically, there are three model parameters—the CEV coefficients β, the volatility-of-volatility *b*, and the correlation ρ—that one can rely on for realistic replication of volatility skews and smiles.

Figure 11. Calibration of a Combined CEV Model to an Actual Black Volatility Pattern Recently Observed for the 3Y5Y Swaptions with *F* = 4.67%; Fitted Model Parameters Are σ = 0.27, β = 0.68, *b* = 0.39 and ρ = - 0.16



The effectiveness of this combined process is illustrated here by an actual calibration of the Black volatility function (15) to a recently observed volatility formation, which happens to have a noticeable embedded smile. Figure 11 shows such a calibration for a set of 3Y5Y swaptions.

EFFECTS ON COMPLEX INSTRUMENTS

A Digital Option

Although the volatility skew of standard options has been extensively studied, it is not immediately clear how it may affect exotic options. Modellers usually try to quantify the effect through skew models applicable to both standard and exotic options. This practice is illustrated in the study of a digital option below.

A digital option, or a binary option as it is sometimes called, is considered the simplest exotic option. It represents a simple win-or-loss option on the underlying being above or below a certain level. For example, a digital swaption pays one percent if and only if the underlying swap rate is above a pre-specified level at expiration or otherwise pays nothing. Conceptually, the value of the digital swaption corresponds to the probability of the swap rate being above the digital level at the expiration. This probability is rather sensitive to the process of the underlying rate and, by extension, sensitive to the related volatility skews. In the following, this sensitivity of the digital option is examined with two distinctive rate processes.

It is known that under the skew-less lognormal process, the unit price of such a digital swaption can be evaluated exactly with the closed form[5]

$$D_{LN} = N\left[\frac{1}{\sigma_{LN}\sqrt{T}}\left(X - \frac{1}{2}\sigma_{LN}^2 T\right)\right]. \tag{16}$$

On the other hand, under the normal model, whose skewed volatility curve is significantly different from the flat skew in the lognormal model, the same digital option can be priced exactly with

$$D_N = N\left(\frac{F-K}{\sigma_N\sqrt{T}}\right). \tag{17}$$

Though precise, the normal result (17) is quite different in appearance from the lognormal result (16). For a straightforward comparison of the digital values under these two skew processes, it is instructive to rewrite (17) approximately as

$$D_N \cong N\left[\frac{1}{\sigma_{LN}\sqrt{T}}\left(X - \frac{1}{2}X^2 + \frac{1}{6}X^3\right)\right], \tag{18}$$

where $\sigma_{LN} = \sigma_N/F$ is the comparable lognormal volatility. This approximation is quite accurate for a range of reasonable digital levels, and, more importantly it affords a direct comparison with the lognormal result (16). While the flat skew results in a linear dependency of the functional argument on the moneyness X, the normal skew induces additional higher order dependency on X. The higher order contribution only vanishes for the at-the-money digital level. In fact, the result (18) assures that the at-the-money digital (X = 0) has an $N(0)$ = 50% chance of being in the money, an outcome expected from the symmetry of a normal process. In contrast, the at-the-money digital with a flat volatility skew does not promise such even chances, which is evident from the constant term in (16).

The valuation of the digital swaption considered here is sensitive to volatility skews because its payoff is highly irregular with respect to the underlying swap rate. However, for payoffs with certain symmetry, the effect of skew may be subdued, and the effect of smile may instead become dominant. Such an elevated smile effect usually occurs in convex payoffs.

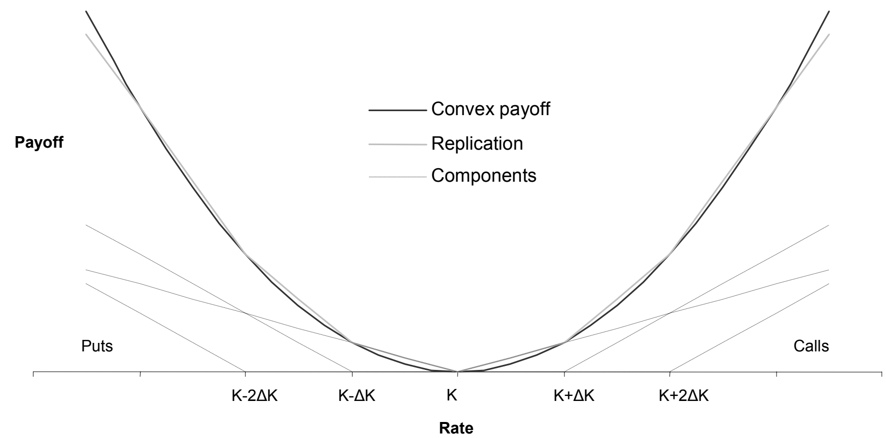
Convexity Pricing and Hedging

A convex payoff is curved and cannot be offset completely with just a linear delta hedge. This mismatch is known as convexity. Specifically, a payoff with positive convexity, as in the floating-rate payment of a constant maturity swap (CMS), increases in value for any movement of the underlying swap rate. A payoff of negative convexity is epitomized by the price profile of a duration-neutral mortgage asset that decreases in value with changing rates.

For the purpose of studying the skew effect on convexity, it is illustrative to duplicate a convex payoff with a portfolio of standard options. Such a payoff replication directly establishes the relationship between the convexity and standard options and explicitly exposes the effect of the volatility skew. Figure 12 demonstrates the replication of a positive convexity.

[5] As in a regular swaption, the full price should also include an overall multiplier equal to the sum of discounted accrual factors.

Figure 12. Convex Payoff Replicated with Vanilla Options at Various Strikes; the Dotted Straight Lines Represent Individual Option Payoffs; the Sum of Them, Represented by the Solid Grey Line, Constitutes the Replication



As seen in the graph, a convex profile can be matched with a series of call and put options struck at certain intervals, with calls placed at ascending high strikes and the puts at descending low strikes. The replication can be made arbitrarily accurate with a large enough number of options positioned at arbitrarily fine intervals. In practice, a compromise between replicating accuracy and manageable size is usually considered.

Since the Black volatilities are directly quoted for these replicating options, the role played by the volatility skew is explicit in the valuation of the convexity. The convexity value is essentially the price sum of all the options employed in the replication and is therefore affected by the Black volatilities at all strikes, weighted variably, though. The replicating weights for the selected strikes are determined by the actual curvature of the convexity around those strikes. A regular quadratic shape, such as a CMS payoff, has a uniform curvature and requires more or less equal weights in replicating options. For the convexity of a mortgage asset, the curvature usually peaks at the optimal refinancing rate and ebbs on both sides of the peak rate, resulting in diminishing replicating weights. Thus, while volatilities of all strikes are relevant to replicate the mortgage convexity, the dependency usually moderates for low and high strikes.

It is expected that a pure sloping volatility skew does not have much of an effect on the valuation of a symmetric convexity. The matching weights between low and high strikes usually balance out the asymmetry in the skew, negating most of its effect. In contrast, the symmetry of a volatility smile superimposed on the even replicating weights maximizes the smile effect in the determination of a convexity value.

The convexity replication described above underscores the significance of the volatility smile to the valuation of convex payoffs. Though conceptually advantageous, it is sometimes cumbersome to actually implement the replication based valuation due to large number of options involved. The skew models discussed previously are usually more efficient computationally. Simply speaking, the value of convexity can be obtained with a mathematical integration of the underlying distribution over the convex payoff. For certain convex profiles, it can be carried out conveniently. For instance, under a normal skew model, the value of the CMS convexity (apart from an accrual factor) can be written approximately as

$$C_N \cong \frac{1}{2} c \sigma_N^2 T, \tag{19}$$

where c is an overall coefficient and its value depends exclusively on the structure of the underlying swap. The simple expression (19) is used ubiquitously by practitioners as a first-order estimate for the CMS convexity. Its simplicity has earned it popularity and has led it to become an ad hoc market practice, to the extent that most users are no longer aware of its origin.

This indifference, however, overlooks the effect of volatility smiles on the convexity value. Suppose the same CMS is valued under the lognormal model, its convexity value is instead given by

$$C_{LN} \cong \frac{1}{2} c \sigma_N^2 T \left(1 + \frac{1}{2} \sigma_{LN}^2 T \right), \tag{20}$$

where $\sigma_N = F\sigma_{LN}$ is a comparable basis point volatility. The lognormal result differs from the normal result (19) by a percentage factor that is quadratic in volatility σ_{LN}. This is expected because the linear term, which would have been from the skew, cancels due to symmetry. The result shows that the CMS convexity is, indeed, sensitive to volatility smiles. The difference between (19) and (20) results from the embedded smile of the normal skew curve.

The above comparison affirms that the convexity value is shielded from the influence of volatility skews, but not from smiles. This is generally true and can be further demonstrated with more advanced skew models. For instance, under the volatility of volatility model considered before, the value of the CMS convexity takes on the form

$$C_{vol} \cong \frac{1}{2} c \sigma_N^2 T \left[1 + \frac{1}{2} \sigma^2 T \left(1 + 4\tau\rho + \tau^2 \right) \right],$$

where $\sigma_N = F\sigma$ is a basis point volatility comparable to those in (19) and (20). Again, the result differs from others only by a quadratic term in volatility, which is attributed to the smile portion of the volatility curve. In the case where the smile component is significant (say, with very large τ,), the effect on the overall convexity value also becomes significant. Hence, in an environment in which a prominent smile is present in the Black volatility curve, the CMS convexity value estimated with the ad hoc expression (19) is no longer adequate. It could lead to substantial underestimation.

Although the analysis above is carried out for the CMS convexity, similar studies can be applied to the negative convexity of mortgages. As opposed to evaluating the entire convexity as in the CMS case, mortgage investors often concentrate on certain parts of the negative convexity risks. Often, they are concerned about certain large directional moves in rates and try to purchase options to neutralize these moves. Despite the difference in objectives, the replication scheme is still applicable. In fact, the replication can be tailored to match a particular part of the negative convexity profile deemed risky by hedgers. This partial hedging strategy is cost effective in managing convexity exposures, but additional care needs to be taken in identifying the relevant portion of the volatility skew curve.

BERMUDAN SKEWS

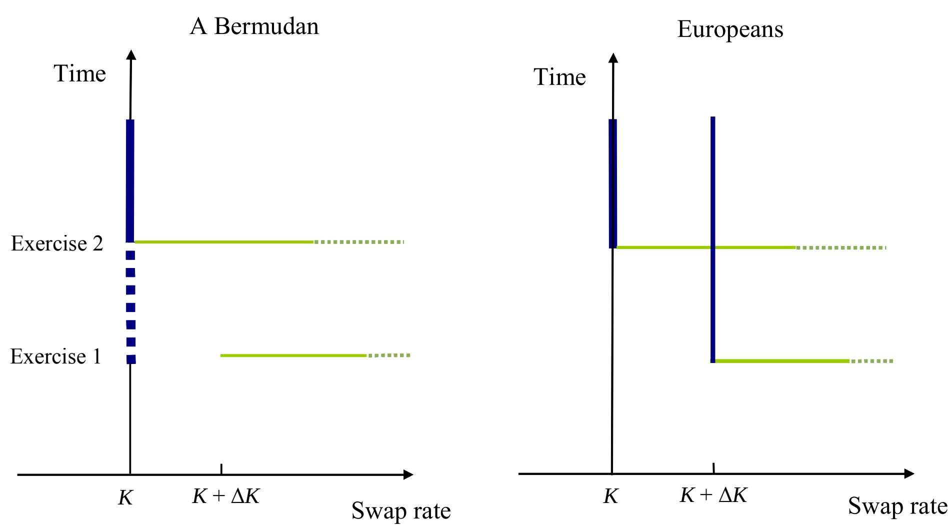
Discussion of the skew effect on complex products would be hardly satisfactory without explaining its role in Bermudan swaptions. Although Bermudans have traditionally been considered exotic, they are now viewed as standard products because of their widespread presence in the marketplace. They are routinely traded in the volatility market and are a major part of flow volume. Their liquidity has improved noticeably in recent years and is, in fact, currently not that far behind that of European swaptions. As volumes have increased, bid-offer spreads have been squeezed. As a result, pricing and hedging Bermudans has become increasingly delicate and has demanded unprecedented precisions. Due to their close affinity with European counterparts, Bermudans are examined in relationship to European volatilities. It is imperative that the interplay of volatility skews between them be understood.

A Bermudan swaption differs from the corresponding European swaption in that it allows for multiple exercises over the life of the underlying swap. There is no sure option expiration as in the case of an European swaption. With variable probability, potential expirations spread across the term of the underlying swap. Because of this structural variation, Bermudan swaptions are normally studied with term-structure interest rate models that serve as a bridge to European swaptions. In order to take into account the skewed volatilities observed in Europeans, these models are required not only to fit to the at-the-money vol-grid but also to match the off-the-money volatility skews. A Bermudan price, computed with a properly calibrated term-structure model, is then consistent with observed European skews and is considered to reflect the fair costs of its European hedges.

Term-structure interest rate models are a systematic and proven means in dealing with actual Bermudans. However, in order to understand the skew effect on a Bermudan swaption, it is more intuitive to examine the Bermudan payoff in relation to those of hedging Europeans. Similar to the convexity analysis carried out before, the replication technique can be utilized for this purpose, though it has to be extended to accommodate multiple expirations.

For the sake of visualizing such a multi-expiry replication, a hypothetic Bermudan payer swaption with only two allowable exercises is considered here. This swaption also represents a call option on the swap rate. A schematic setup of this trial Bermudan, along with its European replication, is depicted in Figure 13.

Figure 13. Two-Expiry Bermudan Payer Swaption with Strike *K* Replicated with Two European Payers Struck at *K* and *K* + Δ*K*, Respectively



In the Bermudan part of the diagram, the thick vertical line represents the underlying swap and is placed at its strike K. Because of the second exercise opportunity, the front portion of the swap may not materialize, which is indicated with line breaks. The two horizontal lines reside over the exercise dates and mark the range of the swap rates over which the Bermudan is to be exercised. For instance, the line at the second exercise indicates that as long as the swap is in the money on that date the option should be exercised, because further waiting would have resulted in no value at all. On the first exercise date, however, it is only advantageous to exercise when the underlying rate exceeds the strike K by a certain margin ΔK because if the rate is too close to the strike K, the in-the-money amount from the immediate exercise would have been too small to justify giving up the remaining exercise opportunity. It is optimal to delay the exercise in this situation for a statistically higher payoff at the second opportunity. The dotted part on the right of the exercise lines signifies that the exercise range extends into infinity with diminishing probability.

The payoff profile of this experimental Bermudan can be effectively duplicated with two European payers, as shown on the right of the drawing. The two underlying swaps are respectively represented by the two vertical lines, one long and the other short. The long one coincides with the swap that would result from the first Bermudan exercise; while the short one mirrors the swap from the second exercise. The latter is intended to hedge the second Bermudan exercise. Since there is no further exercise opportunity after the second date, the notional of the European with the short underlying has to be the same as the Bermudan in order match the two payoffs on the second date. However, the notional of the long underlying in the replication is expected to be reduced because the intrinsic value from exercising the Bermudan on the first date would be partially offset by the value of the European used to hedge the second exercise. This lesser amount is indicated by a thinner width for the long underlying. Practically speaking, if the Bermudan shall be exercised on the first date, its payout would be properly hedged by exercising the expiring European and selling the second European for its then market value.

In spite of its simplification, this artificial structure already contains essential features of an actual Bermudan. The setup is insightful in deciphering the interdependency between a Bermudan and its offsetting Europeans. In this example, the hedging portfolio contains two Europeans of different option expiries and swap tenors as well as at different strikes, a rich display of the interplay between the Bermudan and the related Europeans. In particular, the skew effect on the Bermudan can be readily analyzed. Depending on both the overall and relative levels of the two swap rates involved, the European payers could be either low or high strike, not just dependent on the overall strike K. For example, if the swap curve is upward sloping and the original strike K is at the money, namely, the same as the swap rate corresponding to the first exercise, the European with the long underlying is then at a high strike by the amount of ΔK while the one with the short underlying is at a low strike by the difference in the two swap rates. Therefore, the Bermudan under consideration is influenced by both low and high strikes, and the overall effect depends on the relative weights of two Europeans.

The European replication of a Bermudan swaption therefore enables one to pinpoint the effective option strikes and to identify the part of the volatility skew that is relevant to the Bermudan valuation. The choice of the simplified structure in the above analysis is intended to accentuate the decomposition scheme and to bring to light the European skew effect on a Bermudan. The approach can nevertheless be extended to analyze realistic Bermudan structures. In Figure 14, an actual 3Y7Y Bermudan payer is decomposed in the same spirit, and its European skew distribution is summarized.

The table describes the Bermudan in the first row and lists afterwards the corresponding Europeans required in the replication. Their maturities are identified in the first column as NxM, meaning an N-month option on an M-month swap and, in the case of the Bermudan, a non-exercise period of N months. The potential exercise dates are listed in the second column, followed by the forward swap rates for the corresponding underlying swaps. The fourth column shows the strikes of the Europeans resulted in the replication, and the fifth column displays the hedge amounts. The last column provides the individual option premia, with the European total matching the Bermudan value.

Figure 14. 3Y7Y Bermudan Priced as of June 16, 2003; Underlying Forward Swap Covers a Term from June 19, 2006, to June 19, 2013; Bermudan Is Struck at the Money and Is Exercisable Semi-annually

Swaption	Exercise	Forward	Strike	Notional	Premium
36x84 Bermudan payer	20??????	4.44	4.44	10000	590.53
36x84 European payer	20060619	4.44	6.47	1556	16.03
42x78 European payer	20061219	4.56	6.17	1458	22.14
48x72 European payer	20070619	4.67	5.99	1578	29.58
54x66 European payer	20071219	4.75	5.75	1554	34.75
60x60 European payer	20080619	4.83	5.67	1673	38.69
66x54 European payer	20081219	4.90	5.57	1791	41.95
72x48 European payer	20090619	4.96	5.45	1946	45.47
78x42 European payer	20091221	5.02	5.32	2030	46.21
84x36 European payer	20100621	5.07	5.21	2244	48.18
90x30 European payer	20101220	5.13	5.11	2563	49.82
96x24 European payer	20110620	5.19	4.91	2974	52.12
102x18 European payer	20111219	5.24	4.81	3835	54.02
108x12 European payer	20120619	5.29	4.62	5388	55.74
114x6 European payer	20121219	5.32	4.44	10000	55.83

As shown in the table, the strikes of the hedging Europeans start at 4.44% on the last exercise date and gradually move higher, reaching 6.47% for the first expiry. The required notional in the hedging portfolio starts from the original amount (10000) for the last European and gradually drops to 1156 for the first one. The moneyness of the Europeans is relative to the shape of the swap curve, which is upward sloping in this case. The forward swap rates range from 4.44% for the first exercise date to 5.32% for the last exercise date. As a result, the European expiring on the first date is high strike while the one expiring on the last date is low strike, with varying moneyness in between. As seen in this real example, the skew dependency of a Bermudan swaption is rather intricate. Nonetheless, it can be explicitly decomposed and analyzed.

Acknowledgement

Various comments received from individuals of the Lehman Brothers' liquid market sales, trading and research are greatly appreciated, with special gratitude extended to Kanav Bhagat and Fredrik Akesson for carefully reviewing the script to ensure accuracy and readability.

Lehman Brothers Fixed Income Research analysts produce proprietary research in conjunction with firm trading desks that trade as principal in the instruments mentioned herein, and hence their research is not independent of the proprietary interests of the firm. The firm's interests may conflict with the interests of an investor in those instruments.

Lehman Brothers Fixed Income Research analysts receive compensation based in part on the firm's trading and capital markets revenues. Lehman Brothers and any affiliate may have a position in the instruments or the company discussed in this report.

The views expressed in this report accurately reflect the personal views of Fei Zhou, the primary analyst responsible for this report, about the subject securities or issuers referred to herein, and no part of such analyst's compensation was, is or will be directly or indirectly related to the specific recommendations or views expressed herein.

The research analysts responsible for preparing this report receive compensation based upon various factors, including, among other things, the quality of their work, firm revenues, including trading and capital markets revenues, competitive factors and client feedback.

Any reports referenced herein published after 14 April 2003 have been certified in accordance with Regulation AC. To obtain copies of these reports and their certifications, please contact Larry Pindyck (lpindyck@lehman.com; 212-526-6268) or Valerie Monchi (vmonchi@lehman.com; 44-(0)207-011-8035).

Publications—L. Pindyck, B. Davenport, W. Lee, D. Kramer, R. Madison, A. Acevedo, T. Wan, V. Monchi, C. Rial, K. Banham, G. Garnham, D. Cappadona