Equity Derivatives

Europe

Madrid, April 16, 2012

VOLATILITY TRADING

Trading Volatility, Correlation, Term Structure and Skew

Colin Bennett
Head of Derivatives Strategy
(+34) 91 289 3056
cdbennett@gruposantander.com

Miguel A. Gil
Equity Derivatives Strategy
(+34) 91 289 5515
mgil@gruposantander.com

Santander GLOBAL BANKING & MARKETS

CONTENTS

While there are many different aspects to volatility trading, not all of them are suitable for all investors. In order to allow easy navigation, we have combined the sections into six chapters (plus Appendix) that are likely to appeal to different parts of the equity derivatives client base. The earlier chapters are most suited to equity investors, while the later chapters are aimed at hedge funds and proprietary trading desks. We have tried to explain the topics assuming the least amount of prior knowledge, but swiftly move onto advanced areas in order to add the most value.

Click on section title below to navigate **Page**

EXECUTIVE SUMMARY

DIRECTIONAL VOLATILITY TRADING

Directional investors can use options to replace a long position in the underlying, to enhance the yield of a position through call overwriting, or to provide protection from declines. We evaluate these strategies and explain how to choose an appropriate strike and expiry. We show the difference between delta and the probability that an option expires in the money and explain when an investor should convert an option before maturity.

- **Option trading in practice.** Using options to invest has many advantages over investing in cash equity. Options provide leverage and an ability to take a view on volatility as well as equity direction. However, investing in options is more complicated than investing in equity, as a strike and expiry need to be chosen. This can be seen as an advantage, as it enforces investor discipline in terms of anticipated return and ensures a position is not held longer than it should be. We examine how investors can choose the appropriate strategy, strike and expiry. We also explain the hidden risks, such as dividends, and the difference between delta and the probability an option ends up in-the-money.

- **Maintenance of option positions.** During the life of an American option, many events can occur where it might be preferable to own the underlying shares (rather than the option) and exercise early. In addition to dividends, an investor might want the voting rights or, alternatively, might want to sell the option to purchase another option (rolling the option). We investigate these life-cycle events and explain when it is in an investor's interest to exercise, or roll, an option before expiry.

- **Call overwriting.** For a directional investor who owns a stock (or index), call overwriting by selling an OTM call is one of the most popular methods of yield enhancement. Historically, call overwriting has been a profitable strategy due to implied volatility usually being overpriced. However, call overwriting does underperform in volatile, strongly rising equity markets. Overwriting with the shortest maturity is best, and the strike should be slightly OTM for optimum returns.

- **Protection strategies using options.** For both economic and regulatory reasons, one of the most popular uses of options is to provide protection against a long position in the underlying. The cost of buying protection through a put is lowest in calm, low volatility markets but, in more turbulent markets, the cost can be too high. In order to reduce the cost of buying protection in volatile markets (which is often when protection is in most demand), many investors sell an OTM put and/or an OTM call to lower the cost of the long put protection bought.

- **Option structures trading.** While a simple view on both volatility and equity market direction can be implemented via a long or short position in a call or put, a far wider set of payoffs is possible if two or three different options are used. We investigate strategies using option structures (or option combos) that can be used to meet different investor needs.

VOLATILITY AND CORRELATION TRADING

We investigate the benefits and disadvantages of volatility trading via options, volatility swaps, variance swaps and gamma swaps. We also show how these products, correlation swaps, basket options and covariance swaps can give correlation exposure. Recently, options on alternative underlyings have been created, such as options on variance and dividends. We show how the distribution and skew for these underlyings is different from those for equities.

- **Volatility trading using options.** While directional investors typically use options for their equity exposure, volatility investors delta hedge their equity exposure. A delta hedged option (call or put) is not exposed to equity markets, but only to volatility markets. We demonstrate how volatility investors are exposed to dividend and borrow cost risk and how volatility traders can 'pin' a stock approaching expiry.

- **Variance is the key, not volatility.** Partly due to its use in Black-Scholes, volatility has historically been used as the measure of deviation for financial assets. However, the correct measure of deviation is variance (or volatility squared). Volatility should be considered to be a derivative of variance. The realisation that variance should be used instead of volatility-led volatility indices, such as the VIX, to move away from ATM volatility (VXO index) towards a variance-based calculation.

- **Volatility, variance and gamma swaps.** In theory, the profit and loss from delta hedging an option is fixed and based solely on the difference between the implied volatility of the option when it was purchased and the realised volatility over the life of the option. In practice, with discrete delta hedging and unknown future volatility, this is not the case, which has led to the creation of volatility, variance and gamma swaps. These products also remove the need to continuously delta hedge, which can be very labour-intensive and expensive. Until the credit crunch, variance swaps were the most liquid of the three, but now volatility swaps are more popular for single stocks.

- **Options on variance.** As the liquidity of the variance swap market improved in the middle of the last decade, market participants started to trade options on variance. As volatility is more volatile at high levels, the skew is positive (the inverse of the negative skew seen in the equity market). In addition, volatility term structure is inverted, as volatility mean reverts and does not stay elevated for long periods of time.

- **Correlation trading.** The volatility of an index is capped at the weighted average volatility of its constituents. Due to diversification (or less than 100% correlation), the volatility of indices tends to trade significantly less than its constituents. The flow from both institutions and structured products tends to put upward pressure on implied correlation, making index-implied volatility expensive. Hedge funds and proprietary trading desks try to profit from this anomaly either by selling correlation swaps or through dispersion trading (going short index implied and long single stock implied). Basket options and covariance swaps can also be used to trade correlation.

- **Dividend volatility trading.** If a constant dividend yield is assumed, then the volatility surface for options on realised dividends should be identical to the volatility surface for equities. However, as companies typically pay out less than 100% of earnings, they have the ability to reduce the volatility of dividend payments. In addition to lowering the volatility of dividends to between ½ and ⅔ of the volatility of equities, companies are reluctant to cut dividends. This means that skew is more negative than for equities, as any dividend cut is sizeable.

OPPORTUNITIES, IMBALANCES AND MYTHS

The impact of hedging both structured products and variable annuity products can cause imbalances in the volatility market. These distortions can create opportunities for investors willing to take the other side. We examine the opportunities from imbalances and dispel the myths of overpriced volatility and using volatility as an equity hedge.

- **Overpricing of vol is partly an illusion.** Selling implied volatility is one of the most popular trading strategies in equity derivatives. Empirical analysis shows that implied volatility or variance is, on average, overpriced. However, as volatility is negatively correlated to equity returns, a short volatility (or variance) position is implicitly long equity risk. As equity returns are expected to return an equity risk premium over the risk-free rate (which is used for derivative pricing), this implies short volatility should also be abnormally profitable. Therefore, part of the profits from short volatility strategies can be attributed to the fact equities are expected to deliver returns above the risk-free rate.

- **Long volatility is a poor equity hedge.** An ideal hedging instrument for a security is an instrument with -100% correlation to that security and zero cost. As the return on variance swaps can have up to a c-70% correlation with equity markets, adding long volatility positions (either through variance swaps or futures on volatility indices such as VIX or vStoxx) to an equity position could be thought of as a useful hedge. However, as volatility is on average overpriced, the cost of this strategy far outweighs any diversification benefit.

- **Variable annuity hedging lifts long term vol.** Since the 1980s, a significant amount of variable annuity products have been sold, particularly in the US. The size of this market is now over US$1trn. From the mid-1990s, these products started to become more complicated and offered guarantees to the purchaser (similar to being long a put). The hedging of these products increases the demand for long-dated downside strikes, which lifts long-dated implied volatility and skew.

- **Structured products vicious circle.** The sale of structured products leaves investment banks with a short skew position (eg, short an OTM put in order to provide capital-protected products). Whenever there is a large decline in equities, this short skew position causes the investment bank to be short volatility (eg, as the short OTM put becomes more ATM, the vega increases). The covering of this short vega position lifts implied volatility further than would be expected. As investment banks are also short vega convexity, this increase in volatility causes the short vega position to increase in size. This can lead to a ‘structured products vicious circle’ as the covering of short vega causes the size of the short position to increase. Similarly, if equity markets rise and implied volatility falls, investment banks become long implied volatility and have to sell. Structured products can therefore cause implied volatility to undershoot in a recovery, as well as overshoot in a crisis.

FORWARD STARTING PRODUCTS AND LIGHT EXOTICS

Advanced investors can make use of more exotic equity derivatives. Some of the most popular are forward starting products and light exotics, such as barriers, worst of/best of options, outperformance options, look-back options, contingent premium options, composite options and quanto options.

- **Forward starting products.** Forward starting options are a popular method of trading forward volatility and term structure as there is no exposure to near-term volatility and, hence, zero theta (until the start of the forward starting option). Recently, trading forward volatility via VIX and vStoxx forwards have become increasingly popular. However, as is the case with forward starting options, there are modelling issues. Forward starting variance swaps are easier to price as the price is determined by two variance swaps.

- **Barrier options.** Barrier options are the most popular type of light exotic product as they are used within structured products or to provide cheap protection. The payout of a barrier option knocks in or out depending on whether a barrier is hit. There are eight types of barrier option, but only four are commonly traded, as the remaining four have a similar price to vanilla options. Barrier puts are more popular than calls (due to structured product and protection flow), and investors like to sell visually expensive knock-in options and buy visually cheap knock-out options.

- **Worst-of/best-of options.** Worst-of (or best-of) options give payouts based on the worst (or best) performing asset. They are the second most popular light exotic due to structured product flow. Correlation is a key factor in pricing these options, and investor flow typically buys correlation (making uncorrelated assets with low correlation the most popular underlyings). The underlyings can be chosen from different asset classes (due to low correlation), and the number of underlyings is typically between three and 20.

- **Outperformance options.** Outperformance options are an option on the difference between returns on two different underlyings. They are a popular method of implementing relative value trades, as their cost is usually cheaper than an option on either underlying. The key unknown parameter for pricing outperformance options is implied correlation, as outperformance options are short correlation.

- **Look-back options.** There are two types of look-back options, strike look-back and payout look-back, and both are usually multi-year options. Strike reset (or look-back) options have their strike set to the highest, or lowest, value within an initial look-back period (of up to three months). These options are normally structured so the strike moves against the investor in order to cheapen the cost. Payout look-back options conversely tend to be more attractive and expensive than vanilla options, as the value for the underlying used is the best historical value.

- **Contingent premium options.** Contingent premium options are initially zero premium and only require a premium to be paid if the option becomes ATM on the close. The contingent premium to be paid is, however, larger than the initial premium would be, compensating for the fact that it might never have to be paid. Puts are the most popular, giving protection with zero initial premium.

- **Composite and quanto options.** There are two types of options involving different currencies. The simplest is a composite option, where the strike (or payoff) currency is in a different currency to the underlying. A slightly more complicated option is a quanto option, which is similar to a composite option but the exchange rate of the conversion is fixed.

ADVANCED VOLATILITY TRADING

Advanced investors often use equity derivatives to gain different exposures; for example, relative value or the jumps on earnings dates. We demonstrate how this can be done and also reveal how profits from equity derivatives are both path dependent and dependent on the frequency of delta hedging.

- **Relative value trading.** Relative value is the name given to a variety of trades that attempt to profit from the mean reversion of two related assets that have diverged. The relationship between the two securities chosen can be fundamental (different share types of same company or significant cross-holding) or statistical (two stocks in same sector). Relative value can be carried out via cash (or delta-1), options or outperformance options.

- **Trading earnings announcements/jumps.** From the implied volatilities of near-dated options it is possible to calculate the implied jump on key dates. Trading these options in order to take a view on the likelihood of unanticipated (low or high) volatility on reporting dates is a very common strategy. We examine the different methods of calculating the implied jump and show how the jump calculation should normalise for index term structure.

- **Stretching Black-Scholes assumptions.** The Black-Scholes model assumes an investor knows the future volatility of a stock, in addition to being able to continuously delta hedge. In order to discover what the likely profit (or loss) will be in reality, we stretch these assumptions. If the future volatility is unknown, the amount of profit (or loss) will vary depending on the path, but buying cheap volatility will always show some profit. However, if the position is delta-hedged discretely, the purchase of cheap volatility may reveal a loss. The variance of discretely delta-hedged profits can be halved by hedging four times as frequently. We also show why traders should hedge with a delta calculated from expected – not implied – volatility, especially when long volatility.

SKEW AND TERM STRUCTURE TRADING

We examine how skew and term structure are linked and the effect on volatility surfaces of the square root of time rule. The correct way to measure skew and smile is examined, and we show how skew trades only breakeven when there is a static local volatility surface.

- **Skew and term structure are linked.** When there is an equity market decline, there is normally a larger increase in ATM implied volatility at the near end of volatility surfaces than the far end. Assuming sticky strike, this causes near-dated skew to be larger than far-dated skew. The greater the term structure change for a given change in spot, the higher skew is. Skew is also positively correlated to term structure (this relationship can break down in panicked markets). For an index, skew (and potentially term structure) is also lifted by the implied correlation surface. Diverse indices tend to have higher skew for this reason, as the ATM correlation is lower (and low strike correlation tends to 100% for all indices).

- **Square root of time rule can compare different term structures and skews.** When implied volatility changes, typically the change in ATM volatility multiplied by the square root of time is constant. This means that different $(T_2\text{-}T_1)$ term structures can be compared when multiplied by $\sqrt{(T_2T_1)}/(\sqrt{T_2}\text{-}\sqrt{T_1})$, as this normalises against 1Y-3M term structure. Skew weighted by the square root of time should also be constant. Looking at the different term structures and skews, when normalised by the appropriate weighting, can allow us to identify calendar and skew trades in addition to highlighting which strike and expiry is the most attractive to buy (or sell).

- **How to measure skew and smile.** The implied volatilities for options of the same maturity, but of different strike, are different from each other for two reasons. Firstly, there is skew, which causes low-strike implieds to be greater than high-strike implieds due to the increased leverage and risk of bankruptcy. Secondly, there is smile (or convexity/kurtosis), when OTM options have a higher implied than ATM options. Together, skew and smile create the 'smirk' of volatility surfaces. We look at how skew and smile change by maturity in order to explain the shape of volatility surfaces both intuitively and mathematically. We also examine which measures of skew are best and why.

- **Skew trading.** The profitability of skew trades is determined by the dynamics of a volatility surface. We examine sticky delta (or 'moneyness'), sticky strike, sticky local volatility and jumpy volatility regimes. Long skew suffers a loss in both a sticky delta and sticky strike regimes due to the carry cost of skew. Long skew is only profitable with jumpy volatility. We also show how the best strikes for skew trading can be chosen.

APPENDIX

This includes technical detail and areas related to volatility trading that do not fit into earlier sections.

- **Local volatility.** While Black-Scholes is the most popular method for pricing vanilla equity derivatives, exotic equity derivatives (and ITM American options) usually require a more sophisticated model. The most popular model after Black-Scholes is a local volatility model as it is the only completely consistent volatility model. A local volatility model describes the instantaneous volatility of a stock, whereas Black-Scholes is the average of the instantaneous volatilities between spot and strike.
- **Measuring historical volatility.** We examine different methods of historical volatility calculation, including close-to-close volatility and exponentially weighted volatility, in addition to advanced volatility measures such as Parkinson, Garman-Klass (including Yang-Zhang extension), Rogers and Satchell and Yang-Zhang. We also show that it is best to assume a zero drift assumption for close-to-close volatility and that under this condition variance is additive.
- **Proof variance swaps can be hedged by a log contract (= $1/K^2$).** A log contract is a portfolio of options of all strikes (K) weighted by $1/K^2$. When this portfolio of options is delta hedged on the close, the payoff is identical to the payoff of a variance swap. We prove this relationship and hence show that the volatility of a variance swap can be hedged with a static position in a log contract.
- **Modelling volatility surfaces.** There are a variety of constraints on the edges of a volatility surface, and this section details some of the most important constraints from both a practical and theoretical point of view. We examine the considerations for very short-dated options (a few days or weeks), options at the wings of a volatility surface and very long-dated options.
- **Black-Scholes formula.** The most popular method of valuing options is the Black-Scholes-Merton model. We show the key formulas involved in this calculation. The assumptions behind the model are also discussed.
- **Greeks and their meaning.** Greeks is the name given to the (usually) Greek letters used to measure risk. We give the Black-Scholes formula for the key Greeks and describe which risk they measure.
- **Advanced (practical or shadow) Greeks.** How a volatility surface changes over time can impact the profitability of a position. Two of the most important are the impact of the passage of time on skew (volatility slide theta) and the impact of a movement in spot on OTM options (anchor delta).
- **Shorting stock by borrowing shares.** The hedging of equity derivatives assumes you can short shares by borrowing them. We show the processes involved in this operation. The disadvantages – and benefits – for an investor who lends out shares are also explained.
- **Sortino ratio.** If an underlying is distributed normally, standard deviation is the perfect measure of risk. For returns with a skewed distribution, such as with option trading or call overwriting, there is no one perfect measure of risk; hence, several measures of risk should be used. The Sortino is one of the most popular measures of skewed risk, as it only takes into account downside risk.
- **Capital structure arbitrage.** The high levels of volatility and credit spreads during the bursting of the TMT bubble demonstrated the link between credit spreads, equity, and implied volatility. We examine four models that demonstrate this link (Merton model, jump diffusion, put vs CDS, and implied no-default volatility).

DIRECTIONAL VOLATILITY TRADING

OPTION TRADING IN PRACTICE

Using options to invest has many advantages over investing in cash equity. Options provide leverage and an ability to take a view on volatility as well as equity direction. However, investing in options is more complicated than investing in equity, as a strike and expiry need to be chosen. This can be seen as an advantage, as it enforces investor discipline in terms of anticipated return and ensures a position is not held longer than it should be. We examine how investors can choose the appropriate strategy, strike and expiry. We also explain hidden risks, such as dividends and the difference between delta and the probability an option ends up in-the-money.

HISTORY OF VOLATILITY TRADING

Options were first traded in London from 1690

While standardised exchange traded options only started trading in 1973 when the CBOE (Chicago Board Options Exchange) opened, options were first traded in London from 1690. Pricing was made easier by the Black-Scholes-Merton formula (usually shortened to Black-Scholes), which was invented in 1970 by Fischer Black, Myron Scholes and Robert Merton. The derivatives explosion in the 1990s was partly due to the increasing popularity of hedge funds, which led to volatility becoming an asset class in its own right. New volatility products such as volatility swaps and variance swaps were created, and a decade later futures on volatility indices gave investors listed instruments to trade volatility. In this section we shall concentrate on option trading.

LONG OR SHORT STRATEGIES ARE POSSIBLE WITH OPTION TRADING

A European call is a contract that gives the investor the right (but not the obligation) to buy a security at a certain strike price on a certain expiry date (American options can be exercised before expiry). A put is identical except it is the right to sell the security. A call option profits when markets rise (as exercising the call means the investor can buy the underlying security cheaper than it is trading, and then sell it at a profit). A put option profits when markets fall (as you can buy the underlying security for less, exercise the put and sell the security for a profit). Options therefore allow investors to put on long (profit when prices rise) or short (profit when prices fall) strategies.

Option trading allows investors to take a long or short position on volatility

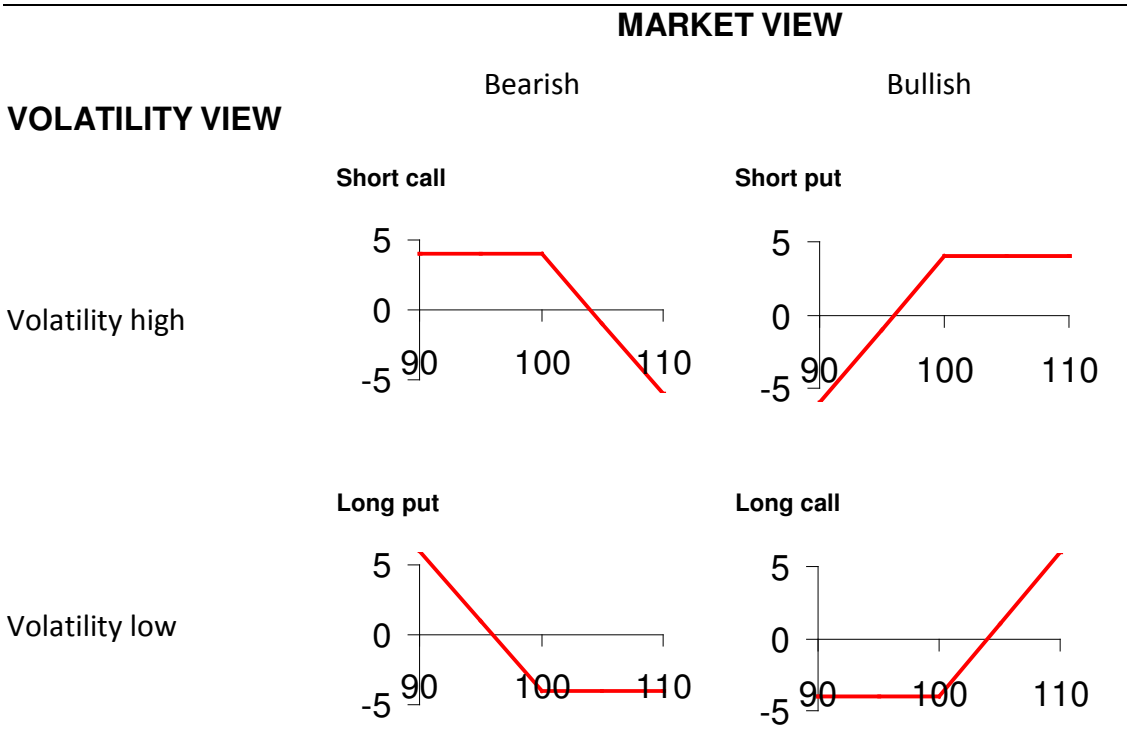
Options increase in value as volatility rises

If the volatility of an underlying is zero, then the price will not move and an option's payout is equal to the intrinsic value. Intrinsic value is the greater of zero and the 'spot – strike price' for a call and is the greater of zero and 'strike price – spot' for a put. Assuming that stock prices can move, the value of a call and put will be greater than intrinsic due to the time value (price of option = intrinsic value + time value). If an option strike is equal to spot (or is the nearest listed strike to spot) it is called at-the-money (ATM). If volatility is zero, an ATM option has a price of zero (as intrinsic is zero). However, if we assume a stock is €50 and has a 50% chance of falling to €40 and 50% chance of rising to €60, it has a volatility above zero. In this example, an ATM call option with strike €50 has a 50% chance of making €10 (if the price rises to €60 the call can be exercised to buy the stock at €50, which can be sold for €10 profit). The fair value of the ATM option is therefore €5 (50% × €10); hence, as volatility rises the value of a call rises (a similar argument can be used for puts). An ATM option has the greatest time value. This can be seen in the same example by looking at an out-of-the-money (OTM) call option of strike €60 (an OTM option has strike far away from spot and zero intrinsic value). This OTM €60 call option would be worth zero, as the stock in this example cannot rise above €60.

Both an equity and volatility view is needed to trade options

Option trading allows a view on equity and volatility markets to be taken. The appropriate strategy for a one leg option trade is shown in Figure 1 below. Multiple leg (combos) are dealt with in the section *Option Structures Trading*.

Figure1. Option Strategy for Different Market and Volatility Views



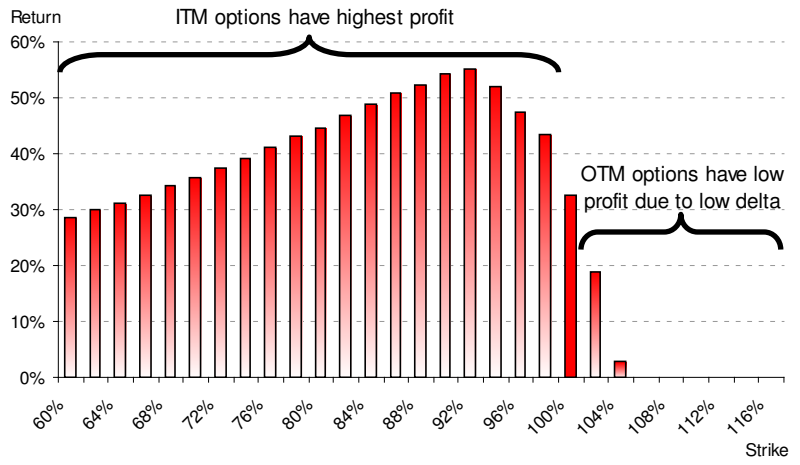
Source: Santander Investment Bolsa.

CHOOSING THE STRIKE OF AN OPTION STRATEGY IS NOT TRIVIAL

If an investor is certain of market direction (counter intuitively), the best strike is ITM

While it is relatively simple to pick the option strategy, choosing the strike and expiry is the most difficult part of an options strategy. Choosing the maturity of the option is easier if there is a specific event (eg, an earnings date) that is anticipated to be a driver for the stock. Choosing the strike of the trade is not trivial either. Investors could choose ATM to benefit from greatest liquidity. Alternatively, they could look at the highest expected return (option payout less the premium paid, as a percentage of the premium paid). While choosing a cheap OTM option might be thought of as giving the highest return, Figure 2 below shows that, in fact, the highest returns come from in-the-money (ITM) options (ITM options have a strike far away from spot and have intrinsic value). This is because an ITM option has a high delta (sensitivity to equity price); hence, if an investor is relatively confident of a specific return, an ITM option has the highest return (as trading an ITM option is similar to trading a forward).

Figure 2. Profit of 12 Month Options if Markets Rise 10% by Expiry



Source: Santander Investment Bolsa.

Forwards (or futures) are better than options for pure directional plays

A forward is a contract that obliges the investor to buy a security on a certain expiry date at a certain strike price. A forward has a delta of 100%. An ITM call option has many similarities with being long a forward, as it has a relatively small time value (compared to ATM) and a delta close to 100%. While the intrinsic value does make the option more expensive, this intrinsic value is returned at expiry. However, for an ATM option, the time value purchased is deducted from the returns. ATM or OTM options are only the best strike (if an investor is very confident of the eventual return) if the anticipated return is very large (as leverage boosts the returns). For pure directional plays, forwards (or futures, their listed equivalent) are more profitable than options. The advantage of options is in offering convexity: if markets move against the investor the only loss is the premium paid, whereas a forward has a virtually unlimited loss.

OPTION LIQUIDITY CAN BE A FACTOR IN IMPLEMENTING TRADES

If an underlying is relatively illiquid, or if the size of the trade is large, an investor should take into account the liquidity of the maturity and strike of the option. Typically, OTM options are more liquid than ITM options as ITM options tie up a lot of capital. This means that for strikes less than spot, puts are more liquid than calls and vice versa. We note that as low-strike puts have a higher implied than high-strike calls, their value is greater and, hence, traders are more willing to use them. Low strike put options are therefore usually more liquid than high-strike call options. In addition, demand for protection lifts liquidity for low strikes compared with high strikes.

Single stock liquidity is limited for maturities of two years or more

For single stock options, liquidity starts to fade after one year and options rarely trade over two years. For indices, longer maturities are liquid, partly due to the demand for long-dated hedges and their use in structured products. While structured products can have a maturity of five to ten years, investors typically lose interest after a few years and sell the product back. The hedging of a structured product, therefore, tends to be focused on more liquid maturities of around three years. Hedge funds tend to focus around the one-year maturity, with two to three years being the longest maturity they will consider. The two-to-three year maturity is where there is greatest overlap between hedge funds and structured desks.

DELTA IS THE DIVIDEND RISK, AS WELL AS THE EQUITY RISK

Delta measures dividend sensitivity

The delta of the option is the amount of equity market exposure an option has. As a stock price falls by the dividend amount on its ex-date, delta is equal to the exposure to dividends that go ex before expiry. The dividend risk is equal to the negative of the delta. For example, if you have a call of positive delta, if (expected or actual) dividends rise, the call is worth less (as the stock falls by the dividend amount).

If a dividend is substantial, it could be in an investor's interest to exercise early. For more details, see the section *Maintenance of Option Positions*.

DIFFERENCE BETWEEN DELTA AND PROBABILITY EXPIRES ITM

A digital call option is an option that pays 100% if spot expires above the strike price (a digital put pays 100% if spot is below the strike price). The probability of such an option expiring ITM is equal to its delta, as the payoff only depends on it being ITM or not (the size of the payment does not change with how much ITM spot is). For a vanilla option this is not the case; hence, there is a difference between the delta and the probability of being ITM. This difference is typically small unless the maturity of the option is very long.

Delta takes into account the amount an option can be ITM

While a call can have an infinite payoff, a put's maximum value is the strike (as spot cannot go below zero). The delta hedge for the option has to take this into account, so a call delta must be greater than the probability of being ITM. Similarly, the absolute value (as put deltas are negative) of the put delta must be less than the probability of expiring ITM. A more mathematical explanation (for European options) is given below:

Call delta > Probability call ends up ITM

Abs (Put delta) < Probability put ends up ITM

Mathematical proof option delta is different from probability of being ITM at expiry

Call delta	$= N(d_1)$	Put delta	$= N(d_1) - 1$
Call probability ITM	$= N(d_2)$	Put probability ITM	$= 1 - N(d_2)$

where:

Definition of d_1 is the standard Black-Scholes formula for d_1.

$d_2 = d_1 - \sigma\sqrt{T}$

σ = implied volatility

T = time to expiry

N(z) = cumulative normal distribution

Difference between delta and ITM is greatest for long-dated options with high volatility

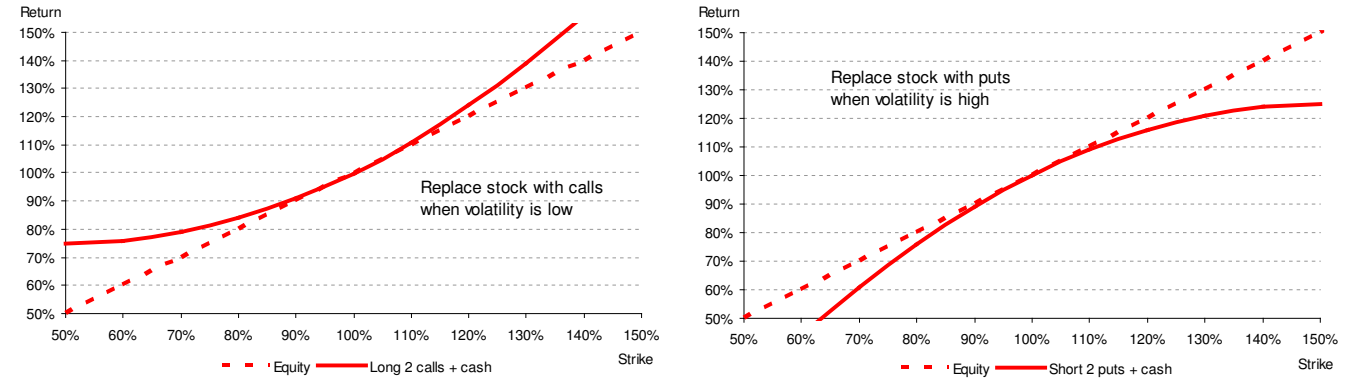
As d_2 is less than d_1 (see above) and N(z) is a monotonically increasing function, this means that $N(d_2)$ is less than $N(d_1)$. Hence, the probability of a call being in the money = $N(d_2)$ is less than the delta = $N(d_1)$. As the delta of a put = delta of call – 1, and the sum of call and put being ITM = 1, the above results for a put must be true as well.

The difference between delta and probability being ITM at expiry is greatest for long-dated options with high volatility (as the difference between d_1 and d_2 is greatest for them).

STOCK REPLACING WITH LONG CALL OR SHORT PUT

As a stock has a delta of 100%, the identical exposure to the equity market can be obtained by purchasing calls (or selling puts) whose total delta is 100%. For example, one stock could be replaced by two 50% delta calls, or by going short two -50% delta puts. Such a strategy can benefit from buying (or selling) expensive implied volatility. There can also be benefits from a tax perspective and, potentially, from any embedded borrow cost in the price of options (price of positive delta option strategies is improved by borrow cost). As the proceeds from selling the stock are typically greater than the cost of the calls (or margin requirement of the short put), the difference can be invested to earn interest. It is important to note that the dividend exposure is not the same, as only the owner of a stock receives dividends. While the option owner does not benefit directly, the expected dividend will be used to price the option fairly (hence investors only suffer/benefit if dividends are different from expectations).

Figure 3. Stock Replacing with Calls **Stock Replacing with Puts**



Source: Santander Investment Bolsa estimates.

Stock replacing via calls benefits from convexity

Stock replacing with calls suffers if underlying range trades

As a call option is convex, this means that the delta increases as spot increases and vice versa. If a long position in the underlying is sold and replaced with calls of equal delta, then if markets rise the delta increases and the calls make more money than the long position would have. Similarly, if markets fall the delta decreases and the losses are reduced. This can be seen in Figure 3 above as the portfolio of cash (proceeds from sale of the underlying) and call options is always above the long underlying profile. The downside of using calls is that the position will give a worse profile than the original long position if the underlying does not move much (as call options will fall each day by the theta if spot remains unchanged). Using call options is best when implied volatility is cheap and the investor expects the stock to move by more than currently implied.

Put underwriting benefits from selling expensive implied volatility

Gaining equity exposure (or stock replacing) via puts is known as put underwriting

Typically the implied volatility of options trades slightly above the expected realised volatility of the underlying over the life of the option (due to a mismatch between supply and demand). Stock replacement via put selling therefore benefits from selling (on average) expensive volatility. Selling a naked put is known as put underwriting, as the investor has effectively underwritten the stock (in the same way investment banks underwrite a rights issue). The strike should be chosen at the highest level at which the investor would wish to purchase the stock, which allows an investor to earn a premium from taking this view (whereas normally the work done to establish an attractive entry point would be wasted if the stock did not fall to that level). This strategy has been used significantly recently by asset allocators who are underweight equities and are waiting for a better entry point to re-enter the equity market (earning the premium provides a buffer should equities rally). If an investor does not wish to own the stock and only wants to earn the premium, then an OTM strike should be chosen at a support level that is likely to remain firm.

If OTM puts are used, put underwriting benefits from selling skew

Put underwriting gives a similar profile to a long stock, short call profile, otherwise known as call overwriting. One difference between call overwriting and put underwriting is that if OTM options are used, then put underwriting benefits from selling skew (which is normally overpriced). For more details on the benefits of selling volatility, see the section *Call Overwriting*.

MAINTENANCE OF OPTION POSITIONS

During the life of an American option, many events can occur in which it might be preferable to own the underlying shares (rather than the option) and exercise early. In addition to dividends, an investor might want the voting rights, or alternatively might want to sell the option to purchase another option (rolling the option). We investigate these life cycle events and explain when it is in an investor's interest to exercise, or roll, an option before expiry.

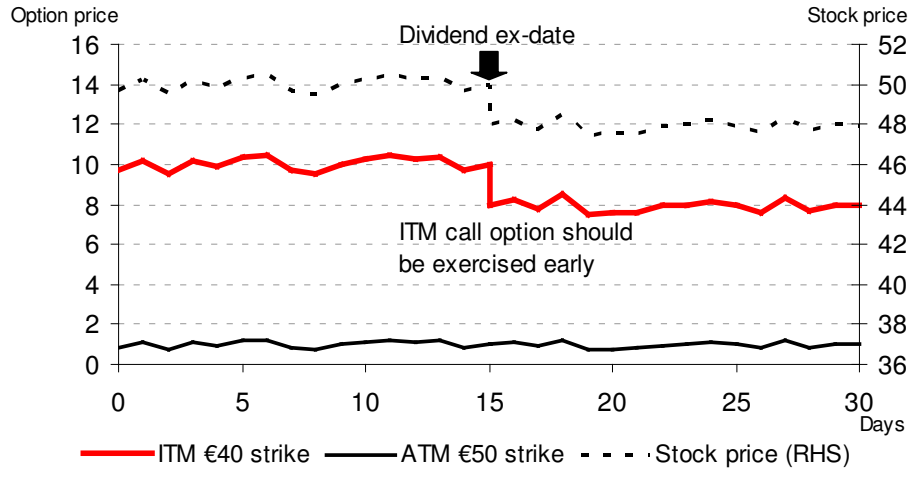
CONVERTING OPTIONS EARLY IS RARE, BUT SOMETIMES NECESSARY

The decision to exercise early depends on the net benefit vs the time value of the options

Options on indices are usually European, which means they can only be exercised at maturity. The inclusion of automatic exercise, and the fact it is impossible to exercise before maturity, means European options require only minimal maintenance. Single stock options, however, are typically American (apart from emerging market underlyings). While American options are rarely exercised early, there are circumstances when it is in an investor's interest to exercise an ITM option early. For both calls and puts the correct decision for early exercise depends on the net benefit of doing so (ie, the difference between earning the interest on the strike and net present value of dividends) versus the time value of the option.

- **Calls should be exercised just before the ex-date of a large unadjusted dividend.** In order to exercise a call, the strike price needs to be paid. The interest on this strike price normally makes it unattractive to exercise early. However, if there is a large unadjusted dividend that goes ex before expiry, it might be in an investor's interest to exercise an ITM option early (see Figure 4 below). In this case, the time value should be less than the dividend NPV (net present value) less total interest r ($=e^{rfr \times T}-1$) earned on the strike price K. In order to maximise 'dividend NPV– Kr', it is best to exercise just before an ex-date (as this maximises 'dividend NPV' and minimises the total interest r).
- **Puts should be exercised early (preferably just after ex-date) if interest rates are high.** If interest rates are high, then the interest r from putting the stock back at a high strike price K (less dividend NPV) might be greater than the time value. In this case, a put should be exercised early. In order to maximise 'Kr – dividend NPV', a put should preferably be exercised just after an ex-date.

Figure 4. Price of ITM and ATM Call Option with Stock Price over Ex-Date of Dividend



Source: Santander Investment Bolsa.

Calls should only be exercised early if there is an unadjusted dividend

The payout profile of a long call is similar to the payout of a long stock + long put of the same strike[1]. As only ITM options should be exercised and as the strike of an ITM call means the put of the same strike is OTM, we shall use this relationship to calculate when an option should be exercised early.

An American call should only be exercised if it is in an investor's interest to exercise the option and buy a European put of the same strike (a European put of same strike will have the same time value as a European call if intrinsic value is assumed to be the forward).

- **Choice A:** Do not exercise. In this case there is no benefit or cost.
- **Choice B:** Borrow strike K at interest r ($=e^{rfr \times T}-1$) in order to exercise the American call. The called stock will earn the dividend NPV and the position has to be hedged with the purchase of a European put (of cost equal to the time value of a European call).

An investor will only exercise early if choice B > choice A.

⇨ -Kr + dividend NPV – time value > 0

⇨ dividend NPV - Kr > time value for American call to be exercised

Puts should only be exercised if interest earned (less dividends) exceeds time value

For puts, it is simplest to assume an investor is long stock and long an American put. This payout is similar to a long call of the same strike. An American put should only be exercised against the long stock in the same portfolio if it is in an investor's interest to exercise the option and buy a European call of the same strike.

- **Choice A:** Do not exercise. In this case the portfolio of long stock and long put benefits from the dividend NPV.
- **Choice B:** Exercise put against long stock, receiving strike K, which can earn interest r ($=e^{rfr \times T}-1$). The position has to be hedged with the purchase of a European call (of cost equal to the time value of a European put).

An investor will only exercise early if choice B > choice A

⇨ Kr – time value > dividend NPV

⇨ Kr – dividend NPV > time value for American put to be exercised

Selling ITM options that should be exercised early can be profitable

Investors need to take into account trading costs and taxation

There have been occasions when traders deliberately sell ITM options that should be exercised early, hoping that some investors will forget. Even if the original counterparty is aware of this fact, exchanges randomly assign the counterparty to exercised options. As it is unlikely that 100% of investors will realise in time, such a strategy can be profitable.

[1] But not identical due to the difference between spot and forward.

ITM OPTIONS TEND TO BE EXERCISED AT EXPIRY TO PREVENT LOSSES

Hurdle for automatic exercise is usually above ATM to account for trading costs

In order to prevent situations where an investor might suffer a loss if they do not give notice to exercise an ITM option in time, most exchanges have some form of automatic exercise. If an investor (for whatever reason) does not want the option to be automatically exercised, he must give instructions to that effect. The hurdle for automatic exercise is usually above ATM in order to account for any trading fees that might be incurred in selling the underlying post exercise.

Eurex automatic exercise has a higher hurdle than Euronext-Liffe or CBOE

For the CBOE, options are automatically exercised if they are US$0.01 or more ITM (reduced in June 2008 from US$0.05 or more), which is in line with Euronext-Liffe rules of a €0.01 or GBP0.01 minimum ITM hurdle. Eurex has a higher automatic hurdle, as a contract price has to be ITM by 99.99 or more (eg, for a euro-denominated stock with a contract size of 100 shares this means it needs to be at €0.9999 or more). Eurex does allow an investor to specify an automatic exercise level lower than the automatic hurdle, or a percentage of exercise price up to 9.99%.

CORPORATE ACTIONS CAN CAUSE STRIKE TO BE ADJUSTED

Principal of unchanged contract values ensures investor does not suffer from special dividends or corporate actions

While options do not adjust for ordinary dividends[2], they do adjust for special dividends. Different exchanges have different definitions of what is a special dividend, but typically it is considered special if it is declared as a special dividend, or is larger than a certain threshold (eg, 10% of the stock price). In addition, options are adjusted in the event of a corporate action, for example, a stock split or rights issue. Options on equities and indices can treat bonus share issues differently. A stock dividend in lieu of an ordinary dividend is considered an ordinary dividend for options on an equity (hence is not adjusted) but is normally adjusted by the index provider. For both special dividends and corporate actions, the adjustment negates the impact of the event (principal of unchanged contract values), so the theoretical price of the options should be able to ignore the event. As the strike post adjustment will be non-standard, typically exchanges create a new set of options with the normal strikes. While older options can still trade, liquidity generally passes to the new standard strike options (particularly for longer maturities which do not have much open interest).

M&A AND SPINOFFS CAN CAUSE PROBLEMS

If a company spins off a subsidiary and gives shareholders shares in the new company, the underlying for the option turns into a basket of the original equity and the spun-off company. New options going into the original company are usually created, and the liquidity of the options into the basket is likely to fade. For a company that is taken over, the existing options in that company will convert into whatever shareholders were offered. If the acquisition was for stock, then the options convert into shares, but if the offer is partly in cash, then options can lose a lot of value as the volatility of cash is zero.

OPTIONS OFTEN ROLLED BEFORE EXPIRY TO REDUCE TIME DECAY

The time value of an option decays quicker for short-dated options than for far-dated options. To reduce the effect of time decay, investors often roll before expiry. For example, an investor could buy a one-year option and roll it after six months to a new one-year option.

[2] Some option markets adjust for all dividends.

CALL OVERWRITING

For a directional investor who owns a stock (or index), call overwriting by selling an OTM call is one of the most popular methods of yield enhancement. Historically, call overwriting has been a profitable strategy due to implied volatility usually being overpriced. However, call overwriting does underperform in volatile, strongly rising equity markets. Overwriting with the shortest maturity is best, and the strike should be slightly OTM for optimum returns.

OPTION IMPLIED VOLATILITY IS USUALLY OVERPRICED

The implied volatility of options is on average 1-2pts above the volatility realised over the life of the option. This 'implied volatility premium' is usually greater for indices than for single stocks. As we can see no reason why these imbalances will fade, we expect call overwriting to continue to outperform on average. The key imbalances are:

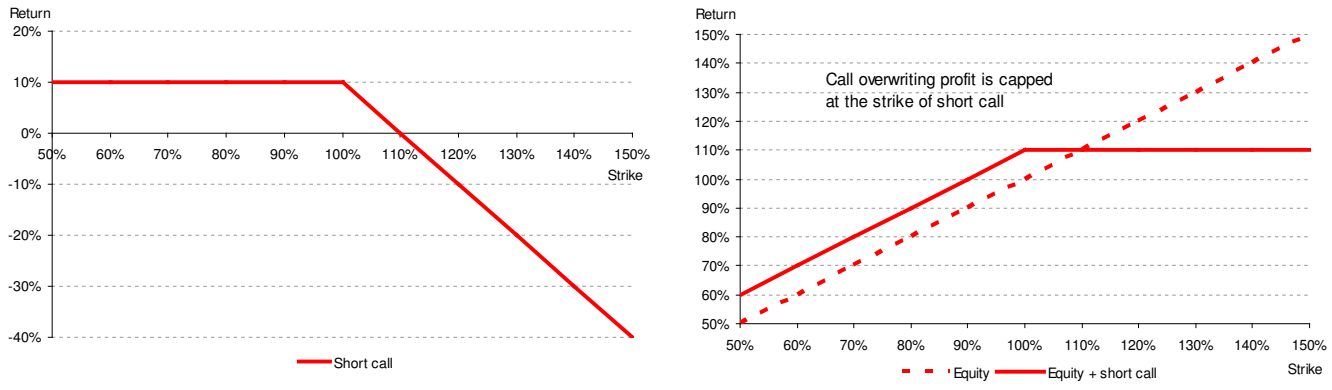
- **Option buying for protection**.
- **Unwillingness to sell low premium options causes market makers to raise their prices** (selling low premium options, like selling lottery tickets, has to be done on a large scale to be attractive).
- **High gamma of near-dated options has a gap risk premium** (risk of stock jumping, either intraday or between closing and opening prices).
- **Index implieds lifted by structured products**.

CALL OVERWRITING BENEFITS FROM SELLING EXPENSIVE VOLATILITY

Best to overwrite with slightly OTM option to reduce probability of expiring in ITM

Short-dated implied volatility has historically been overpriced[3] due to the above supply and demand imbalances. In order to profit from this characteristic, a long investor can sell a call against a long position in the underlying of the option. Should the underlying perform well and the call be exercised, the underlying can be used to satisfy the exercise of the call. As equities should be assumed to have, on average, a positive return, it is best to overwrite with a slightly OTM option to reduce the probability of the option sold expiring ITM.

Figure 5. Short Call Call Overwriting (or Buy Write)



Source: Santander Investment Bolsa estimates.

[3] We note that implied volatility is not necessarily as overpriced as would first appear. For more detail, see the section *Overpricing of Vol Is Partly an Illusion.*

Call overwriting is a useful way to gain yield in range trading markets

If markets are range trading, or are approaching a technical resistance level, then selling a call at the top of the range (or resistance level) is a useful way of gaining yield. Such a strategy can be a useful tactical way of earning income on a core strategic portfolio, or potentially could be used as part of an exit strategy for a given target price.

Selling at target price enforces disciplined investing

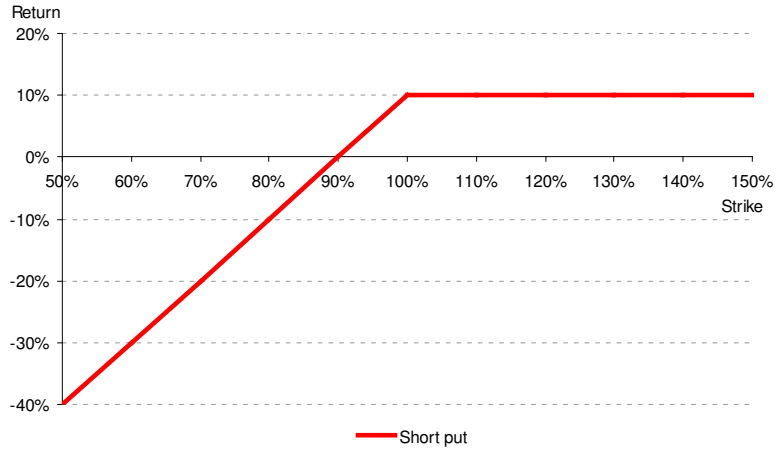
As there are typically more Buy ideas than Sell ideas, call overwriting is useful to enforce an exit strategy

If a stock reaches the desired target price, there is the temptation to continue to own the strong performer. Over time a portfolio can run the risk of being a collection of stocks that had previously been undervalued, but are now at fair value. To prevent this inertia diluting the performance of a fund, some fund managers prefer to call overwrite at their target price to enforce disciplined investing, (as the stock will be called away when it reaches the target). As there are typically more Buy recommendations than Sell recommendations, call overwriting can ensure a better balance between the purchase and (called away) sale of stocks.

CALL OVERWRITING PROFILE IS SIMILAR TO PUT UNDERWRITING

Figure 5 shows the profiles of a short call and of a long equity with an overwritten call. The resulting profile of call overwriting is similar to that of a short put (Figure 6); hence, call overwriting could be considered similar to stock replacement with a short put (or put underwriting). Both call overwriting and put underwriting attempt to profit from the fact that implied volatility, on average, tends to be overpriced. While selling a naked put is seen as risky, due to the near infinite losses should stock prices fall, selling a call against a long equity position is seen as less risky (as the equity can be delivered against the exercise of the call).

Figure 6. Put Underwriting

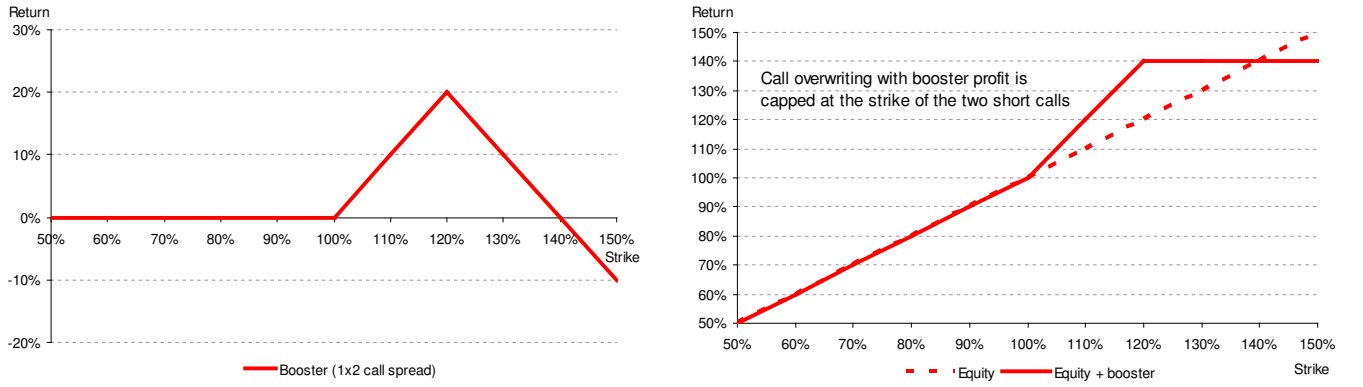


Source: Santander Investment Bolsa.

Boosters (1×2 call spreads) are useful when a bounce-back is expected

If a near zero cost 1×2 call spread (long 1×ATM call, short 2×OTM calls) is overlaid on a long stock position, the resulting position offers the investor twice the return for equity increases up to the short upper strike. For very high returns the payout is capped, in a similar way as for call overwriting. Such positioning is useful when there has been a sharp drop in the markets and a limited bounce back to earlier levels is anticipated. The level of the bounce back should be in line with or below the short upper strike. Typically, short maturities are best (less than three months) as the profile of a 1×2 call spread is similar to a short call for longer maturities.

Figure 7. Booster (1×2 Call Spread) **Call Overwriting with Booster**



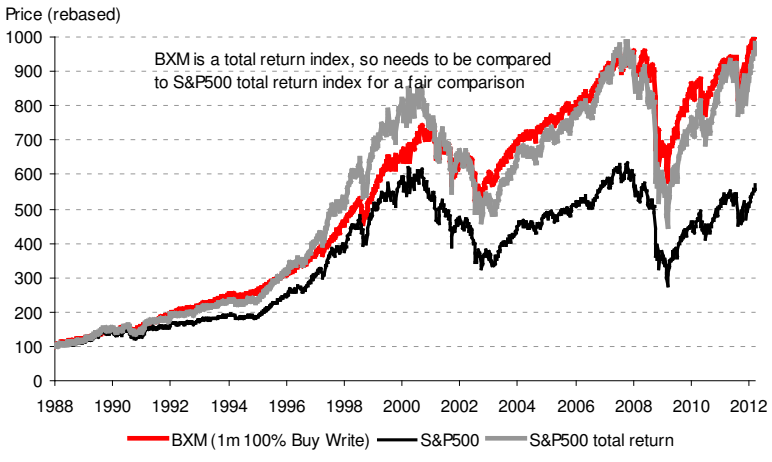
Source: Santander Investment Bolsa estimates.

CALL OVERWRITING IS BEST DONE ON AN INDEX

Need to compare BXM index to S&P500 total return index (as BXM is total return)

Many investors call overwrite on single stocks. However, single-stock implied volatility trades more in line with realised volatility than index implieds. The reason why index implieds are more overpriced than single-stock implieds is due to the demand from hedgers and structured product sellers. Call overwriting at the index level also reduces trading costs (due to the narrower bid-offer spread). The CBOE has created a one-month call overwriting index on the S&P500 (BXM index), which is the longest call overwriting time series available. It is important to note that the BXM is a total return index; hence, it needs to be compared to the S&P500 total return index (SPXT Bloomberg code) not the S&P500 price return (SPX Bloomberg code). As can be seen in Figure 8 below, comparing the BXM index to the S&P500 price return index artificially flatters the performance of call overwriting.

Figure 8. S&P500 and S&P500 1M ATM Call Overwriting Index (BXM)

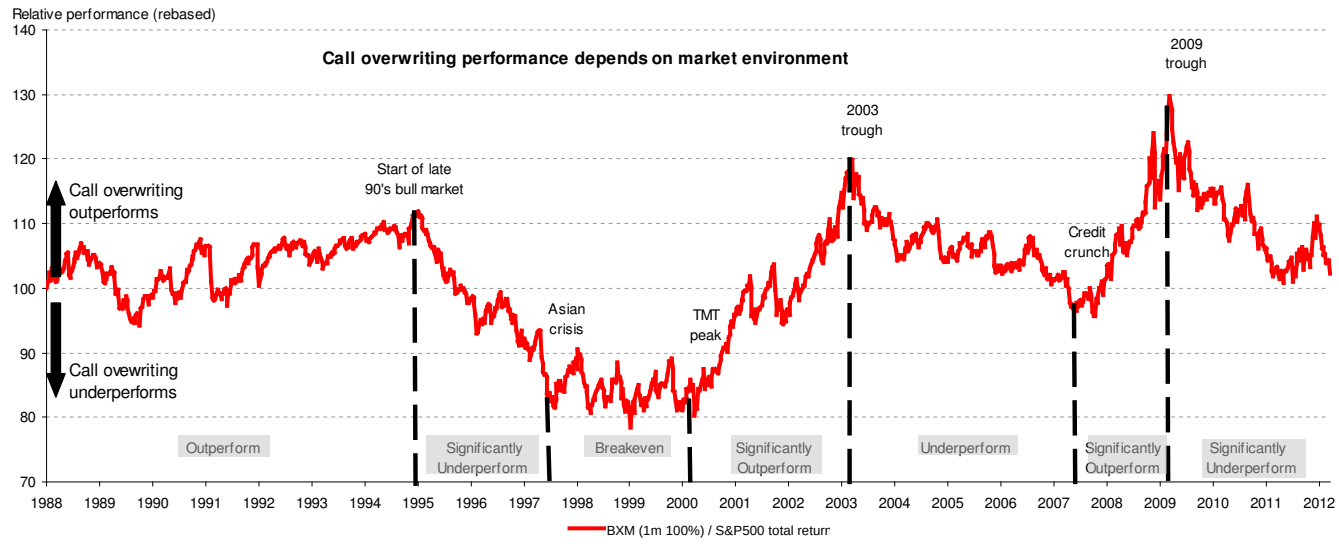


Source: Santander Investment Bolsa.

Call overwriting performance varies according to equity and volatility market conditions

On average, call overwriting has been a profitable strategy. However, there have been periods of time when it is has been unprofitable. The best way to examine the returns under different market conditions is to divide the BXM index by the total return S&P500 index (as the BXM is a total return index).

Figure 9. S&P500 1M ATM Call Overwriting (BXM) Divided by S&P500 Total Return Index



Source: Santander Investment Bolsa estimates.

Call overwriting underperforms in strong bull markets with low volatility

It is difficult to identify in advance periods when call overwriting will be successful

Since the BXM index was created, there have been seven distinct periods (see Figure 9 above), each with different equity and volatility market conditions. Of the seven periods, the two in which returns for call overwriting are negative are the bull markets of the mid-1990s and middle of the last decade. These were markets with very low volatility, causing the short call option sold to earn insufficient premium to compensate for the option being ITM. It is important to note that call overwriting can outperform in slowly rising markets, as the premium earned is in excess of the amount the option ends up ITM. This was the case for the BXM between 1986 and the mid-1990s. It is difficult to identify these periods in advance as there is a very low correlation between BXM outperformance and the earlier historical volatility.

LOWER DELTA REDUCES BENEFIT OF EQUITY RISK PREMIUM

We note that while profits should be earned from selling an expensive call, the delta (or equity sensitivity) of the long underlying short call portfolio is significantly less than 100% (even if the premium from the short call is reinvested into the strategy). Assuming that equities are expected to earn more than the risk free rate (ie, have a positive equity risk premium), this lower delta can mean more money is lost by having a less equity-sensitive portfolio than is gained by selling expensive volatility. On average, call overwriting appears to be a successful strategy, and its success has meant that it is one of the most popular uses of trading options.

OVERWRITING WITH NEAR-DATED OPTIONS HAS HIGHEST RETURN

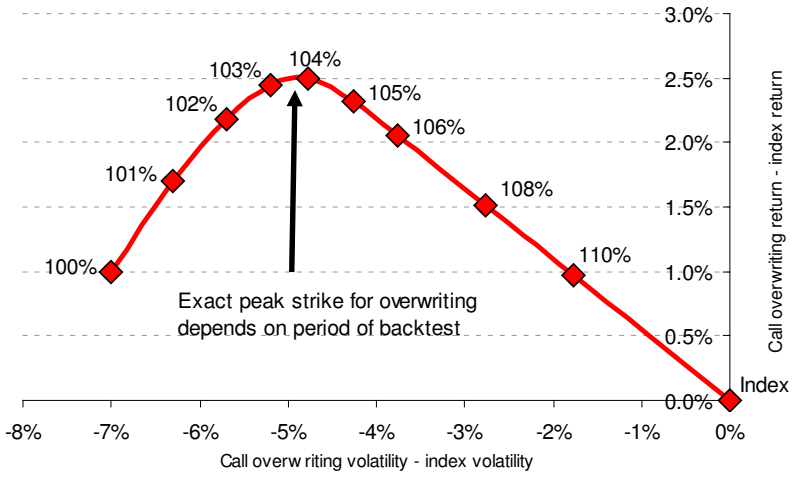
Overwriting with near-dated options has highest return, and highest risk

Near-dated options have the highest theta, so an investor earns the greatest carry from call overwriting with short-dated options. It is possible to overwrite with 12 one-month options in a year, as opposed to four three-month options or one 12-month option. While overwriting with the shortest maturity possible has the highest returns on average, the strategy does have potentially higher risk. If a market rises one month, then retreats back to its original value by the end of the quarter, a one-month call overwriting strategy will have suffered a loss on the first call sold but a three-month overwriting strategy will not have had a call expire ITM. However, overwriting with far-dated expiries is more likely to eliminate the equity risk premium the investor is trying to earn (as any outperformance above a certain level will be called away).

BEST RETURNS FROM OVERWRITING WITH SLIGHTLY OTM OPTIONS

While overwriting with near-dated expiries is clearly superior to overwriting with far-dated expiries, the optimal choice of strike to overwrite with depends on the market environment. As equities are expected, on average, to post a positive return, overwriting should be done with slightly OTM options. However, if a period of time where equities had a negative return is chosen for a back-test, then a strike below 100% could show the highest return. Looking at a period of time where the SX5E had a positive return shows that for one-month options a strike between 103%-104% is best (see Figure 10 below). For three-month options, the optimal strike is a higher 107%-108%, but the outperformance is approximately half as good as for one-month options. These optimal strikes for overwriting could be seen to be arguably high, as recently there have been instances of severe declines (TMT bubble bursting, Lehman bankruptcy), which were followed by significant price rises afterwards. For single-stock call overwriting, these strikes could seem to be low, as single stocks are more volatile. For this reason, many investors use the current level of volatility to determine the strike or choose a fixed delta option (eg, 25%).

Figure 10. Call Overwriting SX5E with One-Month Calls of Different Strikes



Source: Santander Investment Bolsa.

OVERWRITING REDUCES VOLATILITY AND INCREASES RETURNS

While selling an option could be considered risky, the volatility of returns from overwriting a long equity position is reduced by overwriting. This is because the payout profile is capped for equity prices above the strike. An alternative way of looking at this is that the delta of the portfolio is reduced from 100% (solely invested in equity) to 100% less the delta of the call (c50% depending on strike). The reduced delta suppresses the volatility of the portfolio.

Benefit of risk reduction is less impressive if Sortino ratios are used to measure risk

Sortino ratio is a better measure of risk for skewed distributions (such as returns from call overwriting)

We note that the low call overwriting volatility is due to the lack of volatility to the upside, as call overwriting has the same downside risk as a long position. For this reason, using the Sortino ratio (for more details, see the section *Sortino Ratio* in the *Appendix*) is likely to be a fairer measure of call overwriting risk than standard deviation, as standard deviation is not a good measure of risk for skewed distributions. Sortino ratios show that the call overwriting downside risk is identical to a long position; hence, call overwriting should primarily be done to enhance returns and is not a viable strategy for risk reduction.

We expect optimal strike for overwriting to be similar for single stocks and indices

While this analysis is focused on the SX5E, the analysis can be used to guide single-stock call overwriting (although the strike could be adjusted higher by the single-stock implied divided by SX5E implied).

ENHANCED CALL OVERWRITING IS POSSIBLE BUT DIFFICULT

Enhanced call overwriting is the term given when call overwriting is only done opportunistically or the parameters (strike or expiry) are varied according to market conditions. On the index level, the returns from call overwriting are so high that enhanced call overwriting is difficult, as the opportunity cost from not always overwriting is too high. For single stocks, the returns for call overwriting are less impressive; hence, enhanced call overwriting could be more successful. An example of single-stock enhanced call overwriting is to only overwrite when an implied is high compared to peers in the same sector. We note that even with enhanced single-stock call overwriting, the wider bid-offer cost and smaller implied volatility premium to realised means returns can be lower than call overwriting at the index level.

Enhanced call overwriting returns is likely to be arbitraged away

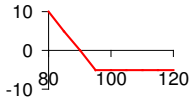
Should a systematic way to enhance call overwriting be viable, this method could be applied to volatility trading without needing an existing long position in the underlying. Given the presence of statistical arbitrage funds and high frequency traders, we believe it is unlikely that a simple automated enhanced call overwriting strategy on equity or volatility markets is likely to outperform vanilla call overwriting on an ongoing basis.

PROTECTION STRATEGIES USING OPTIONS

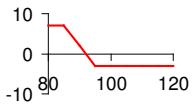
For both economic and regulatory reasons, one of the most popular uses of options is to provide protection against a long position in the underlying. The cost of buying protection through a put is lowest in calm, low-volatility markets, but in more turbulent markets the cost can be too high. In order to reduce the cost of buying protection in volatile markets (which is often when protection is in most demand), many investors sell an OTM put and/or an OTM call to lower the cost of the long put protection bought.

CHEAPEN PUT PROTECTION BY SELLING OTM PUTS AND CALLS

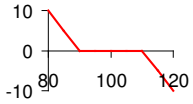
Buying a put against a long position gives complete and total protection for underlying moves below the strike (as the investor can simply put the long position back for the strike price following severe declines). The disadvantage of a put is the relatively high cost, as an investor is typically unwilling to pay more than 1%-2% for protection (as the cost of protection usually has to be made up through alpha to avoid underperforming if markets do not decline). The cost of the long put protection can be cheapened by selling an OTM put (turning the long put into a long put spread), by selling an OTM call (turning put protection into a collar), or both (resulting in a put spread vs call, or put spread collar). The strikes of the OTM puts and calls sold can be chosen to be in line with technical supports or resistance levels.



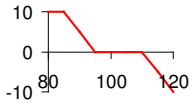
- **Puts give complete protection without capping performance.** As puts give such good protection, their cost is usually prohibitive unless the strike is low. For this reason, put protection is normally bought for strikes around 90%. Given that this protection will not kick in until there is a decline of 10% or more, puts offer the most cost-effective protection only during a severe crash (or if very short-term protection is required).



- **Put spreads only give partial protection but are cost effective.** While puts give complete protection, often only partial protection is necessary, in which case selling an OTM put against the long put (a put spread) can be an attractive protection strategy. The value of the put sold can be used to either cheapen the protection or lift the strike of the long put.



- **Collars can be zero cost as they give up some upside.** While investors appreciate the need for protection, the cost needs to be funded through reduced performance (or less alpha) or by giving up some upside. Selling an OTM call to fund a put (a collar) results in a cap on performance. However, if the strike of the call is set at a reasonable level, the capped return could still be attractive. The strike of the OTM call is often chosen to give the collar a zero cost. Collars can be a visually attractive low (or zero) cost method of protection as returns can be floored at the largest tolerable loss and capped at the target return. A collar is unique among protection strategies in not having significant volatility exposure, as its profile is similar to a short position in the underlying. Collars are, however, exposed to skew.

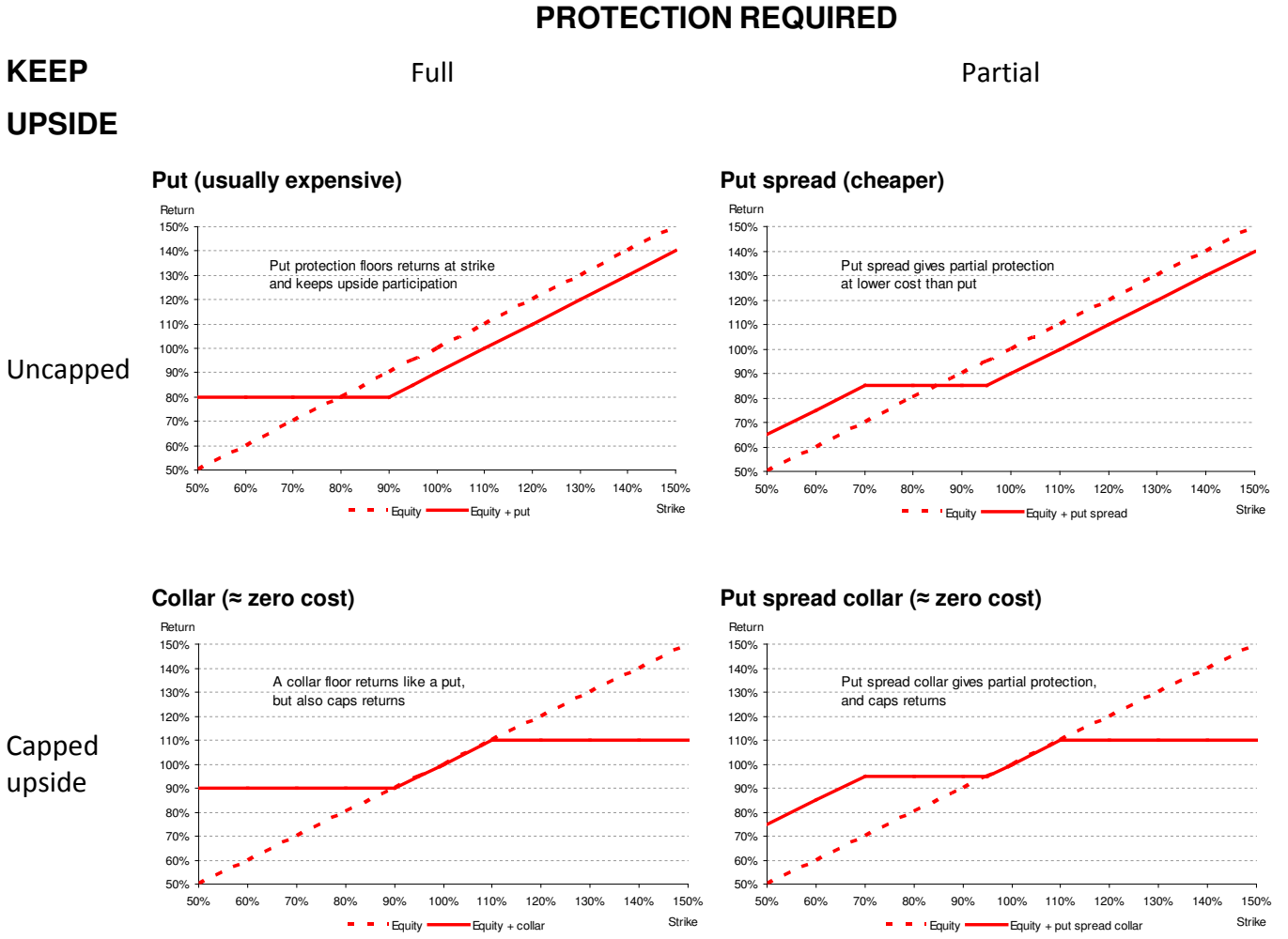


- **Put spread collars best when volatility is high, as two OTM options are sold.** Selling both an OTM put and OTM call against a long put (a put spread collar) is typically attractive when volatility is high, as this lifts the value of the two OTM options sold more than the long put bought. If equity markets are range bound, a put spread collar can also be an attractive form of protection.

Portfolio protection is usually done via indices for lower costs and macro exposure

While an equity investor will typically purchase individual stocks, if protection is bought then this is usually done at the index level. This is because the risk the investor wishes to hedge against is the general equity or macroeconomic risk. If a stock is seen as having excessive downside risk, it is usually sold rather than a put bought against it. An additional reason why index protection is more common than single stock protection is the fact that bid-offer spreads for single stocks are wider than for an index.

Figure 11. Protection Strategies



Source: Santander Investment Bolsa estimates.

Partial protection can give a more attractive risk reward profile than full protection

For six-month maturity options, the cost of a 90% put is typically in line with a 95%-85% put spread (except during periods of high volatility, when the cost of a put is usually more expensive). Put spreads often have an attractive risk-reward profile for protection of the same cost, as the strike of the long put can be higher than the long put of a put spread. Additionally, if an investor is concerned with outperforming peers, then a c10% outperformance given by a 95%-85% put spread should be sufficient to attract investors (there is little incremental competitive advantage in a greater outperformance).

Implied volatility is far more important than skew for put-spread pricing

A rule of thumb is that the value of the OTM put sold should be approximately one-third the value of the long put (if it were significantly less, the cost saving in moving from a put to a put spread would not compensate for giving up complete protection). While selling an OTM put against a near-ATM put does benefit from selling skew (as the implied volatility of the OTM put sold is higher than the volatility of the near ATM long put bought), the effect of skew on put spread pricing is not normally that significant (far more significant is the level of implied volatility).

Collars are more sensitive to skew than implied volatility

Selling a call against a long put suffers from buying skew. The effect of skew is greater for a collar than for a put spread, as skew affects both legs of the structure the same way (whereas the effect of skew on the long and short put of a put spread partly cancels). If skew was flat, the cost of a collar typically reduces by 1% of spot. The level of volatility for near-zero cost collars is not normally significant, as the long volatility of the put cancels the short volatility of the call.

Choice of protection strategy is determined by magnitude of expected decline

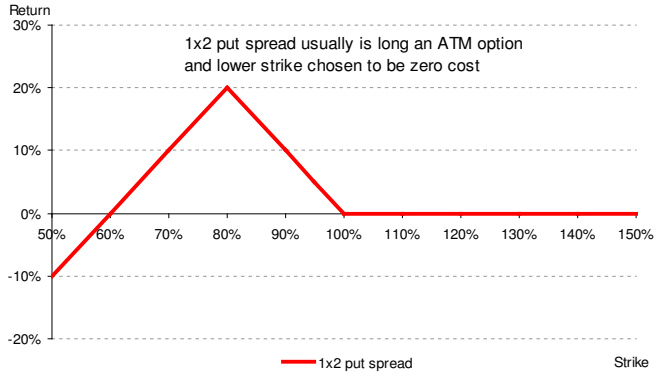
Capping performance should only be used when a long-lasting rally is unlikely

A collar or put spread collar caps the performance of the portfolio at the strike of the OTM call sold. They should only therefore be used when the likelihood of a strong, long-lasting rally (or significant bounce) is perceived to be relatively small.

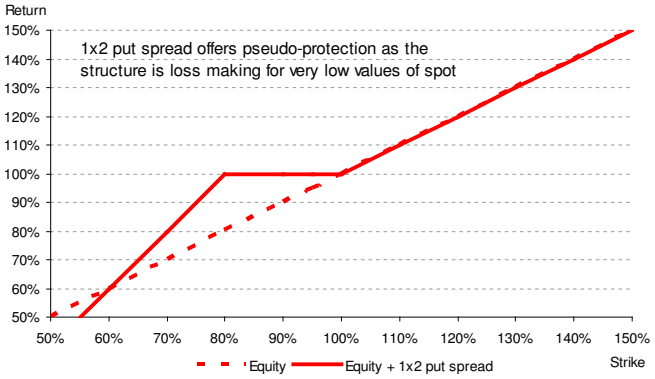
Bullish investors could sell two puts against long put (=pseudo-protection 1×2 put spread)

If an investor is bullish on the equity market, then a protection strategy that caps performance is unsuitable. Additionally, as the likelihood of substantial declines is seen to be small, the cost of protection via a put or put spread is too high. In this scenario, a zero cost 1×2 put spread could be used as a pseudo-protection strategy. The long put is normally ATM, which means the portfolio is 100% protected against falls up to the lower strike. We do not consider it to offer true protection, as during severe declines a 1×2 put spread will suffer a loss when the underlying portfolio is also heavily loss making. The payout of 1×2 put spreads for maturities of around three months or more is initially similar to a short put, so we consider it to be a bullish strategy. However, for the SX5E a roughly six-month zero-cost 1×2 put spread, whose upper strike is 95%, has historically had a breakeven below 80% and declines of more than 20% in six months are very rare. As 1×2 put spreads do not provide protection when you need it most, they could be seen as a separate long position rather than a protection strategy.

Figure 12. 1×2 Put Spread



Pseudo-Protection with 1×2 Put Spread



Source: Santander Investment Bolsa estimates.

PROTECTION MUST BE PAID FOR: THE QUESTION IS HOW?

If an investor seeks protection, the most important decision that has to be made is how to pay for it. The cost of protection can be paid for in one of three ways. Figure 13 below shows when this cost is suffered by the investor, and when the structure starts to provide protection against declines.

Premium. The simplest method of paying for protection is through premium. In this case, a put or put spread should be bought.

Loss of upside. If the likelihood of extremely high returns is small, or if a premium cannot be paid, then giving up upside is the best method of paying for protection. Collars and put spread collars are therefore the most appropriate method of protection if a premium cannot be paid.

Potential losses on extreme downside. If an investor is willing to tolerate additional losses during extreme declines, then a 1×2 put spread can offer a zero cost way of buying protection against limited declines in the market.

Figure 13. Protection Strategy Comparison

Equity Performance	Put	Put Spread	Collar	Put Spread Collar	1×2 Put Spread
Bull markets (+10% or more)	Loss of premium	Loss of premium	Loss of upside	Loss of upside	–
Flat markets (±5%)	Loss of premium	Loss of premium	–	–	–
Moderate dip (c-10%)	Loss of premium	Protected	–	Protected	Protected
Correction (c-15%)	Protected	Protected	Protected	Protected	Protected
Bear market (c-20% or worse)	Protected	Partially protected	Protected	Partially protected	Severe loss

Source: Santander Investment Bolsa estimates.

STRATEGY ATTRACTIVENESS DETERMINED BY LEVEL OF VOLATILITY

The level of volatility can determine the most suitable protection strategy an investor needs to decide how bullish and bearish they are on the equity and volatility markets. If volatility is low, then puts should be affordable enough to buy without offsetting the cost by selling an OTM option. For low to moderate levels of volatility, a put spread is likely to give the best protection that can be easily afforded. As a collar is similar to a short position with limited volatility exposure, it is most appropriate for a bearish investor during average periods of volatility (or if an investor does not have a strong view on volatility). Put spreads collars (or 1×2 put spreads) are most appropriate during high levels of volatility (as two options are sold for every option bought).

MATURITY DRIVEN BY SEVERITY AND DURATION OF LIKELY DECLINE

Declines can be categorised as a crash, correction or a bear market

The choice of protection strategy is typically driven by an investor's view on equity and volatility markets. Similarly the choice of strikes is usually restricted by the premium an investor can afford. Maturity is potentially the area where there is most choice, and the final decision will be driven by an investor's belief in the severity and duration of any decline. If he wants protection against a sudden crash, a short-dated put is the most appropriate strategy. However, for a long drawn out bear market, a longer maturity is most appropriate.

Figure 14. Types of DAX Declines (of 10% or more) since 1960

Type	Average Decline	Decline Range	Average Duration	Duration Range
Crash	31%	19% to 39%	1 month	0 to 3 months
Correction	14%	10% to 22%	3 months	0 to 1 year
Bear market	44%	23% to 73%	2.5 years	1 to 5 years

Source: Santander Investment Bolsa.

Median maturity of protection bought is c4 months but can be more than one year

Average maturity of protection is c6 months, boosted by a few long dated hedges

The average choice of protection is c6 months, but this is skewed by a few long-dated hedges. The median maturity is c4 months. Protection can be bought for maturities of one week to over a year. Even if an investor has decided how long he needs protection, he can implement it via one far-dated option or multiple near-dated options. For example, one-year protection could be via a one-year put or via the purchase of a three-month put every three months (four puts over the course of a year). The typical cost of ATM puts for different maturities is given below.

Figure 15. Cost of ATM Put on SX5E

Cost	1 Month	2 Months	3 Months	6 Months	1 Year
Individual premium	2.3%	3.3%	4.0%	5.7%	8%
Cost for year rolling protection	27.7%	19.6%	16%	11.3%	8%

Source: Santander Investment Bolsa.

Short-dated puts offer greatest protection but highest cost

If equity markets fall 20% in the first three months of the year and recover to the earlier level by the end of the year, then a rolling three-month put strategy will have a positive payout in the first quarter but a one-year put will be worth nothing at expiry. While rolling near-dated puts will give greater protection than a long-dated put, the cost is higher (see Figure 15 above).

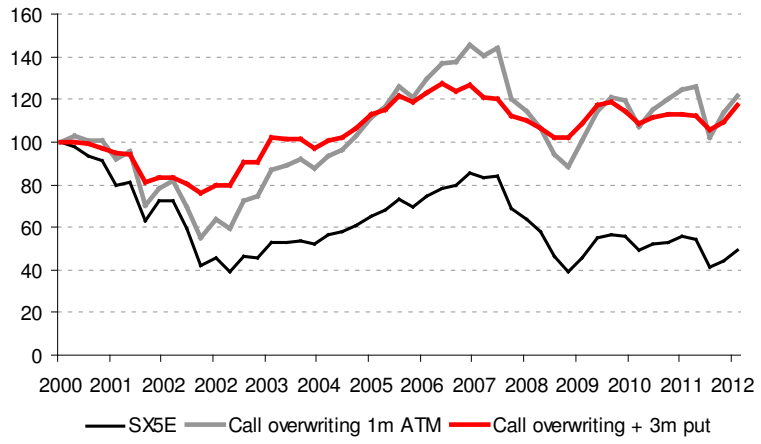
MULTIPLE EXPIRY PROTECTION STRATEGIES

Typically, a protection strategy involving multiple options has the same maturity for all of the options. However, some investors choose a nearer maturity for the options they are short, as more premium can be earned selling a near-dated option multiple times (as near-dated options have higher theta). For more details, see the section *Greeks and Their Meaning* in the *Appendix*. These strategies are most successful when term structure is inverted, as the volatility for the near-dated option sold is higher. Having a nearer maturity for the long put option and longer maturity for the short options makes less sense, as this increases the cost (assuming the nearer-dated put is rolled at expiry).

Calendar collar effectively overlays call overwriting on a long put position

If the maturity of the short call of a collar is closer than the maturity of the long put, then this is effectively the combination of a long put and call overwriting. For example, the cost of a three-month put can be recovered by selling one-month calls. This strategy outperforms in a downturn and also has a lower volatility (see Figure 16).

Figure 16. Performance of 3M Put vs 1M Call Overwriting



Source: Santander Investment Bolsa.

Calendar put spread collar effectively sells short-dated volatility against long put

Selling near-dated volatility can help pay for far dated protection

For a calendar put spread collar, if the maturity of the short put is identical to the long put, then the results are similar to a calendar collar above. If the maturity of the short put is the same as the maturity of the short near-dated put, then, effectively, this position funds the long put by selling short-dated volatility. This type of calendar put spread collar is similar to a long far-dated put and short near-dated straddle (as the payoff of a short strangle and straddle are similar, we shall assume the strikes of the short call and short put are identical). For an investor who is able to trade OTC, a similar strategy involves long put and short near-dated variance swaps.

OPTION STRUCTURES TRADING

While a simple view on both volatility and equity market direction can be implemented via a long or short position in a call or put, a far wider set of payoffs is possible if two or three different options are used. We investigate strategies using option structures (or option combos) that can be used to meet different investor needs.

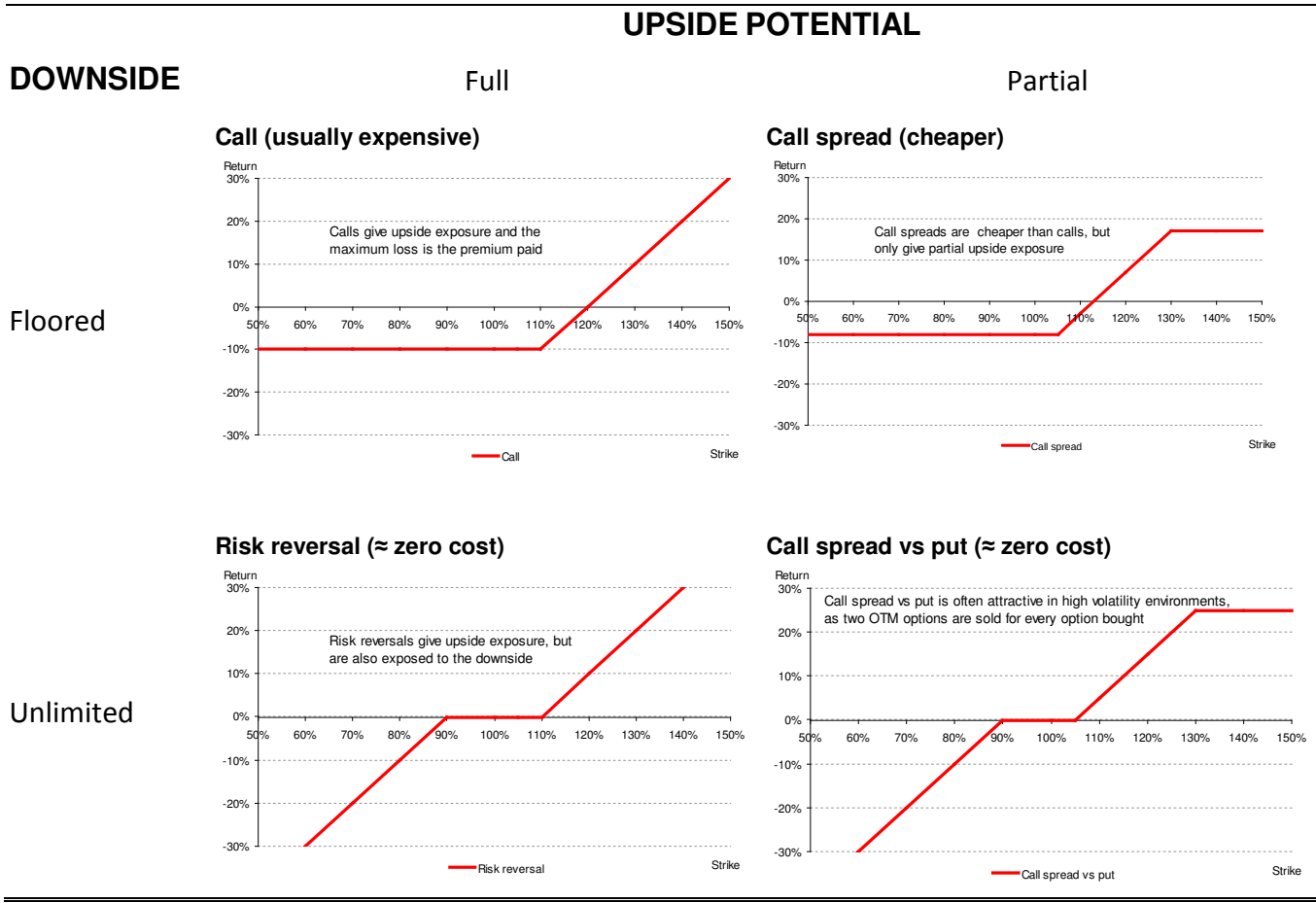
BULLISH COMBOS ARE MIRROR IMAGE OF PROTECTION STRATEGIES

Long call can be cheapened by selling an OTM call and/or and OTM put

Using option structures to implement a bearish strategy has already been discussed in the section *Protection Strategies Using Options*. In the same way a long put protection can be cheapened by selling an OTM put against the put protection (to create a put spread giving only partial protection), a call can be cheapened by selling an OTM call (to create a call spread offering only partial upside). Similarly, the upside exposure of the call (or call spread) can be funded by put underwriting (just as put or put spread protection can be funded by call overwriting). The four option structures for bullish strategies are given below.

- **Calls give complete upside exposure and floored downside.** Calls are the ideal instrument for bullish investors as they offer full upside exposure and the maximum loss is only the premium paid. Unless the call is short dated or is purchased in a period of low volatility, the cost is likely to be high.
- **Call spreads give partial upside but are cheaper.** If an underlying is seen as unlikely to rise significantly, or if a call is too expensive, then selling an OTM call against the long call (to create a call spread) could be the best bullish strategy. The strike of the call sold could be chosen to be in line with a target price or technical resistance level. While the upside is limited to the difference between the two strikes, the cost of the strategy is normally one-third cheaper than the cost of the call.
- **Risk reversals (short put, long call of different strikes) benefit from selling skew.** If a long call position is funded by selling a put (to create a risk reversal), the volatility of the put sold is normally higher than the volatility of the call bought. The higher skew is, the larger this difference and the more attractive this strategy is. Similarly, if interest rates are low, then the lower the forward (which lifts the value of the put and decreases the value of the call) and the more attractive the strategy is. The profile of this risk reversal is similar to being long the underlying.
- **Call spread vs put is most attractive when volatility is high.** A long call can be funded by selling an OTM call and OTM put. This strategy is best when implied volatility is high, as two options are sold.

Figure 17. Upside Participation Strategies

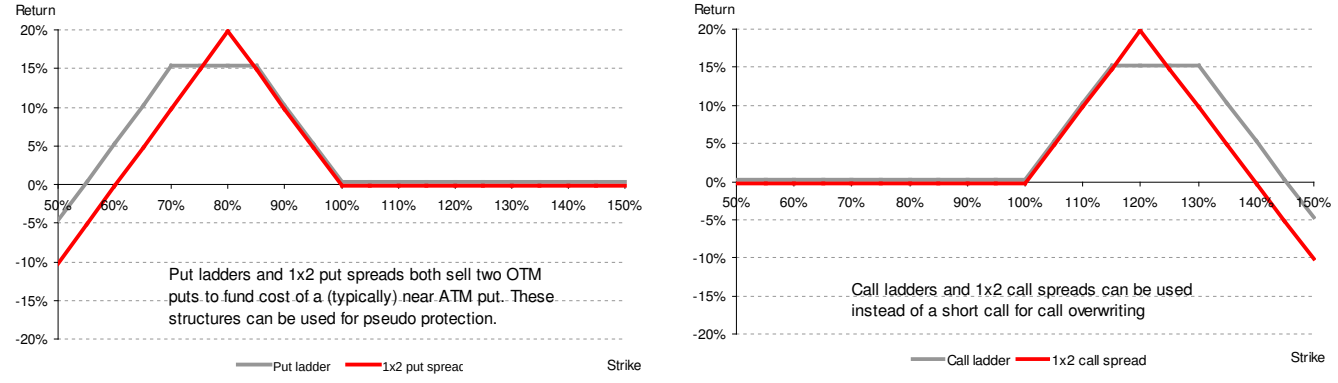


Source: Santander Investment Bolsa estimates.

LADDERS HAVE A SIMILAR PROFILE TO 1×2 SPREADS

With a 1×2 call or put spread, two OTM options of the same strike are sold against one (usually near ATM) long option of a different strike. A ladder has exactly the same structure, except the two short OTM options have a different strike.

Figure 18. Put Ladders and 1×2 Put Spreads **Call Ladders and 1×2 Call Spreads**



Source: Santander Investment Bolsa estimates.

STRADDLES, STRANGLES AND BUTTERFLIES ARE SIMILAR

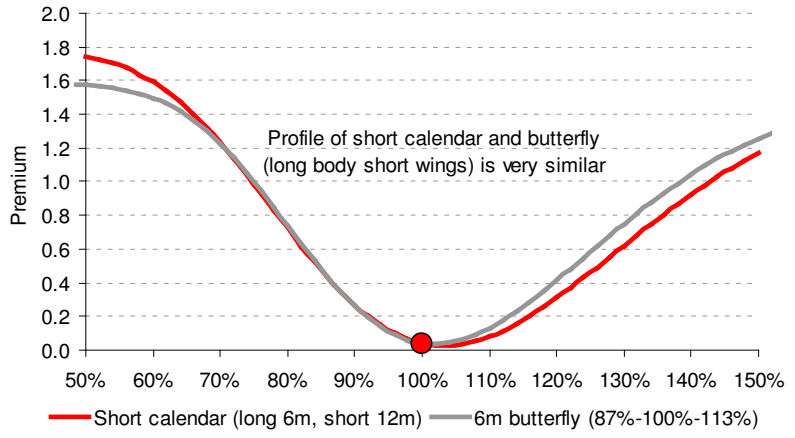
Using option structures allows a straddle (long call and put of same strike) or strangle (long call and put of different strikes) to be traded. These structures are long volatility, but do not have any exposure to the direction of the equity market. For more details, see next section *Volatility Trading Using Options*. Butterflies combine a short straddle with a long strangle, which floors the losses.

1X1 CALENDAR TRADES ARE SIMILAR TO TRADING A BUTTERFLY

Calendars are useful as a butterfly substitute when OTM options are illiquid

We note the theoretical profile of a short calendar trade is similar to trading a butterfly (see Figure 19 below). If an underlying does not have liquid OTM options, then a calendar can be used as a butterfly substitute (although this approach does involve term structure risk, which a butterfly does not have). A long calendar (short near-dated, long far-dated) is therefore short gamma (as the short near-dated option has more gamma than the far-dated option).

Figure 19. Theoretical Value of Butterfly and Short Calendar

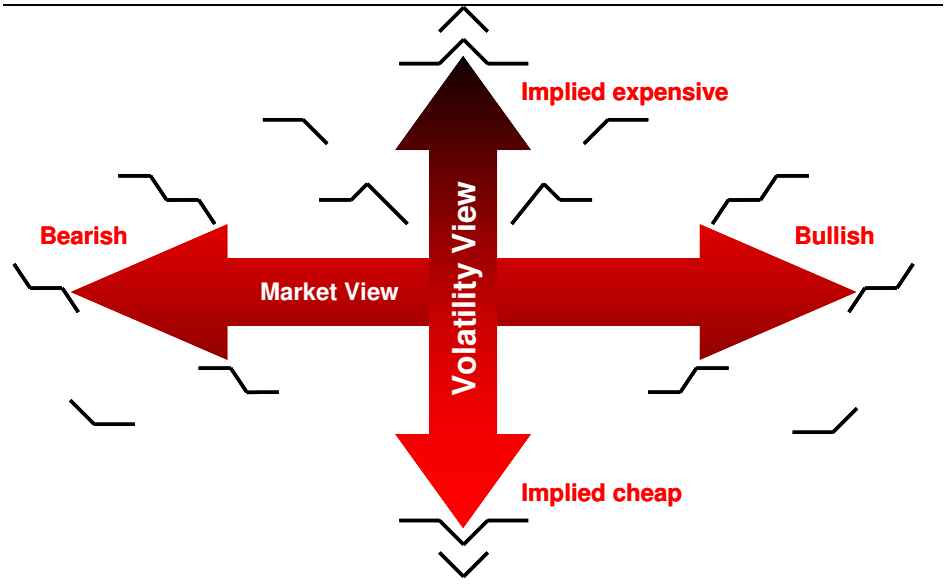


Source: Santander Investment Bolsa.

OPTION STRUCTURES ALLOW A RANGE OF VIEWS TO BE TRADED

Figure 20 shows the most common structures that can be traded with up to three different options in relation to a view on equity and volatility markets. For simplicity, strangles and ladders are not shown, but they can be considered to be similar to straddles and 1×2 ratio spreads, respectively.

Figure 20. Option Structures



Source: Santander Investment Bolsa.

This page left blank intentionally.

VOLATILITY AND CORRELATION TRADING

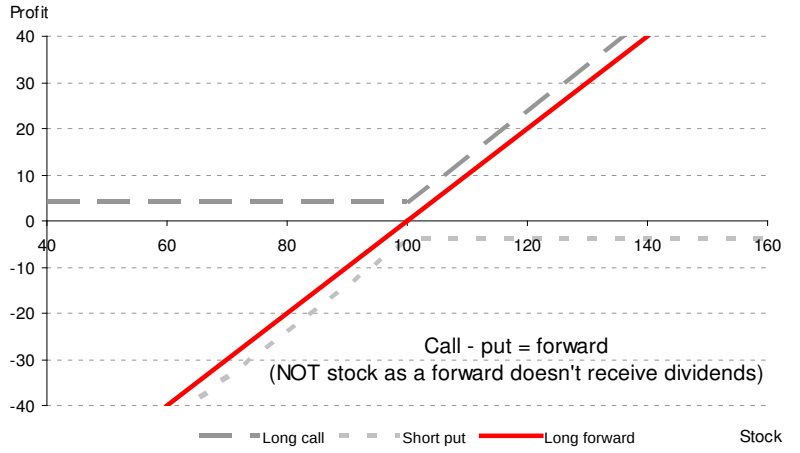
VOLATILITY TRADING USING OPTIONS

While directional investors typically use options for their equity exposure, volatility investors delta hedge their equity exposure. A delta-hedged option (call or put) is not exposed to equity markets, but only to volatility markets. We demonstrate how volatility investors are exposed to dividend and borrow cost risk and how volatility traders can 'pin' a stock approaching expiry.

VOL TRADING VIA CALLS AND PUTS IS IDENTICAL (PUT-CALL PARITY)

A forward is a contract that obliges the investor to buy (or sell if you have sold the forward) a security on a certain expiry date (but not before) at a certain strike price. A portfolio of a long European call and a short European put of identical expiry and strike is the same as a forward of that expiry and strike, as shown in Figure 21. This means that if a call, a put or a straddle is delta hedged with a forward contract (not stock), the end profile is identical. We note put-call parity is only true for European options, as American options can be exercised before expiry (although in practice they seldom are).

Figure 21. Put-Call Parity: Call - Put = Long Forward (not long stock)



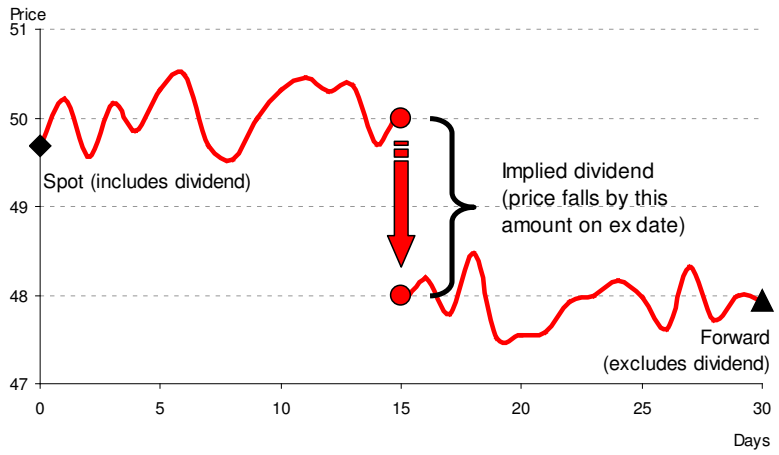
Source: Santander Investment Bolsa.

Delta hedging must be done with forward of identical maturity for put call parity

Dividend risk is equal to delta

It is important to note that the delta hedging must be done with a forward of identical maturity to the options. If it is done with a different maturity, or with stock, there will be dividend risk. This is because a forward, like a European call or put, gives the right to a security at maturity but does not give the right to any benefits such as dividends that have an ex date before expiry. A long forward position is therefore equal to long stock and short dividends that go ex before maturity (assuming interest rates and borrow cost are zero or are hedged). This can be seen from the diagram below, as a stock will fall by the value of the dividend (subject to a suitable tax rate) on the ex date. The dividend risk of an option is therefore equal to the delta.

Figure 22. Why Forwards Are Short Dividends



Source: Santander Investment Bolsa.

BORROW COST IMPACT ON OPTION PRICING

From a derivative pricing point of view, borrow cost (or repo) can be added to the dividend. This is because it is something that the owner of the shares receives and the owner of a forward does not. While the borrow cost should, in theory, apply to both the bid and offer of calls and puts, in practice an investment bank's stock borrow desk is usually separate from the volatility trading desk (or potentially not all of the long position can be lent out). If the traders on the volatility trading desk do not get an internal transfer of the borrow cost, then only one side of the trade (the side that has positive delta for the volatility trading desk, or negative delta for the client) usually includes the borrow cost. This is shown in Figure 23 below. While the borrow cost is not normally more than 40bp for General Collateral (GC) names, it can be more substantial for emerging market (EM) names. If borrow cost is only included in one leg of pricing, it creates a bid-offer arbitrage channel.

Figure 23. When Borrow Cost Is Usually Included in Implied Volatility Calculations

Option	Bid	Ask
Calls	Include borrow	–
Puts	–	Include borrow

Source: Santander Investment Bolsa.

Zero delta straddles still need to include borrow cost on one leg of the straddle

Like dividends, the exposure to borrow cost is equal to the delta. However, a zero delta straddle still has exposure to borrow cost because it should be priced as the sum of two separate trades, one call and one put. As one of the legs of the trade should include borrow, so does a straddle. This is particularly important for EM or other high borrow cost names.

Zero delta straddles have strike above spot

Zero delta straddles have strike greater than spot (or forward)

A common misperception is that ATM options have a 50% delta; hence, an ATM straddle has to be zero delta. In fact, a zero delta straddle has to have a strike above spot (an ATM straddle has negative delta). The strike of a zero delta straddle is given below.

$$\text{Strike (\%) of zero delta straddle} = e^{\left(r+\sigma^2/2\right)T}$$

where r = interest rate, σ = volatility and T = time.

DELTA HEDGING AN OPTION REMOVES EQUITY RISK

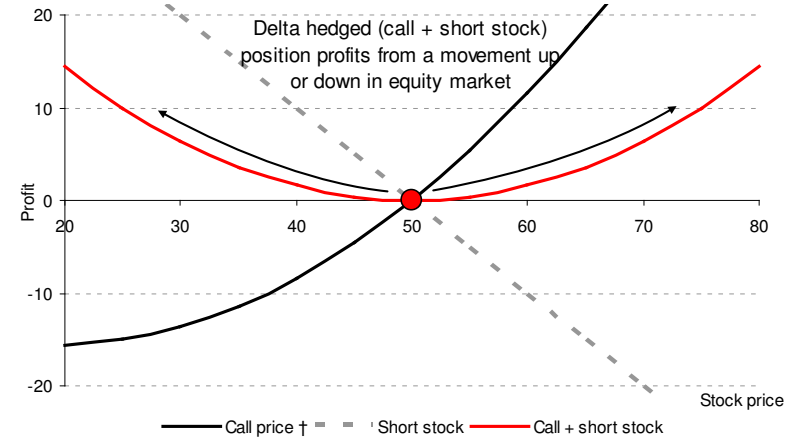
Delta-hedged option gives a position whose profitability is determined by volatility

If an option is purchased at an implied volatility that is lower than the realised volatility over the life of the option, then the investor, in theory, earns a profit from buying cheap volatility. However, the effect of buying cheap volatility is dwarfed by the profit or loss from the direction of the equity market. For this reason, directional investors are usually more concerned with premium rather than implied volatility. Volatility investors will, however, hedge the equity exposure. This will result in a position whose profitability is solely determined by the volatility (not direction) of the underlying. As delta measures the equity sensitivity of an option, removing equity exposure is called delta hedging (as a portfolio with no equity exposure has delta = 0).

Delta hedging example

As the delta of a portfolio is equal to the sum of the deltas of the securities in the portfolio, a position can be delta hedged by purchasing, or going short, a number of shares (or futures in the case of an index) equal to the delta. For example, if ten call options have been bought with a delta of 40%, then four shares (10 × 40% = 4) have to be shorted to create a portfolio of zero delta. The shares have to be shorted as a call option has positive delta; hence, the delta hedge has to be negative for the sum of the two positions to have zero delta. If we were long a put (which has negative delta), then we would have to buy stock to ensure the overall delta was zero.

Figure 24. Delta-Hedged Call



Source: Santander Investment Bolsa.

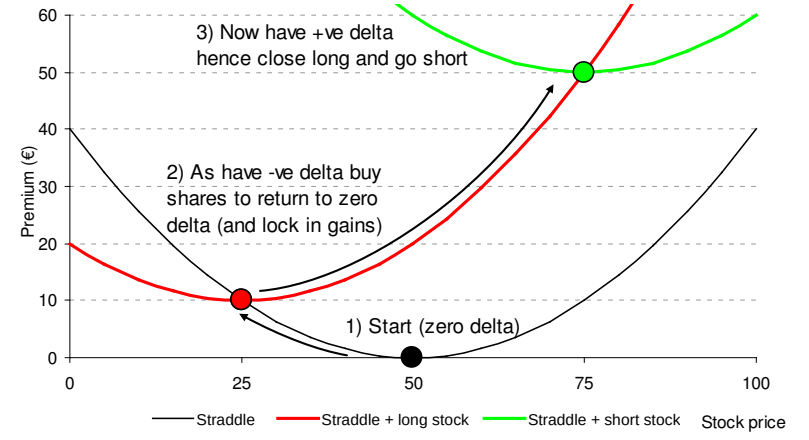
Constant delta hedging is called gamma scalping

The rate delta changes as spot moves is called gamma; hence, gamma is the convexity of the payout. As the delta changes, a volatility investor has to delta hedge in order to ensure there is no equity exposure. Constantly delta hedging in this way is called gamma scalping, as it ensures a long volatility position earns a profit as spot moves.

Gamma scalping (delta re-hedging) locks in profit as underlying moves

We shall assume an investor has purchased a zero delta straddle (or strangle), but the argument will hold for long call or put positions as well. If equity markets fall (from position 1 to position 2 in the chart) the position will become profitable and the delta will decrease from zero to a negative value. In order to lock in the profit, the investor must buy stock (or futures) for the portfolio to return to zero delta. Now that the portfolio is equity market neutral, it will profit from a movement up or down in the equity market. If equity markets then rise, the initial profit will be kept and a further profit earned (movement from position 2 to position 3). At position 3 the stock (or futures) position is sold and a short position initiated to return the position to zero delta.

Figure 25. Locking in Gains through Delta Hedging



Source: Santander Investment Bolsa.

Long gamma position can sit on the bid and offer

Short gamma positions have to cross the underlying bid-offer spread (when delta hedging)

As shown above, a long gamma (long volatility) position has to buy shares if they fall, and sell them if they rise. Buying low and selling high earns the investor a profit. Additionally, as a gamma scalper can enter bids and offers away from current spot, there is no need to cross the spread (as a long gamma position can be delta hedged by sitting on the bid and offer). A short gamma position represents the reverse situation, and requires crossing the spread to delta hedge. While this hidden cost is small, it could be substantial over the long term for underlyings with relatively wide bid-offer spreads.

Best to delta hedge on key dates or on turn of market

If markets have a clear direction (ie, they are trending), it is best to delta hedge less frequently. However, in choppy markets that are range bound it is best to delta hedge very frequently. For more detail on how hedging frequency affects returns and the path dependency of returns, see the section *Stretching Black-Scholes Assumptions*. If there is a key announcement (either economic or earnings-related to affect the underlying), it is best to delta hedge just before the announcement to ensure that profit is earned from any jump (up or down) that occurs.

GAMMA HEDGING CAN 'PIN' A STOCK APPROACHING EXPIRY

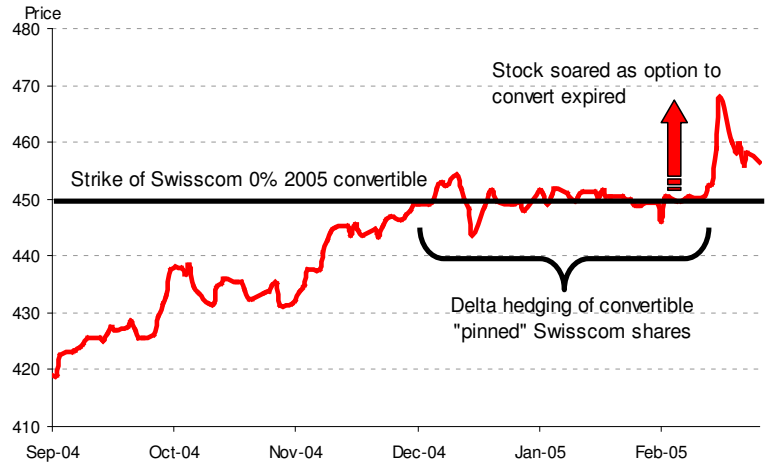
Single stocks (but not indices) can be pinned approaching expiry

As an investor who is long gamma can delta hedge by sitting on the bid and offer, this trade can pin an underlying to the strike. This is a side effect of selling if the stock rises above the strike, and buying if the stock falls below the strike. The amount of buying and selling has to be significant compared with the traded volume of the underlying, which is why pinning normally occurs for relatively illiquid stocks or where the position is particularly sizeable. Given the high trading volume of indices, it is difficult to pin a major index. Pinning is more likely to occur in relatively calm markets, where there is no strong trend to drive the stock away from its pin.

Large size of Swisscom convertible pinned underlying for many months

One of the most visible examples of pinning occurred in late 2004/early 2005, due to a large Swiss government debt issue, (Swisscom 0% 2005) convertible into the relatively illiquid Swisscom shares. As the shares traded close to the strike approaching maturity, the upward trend of the stock was broken. Swisscom was pinned for two to three months until the exchangeable expired. After expiration, the stock snapped back to where it would have been if the upward trend had not been paused. A similar event occurred to AXA in the month preceding the Jun05 expiry, when it was pinned close to €20 despite the broader market rising (after expiry AXA rose 4% in four days to make up for its earlier underperformance).

Figure 26. Pinning of Swisscom Stock Approaching Convertible Expiry



Source: Santander Investment Bolsa.

OPTION TRADING RULES OF THUMB

To calculate option premiums and volatility exactly is typically too difficult to do without the aid of a tool. However, there are some useful rules of thumb that can be used to give an estimate. These are a useful sanity check in case an input to a pricing model has been entered incorrectly.

- **Profit from delta hedging is proportional to square of return.** Due to the convexity of an option, if the volatility is doubled the profits from delta hedging are multiplied by a factor of four. For this reason, variance (which looks at squared returns) is a better measure of deviation than volatility.
- **ATM option premium in percent is roughly 0.4 × volatility × square root of time.** If one assumes zero interest rates and dividends, then the formula for the premium of an ATM call or put option simplifies to $0.4 \times \sigma \times \sqrt{T}$. Therefore, a one-year ATM option on an underlying with 20% implied is worth c8% ($= 0.4 \times 20\% \times \sqrt{1}$). OTM options can be calculated from this estimate using an estimated 50% delta.
- **Historical annualised volatility roughly equal to 16 × percentage daily move.** Historical volatility can be estimated by multiplying the typical return over a period by the square root of the number of periods in a year (eg, 52 weeks or 12 months in a year). Hence, if a security moves 1% a day, it has an annualised volatility of 16% (as $16 \approx \sqrt{252}$ and we assume there are 252 trading days).

Historically, annualised volatility is roughly equal to 16 × percentage daily move

A stock that moves 1% a day has 16% volatility, and a stock that moves 2% has 32% volatility

Volatility is defined as the annualised standard deviation of log returns (where return = P_i / P_{i-1}). As returns are normally close to 1 (=100%) the log of returns is very similar to 'return – 1' (which is the percentage change of the price). Hence, to calculate the annualised volatility for a given percentage move, all that is needed is to annualise the percentage change in the price. This is done by multiplying the percentage move by the square root of the number of samples in a year (as volatility is the square root of variance). For example, market convention is to assume there are approximately 252 trading days a year. If a stock moves 1% a day, then its volatility is $1\% \times \sqrt{252}$, which is approximately $1\% \times 16 = 16\%$ volatility. Similarly, if a stock moves 2% a day it has 32% volatility.

Number of trading days in year	= 252	=> Multiply daily returns by $\sqrt{252}$	≈ 16
Number of weeks in year	= 52	=> Multiply weekly returns by $\sqrt{52}$	≈ 7
Number of months in year	= 12	=> Multiply monthly returns by $\sqrt{12}$	≈ 3.5

ATM option premium in percent is roughly 0.4 × volatility × square root of time

Call price = $S\,N(d_1) - K\,N(d_2)\,e^{-rT}$

Assuming zero interest rates and dividends (r = 0)

- ⇨ ATM call price = $S\,N(\sigma \times \sqrt{T} / 2) - S\,N(-\sigma \times \sqrt{T} / 2)$ as K=S (as ATM)
- ⇨ ATM call price = $S \times \sigma \times \sqrt{T} / \sqrt{(2\pi)}$
- ⇨ ATM call price = $\sigma \times \sqrt{T} / \sqrt{(2\pi)}$ in percent
- ⇨ ATM call price ≈ $0.4 \times \sigma \times \sqrt{T}$ in percent

where:

Definition of d_1 and d_2 is the standard Black-Scholes formula.

σ = implied volatility

S = spot

K = strike

R = interest rate

T = time to expiry

N(z) = cumulative normal distribution

Example 1

1Y ATM option on an underlying with 20% implied is worth c.8% (=0.4 × 20% × √1)

Example 2

3M ATM option on an underlying with 20% implied is worth c.4% (=0.4 × 20% ×√0.25 =0.4 × 20% × 0.5)

OTM options can be calculated by assuming 50% delta

If an index is 3000pts and has a 20% implied then the price of a 3M ATM option is approximately 240pts (3000×8% as calculated above). A 3200 call is therefore approximately 240 – 50% (3200-3000) = 140pts assuming a 50% delta. Similarly, a 3200 put is approximately 340pts.

Profit from delta hedging is proportional to percentage move squared

Due to the convexity of an option, if the volatility is doubled, the profits from delta hedging are multiplied by a factor of four. For this reason, variance (which looks at squared returns) is a better measure of deviation than volatility. Assuming constant volatility, zero interest rates and dividend, the daily profit and loss (P&L) from delta hedging an option is given below.

Daily P&L from option = Delta P&L + Gamma P&L + Theta P&L

- ⇨ Daily P&L from option = $S\delta + S^2\gamma/2 + t\theta$ where S is change in Stock and t is time
- ⇨ Daily P&L from option - $S\delta$ = + $S^2\gamma/2 + t\theta$ = Delta hedged P&L from option
- ⇨ Delta hedged P&L from option = $S^2\gamma/2$ + cost term ($t\theta$ does not depend on stock price)

where:

δ = delta

γ = gamma

t = time

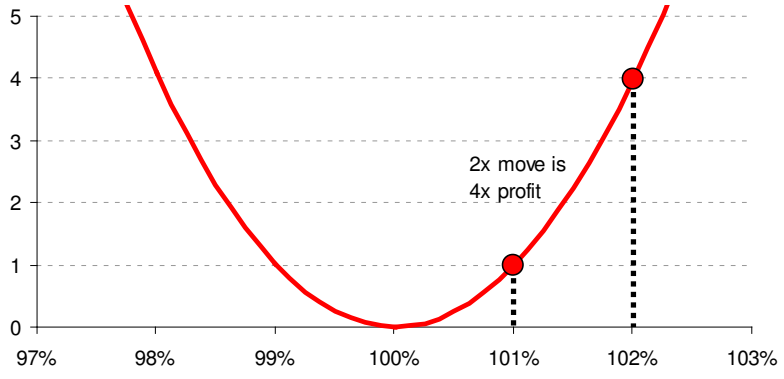
θ = theta

Hedge investors prefer occasional large moves to constant small moves

If the effect of theta is ignored (as it is a cost that does not depend on the size of the stock price movement), the profit of a delta hedged option position is equal to a scaling factor (gamma/2) multiplied by the square of the return. This means that the profit from a 2% move in a stock price is four times (2^2=4) the profit from a 1% move in stock price.

This can also be seen from Figure 27 below, as the additional profit from the move from 1% to 2% is three times the profit from 0% to 1% (for a total profit four times the profit for a 1% move).

Figure 27. Profile of a Delta-Hedged Option

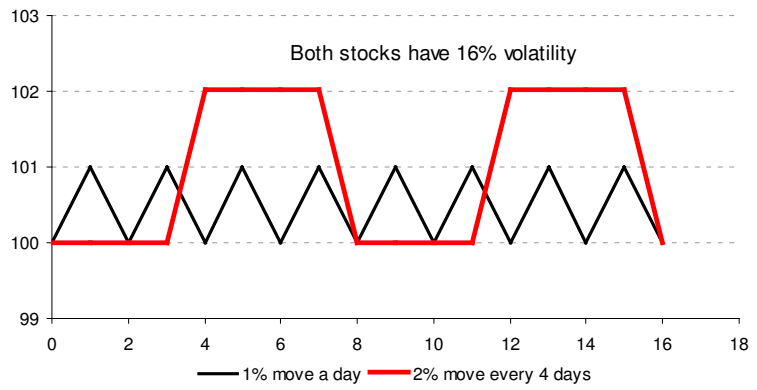


Source: Santander Investment Bolsa.

Example: make same delta hedge profit with 1% a day move as 2% every four days

Let's assume there are two stocks: one of them moves 1% a day and the other 2% every four days (see Figure 28 below). Both stocks have the same 16% volatility and delta hedging them earns the same profit (as four times as much profit is earned on the days the stock moves 2% as when it moves 1%).

Figure 28. Two Stocks with the Same Volatility



Source: Santander Investment Bolsa.

VARIANCE IS THE KEY, NOT VOLATILITY

Partly due to its use in Black-Scholes, historically, volatility has been used as the measure of deviation for financial assets. However, the correct measure of deviation is variance (or volatility squared). Volatility should be considered to be a derivative of variance. The realisation that variance should be used instead of volatility led volatility indices, such as the VIX, to move away from ATM volatility (VXO index) towards a variance-based calculation.

VARIANCE, NOT VOLATILITY, IS CORRECT MEASURE FOR DEVIATION

Variance swaps are quoted in volatility terms due to greater user familiarity

There are three reasons why variance, not volatility, should be used as the correct measure for volatility. However, despite these reasons, even variance swaps are normally quoted as the square root of variance for an easier comparison with the implied volatility of options (but we note that skew and convexity mean the fair price of variance should always trade above ATM options).

- **Variance takes into account implied volatility at all stock prices.** Variance takes into account the implied volatility of all strikes with the same expiry (while ATM implied will change with spot, even if volatility surface does not change).
- **Deviations need to be squared to avoid cancelling.** Mathematically, if deviations were simply summed then positive and negative deviations would cancel. This is why the sum of squared deviations is taken (variance) to prevent the deviations from cancelling. Taking the square root of this sum (volatility) should be considered a derivative of this pure measure of deviation (variance).
- **Profit from a delta-hedged option depends on the square of the return.** Due to the convexity of an option, if the volatility is doubled, the profits from delta hedging are multiplied by a factor of four. For this reason, variance (which looks at squared returns) is a better measure of deviation than volatility.

(1) VARIANCE TAKES INTO ACCOUNT VOLATILITY AT ALL STOCK PRICES

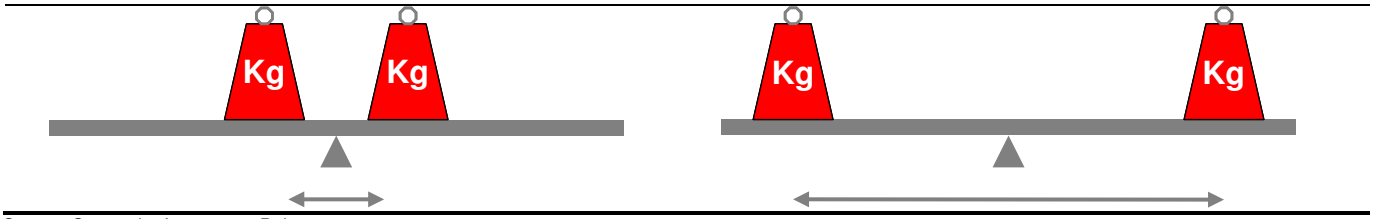
When looking at how rich or cheap options with the same maturity are, rather than looking at the implied volatility for a certain strike (ie, ATM or another suitable strike) it is better to look at the implied variance as it takes into account the implied volatility of all strikes. For example, if an option with a fixed strike that is initially ATM is bought, then as soon as spot moves it is no longer ATM. However, if a variance swap (or log contract[4] of options in the absence of a variance swap market) is bought, then its traded level is applicable no matter what the level of spot. The fact a variance swap (or log contract) payout depends only on the realised variance and is not path dependent makes it the ideal measure for deviation.

[4] For more details, see the section *Volatility, Variance and Gamma Swaps*.

(2) DEVIATIONS NEED TO BE SQUARED TO AVOID CANCELLING

If a seesaw has two weights on it and the weights are the same distance either side from the pivot, the weights are balanced as the centre of the mass is in line with the pivot (see graph on left hand side below). If the weights are further away from the pivot the centre of the mass (hence the average/expected distance of the weights) is still in line with the pivot (see graph on right hand side below). If the deviation of the two weights from the pivot is summed together, in both cases they would be zero (as one weight's deviation from the pivot is the negative of the other). In order to avoid the deviation cancelling this way, the square of the deviation (or variance) is taken, as the square of a number is always positive.

Figure 29. Low Deviation **High Deviation**



Source: Santander Investment Bolsa.

(3) PROFIT FROM DELTA HEDGING PROPORTIONAL TO RETURN SQUARED

Assuming constant volatility, zero interest rates and dividend, the daily profit and loss (P&L) from delta hedging an option is given below:

Delta-hedged P&L from option = $S^2\gamma$ /2 + cost term

where:

γ = gamma

This can also be seen from Figure 27 *Profile of a Delta-Hedged Option* in the previous section (page 45), as the additional profit from the move from 1% to 2% is three times the profit from 0% to 1% (for a total profit four times the profit for a 1% move).

VOLATILITY SHOULD BE CONSIDERED A DERIVATIVE OF VARIANCE

The three examples above show why variance is the natural measure for deviation. Volatility, the square root of variance, should be considered a derivative of variance rather than a pure measure of deviation. It is variance, not volatility, that is the second moment of a distribution (the first moment is the forward or expected price). For more details on moments, read the section *How to Measure Skew and Smile*.

VIX AND VDAX MOVED FROM OLD ATM CALCULATION TO VARIANCE

VIMEX is one of the few volatility indices not to use a variance swap based calculation

Due to the realisation that variance, not volatility, was the correct measure of deviation, on Monday, September 22, 2003, the VIX index moved away from using ATM implied towards a variance-based calculation. Variance-based calculations have also been used for by other volatility index providers. The old VIX, renamed VXO, took the implied volatility for strikes above and below spot for both calls and puts. As the first two-month expiries were used, the old index was an average of eight implied volatility measures as 8 = 2 (strikes) × 2 (put/call) × 2 (expiry). We note that the use of the first two expiries (excluding the front month if it was less than eight calendar days) meant the maturity was on average 1.5 months, not one month as for the new VIX.

Similarly, the VDAX index, which was based on 45-day ATM-implied volatility, has been superseded by the V1X index, which, like the new VIX, uses a variance swap calculation. All recent volatility indices, such as the vStoxx (V2X), VSMI (V3X), VFTSE, VNKY and VHSI, use a variance swap calculation, although we note the recent VIMEX index uses a similar methodology to the old VIX (potentially due to illiquidity of OTM options on the Mexbol index).

VARIANCE TERM STRUCTURE IS NOT ALWAYS FLAT

While average variance term structure should be flat in theory, in practice supply and demand imbalances can impact variance term structure. The buying of protection at the long end should mean that variance term structure is on average upward sloping, but in turbulent markets it is usually inverted.

VOLATILITY, VARIANCE AND GAMMA SWAPS

In theory, the profit and loss from delta hedging an option is fixed and is based solely on the difference between the implied volatility of the option when it was purchased and the realised volatility over the life of the option. In practice, with discrete delta hedging and unknown future volatility, this is not the case, leading to the creation of volatility, variance and gamma swaps. These products also remove the need to continuously delta hedge, which can be very labour-intensive and expensive. Until the credit crunch, variance swaps were the most liquid of the three, but now volatility swaps are more popular for single stocks.

VOLATILITY, VARIANCE & GAMMA SWAPS GIVE PURE VOL EXPOSURE

Naming of volatility swaps, variance swaps and gamma swaps is misleading, as they are in fact forwards

As spot moves away from the strike of an option the gamma decreases, and it becomes more difficult to profit via delta hedging. Second-generation volatility products, such as volatility swaps, variance swaps and gamma swaps, were created to give volatility exposure for all levels of spot and also to avoid the overhead and cost of delta hedging. While volatility and variance swaps have been traded since 1993, they became more popular post-1998, when Russia defaulted on its debts and Long-Term Capital Management (LTCM) collapsed. The naming of volatility swaps, variance swaps and gamma swaps is misleading, as they are in fact forwards. This is because their payoff is at maturity, whereas swaps have intermediate payments.

- **Volatility swaps.** Volatility swaps were the first product to be traded significantly and became increasingly popular in the late 1990s until interest migrated to variance swaps. Following the collapse of the single-stock variance market in the credit crunch, they are having a renaissance due to demand from dispersion traders. A theoretical drawback of volatility swaps is the fact that they require a volatility of volatility (vol of vol) model for pricing, as options need to be bought and sold during the life of the contract (which leads to higher trading costs). However, in practice, the vol of vol risk is small and volatility swaps trade roughly in line with ATM forward (ATMf) implied volatility.
- **Variance swaps.** The difficulty in hedging volatility swaps drove liquidity towards the variance swap market, particularly during the 2002 equity collapse. As variance swaps can be replicated by delta hedging a static portfolio of options, it is not necessary to buy or sell options during the life of the contract. The problem with this replication is that it assumes options of all strikes can be bought, but in reality very OTM options are either not listed or not liquid. Selling a variance swap and only hedging with the available roughly ATM options leaves the vendor short tail risk. As the payout is on variance, which is volatility squared, the amount can be very significant. For this reason, liquidity on single-stock variance disappeared in the credit crunch.
- **Gamma swaps.** Dispersion traders profit from overpriced index-implied volatility by going long single-stock variance and short index variance. The portfolio of variance swaps is not static; hence, rebalancing trading costs are incurred. Investment banks attempted to create a liquid gamma swap market, as dispersion can be implemented via a static portfolio of gamma swaps (and, hence, it could better hedge the exposure of their books from selling structured products). However, liquidity never really took off due to limited interest from other market participants.

VOLATILITY SWAP ≤ GAMMA SWAP ≤ VARIANCE SWAP

Variance swaps are quoted as the square root of variance (to allow easier comparison with implied volatility)

Variance and gamma swaps are normally quoted as the square root of variance to allow easier comparison with the options market. However, typically variance swaps trade in line with the 30 delta put (if skew is downward sloping as normal). The square root of the variance strike is always above volatility swaps (and ATMf implied as volatility swaps ≈ ATMf implied). This is due to the fact a variance swap payout is convex (hence, will always be greater than or equal to volatility swap payout of identical vega, which is explained later in the section). Only for the unrealistic case of no vol of vol (ie, future volatility is constant and known) will the price of a volatility swap and variance swap (and gamma swap) be the same[5]. The fair price of a gamma swap is between volatility swaps and variance swaps.

(1) VOLATILITY SWAPS

Only for flat skew will the price of a volatility swap and variance swap (and gamma swap) be the same

The payout of a volatility swap is simply the notional, multiplied by the difference between the realised volatility and the fixed swap volatility agreed at the time of trading. As can be seen from the payoff formula below, the profit and loss is completely path independent as it is solely based on the realised volatility. Volatility swaps were previously illiquid, but are now more popular with dispersion traders, given the single stock variance market no longer exists post credit crunch. Unless packaged as a dispersion, volatility swaps rarely trade. As dispersion is short index volatility, long single stock volatility, single stock volatility swaps tend to be bid only (and index volatility swaps offered only).

Volatility swap payoff

$(\sigma_F - \sigma_S) \times$ volatility notional

where:

σ_F = future volatility (that occurs over the life of contract)

σ_S = swap rate volatility (fixed at the start of contract)

Volatility notional = Vega = notional amount paid (or received) per volatility point

(2) VARIANCE SWAPS

Variance swaps are identical to volatility swaps except their payout is based on variance (volatility squared) rather than volatility. Variance swaps are long skew (more exposure to downside put options than upside calls) and convexity (more exposure to OTM options than ATM). One-year variance swaps are the most frequently traded.

Variance swap payoff

$(\sigma_F^2 - \sigma_S^2) \times$ Variance notional

where:

Variance notional = notional amount paid (or received) per variance point

NB: Variance notional = Vega / $(2 \times \sigma_S)$ where σ_S = current variance swap price

[5] A variance swap payout is based on cash return assuming zero mean, whereas a delta-hedged option variance payout is based on a forward. Hence, a variance swap fair price will be slightly above a constant and flat volatility surface if the drift is non-zero (as close-to-close cash returns will be lifted by the drift).

VARIANCE SWAPS CAPS ARE EFFECTIVELY SHORT OPTION ON VAR

Variance swaps on single stocks and emerging market indices are normally capped at 2.5 times the strike, in order to prevent the payout from rising towards infinity in a crisis or bankruptcy. A cap on a variance swap can be modelled as a vanilla variance swap less an option on variance whose strike is equal to the cap. More details can be found in the section *Options on Variance*.

Capped variance should be hedged with OTM calls, not OTM puts

The presence of a cap on a variance swap means that if it is to be hedged by only one option it should be a slightly OTM call, not an OTM (approx delta 30) put. This is to ensure the option is so far OTM when the cap is hit that the hedge disappears. If this is not done, then if a trader is long a capped variance swap he would hedge by going short an OTM put. If markets fall with high volatility hitting the cap, the trader would be naked short a (now close to ATM) put. Correctly hedging the cap is more important than hedging the skew position.

S&P500 variance market is increasing in liquidity, while SX5E has become less liquid

The payout of volatility swaps and variance swaps of the same vega is similar for small payouts, but for large payouts the difference becomes very significant due to the quadratic (ie, squared) nature of variance. The losses suffered in the credit crunch from the sale of variance swaps, particularly single stock variance (which, like single stock volatility swaps now, was typically bid), have weighed on their subsequent liquidity. Now variance swaps only trade for indices (usually without cap, but sometimes with). The popularity of VIX futures has raised awareness of variance swaps, which has helped S&P500 variance swaps become more liquid than they were before the credit crunch. S&P500 variance swaps now trade with a bid-offer spread of c30bp and sizes of approximately US$5mn vega can regularly trade every day. However, SX5E variance swap liquidity is now a fraction of its pre-credit-crunch levels, with bid-offer spreads now c80bp compared with c30bp previously.

CORRIDOR VARIANCE SWAPS ARE NOT LIQUID

Corridor variance swaps give exposure to volatility, only when spot is in a certain range

As volatility and spot are correlated, volatility buyers would typically only want exposure to volatility levels for low values of spot. Conversely, volatility sellers would only want exposure for high values of spot. To satisfy this demand, corridor variance swaps were created. These only have exposure when spot is between spot values A and B. If A is zero, then it is a down variance swap. If B is infinity, it is an up variance swap. There is only a swap payment on those days the spot is in the required range, so if spot is never in the range there is no payment. Because of this, a down variance swap and up variance swap with the same spot barrier is simply a vanilla variance swap. The liquidity of corridor variance swaps was always far lower than for variance swaps and, since the credit crunch, they are rarely traded.

Corridor variance swap payoff

$(\sigma_{F \text{ when in corridor}}^2 - \sigma_S^2) \times$ variance notional $\times$ percentage of days spot is within corridor

where:

$\sigma_{F \text{ when in corridor}}$ = future volatility (of returns P_i/P_{i-1} which occur when $B_L < P_{i-1} \leq B_H$)
B_L and B_H, are the lower and higher barriers, where B_L could be 0 and B_H could be infinity.

(3) GAMMA SWAPS

The payout of gamma swaps is identical to that of a variance swap, except the daily P&L is weighted by spot ($price_n$) divided by the initial spot ($price_0$). If spot range trades after the position is initiated, the payouts of a gamma swap are virtually identical to the payout of a variance swap. Should spot decline, the payout of a gamma swap decreases. Conversely, if spot increases, the payout of a gamma swap increases. This spot-weighting of a variance swap payout has the following attractive features:

- Spot weighting of variance swap payout makes it unnecessary to have a cap, even for single stocks (if a company goes bankrupt with spot dropping close to zero with very high volatility, multiplying the payout by spot automatically prevents an excessive payout).

Gamma swaps are ideal for trading dispersion as it is 'fire and forget'

- If a dispersion trade uses gamma swaps, the amount of gamma swaps needed does not change over time (hence, the trade is 'fire and forget', as the constituents do not have to be rebalanced as they would if variance swaps were used).
- A gamma swap can be replicated by a static portfolio of options (although a different static portfolio to variance swaps), which reduces hedging costs. Hence, no volatility of volatility model is needed (unlike volatility swaps).

Gamma swap market has never had significant liquidity

A number of investment banks attempted to kick start a liquid gamma swap market, partly to satisfy potential demand from dispersion traders and partly to get rid of some of the exposure from selling structured products (if the product has less volatility exposure if prices fall, then a gamma swap better matches the change in the vega profile when spot moves). While the replication of the product is as trivial as for variance swaps, it was difficult to convince other market participants to switch to the new product and liquidity stayed with variance swaps (although after the credit crunch, single-stock variance liquidity moved to the volatility swap market). If the gamma swap market ever gains liquidity, long skew trades could be put on with a long variance-short gamma swap position (as this would be long downside volatility and short upside volatility, as a gamma swap payout decreases/increases with spot).

Gamma swap payoff

$(\sigma_G^2 - \sigma_S^2) \times$ variance notional

where:

σ_G^2 = future spot weighted (ie, multiplied by $\frac{price_n}{price_0}$) variance

σ_S^2 = swap rate variance (fixed at the start of contract)

PAYOUT OF VOLATILITY, VARIANCE AND GAMMA SWAPS

The payout of volatility swaps, variance swaps and gamma swaps is the difference between the fixed and floating leg, multiplied by the notional. The calculation for volatility assumes zero mean return (or zero drift) to make the calculation easier and to allow the variance calculation to be additive.

- **Fixed leg.** The cost (or fixed leg) of going long a volatility, variance or gamma swap is always based on the swap price, σ_S (which is fixed at inception of the contract). The fixed leg is σ_S for volatility swaps, but is σ_S^2 for variance and gamma swaps).

- **Floating leg.** The payout (or floating leg) for volatility and variance swaps is based on the same variable σ_F (see equation below). The only difference is that a volatility swap payout is based on σ_F, whereas for a variance swap it is σ_F^2. The gamma swap payout is based on a similar variable $\sigma_G^{2,}$ which is σ_F^2 multiplied by $\text{price}_n/\text{price}_{0.}$

$$\sigma_F = 100 \times \sqrt{\frac{\sum_{i=1}^{T} [\text{Ln}(\text{return}_i)]^2}{T_{\text{exp}}}} \times \text{number business days in year}$$

$$\sigma_G = 100 \times \sqrt{\frac{\sum_{i=1}^{T} \frac{\text{price}_i}{\text{price}_0} [\text{Ln}(\text{return}_i)]^2}{T_{\text{exp}}}} \times \text{number business days in year}$$

$$\text{return}_i = \frac{\text{price}_i}{\text{price}_{i-1}} \text{ for indices}$$

$$\text{return}_i = \frac{\text{price}_i + \text{dividend}_i}{\text{price}_{i-1}} \text{ for single stocks (dividend}_i \text{ is dividend going ex on day n)}$$

where:

number of business days in year = 252 (usual market practice)

T_{exp} = Expected value of N (if no market disruption occurs). A market disruption is when shares accounting for at least 20% of the index market cap have not traded in the last 20 minutes of the trading day.

Variance is additive with zero mean assumption

As variance is additive, payout is not path dependent and no vol of vol model is needed

Normally, standard deviation or variance looks at the deviation from the mean. The above calculations assume a zero mean, which simplifies the calculation (typically, one would expect the mean daily return to be relatively small). With a zero mean assumption, variance is additive. A mathematical proof of the formula below is given in the section *Measuring Historical Volatility* in the *Appendix*.

Past variance + future variance = total variance

Lack of dividend adjustment for indices affects pricing

The return calculation for a variance swap on an index does not adjust for any dividend payments that go ex. This means that the dividend modelling method can affect the pricing. Near-dated and, hence, either known or relatively certain dividends should be modelled discretely rather than as a flat yield. The changing exposure of the variance swap to the volatility on the ex date can be as large as 0.5 volatility points for a three-year variance swap (if all other inputs are kept constant, discrete (ie, fixed) dividends lift the value of both calls and puts, as proportional dividends simply reduce the volatility of the underlying by the dividend yield).

Calculation agents might have discretion as to when a market disruption event occurs

Normally, the investment bank is the calculation agent for any variance swaps traded. As the calculation agent normally has some discretion over when a market disruption event occurs, this can lead to cases where one calculation agent believes a market disruption occurs and another does not. This led to a number of disputes in 2008, as it was not clear if a market or exchange disruption had occurred. Similarly, if a stock is delisted, the estimate of future volatility for settlement prices is unlikely to be identical between firms, which can lead to issues if a client is long and short identical products at different investment banks. These problems are less of an issue if the counterparties are joint calculation agents.

HEDGING OF VARIANCE SWAPS CAN IMPACT EQUITY & VOL MARKET

Hedging volatility, variance and gamma swaps always involve the trading of a strip of options of all strikes and delta hedging at the close. The impact the hedging of all three products has on equity and volatility markets is similar, but we shall use the term variance swaps, as it has by far the most impact of the three (the same arguments will apply for volatility swaps and gamma swaps).

Short end of volatility surfaces is now pinned to realised

If there is a divergence between short-dated variance swaps and realised volatility, hedge funds will put on variance swap trades to profit from this divergence. This puts pressure on the short-dated end of volatility surfaces to trade close to the current levels of realised volatility. Due to the greater risk of unexpected events, it is riskier to attempt a similar trade at the longer-dated end of volatility surfaces.

Skew levels affected by direction of volatility trading

As variance swaps are long skew, momentum of volatility trading moves skew the same way

As variance swaps became a popular way to express a view of the direction of implied volatility, they impacted the levels of skew. This occurred as variance swaps are long skew (explained below) and, if volatility is being sold through variance swaps, this weighs on skew. This occurred between 2003 and 2005, which pushed skew to a multiple-year low. As volatility bottomed, the pressure from variance swap selling abated and skew recovered.

Delta hedge can suppress or exaggerate market moves

Direction of hedging variance swap flow, is in the direction which ensures less money is made

As the payout of variance swaps is based on the close-to-close return, they all have an intraday delta (which is equal to zero if spot is equal to the previous day's close). As this intraday delta resets to zero at the end of the day, the hedging of these products requires a delta hedge at the cash close. A rule of thumb is that the direction of hedging flow is in the direction that makes the trade the least profit (ensuring that if a trade is crowded, it makes less money). This flow can be hundreds of millions of US dollars or euros per day, especially when structured products based on selling short-dated variance are popular (as they were in 2006 and 2007, less so since the high volatility of the credit crunch).

- **Variance buying suppresses equity market moves.** If clients are net buyers of variance swaps, they leave the counterparty trader short. The trader will hedge this short position by buying a portfolio of options and delta hedging them on the close. If spot has risen over the day the position (which was originally delta-neutral) has a positive delta (in the same way as a delta-hedged straddle would have a positive delta if markets rise). The end of day hedge of this position requires selling the underlying (to become delta-flat), which suppresses the rise of spot. Similarly, if markets fall, the delta hedge required is to buy the underlying, again suppressing the market movement.
- **Variance selling exaggerates equity market moves.** Should clients be predominantly selling variance swaps, the hedging of these products exaggerates market moves. The argument is simply the inverse of the argument above. The trader who is long a variance swap (as the client is short) has hedged by selling a portfolio of options. If markets rise, the delta of the position is negative and, as the variance swap delta is reset to zero at the end of the day, the trader has to buy equities at the same time (causing the close to be lifted for underlyings that have increased in value over the day). If markets fall, then the trader has to sell equities at the end of the day (as the delta of a short portfolio of options is positive). Movements are therefore exaggerated, and realised volatility increases if clients have sold variance swaps.

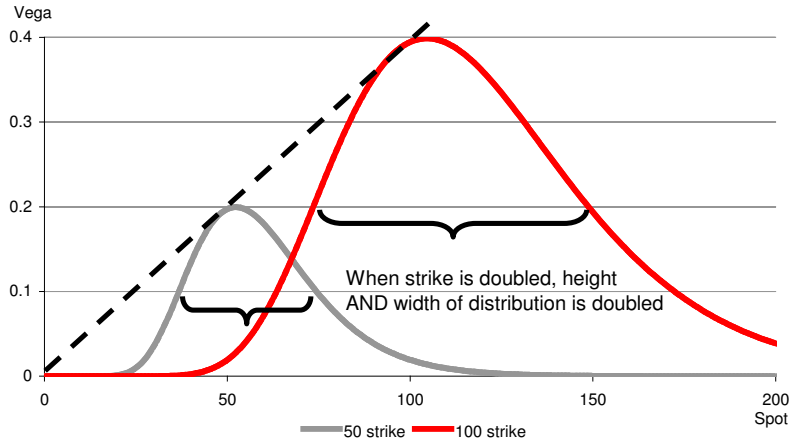
Basis risk between cash and futures can cause traders problems

We note that the payout of variance swaps is based on the cash close, but traders normally delta hedge using futures. The difference between the cash and futures price is called the basis, and the risk due to a change in basis is called basis risk. Traders have to take this basis risk between the cash close and futures close, which can be significant as liquidity in the futures market tends to be reduced after the cash market closes.

HEDGING VOLATILITY, VARIANCE AND GAMMA SWAPS WITH OPTIONS

As volatility, variance and gamma swaps give volatility exposure for all values of spot, they need to be hedged by a portfolio of options of every strike. An equal-weighted portfolio is not suitable, as the vega profile of an option increases in size and width as strike increases (ie, an option of strike 2K has a peak vega double the peak vega of an option of strike K and is also twice the width). This is shown below.

Figure 30. Vega of Options of Different Strikes



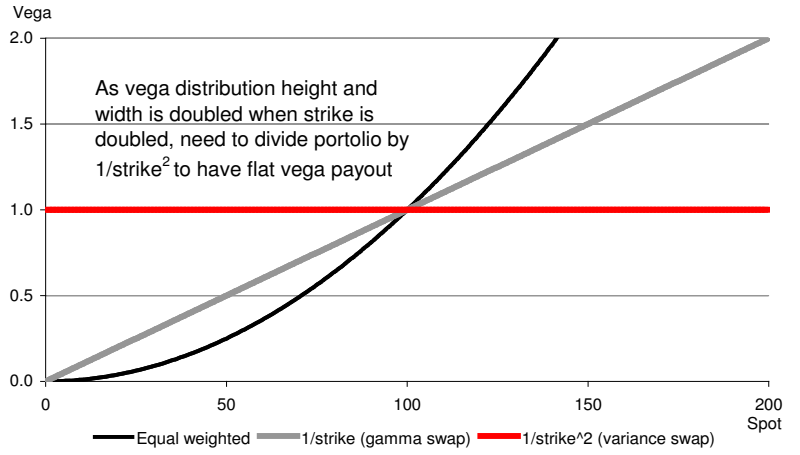
Source: Santander Investment Bolsa.

Variance swaps are hedged with portfolio weighted $1/K^2$

As height and width of vega profile is doubled when strike is doubled, variance swaps hedged with portfolio 1/strike2

Because a variance swap has a flat vega profile, the correct hedge is a portfolio of options weighted $1/K^2$ (where K is the strike of the option, ie, each option is weighted by 1 divided by its own strike squared). The reason why this is the correct weighting is due to the fact the vega profile doubles in height and width if the strike is doubled. The portfolio has to be divided by strike K once, to correct for the increase in height, and again to compensate for the increase in width (for a combined weight of $1/K^2$). A more mathematical proof of why the hedge for a variance swap is a portfolio of options weighted $1/K^2$ (a so-called log contract) is given in the section *Proof Variance Swaps Should Be Hedged by a Log Contract (= $1/K^2$)* in the *Appendix*. As a gamma swap payout is identical to a variance swap multiplied by spot, the weighting is 1/K (multiplying by spot cancels one of the K's on the denominator). The vega profile of a portfolio weighted 1/K and $1/K^2$ is shown below, along with an equal-weighted portfolio for comparison. We note that although the vega profile of a variance swap is a flat line, the value is not constant and it moves with volatility (variance swap vega = variance notional $\times \sigma$). The vega profile of a volatility swap is of course a flat line (as vega is equal to the volatility notional).

Figure 31. Vega of Portfolio of Options of All Strikes



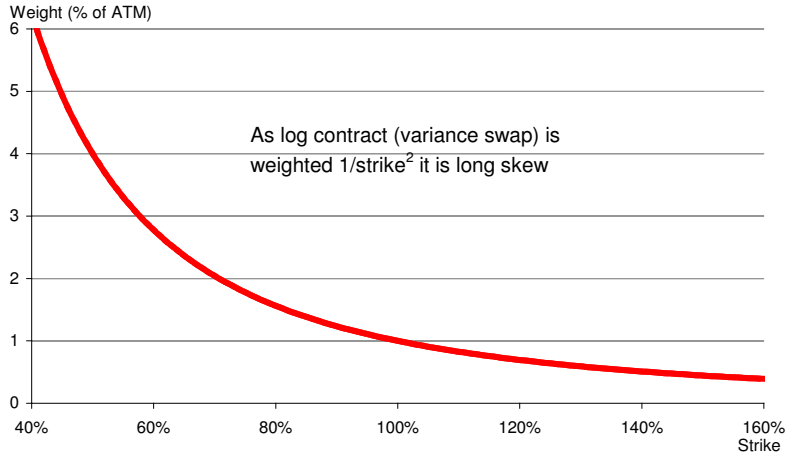
Source: Santander Investment Bolsa.

Variance swaps are long skew and volatility surface curvature

Variance swaps are long skew, long volatility surface curvature and vega convexity

The $1/K^2$ weighting means a larger amount of OTM puts are traded than OTM calls (approx 60% is made up of puts). This causes a log contract (portfolio of options weighted $1/K^2$) to be long skew. The curved nature of the weighting means the wings (very out-of-the-money options) have a greater weighting than the body (near ATM options), which means a log contract is long volatility surface curvature[6].

Figure 32. Weight of Options in Log Contract (Variance Swap)



Source: Santander Investment Bolsa.

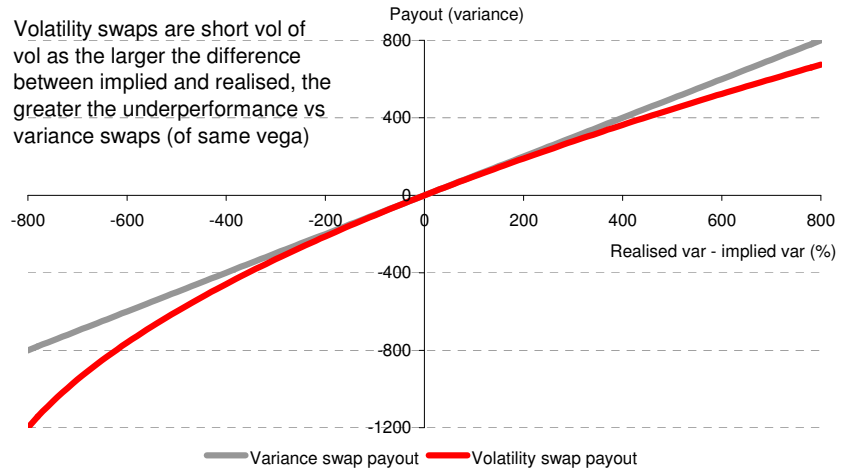
[6] The inclusion of OTM (and hence convex) options mean the log contract is also long volga (or vega convexity), but they are not the same thing. Long OTM (wing) options is long vega convexity, but not volatility surface curvature (unless they are shorting the ATM or body at the same time). The curvature of the volatility surface can be defined as the difference between 90-100 skew and 100-110 skew (ie, the value of 90% + 110% – 2×100% implied volatilities).

VOLATILITY SWAPS CAN BE HEDGED WITH VARIANCE SWAPS

Unlike variance swaps (or gamma swaps), volatility swaps cannot be hedged by a static portfolio of options. Volatility swaps can be hedged with variance swaps as, for small moves, the payout can be similar (see Figure 4 below). The vega of a variance swap is equal to variance notional×2σ. For example, for σ=25 the vega is 2×25 = 50 times the size of the variance swap notional. So, a volatility swap of vega 'V' can be hedged with V/2σ variance notional of a variance swap. As a variance swap is normally quoted in vega, the vega / 2σ formula is used to calculate the variance notional of the trade.

Variance notional = Vega / (2σ)

Figure 33. Payout of Variance Swap and Volatility Swap



Source: Santander Investment Bolsa.

VOLATILITY SWAPS ARE SHORT VOL OF VOL

As vega of variance swap is 1/2σ, this shows that volatility swaps (which have constant vega) are short vol of vol

The graph above shows that the payout of a variance swap is always in excess of the payout of a volatility swap of the same vega. This is why the fair level of a variance swap is usually one or two volatility points above volatility swaps. The negative convexity of the payout (compared to a variance swap) shows that volatility swaps are short vol of vol.

A volatility swap being short vol of vol can also be shown by the fact the identical vega of a variance swap has to be weighted 1/(2σ). If a trader is long a volatility swap and has hedged with a short variance swap position weighted 1/(2σ), then as volatility decreases more variance swaps have to be sold (as σ decreases, 1/(2σ) rises). Conversely, as volatility rises, variance swaps have to be bought (to decrease the short). Having to sell when volatility declines and buy when it rises shows that volatility swaps are short vol of vol.

Difference between variance and volatility swap prices can be approximated

As vol of vol and maturity are not both large at the same time, the approximation is a good one

Given that the difference between variance and volatility swap prices is due to the fact volatility swaps are short vol of vol, it is possible to derive the formula below, which approximates the difference between variance swap and volatility swap prices (as long as the maturity and vol of vol are not both excessive, which tends not to happen as longer maturities have less vol of vol). Using the formula, the price of a volatility swap can be approximated by the price of a variance swap less the convexity adjustment *c*. Using this formula, the difference between variance and volatility swaps is graphed in Figure 34.

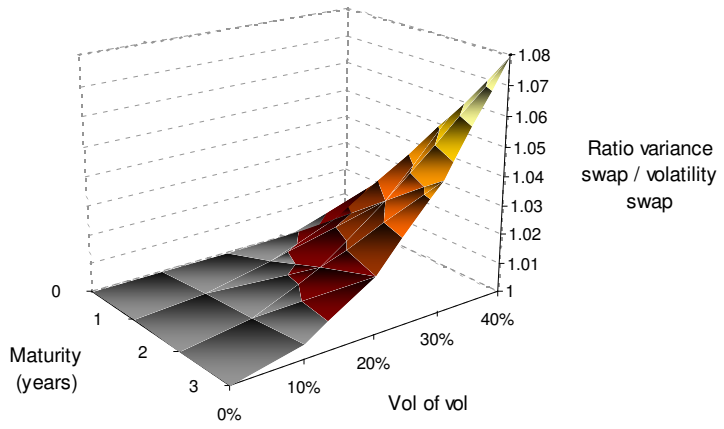
$$c \approx \frac{1}{6}\omega^2 T\sqrt{\text{variance swap price} \times e^{rT}}$$

where:

v = variance swap price

ω = volatility of volatility

Figure 34. Difference between Variance and Volatility Swap Prices



Source: Santander Investment Bolsa.

Model risk of vol of vol is small vs tail risk of variance swap

Hedging vol of vol raises trading costs and introduces model risk. Since the credit crunch, however, single-stock variance no longer trades and dispersion is now quoted using volatility swaps instead. Investment banks are happier taking the small model risk of vol of vol rather than being short the tail risk of a variance swap. As can be seen in Figure 35 below, variance swaps trade one or two volatility points above volatility swaps (for the most popular maturities). A simpler rule of thumb is that volatility swaps trade roughly in line with ATMf implied volatilities.

Figure 35. Typical Values of Vol of Vol and the Effect on Variance and Volatility Swap Pricing

Maturity	3 Month	6 Month	1 Year	2 Year
Vol of vol	85%	70%	55%	40%
Ratio var/vol	1.030	1.041	1.050	1.053
Difference var - vol (for 30% vol)	0.90	1.23	1.51	1.60

Source: Santander Investment Bolsa.

Max loss of variance swap = swap level × vega / 2

The notional of a variance swap trade is vega / $2\sigma_S$ (σ_S is traded variance swap level) and the payoff is (realised^2 - σ_S^2) × Notional. The maximum loss of a variance swap is when realised variance is zero, when the loss is σ_S^2 × Notional = σ_S^2 × vega / $2\sigma_S$ = σ_S × vega / 2.

GREEKS OF VOLATILITY, VARIANCE AND GAMMA SWAPS

'Shadow delta' caused by correlation between spot and implied volatility

As a volatility swap needs a vol of vol model, the Greeks are dependent on the model used. For variance swaps and gamma swaps, there is no debate as to the Greeks. However, practical considerations can introduce 'shadow Greeks'. In theory, a variance swap has zero delta, but in practice it has a small 'shadow delta' due to the correlation between spot and implied volatility (skew). Similarly, theta is not necessarily as constant as it should be in theory, as movements of the volatility surface can cause it to change.

Variance swap vega decays linearly with time

As variance is additive, the vega decays linearly with time. For example, 100K vega of a one year variance swap at inception will have 75K vega after three months, 50K after six months and 25K after nine months.

Variance swaps offer constant cash gamma, gamma swaps have constant share gamma

Share gamma is the number of shares that need to be bought (or sold) for a given change in spot (typically 1%). It is proportional to the Black-Scholes gamma (second derivative of price with respect to spot) multiplied by spot. Cash gamma (or dollar gamma) is the cash amount that needs to be bought or sold for a given movement in spot; hence, it is proportional to share gamma multiplied by spot (ie, proportional to Black-Scholes gamma multiplied by spot squared). Variance swaps offer a constant cash gamma (constant convexity), whereas gamma swaps offer constant share gamma (hence the name gamma swaps).

$\gamma \times S / 100$ = share gamma = number of shares bought (or sold) per 1% spot move

$\gamma \times S^2 / 100$ = cash (or dollar) gamma = notional cash value bought (or sold) per 1% spot move

OPTIONS ON VARIANCE

As the liquidity of the variance swap market improved in the middle of the last decade, market participants started to trade options on variance. As volatility is more volatile at high levels, the skew is positive (the inverse of the negative skew seen in the equity market). In addition, volatility term structure is inverted, as volatility mean reverts and does not stay elevated for long periods of time.

OPTIONS ON VARIANCE EXPIRY = EXPIRY OF UNDERLYING VAR SWAP

The price of an option on variance swap is quoted in volatility points (just like variance swaps)

An option on variance is a European option (like all exotics) on a variance swap whose expiry is the same expiry as the option. As it is an option on variance, a volatility of volatility model is needed in order to price the option. At inception, the underlying is 100% implied variance, whereas at maturity the underlying is 100% realised variance (and in between it will be a blend of the two). As the daily variance of the underlying is locked in every day, the payoff could be considered to be similar to an Asian (averaging) option.

Options on variance are quoted in volatility points

Like a variance swap, the price of an option on variance is quoted in volatility points. The typical 3-month to 18-month maturity of the option is in line with the length of time it takes 3-month realised volatility to mean revert after a crisis. The poor liquidity of options on variance, and the fact the underlying tends towards a cash basket over time, means a trade is usually held until expiry.

Option on variance swap payoff

$\text{Max}(\sigma_F^2 - \sigma_K^2, 0) \times$ Variance notional

where:

σ_F = future volatility (that occurs over the life of contract)

σ_K = strike volatility (fixed at the start of contract)

Variance notional = notional amount paid (or received) per variance point

NB: Variance notional = Vega / $(2\sigma_S)$ where σ_S = variance swap reference (current fair price of variance swap, not the strike)

PUT CALL PARITY APPLIES TO OPTIONS ON VARIANCE

As variance swaps have a convex volatility payout, so do options on variance. As options on variance are European, put call parity applies. The fact a long call on variance and short put on variance (of the same strike) is equal to a forward on variance (or variance swap) gives the following result for options on variance whose strike is not the current level of variance swaps.

$\text{Call Premium}_{\text{variance points}} - \text{Put Premium}_{\text{variance points}} = \text{PV}(\text{Current Variance Price}^2 - \text{Strike}^2)$

where:

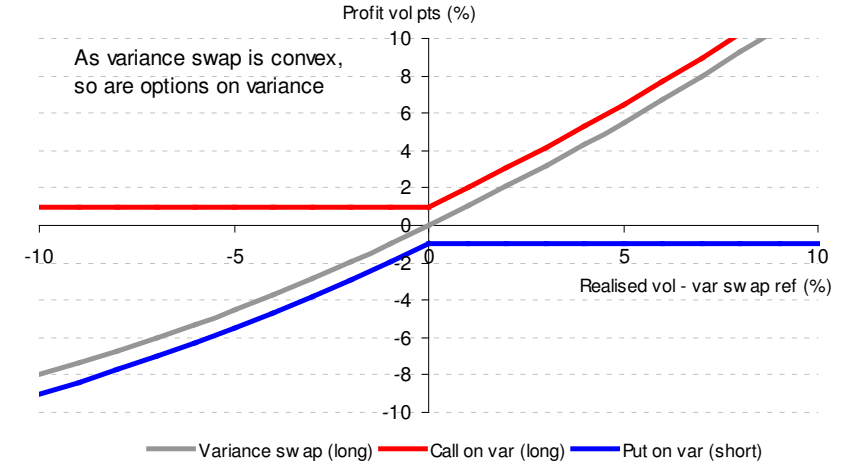
$\text{Premium}_{\text{variance points}} = 2\sigma_S \times \text{Premium}_{\text{volatility points}}$ where σ_S = variance swap reference

PREMIUM PAID FOR OPTION = VEGA × PREMIUM IN VOL POINTS

The premium paid for the option can either be expressed in terms of vega, or variance notional. Both are shown below:

Fixed leg cash flow = Variance notional × $\text{Premium}_{\text{variance points}}$ = Vega × $\text{Premium}_{\text{volatility points}}$

Figure 36. Variance Swap, ATM Call on Variance and ATM Put on Variance



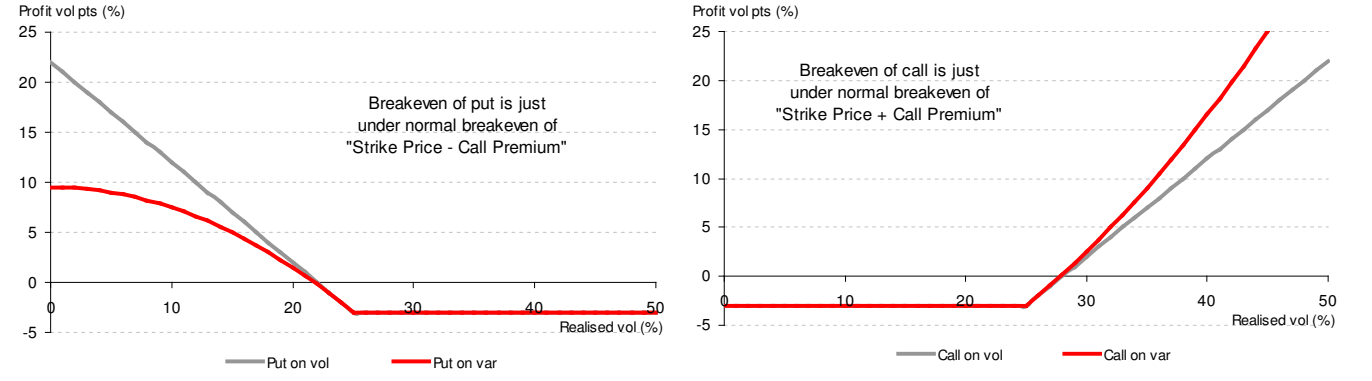
Source: Santander Investment Bolsa.

CONVEX PAYOUT MEANS BREAKEVENS ARE NON-TRIVIAL

Breakeven of options on variance is slightly below normal breakeven

The convexity of a variance swap means that a put on a variance swap has a lower payout than a put on volatility and a call on variance swap has a higher payout than a call on volatility (see Figure 37). Similarly, it also means the maximum payout of a put on variance is significantly less than the strike. This convexity also means the breakevens for option on variance are slightly different from the breakevens for option on volatility (strike – premium for puts, strike + premium for calls).

Figure 37. Put on Variance Swap **Call on Variance Swap**



Source: Santander Investment Bolsa estimates.

Breakevens are similar but not identical to options on volatility

In order to calculate the exact breakevens, the premium paid (premium P in vol points × Vega) must equal the payout of the variance swap.

Premium paid = payout of variance swap

For call option on variance: $\mathrm{P} \times \mathrm{Vega} = (\sigma_{\text{Call Breakeven}}{}^2 - \sigma_K{}^2) \times \dfrac{\mathrm{Vega}}{2\sigma_S}$

⇨ $\sigma_{\text{Call Breakeven}} = \sqrt{\sigma_K{}^2 + 2\sigma_S \mathrm{P}} \leq \sigma_K + \mathrm{P}$ = Call on volatility breakeven

Similarly $\sigma_{\text{Put Breakeven}} = \sqrt{\sigma_K{}^2 - 2\sigma_S \mathrm{P}} \leq \sigma_K - \mathrm{P}$ = Put on volatility breakeven

OPTIONS ON VARIANCE HAVE POSITIVE SKEW

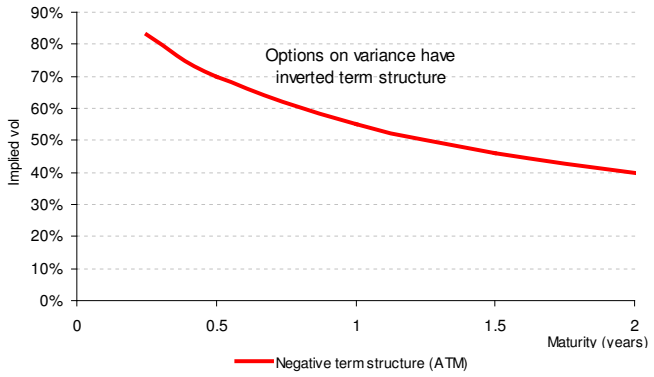
Volatility (and hence variance) is relatively stable when it is low, as calm markets tend to have low and stable volatility. Conversely, volatility is more unstable when it is high (as turbulent markets could get worse with higher volatility, or recover with lower volatility). For this reason, options on variance have positive skew, with high strikes having higher implied volatility than low strikes.

Implied variance term structure is inverted, but not as inverted as realised variance

Skew and term structure of options on variance are opposite to options on equity

As historical volatility tends to mean revert in an eight-month time horizon (on average), the term structure of options on variance is inverted (while volatility can spike and be high for short periods of time, over the long term it trades in a far narrower range). We note that, as the highest volatility occurs due to unexpected events, the peak of implied volatility (which is based on the market's expected future volatility) is lower than the peak of realised volatility. Hence, the volatility of implied variance is lower than the volatility of realised variance, especially for short maturities.

Figure 38. Option on Variance Term Structure



Option on Variance Skew

Source: Santander Investment Bolsa estimates.

CAPPED VARIANCE SWAPS HAVE EMBEDDED OPTION ON VAR

While options on variance swaps are not particularly liquid, their pricing is key for valuing variance swaps with a cap. Capped variance swaps are standard for single stocks and emerging market indices and can be traded on regular indices as well. When the variance swap market initially became more liquid, some participants did not properly model the cap, as it was seen to have little value. The advent of the credit crunch and resulting rise in volatility made the caps more valuable, and now market participants fail to model them at their peril.

Variance Swap with Cap C = Variance Swap - Option on Variance with Cap C

⇨ Option on Variance with Cap C = Variance Swap - Variance Swap with Cap C

While value of cap is small at inception, it can become more valuable as market moves

A capped variance swap can be modelled as a vanilla variance swap less an option on variance, whose strike is the cap. This is true as the value of an option on variance at the cap will be equal to the difference between the capped and uncapped variance swaps. Typically, the cap is at 2.5× the strike and, hence, is not particularly valuable at inception. However, as the market moves, the cap can become closer to the money and more valuable.

OPTIONS ON VAR STRATEGIES ARE SIMILAR TO VANILLA OPTIONS

Options on variance can be used as relative value trade vs VIX options or volatility swaps

Strategies that are useful for vanilla options have a read-across for options on variance. For example, a long variance position can be protected or overwritten. The increased liquidity of VIX options allows relative value trades to be put on.

Selling straddles on options on variance can also be a popular strategy, as volatility can be seen to have a floor above zero. Hence, strikes can be chosen so that the lower breakeven is in line with the perceived floor to volatility.

Options on variance can also be used to hedge a volatility swap position, as an option on variance can offset the vol of vol risk embedded in a volatility swap.

CORRELATION TRADING

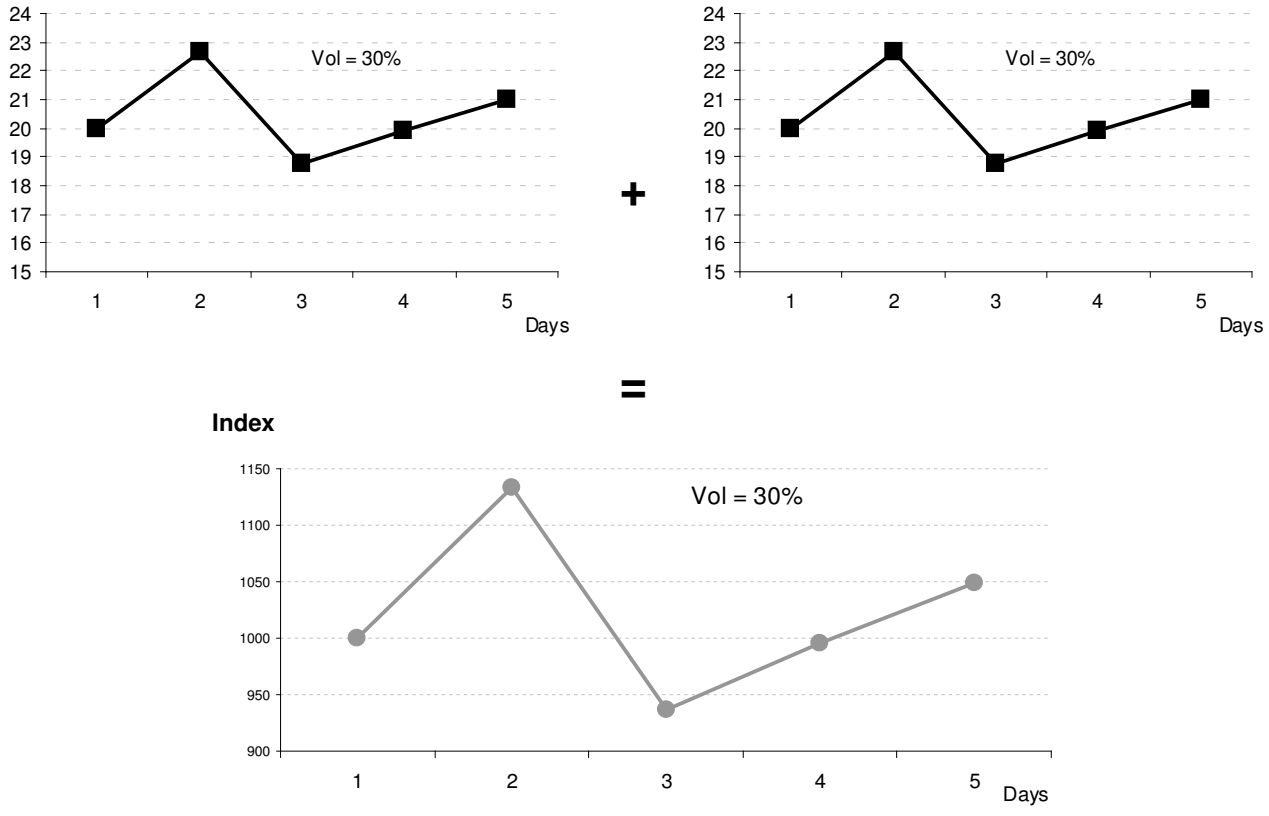
The volatility of an index is capped at the weighted average volatility of its constituents. Due to diversification (or less than 100% correlation), the volatility of indices tends to trade significantly less than its constituents. The flow from both institutions and structured products tends to put upward pressure on implied correlation, making index implied volatility expensive. Hedge funds and proprietary trading desks try to profit from this anomaly by either selling correlation swaps, or through dispersion trading (going short index implied volatility and long single stock implied volatility). Selling correlation became an unpopular strategy following losses during the credit crunch, but demand is now recovering.

INDEX IMPLIED LESS THAN SINGLE STOCKS DUE TO DIVERSIFICATION

Index will only be as volatile as its members if they are 100% correlated

The volatility of an index is capped by the weighted average volatility of its members. In order to show this we shall construct a simple index of two equal weighted members who have the same volatility. If the two members are 100% correlated with each other, then the volatility of the index is equal to the volatility of the members (as they have the same volatility and weight, this is the same as the weighted average volatility of the constituents).

Figure 39. Stock 1 **Stock 2 (100% correlation to stock 1)**

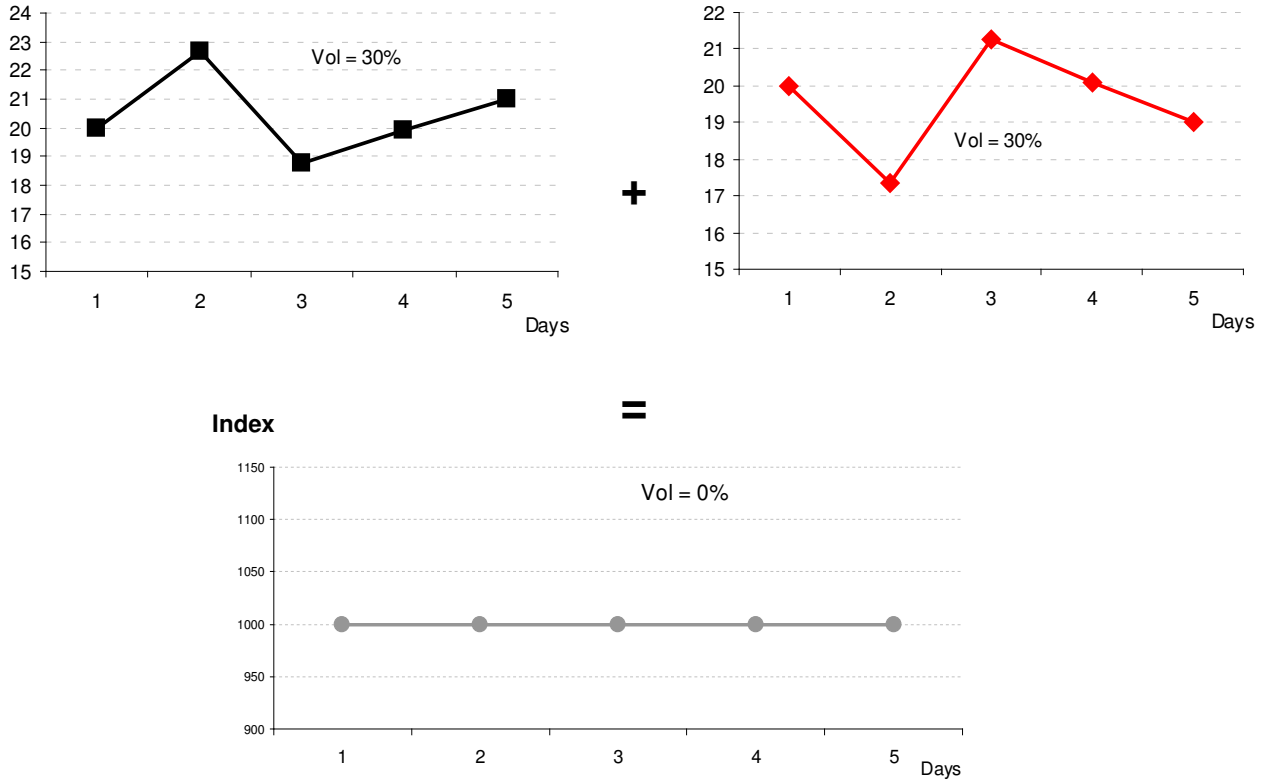


Source: Company data and Santander Investment Bolsa estimates.

Volatility of index has floor at zero when there is very low correlation

If we take a second example of two equal weighted index members with the same volatility, but with a negative 100% correlation (ie, as low as possible), then the index is a straight line with zero volatility.

Figure 40. Stock 1 **Stock 2 (-100% correlation to stock 1)**



Source: Company data and Santander Investment Bolsa estimates.

Index volatility is bounded by zero and weighted average single stock volatility

An implied volatility surface can be calculated from index and single-stock volatility surfaces

While the simple examples above have an index with only two members, results for a bigger index are identical. Therefore, the equation below is true. While we are currently examining historical volatility, the same analysis can be applied to implied volatility. In this way, we can get an implied correlation surface from the implied volatility surfaces of an index and its single-stock members. However, it is usually easiest to look at variance swap levels rather than implied volatility to remove any strike dependency.

$$0 \le \sigma_I^{\,2} \le \sum_{i=1}^{n} w_i^{\,2} \sigma_i^{\,2}$$

where

σ_I = index volatility

σ_i = single stock volatility (of i^{th} member of index)

w_i = single stock weight in index (of i^{th} member of index)

n = number of members of index

CORRELATION OF INDEX CAN BE ESTIMATED FROM VARIANCE

Under reasonable conditions implied correlation is equal to index variance divided by single-stock variance

If the correlation of all the different members of an index is assumed to be identical (a heroic assumption, but a necessary one if we want to have a single measure of correlation), the correlation implied by index and single-stock implied volatility can be estimated as the variance of the index divided by the weighted average single-stock variance. This measure is a point or two higher than the actual implied correlation but is still a reasonable approximation.

$$\rho_{imp} = \frac{\sigma_I^{\,2}}{\sum_{i=1}^{n} w_i \sigma_i^{\,2}}$$

where

ρ_{imp} = implied correlation (assumed to be identical between all index members)

Proof implied correlation can be estimated by index variance divided by single stock variance

The formula for calculating the index volatility from the members of the index is given below.

$$\sigma_I^{\,2} = \sum_{i=1}^{n} w_i^{\,2} \sigma_i^{\,2} + \sum_{i=1,\, j \neq i}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij}$$

where

ρ_{ij} = correlation between single stock i and single stock j

If we assume the correlations between each stock are identical, then this correlation can be implied from the index and single stock volatilities.

$$\rho_{imp} = \frac{\sigma_I^{\,2} - \sum_{i=1}^{n} w_i^{\,2} \sigma_i^{\,2}}{\sum_{i=1,\, j \neq i}^{n} w_i w_j \sigma_i \sigma_j}$$

Assuming reasonable conditions (correlation above 15%, c20 members or more, reasonable weights and implied volatilities), this can be rewritten as the formula below.

$$\rho_{imp} = \frac{\sigma_I^{\,2}}{\left(\sum_{i=1}^{n} w_i \sigma_i\right)^2}$$

This can be approximated by the index variance divided by the weighted average single-stock variance.

$$\rho_{imp} \approx \frac{\sigma_I^{\,2}}{\sum_{i=1}^{n} w_i \sigma_i^{\,2}}$$ eg, if index variance=20% and members average variance=25%, $\rho \approx 64\%$.

This approximation is slightly too high (c2pts) due to Jensen's inequality (shown below).

$$\left(\sum_{i=1}^{n} w_i \sigma_i\right)^2 \leq \sum_{i=1}^{n} w_i \sigma_i^{\,2}$$

STRUCTURED PRODUCTS LIFT IMPLIED CORRELATION

2 correlation points are equivalent to 0.3-0.5 volatility points

Using correlation to visually cheapen payouts through worst of/best of options is common practice for structured products. Similarly, the sale of structured products, such as Altiplano (which receives a coupon provided none of the assets in the basket has fallen), Everest (payoff on the worst performing) and Himalayas (performance of best share of index), leave their vendors short implied correlation. This buying pressure tends to lift implied correlation above fair value. We estimate that the correlation exposure of investment banks totals c€200mn per percentage point of correlation. The above formulae can show that two correlation points is equivalent to 0.3 to 0.5 (single-stock) volatility points. Similarly, the fact that institutional investors tend to call overwrite on single stocks but buy protection on an index also leads to buying pressure on implied correlation. The different methods of trading correlation are shown below.

Correlation is normally traded through dispersion

- **Dispersion trading.** Going short index implied volatility and going long single-stock implied volatility is known as a dispersion trade. As a dispersion trade is short Volga, or vol of vol, the implied correlation sold should be c10pts higher value than for a correlation swap. A dispersion trade was historically put on using variance swaps, but the large losses from being short single stock variance led to the single stock market becoming extinct. Now dispersion is either put on using straddles, or volatility swaps. Straddles benefit from the tighter bid-offer spreads of ATM options (variance swaps need to trade a strip of options of every strike). Using straddles does imply greater maintenance of positions, but some firms offer delta hedging for 5-10bp. A disadvantage of using straddles is that returns are path dependent. For example, if half the stocks move up and half move down, then the long single stocks are away from their strike and the short index straddle is ATM.

- **Correlation swaps.** A correlation swap is simply a swap between the (normally equal weighted) average pairwise correlation of all members of an index and a fixed amount determined at inception. Market value-weighted correlation swaps are c5 correlation points above equal weighted correlation, as larger companies are typically more correlated than smaller companies. While using correlation swaps to trade dispersion is very simple, the relative lack of liquidity of the product is a disadvantage. We note the levels of correlation sold are typically c5pts above realised correlation.

- **Covariance swaps.** While correlation swaps are relatively intuitive and are very similar to trading correlation via dispersion, the risk is not identical to the covariance risk of structured product sellers (from selling options on a basket). Covariance swaps were invented to better hedge the risk on structure books, and they pay out the correlation multiplied by the volatility of the two assets.

- **Basket options.** Basket options (or options on a basket) are similar to an option on an index, except the membership and weighting of the members does not change over time. The most popular basket option is a basket of two equal weighted members, usually indices.

- **Worst-of/best-of option.** The pricing of worst-of and best-of options has a correlation component. These products are discussed in the section *Worst-of/Best-of Options* in the *Forward Starting Products and Light Exotics* chapter.

- **Outperformance options.** Outperformance options pricing has as an input the correlation between the two assets. These products are also discussed in the section *Outperformance Options* in the *Forward Starting Products and Light Exotics* chapter.

Implied correlation of dispersion and level of correlation swap are not the same measure

Dispersion trades c10pts above realised correlation as short vol of vol

We note that the profit from theta-weighted (explained later in section) variance dispersion is roughly the difference between implied and realised correlation multiplied by the average single-stock volatility. As correlation is correlated to volatility, this means the payout when correlation is high is increased (as volatility is high) and the payout when correlation is low is decreased (as volatility is low). A short correlation position from going long dispersion (short index variance, long single-stock variance) will suffer from this as profits are less than expected and losses are greater. Dispersion is therefore short vol of vol; hence, implied correlation tends to trade c10 correlation points more than correlation swaps (which is c5 points above realised correlation). We note this does not necessarily mean a long dispersion trade should be profitable (as dispersion is short vol of vol, the fair price of implied correlation is above average realised correlation).

Implied vs realised correlation increases for low levels of correlation

For example, in normal market conditions the SX5E and S&P500 will have an implied correlation of 50-70 and a realised of 30-60. If realised correlation is 30, implied will tend to be at least 50 (as investors price in the fact correlation is unlikely to be that low for very long; hence, the trade has more downside than upside). The NKY tends to have correlation levels ten points below the SX5E and SPX.

Figure 41. Different Types of Correlation

Traded level (c5pts above payout)	Payout	Formula
Implied correlation (market value weighted)	Realised correlation (using realised vol rather than implied vol in formula)	$\frac{\sigma_I^{\,2} - \sum_{i=1}^{n} w_i^{\,2}\sigma_i^{\,2}}{\sum_{i=1,\, j\neq i}^{n} w_i w_j \sigma_i \sigma_j} \approx \frac{\sigma_I^{\,2}}{\sum_{i=1}^{n} w_i \sigma_i^{\,2}}$
c10pts above implied correlation (as not short volga)		
Correlation swap (market value weighted)	Pairwise realised correlation (market value weighted)	$\frac{\sum_{i=1,\, j>i}^{n} w_i w_j \rho_{ij}}{\sum_{i=1,\, j>i}^{n} w_i w_j}$
c5pts above market value weighted correlation		
Correlation swap (equal weighted)	Pairwise realised correlation (equal weighted)	$\frac{2}{n(n-1)}\rho_{ij}$

Source: Santander Investment Bolsa.

SIZE OF DISPERSION MARKET SHRANK AFTER THE CREDIT CRUNCH

Dispersion trading is less popular than it was

Selling correlation led to severe losses when the market collapsed in 2008, as implied correlation spiked to c90%, which led many investors to cut back exposures or leave the market. Similar events occurred in the market during the May 2010 correction. The amount of crossed vega has been reduced from up to €100mn at some firms to €5-20mn now (crossed vega is the amount of offsetting single-stock and index vega, ie, €10mn crossed vega is €10mn on single stock and €10mn on index). Similarly, the size of trades has declined from a peak of €2.5mn to €0.5mn vega now.

Figure 42. CBOE Implied Correlation Tickers and Expiries

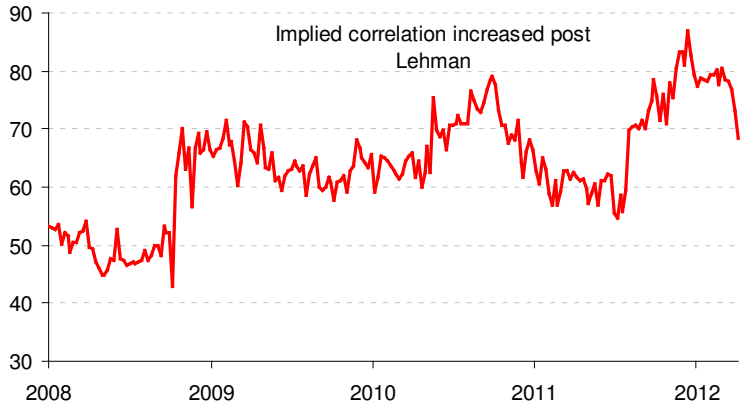
Expiry S&P500	Expiry Top 50 Stocks	Ticker	Start Date	End Date
Dec-09	Jan-10	ICJ	Nov-07	Nov-09
Dec-10	Jan-11	JCJ	Nov-08	Nov-10
Dec-11	Jan-12	KCJ	Nov-09	Nov-11
Dec-12	**Jan-13**	**ICJ**	**Nov-10**	**Nov-12**
Dec-13	**Jan-14**	**JCJ**	**Nov-11**	**Nov-13**
Dec-14	**Jan-15**	**KCJ**	**Nov-12**	**Nov-14**

Source: Santander Investment Bolsa.

CBOE INDICES ALLOW IMPLIED CORRELATION TO BE PLOTTED

There are now correlation indices for calculating the implied correlation of dispersion trades calculated by the CBOE. As there are 500 members of the S&P500, the CBOE calculation only takes the top 50 stocks (to ensure liquidity). There are three correlation indices tickers (ICJ, JCJ and KCJ), but only two correlation indices are calculated at any one time. On any date one correlation index has a maturity up to one year, and another has a maturity between one and two years. The calculation uses December expiry for S&P500 options, and the following January expiry for the top 50 members as this is the only listed expiry (US single stocks tend to be listed for the month after index triple witching expiries). The index is calculated until the previous November expiry, as the calculation tends to be very noisy for maturities only one month to December index expiry. On the November expiry, the one month maturity (to S&P500 expiry) index ceases calculation, and the previously dormant index starts calculation as a two-year (and one-month) maturity index. For the chart below, we use the longest dated available index.

Figure 43. CBOE Implied Correlation (rolling maturity between 1Y and 2Y)



Source: Santander Investment Bolsa.

CORRELATION SWAPS HAVE PURE CORRELATION EXPOSURE

Correlation swaps are normally equal weighted (not market value weighted)

Correlation swaps (which, like variance swaps, are called swaps but are actually forwards) simply have a clean payout of the (normally equal-weighted) correlation between every pair in the basket less the correlation strike at inception. Correlation swaps usually trade on a basket, not an index, to remove the names where a structured product has a particularly high correlation risk. Half of the underlyings are typically European, a third US and the final sixth Asian stocks. The product started trading in 2002 as a means for investment banks to reduce their short correlation exposure from their structured products books. While a weighted pairwise correlation would make most sense for a correlation swap on an index, the calculation is typically equal-weighted as it is normally on a basket.

Equal-weighted correlation is c5 correlation points below market value-weighted correlation

Market value-weighted correlation swaps tends to trade c5 correlation points above realised correlation (a more sophisticated methodology is below). This level is c10 correlation points below the implied correlation of dispersion (as dispersion payout suffers from being short volga). In addition, the correlation levels for equal-weighted correlations tends to be c5 correlation points lower than for market value-weighted, due to the greater weight allocated to smaller – and hence less correlated – stocks. The formula for the payout of a correlation swap is below.

Correlation swap payoff

$(\rho_K - \rho) \times$ Notional

where

Notional= notional paid (or received) per correlation point

ρ_K = strike of correlation swap (agreed at inception of trade)

$\rho = \frac{2}{n(n-1)} \rho_{ij}$ (equal weighted correlation swap)

$\rho = \frac{\sum_{i=1,\, j>i}^{n} w_i w_j \rho_{ij}}{\sum_{i=1,\, j>i}^{n} w_i w_j}$ (market value weighted correlation swap)

n = number of stocks in basket

Correlation swaps tend to trade c5 correlation points above realised

A useful rule of thumb for the level of a correlation swap is that it trades c5 correlation points above realised correlation (either equal-weighted or market value-weighted, depending on the type of correlation swap). However, for very high or very low values of correlation, this formula makes less sense. Empirically, smaller correlations are typically more volatile than higher correlations. Therefore, it makes sense to bump the current realised correlation by a larger amount for small correlations than for higher correlations (correlation swaps should trade above realised due to demand from structured products). The bump should also tend to zero as correlation tends to zero, as having a correlation swap above 100% would result in arbitrage (can sell correlation swap above 100% as max correlation is 100%). Hence, a more accurate rule of thumb (for very high and low correlations) is given by the formula below.

Correlation swap level = $\rho + \alpha (1 - \rho)$

where

ρ = realised correlation

α = bump factor (typical $\alpha = 0.1$)

Maturity of correlation swap is typically between one and three years

Structured products typically have a maturity of 5+ years; however, many investors close their positions before expiry. The fact that a product can also delete a member within the lifetime of the product has led dealers to concentrate on the three-year maturity rather than five-year. As the time horizon of hedge funds is short dated, correlation swaps typically trade between one and three years. The size is usually between €250k and €1,000k.

Correlation swaps suffer from lack of liquidity

The market for correlation swaps has always been smaller than for dispersion. Whereas the variance swap or option market has other market participants who ensure liquidity and market visibility, the investor base for correlation swaps is far smaller. This can be an issue should a position wish to be closed before expiry. It can also cause mark-to-market problems. The correlation swap market grew from 2002 onwards until the credit crunch, when investor appetite for exotic products disappeared. At its peak, it is estimated that some structured derivative houses shed up to c10% of their short correlation risk to hedge funds using correlation swaps.

DISPERSION IS THE MOST POPULAR METHOD OF TRADING CORRELATION

Implied correlation of dispersion trades are short volga (vol of vol)

As the levels of implied correlation are usually overpriced (a side effect of the short correlation position of structured product sellers), index implied volatility is expensive when compared with the implied volatility of single stocks. A long dispersion trade attempts to profit from this by selling index implied and going long single-stock implied[7]. Such a long dispersion trade is short implied correlation. While dispersion is the most common method of trading implied correlation, the payoff is also dependent on the level of volatility. The payout of (theta-weighted) dispersion is shown below. Because of this, and because correlation is correlated to volatility, dispersion trading is short vol of vol (volga).

$$\mathrm{P\,\&\,L}_{\mathrm{dispersion}} = \sum_{i=1}^{n} w_i \sigma_i^{\,2} (\rho_{imp} - \rho)$$

There are four instruments that can be used to trade dispersion:

- **Straddle (or call) dispersion.** Using ATM straddles to trade dispersion is the most liquid and transparent way of trading. Because it uses options, the simplest and most liquid volatility instrument, the pricing is usually the most competitive. Trading 90% strike rather than ATM allows higher levels of implied correlation to be sold. Using options is very labour intensive, however, as the position has to be delta-hedged (some firms offer delta hedging for 5-10bp). In addition, the changing vega of the positions needs to be monitored, as the risks are high given the large number of options that have to be traded. In a worst-case scenario, an investor could be right about the correlation position but suffer a loss from lack of vega monitoring. We believe that using OTM strangles rather than straddles is a better method of using vanilla options to trade dispersion as OTM strangles have a flatter vega profile. This means that spot moving away from strike is less of an issue, but we acknowledge that this is a less practical way of trading.
- **Variance swap dispersion.** Because of the overhead of developing risk management and trading infrastructure for straddle dispersion, many hedge funds preferred to use variance swaps to trade dispersion. With variance dispersion it is easier to see the profits (or losses) from trading correlation than it is for straddles. Variance dispersion suffers from the disadvantage that not all the members of an index will have a liquid variance swap market. Since 2008, the single-stock variance market has disappeared due to the large losses suffered from single-stock variance sellers (as dispersion traders want to go long single-stock variance, trading desks were predominantly short single-stock variance). It is now rare to be able to trade dispersion through variance swaps.
- **Volatility swap dispersion.** Since liquidity disappeared from the single-stock variance market, investment banks have started to offer volatility swap dispersion as an alternative. Excluding dispersion trades, volatility swaps rarely trade.
- **Gamma swap dispersion.** Trading dispersion via gamma swaps is the only 'fire and forget' way of trading dispersion. As a member of an index declines, the impact on the index volatility declines. As a gamma swap weights the variance payout on each day by the closing price on that day, the payout of a gamma swap similarly declines with spot. For all other dispersion trades, the volatility exposure has to be reduced for stocks that decline and increased for stocks that rise. Despite the efforts of some investment banks, gamma swaps never gained significant popularity.

[7] Less liquid members of an index are often excluded, eg, CRH for the Euro STOXX 50 is usually excluded.

Need to decide on weighting scheme for dispersion trades

Vega-weighted assumes parallel move, theta-weighted assumes relative move

While a dispersion trade always involves a short index volatility position and a long single-stock volatility position, there are different strategies for calculating the ratio of the two trade legs. If we assume index implied is initially 20%, if it increases to 30% the market could be considered to have risen by ten volatility points or risen by 50%. If the market is considered to rise by ten volatility points and average single-stock implied is 25%, it would be expected to rise to 50% (vega-weighted). If the market is considered to rise by 50% and average single-stock implied is 30%, it would be expected to rise to 45% (theta- or correlation-weighted). The third weighting, gamma-weighted, is not often used in practice.

- **Vega-weighted.** In a vega-weighted dispersion, the index vega is equal to the sum of the single-stock vega. If both index and single-stock vega rise one volatility point, the two legs cancel and the trade neither suffers a loss or reveals a profit.
- **Theta- (or correlation-) weighted.** Theta weighting means the vega multiplied by √variance (or volatility for volatility swaps) is equal on both legs. This means there is a smaller single-stock vega leg than for vega weighting (as single-stock volatility is larger than index volatility, so it must have a smaller vega for vega × volatility to be equal). Under theta-weighted dispersion, if all securities have zero volatility, the theta of both the long and short legs cancels (and total theta is therefore zero). Theta weighting can be thought of as correlation-weighted (as correlation ≈ index var / average single stock var = ratio of single-stock vega to index vega). If volatility rises 1% (relative move) the two legs cancel and the dispersion breaks even.
- **Gamma-weighted.** Gamma weighting is the least common of the three types of dispersion. As gamma is proportional to vega/vol, then the vega/vol of both legs must be equal. As single-stock vol is larger than index vol, there is a larger single-stock vega leg than for vega-weighted.

Greeks of dispersion trading depend on weighting used

The Greeks of a dispersion trade[8] are very much dependent on the vega weighting of the two legs. The easiest weighting to understand is a vega-weighted dispersion, which by definition has zero vega (as the vega of the short index and long single-stock legs are identical). A vega-weighted dispersion is, however, short gamma and short theta (ie, have to pay theta).

Theta-weighted dispersion needs a smaller long single-stock leg than the index leg (as reducing the long position reduces theta paid on the long single-stock leg to that of the theta earned on the short index leg). As the long single-stock leg is smaller, a theta-weighted dispersion is very short gamma (as it has less gamma than vega-weighted, and vega-weighted is short gamma).

Gamma-weighted dispersion needs a larger long single-stock leg than the index leg (as increasing the long position increases the gamma to that of the short index gamma). As the long single-stock leg is larger, the theta paid is higher than that for vega-weighted.

Figure 44. Greeks of Dispersion Trades with Different Weightings

Greeks	Theta-Weighted	Vega-Weighted	Gamma-Weighted
Theta	0	Short/pay	Very short/pay a lot
Vega	Short	0	Long
Gamma	Very short	Short	0
Total single-stock vega	**Less than index**	**Equal to index**	**More than index**

Source: Santander Investment Bolsa.

[8] The mathematical proof of the Greeks is outside of the scope of this report.

Figure 45. Breakevens for Theta-Weighted, Vega-Weighted and Gamma-Weighted Dispersion

	Theta-Weighted	Vega-Weighted	Gamma-Weighted
Start of trade			
Index vol (vol pts)	20.0	20.0	20.0
Average single-stock vol (vol pts)	25.0	25.0	25.0
Implied correlation (correlation pts)	64.0	64.0	64.0
Trade size			
Index vega (k)	100	100	100
Single-stock vega (k)	80	100	125
End of trade (if P&L = 0, ie, breaks even)			
Index vol (vol pts)	30.0	30.0	30.0
Avg single-stock vol (for trade to break even) (vol pts)	37.5	35.0	33.0
Implied correlation (correlation pts)	64.0	73.5	82.6
Change			
Change in index vol (vol. pts)	10.0	10.0	10.0
Change in single-stock vol (vol pts)	12.5	10.0	8.0
Change in implied correlation (correlation pts)	0.0	9.5	18.6
Change (%)			
Change in index vol (%)	50%	50%	50%
Change in single-stock vol (%)	50%	40%	32%
Change in implied correlation (%)	**0.0%**	14.8%	29.1%

Source: Santander Investment Bolsa.

THETA, VEGA AND GAMMA-WEIGHTED DISPERSION EXAMPLES

Only theta-weighted dispersion gives pure correlation exposure

Figure 45 above shows the different weightings for theta-, vega- and gamma-weighted dispersion. The change in volatility for the different trades to break even is shown. As can be seen, only theta-weighted dispersion gives correlation exposure (ie, if realised correlation is equal to implied correlation, theta-weighted dispersion breaks even).

Theta-weighted dispersion is best weighting for almost pure correlation exposure

The sole factor that determines if theta-weighted dispersion makes a profit or loss is the difference between realised and implied correlation. For timing entry points for theta-weighted dispersion, we believe investors should look at the implied correlation of an index (as theta-weighted dispersion returns are driven by correlation). Note that theta-weighted dispersion breaks even if single stock and index implied moves by the same percentage amount (eg, index vol of 20%, single-stock vol of 25% and both rise 50% to 30% and 37.5%, respectively).

Vega-weighted dispersion gives hedged exposure to mispricing of correlation

Vega-weighted dispersion can be thought of as giving better exposure to correlation mispricing

When a dispersion trade is vega-weighted, it can be thought of as being the sum of a theta-weighted dispersion (which gives correlation exposure), plus a long single-stock volatility position. This volatility exposure can be thought of as a hedge against the short correlation position (as volatility and correlation are correlated); hence, a vega-weighted dispersion gives greater exposure to the mispricing of correlation. When looking at the optimal entry point for vega-weighted dispersion, it is better to look at the difference between average single-stock volatility and index volatility (as this applies an equal weight to both legs, like in a vega-weighted dispersion). Note that vega-weighted dispersion breaks even if single stock and index implied moves by the same absolute amount (eg, index vol of 20%, single-stock vol of 25% and both rise ten volatility points to 30% and 35%, respectively). Empirically, the difference between single-stock and index volatility (ie, vega-weighted dispersion) is not correlated to volatility[9], which supports our view of vega-weighted dispersion being the best.

[9] Single-stock leg is arguably 2%-5% too large; however, slightly over-hedging the implicit short volatility position of dispersion could be seen as an advantage.

Gamma-weighted dispersion is rare, and not recommended

While gamma weighting might appear mathematically to be a suitable weighting for dispersion, in practice it is rarely used. It seems difficult to justify a weighting scheme where more single-stock vega is bought than index (as single stocks have a higher implied than index and, hence, should move more). We include the details of this weighting scheme for completeness, but do not recommend it.

DISPERSION TRADES ARE SHORT VOL OF VOL (VOLGA)

Implied correlation of dispersion trades should be c10 points above correlation swap as short volga

The P&L of a theta-weighted dispersion trade is proportional to the spread between implied and realized market value-weighted correlation (ρ), multiplied by a factor that corresponds to a weighted average variance of the components of the index[10].

$$\text{P \& L}_{\text{theta weighted dispersion}} = \sum_{i=1}^{n} w_i \sigma_i^{\,2} (\rho_{imp} - \rho)$$

where:

ρ = market value weighted correlation

The payout of a theta-weighted dispersion is therefore equal to the difference in implied and realised correlation (market value-weighted pairwise realised correlation) multiplied by the weighted average variance. If vol of vol was zero and volatility did not change, then the payout would be identical to a correlation swap and both should have the same correlation price. If volatility is assumed to be correlated to correlation (as it is, as both volatility and correlation increase in a downturn) and the correlation component is profitable, the profits are reduced (as it is multiplied by a lower volatility). Similarly, if the correlation suffers a loss, the losses are magnified (as it is multiplied by a higher volatility). Dispersion is therefore short volga (vol of vol) as the greater the change in volatility, the worse the payout. To compensate for this short volga position, the implied correlation level of dispersion is c10 correlation points above the level of correlation swaps.

[10] Proof of this result is outside the scope of this publication.

BASKET OPTIONS ARE MOST LIQUID CORRELATION PRODUCT

Basket options are different from an option on an index, as weights and membership do not change

The most common product for trading correlation is a basket option (otherwise known as an option on a basket). If the members of a basket are identical to the members of an index and have identical weights, then the basket option is virtually identical to an option on the index. The two are not completely identical, as the membership and weight of a basket option does not change[11], but it can for an index (due to membership changes, rights issues, etc). The formula for basket options is below.

Basket = $\sum_{i=1}^{n} w_i S_i$ where S_i is the i^{th} security in the basket

Basket call payoff at expiry = Max(0, $\sum_{i=1}^{n} w_i^2 S_i - K$) where K is the strike

BASKET OPTIONS ON TWO INDICES ARE THE MOST POPULAR

While the above formula can be used for all types of basket, the most popular is a basket on two equal weighted indices. In this case the correlation traded is not between multiple members of a basket (or index) but the correlation between only two indices. As the options usually wants the two indices to have identical value, it is easier to define the basket as the equal weighted sum of the two security returns (see the below formula setting n = 2). The previous formula could be used, but the weight w would not be 0.5 (would be 0.5 / $S_{\text{i at inception}}$).

Basket = $\sum_{i=1}^{n} w \frac{S_{\text{i at expiry}}}{S_{\text{i at inception}}}$ where S_i is the i^{th} security in the basket (and w normally = 1/n)

BASKET PAYOFF IS BASED ON COVARIANCE, NOT CORRELATION

The payout of basket options is based on the correlation multiplied by the volatility of the two securities, which is known as covariance. The formula for covariance is shown below. As basket options are typically the payout of structured products, it is better to hedge the exposure using products whose payout is also based on covariance. It is therefore better to use covariance swaps rather than correlation swaps or dispersion to offset structured product risk.

Covariance(A,B) = $\rho \sigma_A \sigma_B$ where ρ is the correlation between A and B

[11] Weighting for Rainbow options is specified at maturity based on the relative performance of the basket members, but discussion of these options is outside of the scope of this publication.

COVARIANCE SWAPS BETTER REPRESENT STRUCTURED PRODUCT RISK

Covariance swap liquidity never took off

The payout of structured products is often based on a basket option. The pricing of an option on a basket involves covariance, not correlation. If an investment bank sells an option on a basket to a customer and hedges through buying correlation (via correlation swaps or dispersion) there is a mismatch[12]. Because of this, attempts were made to create a covariance swap market, but liquidity never took off.

Correlation swap payoff

$[\text{Covariance}(A,B) - K_{\text{covariance}}] \times \text{Notional}$

where

Notional= notional paid (or received) per covariance point

ρ = correlation between A and B

σ_i = volatility of i

$\text{Covariance}(A,B) = \rho\sigma_A\sigma_B$ (note if A = B then covariance = variance as $\rho = 1$)

$K_{\text{covariance}}$ = strike of covariance swap (agreed at inception of trade)

[12] Results in being short cross-gamma. Cross-gamma is the effect a change in the value of one underlying has on the delta of another.

DIVIDEND VOLATILITY TRADING

If a constant dividend yield is assumed, then the volatility surface for options on realised dividends should be identical to the volatility surface for equities. However, as companies typically pay out less than 100% of earnings, they have the ability to reduce the volatility of dividend payments. In addition to lowering the volatility of dividends to between ½ and ⅔ of the volatility of equities, companies are reluctant to cut dividends. This means that skew is more negative than for equities, as any dividend cut is sizeable. Despite the fact that index dividend cuts have historically been smaller than the decline in the index, imbalances in the implied dividend market can cause implied dividends to decline more than spot.

DIVIDEND REALISED VOL IS LOWER THAN EQUITY REALISED VOL

If dividend yields were constant, a dividend volatility surface should be identical to an equity volatility surface

Dividend yields are often thought of as mean reverting, as they cannot rise to infinity nor go below zero. If the dividend yield is constant, then the dividend volatility surface will be identical to the equity volatility surface. However, dividends' volatility tends to be between ½ and ⅔ of the volatility of equities, depending on the time period chosen. This discrepancy is caused by two effects:

- **Dividend volatility suppressed by less than 100% payout ratio.** Companies typically pay out less than 100% of earnings in order to grow the company. As corporates are normally reluctant to cut dividends, they will simply increase the payout ratio in a downturn. This is done both to avoid the embarrassment of cutting a dividend, but also as in a downturn there are likely to be limited opportunities for growth and, hence, little need to reinvest earnings (typically costs and investment are cut in a downturn). If the economy is growing and earnings increasing significantly, then companies will normally increase dividends by less than the jump in earnings. This is done in case the favourable environment does not last or because there are attractive opportunities for investing the retained earnings.

- **Equity volatility is too high compared to fundamentals.** On balance, academic evidence suggests that equity volatility is too high compared to fundamentals such as dividend payouts[13]. Statistical arbitrage funds can normally be expected to eliminate any significant short-term imbalances. However, their investment time horizon is normally not long enough to attempt to reduce the discrepancy between equity and fundamental volatility on a multiple-year time horizon.

REALISED DIVIDENDS DECLINE LESS THAN EQUITIES

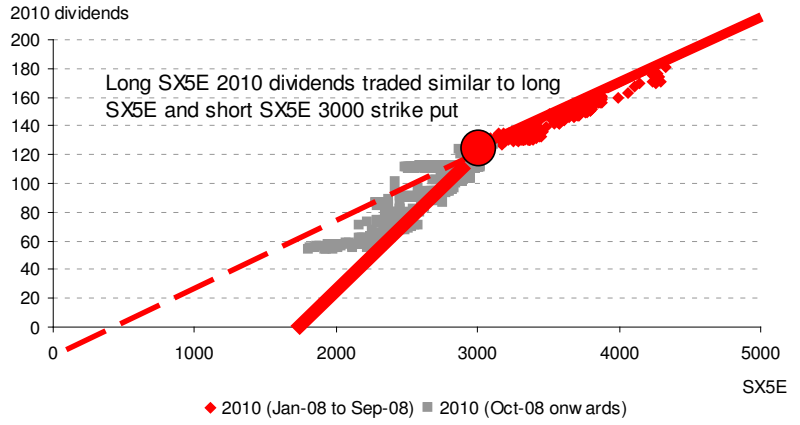
In 140 years of US data, there has never been a larger decline in index realised dividends than the index itself. This is because in a bear market certain sectors are affected more than others, and it is the companies in the worst affected sectors that cut dividends. For example, in the 2000-03 bear market, TMT was particularly affected. Similarly, the credit crunch has hit financials and real estate the hardest. As companies are loath to cut dividends, the remaining sectors tend to resist cutting dividends. This means that, while at a stock level dividend declines can be greater than equity declines (as dividends can be cut to zero while the equity price is above zero), at an index level realised dividends only experience an average decline of half to two-thirds of the equity market decline.

[13] An equity can be modeled as the NPV of future dividend payments.

IMPLIED DIVIDENDS CAN DECLINE MORE THAN EQUITIES

Before the credit crunch, some participants believed that dividends decline less than spot if spot falls, as corporates are reluctant to cut dividends. This is only true at the single-stock level and only for small declines. If a single stock falls by a significant amount, the company will cut dividends by a larger amount than the equity market decline (as dividends will be cut to zero before the stock price reaches zero). A long dividend position is similar to long stock and short put. This can be seen in Figure 46 below. While the diagram below would appear to imply a 'strike' of c3000 for the SX5E, this strike is very dependent on market sentiment and conditions. For severe equity market declines, implied dividends can decline twice as fast as spot. This disconnect between realised and implied dividends occurs when there is a large structured product market (markets such as the USA, which have few structured products, do not act in this way).

Figure 46. SX5E 2010 Dividends vs Spot



Source: Santander Investment Bolsa.

Structured products can cause dividend risk limits to be hit in a downturn

Typically, the sale of structured products not only causes their vendor to be long dividends, but this long position increases as equity markets fall. For example, autocallables are a particularly popular structured product as they give an attractive coupon until they are automatically called (which occurs when the equity market does not fall significantly). As the maturity effectively extends when markets decline (as the product is not called as expected), the vendor becomes long dividends up until the extended maturity. As all investment banks typically have the same position, there are usually few counterparties should a position have to be cut. This effect is most severe during a rapid downturn, as there is limited opportunity for investment banks to reduce their positions in an orderly manner.

IMPLIED DIVIDENDS ARE THE UNDERLYING OF OPTIONS ON DIVIDENDS

Implied dividend volatility, not realised dividend volatility, is the driver for pricing options on dividends

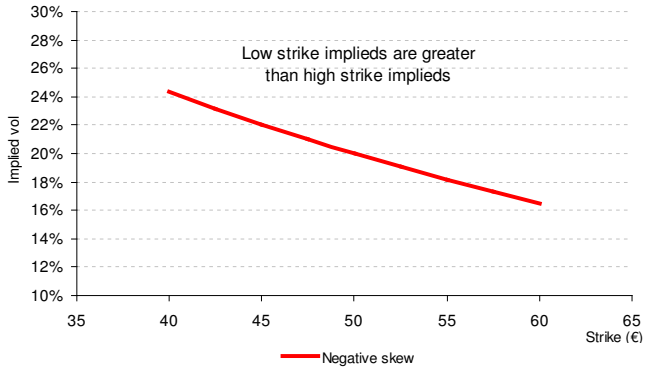
As it is not possible to hedge an option on dividend with realised dividend (they are not traded), the volatility of the underlying implied dividend is the key driver of an option on dividend's value. While realised dividends are less volatile than equities, implied dividends can be more volatile than equities. The volatility of implied dividends is also likely to be time dependent, with greater volatility during the reporting period when dividends are announced, and less volatility at other times (particularly for near dated implied dividends).

DIVIDENDS SHOULD HAVE HIGHER SKEW THAN EQUITIES

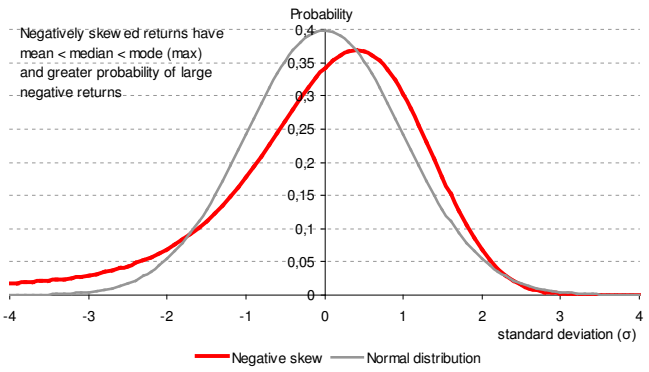
Skew can be measured as the third moment (return is the first moment, variance is the second moment). Equities have a negative skew, which means the volatility surface is downward sloping and the probability distribution has a larger downside tail. The mathematical definition of the third moment is below. Looking at annual US dividend payments over 140 years shows that skew is more negative for dividends than equities. This difference in skew narrows if the third moment for bi-annual periods or longer are examined, potentially as any dividend cuts companies make are swiftly reversed when the outlook improves.

$$\text{Skew} = \text{third moment} = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right]$$

Figure 47. Implied Volatility with Negative Skew



Probability Distribution of Negative Skew



Source: Santander Investment Bolsa estimates.

OPPORTUNITIES, IMBALANCES AND MYTHS

OVERPRICING OF VOL IS PARTLY AN ILLUSION

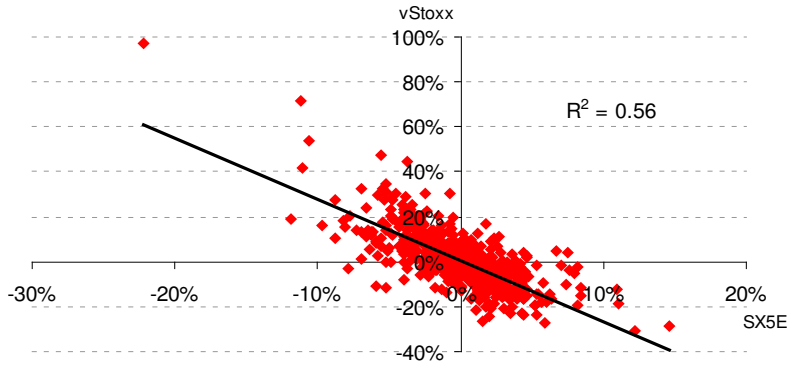
Selling implied volatility is one of the most popular trading strategies in equity derivatives. Empirical analysis shows that implied volatility or variance is, on average, overpriced. However, as volatility is negatively correlated to equity returns, a short volatility (or variance) position is implicitly long equity risk. As equity returns are expected to return an equity risk premium over the risk-free rate (which is used for derivative pricing), this implies short volatility should also be abnormally profitable. Therefore, part of the profits from short volatility strategies can be attributed to the fact equities are expected to deliver returns above the risk-free rate.

SHORT VOLATILITY IS POSITIVELY CORRELATED TO EQUITY RETURNS

As short volatility is positively correlated to equity returns, if equities earn more than the risk free rate so should short volatility

As implied volatility tends to trade at a higher level than realised volatility, a common perception is that implied volatility is overpriced. While there are supply and demand imbalances that can cause volatility to be overpriced, part of the overpricing is due to the correlation between volatility and equity returns. A short volatility position is positively correlated to the equity market (as volatility typically increases when equities decline). As equities' average return is greater than the risk-free rate, this means that the risk-neutral implied volatility should be expected to be above the true realised volatility. Even taking this into account, volatility appears to be overpriced. We believe that implied volatility is overpriced on average due to the demand for hedges.

Figure 48. Correlation vStoxx (volatility) and SX5E



Source: Santander Investment Bolsa.

Far-dated options are most overpriced, due to upward sloping volatility term structure

Volatility selling strategies typically involve selling near-dated volatility (or variance). Examples include call overwriting or selling near-dated variance (until the recent explosion of volatility, this was a popular hedge fund strategy that many structured products copied). As term structure is on average upward sloping, this implies that far-dated implieds are more expensive than near dated implieds. The demand for long-dated protection (eg, from variable annuity providers) offers a fundamental explanation for term structure being upward sloping (see the section *Variable Annuity Hedging Lifts Long-Term Vol*). However, as 12× one month options (or variance swaps) can be sold in the same period of time as 1× one-year option (or variance swap), greater profits can be earned from selling the near-dated product despite it being less overpriced. We note the risk is greater if several near-dated options (or variance swap) are sold in any period.

REASONS WHY VOLATILITY OVERPRICING IS UNLIKELY TO DISAPPEAR

There are several fundamental reasons why volatility, and variance, is overpriced. Since these reasons are structural, we believe that implied volatility is likely to remain overpriced for the foreseeable future. Given variance exposure to overpriced wings (and low strike puts) and the risk aversion to variance post credit crunch, we view variance as more overpriced than volatility.

- **Demand for put protection.** The demand for hedging products, either from investors, structured products or providers of variable annuity products, needs to be offset by market makers. As market makers are usually net sellers of volatility, they charge margin for taking this risk and for the costs of gamma hedging.

- **Demand for OTM options lifts wings.** Investors typically like buying OTM options as there is an attractive risk-reward profile (similar to buying a lottery ticket). Market makers therefore raise their prices to compensate for the asymmetric risk they face. As the price of variance swaps is based on options of all strikes, this lifts the price of variance.

- **Index implieds lifted from structured product demand.** The demand from structured products typically lifts index implied compared to single-stock implied. This is why implied correlation is higher than it should be.

SELLING VOLATILITY SHOULD BE LESS PROFITABLE THAN BEFORE

Hedge funds typically aim to identify mispricings in order to deliver superior returns. However, as both hedge funds and the total hedge fund marketplace grow larger, their opportunities are gradually being eroded. We believe that above-average returns are only possible in the following circumstances:

- **A fund has a unique edge** (eg, through analytics, trading algorithms or proprietary information/analysis).
- **There are relatively few funds in competition**, or it is not possible for a significant number of competitors to participate in an opportunity (either due to funding or legal restrictions, lack of liquid derivatives markets or excessive risk/time horizon of trade).
- **There is a source of imbalance in the markets** (eg, structured product flow or regulatory demand for hedging), causing a mispricing of risk.

All of the above reasons have previously held for volatility selling strategies (eg, call overwriting or selling of one/three-month variance swaps). However, given the abundance of publications on the topic in the past few years and the launch of several structured products that attempt to profit from this opportunity, we believe that volatility selling could be less profitable than before. The fact there remains an imbalance in the market due to the demand for hedging should mean volatility selling is, on average, a profitable strategy. However, we would caution against using a back test based on historical data as a reliable estimate of future profitability.

LONG VOLATILITY IS A POOR EQUITY HEDGE

An ideal hedging instrument for a security is an instrument with -100% correlation to that security and zero cost. As the return on variance swaps have a c-70% correlation with equity markets, adding long volatility positions (either through variance swaps or futures on volatility indices such as VIX or vStoxx) to an equity position could be thought of as a useful hedge. However, as volatility is on average overpriced, the cost of this strategy far outweighs any diversification benefit.

VOLATILITY HAS UP TO A NEGATIVE C70% CORRELATION WITH EQUITY

Equity markets tend to become more volatile when they decline and less volatile when they rise. A fundamental reason for this is the fact that gearing increases as equities decline[14]. As both gearing and volatility are measures of risk, they should be correlated; hence, they are negatively correlated to equity returns. More detailed arguments about the link between equity and volatility are provided in the section *Capital Structure Arbitrage* in the *Appendix*. While short term measures of volatility (eg, vStoxx) only have an R^2 of 50%-60% against the equity market, longer dated variance swaps (purest way to trade volatility) can have up to c70% R^2.

VOL RETURNS MOST CORRELATED TO EQUITY FOR 1-YEAR MATURITY

There are two competing factors to the optimum maturity for a volatility hedge. The longer the maturity, the more likely the prolonged period of volatility will be due to a decline in the market. This should give longer maturities higher equity volatility correlation, as the impact of short-term noise is reduced. However, for long maturities (years), there is a significant chance that the equity market will recover from any downturn, reducing equity volatility correlation. The optimum correlation between the SX5E and variance swaps, is for returns between nine months and one year. This is roughly in line with the c8 months it takes realised volatility to mean revert after a crisis.

SHORT-DATED VOLATILITY INDEX FUTURES ARE A POOR HEDGE

Structured products on VIX or vStoxx have a payout based on term structure as well as volatility

Recently, there have been several products based on rolling VIX or vStoxx futures whose average maturity is kept constant. As these products have to continually buy far-dated futures and sell near-dated futures (to keep average maturity constant as time passes), returns suffer from upward sloping term structure. Since the launch of vStoxx futures, rolling one-month vStoxx futures have had negative returns (see Figure 49 below). This is despite the SX5E also having suffered a negative return, suggesting that rolling vStoxx futures are a poor hedge. For more details on futures on volatility indices, see the section *Forward Starting Products*.

[14] Assuming no rights issues, share buybacks, debt issuance or repurchase/redemption.

Figure 49. SX5E and One-Month Rolling vStoxx Futures

Source: Santander Investment Bolsa.

LONG VOLATILITY HAS NEGATIVE RETURNS ON AVERAGE

Long volatility strategies, on average, have negative returns. This overpricing can be broken down into two components:

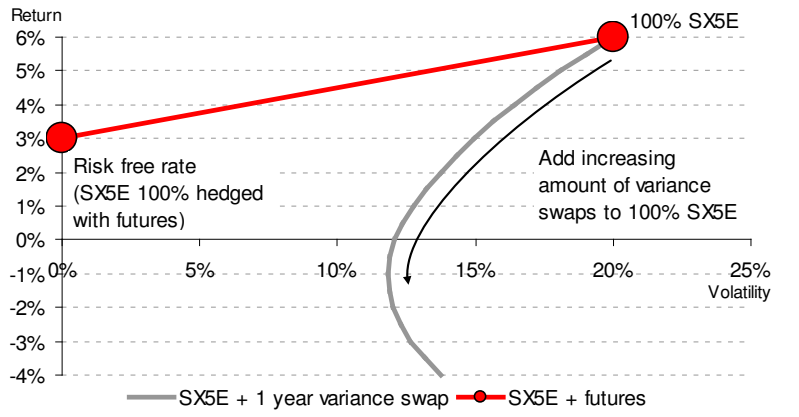
- **Correlation with equity market.** As equity markets are expected to return an equity risk premium over the risk-free rate, strategies that are implicitly long equity risk should similarly outperform (and strategies that are implicitly short equity risk should underperform). As a long volatility strategy is implicitly short equity risk, it should underperform. We note this drawback should affect all hedging instruments, as a hedging instrument by definition has to be short the risk to be hedged.

- **Overpricing of volatility.** Excessive demand for volatility products has historically caused implied volatility to be overpriced. As this demand is not expected to significantly decrease, it is likely that implied volatility will continue to be overpriced (although volatility will probably not be as overpriced as in the past).

VOLATILITY IS A POOR HEDGE COMPARED TO FUTURES

While all hedging instruments can be expected to have a cost (due to being implicitly short equities and assuming a positive equity risk premium), long variance swaps have historically had an additional cost due to the overpricing of volatility. This additional cost makes long variance swaps an unattractive hedge compared to reducing the position (or shorting futures). This is shown in Figure 50 below by adding an additional variance swap position to a 100% investment in equities (we optimistically assume zero margining and other trading costs to the variance swap position).

Figure 50. SX5E hedged with Variance Swaps or Futures



Source: Santander Investment Bolsa.

Long volatility hedge suffers from volatility overpricing, and less than 100% correlation

While the risk of the long equity and long variance swap position initially decreases as the long variance position increases in size, the returns of the portfolio are less than the returns for a reduced equity position of the same risk (we assume the proceeds from the equity sale are invested in the risk-free rate, which should give similar returns to hedging via short futures). Unlike hedging with futures, there comes a point at which increasing variance swap exposure does not reduce risk (and, in fact, increases it) due to the less than 100% correlation with the equity market.

Hedging strategies back-testing period needs to have positive equity returns

While we acknowledge that there are periods of time in which a long volatility position is a profitable hedge, these tend to occur when equity returns are negative (and short futures are usually a better hedge). We believe that the best back-testing periods for comparing hedging strategies are those in which equities have a return above the risk-free rate (if returns below the risk-free rate are expected, then investors should switch allocation away from equities into risk-free debt). For these back-testing periods, long volatility strategies struggle to demonstrate value as a useful hedging instrument. Hence, we see little reason for investors to hedge with variance swaps rather than futures given the overpricing of volatility, and less than 100% correlation between volatility and equity returns.

HEDGING WITH VARIANCE SHOULD NOT BE COMPARED TO PUTS

Puts offer convexity, long variance swaps do not

Due to the lack of convexity of a variance swap hedge, we believe it is best to compare long variance hedges to hedging with futures rather than hedging with puts. Although variance hedges might be cheaper than put hedges, the lack of convexity for long volatility makes this an unfair comparison, in our view.

VARIABLE ANNUITY HEDGING LIFTS LONG-TERM VOL

Since the 1980s, a significant amount of variable annuity products have been sold, particularly in the USA. The size of this market is now over US$1trn. From the mid-1990s, these products started to become more complicated and offered guarantees to the purchaser (similar to being long a put). The hedging of these products increases the demand for long-dated downside strikes, which lifts long-dated implied volatility and skew.

VARIABLE ANNUITY OFTEN GIVES INVESTORS A 'PUT' OPTION

With a fixed annuity, the insurance company that sold the product invests the proceeds and guarantees the purchaser a guaranteed fixed return. Variable annuities, however, allow the purchaser to pick which investments they want to put their funds into. The downside to this flexibility is the unprotected exposure to a decline in the market. To make variable annuities more attractive, from the 1990s many were sold with some forms of downside protection (or put). The different types of protection are detailed below in order of the risk to the insurance company.

- **Return of premium.** This product effectively buys an ATM put in addition to investing proceeds. The investor is guaranteed returns will be no lower than 0%.
- **Roll-up.** Similar to return of premium; however, the minimum guaranteed return is greater than 0%. The hedging of this product buys a put which is ITM with reference to spot, but OTM compared with the forward.
- **Ratchet (or maximum anniversary value).** These products return the highest value the underlying has ever traded at (on certain dates). The hedging of these products involves payout look-back options, more details of which are in the section *Look-Back Options*.
- **Greater of 'ratchet' or 'roll-up'.** This product returns the greater of the 'roll-up' or 'ratchet' protection.

Hedging of variable annuity products lifts index term structure and skew

Variable annuity hedging lifts implied volatility and skew

The hedging of variable annuity involves the purchase of downside protection for long maturities. Often the products are 20+ years long, but as the maximum maturity with sufficient liquidity available on indices can only be 3-5 years, the position has to be dynamically hedged with the shorter-dated option. This constant bid for long-dated protection lifts index term structure and skew, particularly for the S&P500 but also affects other major indices (potentially due to relative value trading). The demand for protection (from viable annuity providers or other investors), particularly on the downside and for longer maturities, could be considered to be the reason why volatility (of all strikes and maturities), skew (for all maturities) and term structure are usually overpriced.

CREDIT CRUNCH HAS HIT VARIABLE ANNUITY PROVIDERS

Until the TMT bubble burst, guarantees embedded in variable annuity products were often seen as unnecessary 'bells and whistles'. The severe declines between 2000 and 2003 made guarantees in variable annuity products more popular. When modelling dynamic strategies, insurance companies need to estimate what implied volatility will be in the future (eg, if hedging short 20-year options with 5-year options). The implied volatility chosen will be based on a confidence interval, say 95%, to give only a 1-in-20 chance that implieds are higher than the level embedded in the security. As the credit crunch caused realised volatility to reach levels that by some measures were higher than in the Great Depression, implied volatility rose to unprecedented heights. This increase in the cost of hedging has weighed on margins.

PROP DESK SPINOFF + MOVE TO EXCHANGE = HEDGE COSTS GO UP

Movement towards listed instruments, and regulation to counter proprietary desks, lifted long-dated skew

The passing of the Dodd-Frank Act in mid-2010 was designed to improve the transparency of derivatives by moving them onto an exchange. However, this would increase the margin requirements of long-dated options, which were previously traded OTC. This made it more expensive to be the counterparty to variable annuity providers. As the act also included the 'Volker Rule', which prohibits proprietary trading, the number of counterparties shrank (as prop desks with attractive funding levels were a common counterparty for the long-dated protection required by variable annuity hedgers). The combination of the spinoff of prop desks, and movement of OTC options onto an exchange caused skew to rise in mid-2010, particularly at the far-dated end of volatility surfaces.

STRUCTURED PRODUCTS VICIOUS CIRCLE

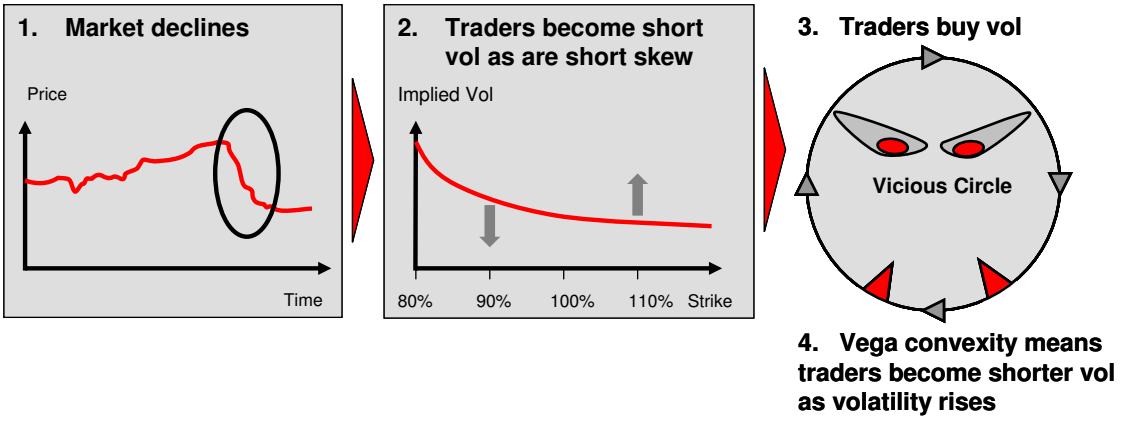
The sale of structured products leaves investment banks with a short skew position (eg, short an OTM put in order to provide capital-protected products). Whenever there is a large decline in equities, this short skew position causes the investment bank to be short volatility (eg, as the short OTM put becomes more ATM, the vega increases). The covering of this short vega position lifts implied volatility further than would be expected. As investment banks are also short vega convexity, this increase in volatility causes the short vega position to increase in size. This can lead to a 'structured products vicious circle' as the covering of short vega causes the size of the short position to increase. Similarly, if equity markets rise and implied volatility falls, investment banks become long implied volatility and have to sell. Structured products can therefore cause implied volatility to undershoot in a recovery, as well as overshoot in a crisis.

IMPLIED VOL OVERSHOOTS IN CRISIS, UNDERSHOOTS IN RECOVERY

Investment banks have a short skew and short vega convexity position

The sale of structured products causes investment banks to have a short skew and short vega convexity position[15]. Whenever there is a significant decline in equities and a spike in implied volatility, or a recovery in equities and a collapse in implied volatility, the position of structured product sellers can exaggerate the movement in implied volatility. This can cause implied volatility to overshoot (in a crisis) or undershoot (in a recovery post-crisis). There are four parts to the 'structured products vicious circle' effect on implied volatilities, which are shown in Figure 51 below.

Figure 51. Four Stages Towards Implied Volatility Overshoot



Source: Santander Investment Bolsa.

(1) EQUITY MARKET DECLINES

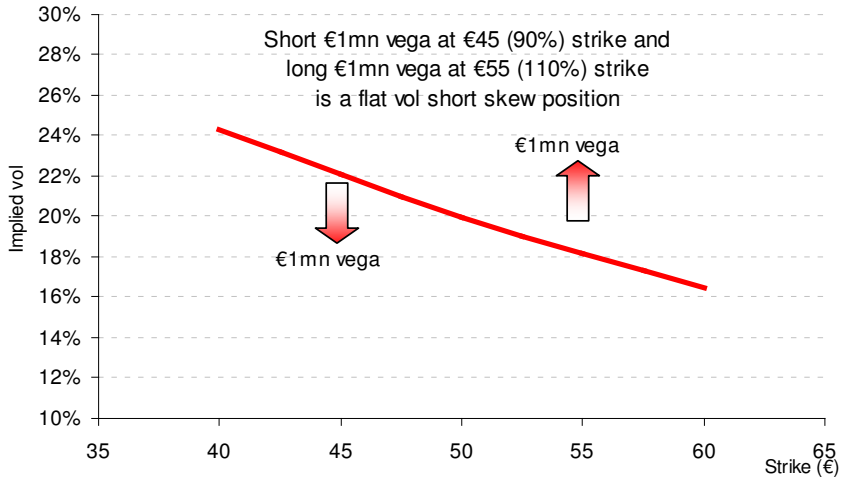
While implied volatility moves – in both directions – are exaggerated, for this example we shall assume that there is a decline in the markets and a rise in implied volatility. If this decline occurs within a short period of time, trading desks have less time to hedge positions, and imbalances in the market become more significant.

[15] There is more detail on the position of structured product sellers at the end of this section.

(2) DESKS BECOME SHORT IMPLIED VOLATILITY (DUE TO SHORT SKEW)

Investment banks are typically short skew from the sale of structured products. This position causes trading desks to become short implied volatility following declines in the equity market. To demonstrate how this occurs, we shall examine a short skew position through a vega flat risk reversal (short 90% put, long 110% call)[16].

Figure 52. Short Skew Position Due to 90%-110% Risk Reversal (initially vega flat)



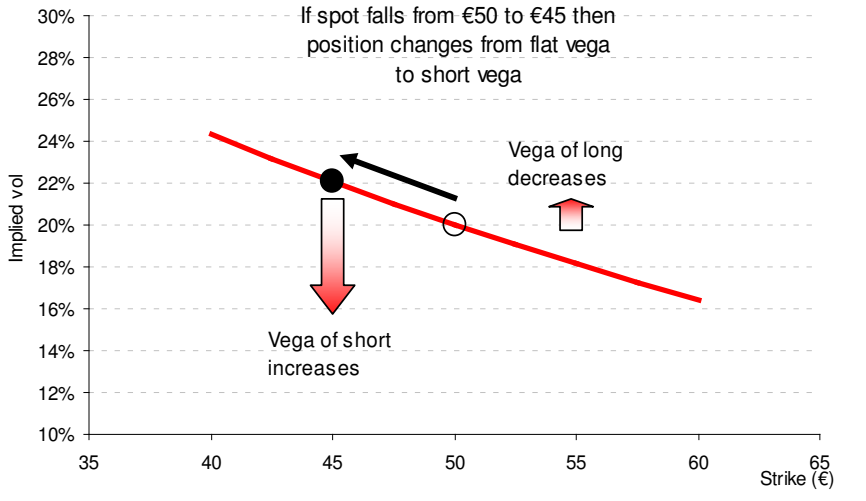
Source: Santander Investment Bolsa.

Short skew + equity markets decline = short vega (ie, short implied volatility)

90% put becomes ATM if equities decline 10%

If there is a 10% decline in equity markets, the 90% put becomes ATM and increases in vega. As the risk reversal is short the 90% put, the position becomes short vega (or short implied volatility). In addition, the 110% call option becomes more OTM and further decreases the vega of the position (increasing the value of the short implied volatility position).

Figure 53. Change in Vega of 90%-110% Risk Reversal If Markets Decline 10%



Source: Santander Investment Bolsa.

[16] This simple example is very different from the position of structured product sellers. We note a vega flat risk reversal is not necessarily 1-1, as the vega of the put is likely to be lower than the vega of the call.

Even if skew was flat, markets declines cause short skew position to become short vega

The above example demonstrates that it is the fact options become more or less ATM that causes the change in vega. It is not the fact downside put options have a higher implied than upside call options. If skew was flat (or even if puts traded at a lower implied than calls), the above argument would still hold. We therefore need a measure of the rate of change of vega for a given change in spot, and this measure is called vanna.

Vanna = dVega/dSpot

Vanna measures size of skew position, skew measures value of skew position

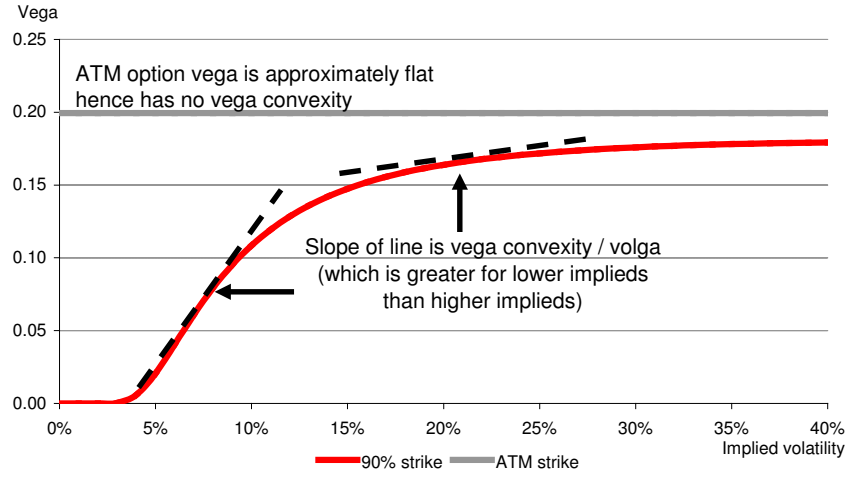
Vanna can be thought of as the size of the skew position (in a similar way that vega is the size of a volatility position), while skew (eg, 90%-100% skew) measures the value of skew (in a similar way that implied volatility measures the value of a volatility position). For more details on different Greeks, including vanna, see the section *Greeks and Their Meaning* in the *Appendix*.

(3) SHORT COVERING OF SHORT VEGA POSITION LIFTS IMPLIED VOL

All trading desks have a similar position

As the size of trading desks' short vega position increases during equity market declines, this position is likely to be covered. As all trading desks have similar positions, this buying pressure causes an increase in implied volatility. This flow is in addition to any buying pressure due to an increase in realised volatility and hence can cause an overshoot in implied volatility.

Figure 54. Vega of ATM and OTM Options Against Implied (Vega Convexity)



Source: Santander Investment Bolsa.

(4) SHORT VEGA POSITION INCREASES DUE TO VEGA CONVEXITY

Options have their peak vega when they are (approximately) ATM. As implied volatility increases, the vega of OTM options increases and converges with the vega of the peak ATM option. Therefore, as implied volatility increases, the vega of OTM options increases (see Figures 54). The rate of change of vega given a change in volatility is called volga (VOL-GAmma) or vomma, and is known as vega convexity.

Volga = dVega/dVol

Vega convexity causes short volatility position to increase

As the vega of options rises as volatility increases, this increases the size of the short volatility position that needs to be hedged. As trading desks' volatility short position has now increased, they have to buy volatility to cover the increased short position, which leads to further gains in implied volatility. This starts a vicious circle of increasing volatility, which we call the 'structured products vicious circle'.

VEGA CONVEXITY IS HIGHEST FOR LOW-TO-MEDIUM IMPLIEDS

As Figure 54 above shows, the slope of vega against volatility is steepest (ie, vega convexity is highest) for low-to-medium implied volatilities. This effect of vega convexity is therefore more important in volatility regimes of c20% or less; hence, the effect of structured products can have a similar effect when markets rise and volatilities decline. In this case, trading desks become long vega, due to skew, and have to sell volatility. Vega convexity causes this long position to increase as volatility declines, causing further volatility sellings. This is typically seen when a market recovers after a volatile decline (eg, in 2003 and 2009, following the end of the tech bubble and credit crunch, respectively).

IMPACT GREATEST FOR FAR-DATED IMPLIEDS

While this position has the greatest impact at the far end of volatility surfaces, a rise in far-dated term volatility and skew tends to be mirrored to a lesser extent for nearer-dated expiries. If there is a disconnect between near- and far-dated implied volatilities, this can cause a significant change in term structure.

STRUCTURED PRODUCT CAPITAL GUARANTEE IS LONG AN OTM PUT

The capital guarantee of many structured products leaves the seller of the product effectively short an OTM put. A short OTM put is short skew and short vega convexity (or volga). This is a simplification, as structured products tend to buy visually cheap options (ie, OTM options) and sell visually expensive options (ie, ATM options), leaving the seller with a long ATM and short OTM volatility position. As OTM options have more volga (or vega convexity) than ATM options (see the section *Greeks and Their Meaning* in the *Appendix*) the seller is short volga. The embedded option in structured products is floored, which causes the seller to be short skew (as the position is similar to being short an OTM put).

FORWARD STARTING PRODUCTS AND LIGHT EXOTICS

FORWARD STARTING PRODUCTS

Forward starting options are a popular method of trading forward volatility and term structure as there is no exposure to near-term volatility and, hence, zero theta (until the start of the forward starting option). As the exposure is to forward volatility rather than volatility, more sophisticated models need to be used to price them than ordinary options. Forward starting options will usually have wider bid-offer spreads than vanilla options, as their pricing and hedging is more complex. Recently, trading forward volatility via VIX and vStoxx forwards has become increasingly popular. However, as is the case with forward starting options, there are modelling issues. Forward starting variance swaps are easier to price as the price is determined by two variance swaps (one expiring at the start and the other at the end of the forward starting variance swap).

ZERO THETA IS AN ADVANTAGE OF FORWARD STARTING PRODUCTS

The main attraction of forward starting products is that they provide investors with long-term volatility (or vega) exposure, without having exposure to short-term volatility (or gamma)[17]. As there is zero gamma until the forward starting product starts, the product does not have to pay any theta. Forward starting products are most appropriate for investors who believe that there is going to be volatility in the future (eg, during a key economic announcement or a reporting date) but that realised volatility is likely to be low in the near term (eg, over Christmas or the summer lull).

Forward starting products are low cost, but also lower payout

While forward starting products have zero theta, they benefit less from any volatility spike

We note that while forward starting products have a lower theta cost than vanilla options, if there is a rise in volatility surfaces before the forward starting period is over, they are likely to benefit less than vanilla options (this is because the front end of volatility surfaces tends to move the most, and this is the area to which forward start has no sensitivity). Forward starting products can therefore be seen as a low-cost, lower-payout method of trading volatility.

TERM STRUCTURE PENALISES FORWARD STARTING PRODUCTS

While forward starting products have zero mathematical theta, they do suffer from the fact that volatility and variance term structure is usually expensive and upward sloping. The average implied volatility of a forward starting product is likely to be higher than a vanilla product, which will cause the long forward starting position to suffer carry as the volatility is re-marked lower[18] during the forward starting period.

THERE ARE THREE MAIN METHODS TO TRADE FORWARD VOLATILITY

Historically, forward volatility could only be traded via forward starting options, which had to be dynamically hedged and, hence, had high costs and wide bid-offer spreads. When variance swaps became liquid, this allowed the creation of forward starting variance swaps (as a forward starting variance can be perfectly hedged by a long and short position in two vanilla variance swaps of different maturity, which is explained later). The client base for trading forward volatility has recently been expanded by the listing of forwards on volatility indices (such as the VIX or vStoxx). The definition of the three main forward starting products is given below:

[17] We shall assume for this section that the investor wishes to be long a forward starting product.

[18] If a 3-month forward starting option is compared to a 3-month vanilla option, then during the forward starting period the forward starting implied volatility should, on average, decline.

(1) **Forward starting options.** A forward starting option is an option whose strike will be determined at the end of the forward starting period. The strike will be quoted as a percentage of spot. For example, a one-year ATM option three-month forward start, bought in September 2012, will turn into a one-year ATM option in December 2012 (ie, expiry will be December 2013 and the strike will be the value of spot in December 2012). Forward starting options are quoted OTC. For flow client requests, the maturity of the forward starting period is typically three months and with an option maturity no longer than a year. The sale of structured products creates significant demand for forward starting products, but of much longer maturity (2-3 years, the length of the structured product). Investment banks will estimate the size of the product they can sell and buy a forward starting option for that size. While the structured product itself does not incorporate a forward start, as the price for the product needs to be fixed for a period of 1-2 months (the marketing period), the product needs to be hedged with a forward start before marketing can begin.

(2) **Futures on volatility index.** A forward on a volatility index works in the same way as a forward on an equity index: they both are listed and both settle against the value of the index on the expiry date. While forwards on volatility indices such as the VIX and vStoxx have been quoted for some time, their liquidity has only recently improved to such an extent that they are now a viable method for trading. This improvement has been driven by increasing structured issuance and by options on volatility indices (delta hedging of these options has to be carried out in the forward market). Current listed maturities for the VIX and vStoxx exist for expiries under a year.

(3) **Forward starting variance swaps.** The easiest forward starting product to trade is a variance swap, as it can be hedged with two static variance swap positions (one long, one short). Like plain variance swaps, these products are traded OTC and their maturities can be up to a similar length (although investors typically ask for quotes up to three years).

HEDGING RISKS INCREASE COST OF FORWARD STARTING PRODUCTS

While forward starting options do not need to be delta hedged before the forward starting period ends, they have to be vega hedged with vanilla, very OTM straddles (as they also have zero delta and gamma). A long OTM straddle has to be purchased on the expiry date of the option, while a short OTM straddle has to be sold on the strike fixing date. As spot moves, the delta of these straddles is likely to move away from zero, requiring re-hedging of the position, which increases costs (which are likely to be passed on to clients) and risks (unknown future volatility and skew) to the trader.

Pricing of forwards on volatility indices tends to be slanted against long investors

Similarly, the hedging of forwards on volatility indices is not trivial, as (like volatility swaps) they require a volatility of volatility model. While the market for forwards on volatility indices has become more liquid, as the flow is predominantly on the buy side, forwards on volatility indices have historically been overpriced. They are a viable instrument for investors who want to short volatility, or who require a listed product.

Forward starting variance swaps have fewer imbalances than other forward products

A forward starting variance swap can be created from two vanilla variance swaps

The price – and the hedging – of a forward starting variance swap is based on two vanilla variance swaps (as it can be constructed from two vanilla variance swaps). The worst-case scenario for pricing is therefore twice the spread of a vanilla variance swap. In practice, the spread of a forward starting variance swap is usually slightly wider than the width of the widest bid-offer of the variance swap legs (ie, slightly wider than the bid-offer of the furthest maturity).

(1) FORWARD STARTING OPTIONS

A forward starting option can be priced using Black-Scholes in a similar way to a vanilla option. The only difference is that the forward volatility (rather than volatility) is needed as an input[19]. The three different methods of calculating the forward volatility, and examples of how the volatility input changes, are detailed below:

- **Sticky delta (or moneyness) and relative time.** This method assumes volatility surfaces never change in relative dimensions (sticky delta and relative time). This is not a realistic assumption unless the ATM term structure is approximately flat.
- **Additive variance rule.** Using the additive variance rule takes into account the term structure of a volatility surface. This method has the disadvantage that the forward skew is assumed to be constant in absolute (fixed) time, which is not usually the case. As skew is normally larger for shorter-dated maturities, it should increase approaching expiry.
- **Constant smile rule.** The constant smile rule combines the two methods above by using the additive variance rule for ATM options (hence, it takes into account varying volatility over time) and applying a sticky delta skew for a relative maturity. It can be seen as 'bumping' the current volatility surface by the change in ATM forward volatility calculated using the additive variance rule.

STICKY DELTA AND RELATIVE TIME USES CURRENT VOL SURFACE

Sticky delta and relative time tends to under price forward starting options

If the relative dimensions of a volatility surface are assumed to never change, then the volatility input for a forward starting option can be priced with the current volatility surface. For example, a three-month 110% strike option forward starting after a period of time T can be priced using the implied volatility of a current three-month 110% strike option (the forward starting time T is irrelevant to the volatility input[20]). As term structure is normally positive, this method tends to underprice forward starting options. An example of a current relative volatility surface, which can be used for pricing forward starting options under this method, is shown below:

Figure 55. Relative Dimensions Implied Volatility Surface

Strike	1 Year	2 Years	3 Years	4 Years
80%	24.0%	23.4%	23.2%	23.0%
90%	22.0%	22.0%	22.0%	22.0%
100%	20.0%	20.6%	20.8%	21.0%
110%	18.0%	19.2%	19.7%	20.0%
120%	16.0%	17.8%	18.5%	19.0%

Source: Santander Investment Bolsa.

[19] Forwards of the other inputs, for example interest rates, are generally trivial to compute.

[20] Hence, the price of the three-month 110% option forward start will only be significantly different from the price of the vanilla three-month 110% option if there is a significant difference in interest rates or dividends.

ADDITIVE VARIANCE RULE (AVR) CALCULATES FORWARD VOLATILITY

As variance time weighted is additive, and as variance is the square of volatility, the forward volatility can be calculated mathematically. Using these relationships to calculate forward volatilities is called the additive variance rule and is shown below.

$$\sigma_2{}^2 T_2 = \sigma_1{}^2 T + \sigma_{12}(T_2 - T_1) \text{ as variance time weighted is additive}$$

$$\Rightarrow \quad \sigma_{12} = \sqrt{\frac{\sigma_2{}^2 T_2 - \sigma_1{}^2 T_1}{T_2 - T_1}} = \text{forward volatility } T_1 \text{ to } T_2$$

where σ_i is the implied volatility of an option of maturity T_i

The above relationship can be used to calculate forward volatilities for the entire volatility surface. This calculation does assume that skew in absolute (fixed) time is fixed. An example, using the previous volatility surface, is shown below.

Figure 56. Current Volatility Surface

Strike	Start End	Now 1 Year	Now 2 Years	Now 3 Years	Now 4 Years
80%		24.0%	23.4%	23.2%	23.0%
90%		22.0%	22.0%	22.0%	22.0%
100%		20.0%	20.6%	20.8%	21.0%
110%		18.0%	19.2%	19.7%	20.0%
120%		16.0%	17.8%	18.5%	19.0%

One Year Additive Variance Rule Forward Vol Surface

Strike	Start End	Now 1 Year	1 Year 2 Years	2 Years 3 Years	3 Years 4 Years
80%		24.0%	22.8%	22.6%	22.5%
90%		22.0%	22.0%	22.0%	22.0%
100%		20.0%	21.2%	21.4%	21.5%
110%		18.0%	20.3%	20.7%	20.9%
120%		16.0%	19.4%	20.0%	20.3%

Source: Santander Investment Bolsa estimates.

ATM ADDITIVE VAR + STICKY DELTA = CONSTANT SMILE RULE (CSR)

Additive variance rule takes into account term structure but forward skew is too low

Using a relative time rule has the advantage of pricing forward skew in a reasonable manner, but it does not price the change in term structure correctly. While pricing using the additive variance rule gives improved pricing for ATM options, for OTM options the skew used is likely to be too low (as the method uses forward skew, which tends to decay by square root of time). The constant smile rule combines the best features of the previous two approaches, with ATM options priced using the additive variance rule and the skew priced using sticky delta.

Figure 57. Current Volatility Surface

Strike	1 Year	2 Years	3 Years	4 Years	1 Year Skew
80%	24.0%	23.4%	23.2%	23.0%	**4.0%**
90%	22.0%	22.0%	22.0%	22.0%	**2.0%**
100%	20.0%	20.6%	20.8%	21.0%	**0.0%**
110%	18.0%	19.2%	19.7%	20.0%	**-2.0%**
120%	16.0%	17.8%	18.5%	19.0%	**-4.0%**

One Year Additive Variance Rule (AVR) Forward Volatility

	Start End	Now 1 Year	1 Year 2 Years	2 Years 3 Years	3 Years 4 Years
100% AVR 1 Year Fwd Volatility		**20.0%**	**21.2%**	**21.4%**	**21.5%**

One Year Constant Smile Rule Forward Volatility Surface

Strike	Start End	Now 1 Year	1 Year 2 Years	2 Years 3 Years	3 Years 4 Years	1 Year skew
80%		24.0%	25.2%	25.4%	25.5%	**4.0%**
90%		22.0%	23.2%	23.4%	23.5%	**2.0%**
100% AVR 1 year Fwd Volatility		**20.0%**	**21.2%**	**21.4%**	**21.5%**	**0.0%**
110%		18.0%	19.2%	19.4%	19.5%	**-2.0%**
120%		16.0%	17.2%	17.4%	17.5%	**-4.0%**

Source: Santander Investment Bolsa.

Constant smile rule bumps sticky delta relative time volatility surface

The above diagrams show how the constant smile rule has the same ATM forward volatilities as the additive variance rule. The static delta (relative time) skew is then added to these ATM options to create the entire surface. An alternative way of thinking of the surface is that it takes the current volatility surface, and shifts (or bumps) each maturity by the exact amount required to get ATM options to be in line with the additive variance rule. The impact of having a relative time skew on a fixed ATM volatility can be measured by volatility slide theta (see the section *Advanced (Practical or Shadow) Greeks* in the *Appendix*).

CONSTANT SMILE RULE IS THE BEST MODEL OF THE THREE

Constant smile rule makes reasonable assumptions for options of all strikes and expiries

Pricing with static delta and relative time usually underprices forward volatility (as volatility term structure is normally upward sloping, and long-dated forward volatility is sold at the lower levels of near-dated implied volatility). While additive variance correctly prices forward volatility, this rule does mean future skew will tend towards zero (as skew tends to decay as maturity increases and the additive variance rule assumes absolute – fixed – time for skew). While this rule has been used in the past, the mispricing of long-dated skew for products such as cliquets has led traders to move away from this model. The constant smile rule would appear to be the most appropriate.

FORWARD STARTING OPTIONS HAVE SHADOW DELTA

While forward starting options have a theoretical delta of zero, they can have a shadow delta caused by fixed dividends and skew. If a dividend is fixed, then the dividend yield tends to zero as spot tends to infinity, which causes a shadow delta (which is positive for calls and negative for puts).

(2) FUTURE ON VOLATILITY INDEX

Recently, futures on the VIX and vStoxx have become more liquid due to increased structured product activity. While the calculation of these volatility indices is similar to a variance swap calculation, as the payout is based on the square root of variance their payout is linear in volatility not variance. They are therefore short volatility of volatility, just like volatility swaps.

PRICE IS IN BETWEEN VAR AND VOL

A future on a volatility index functions in exactly the same way as a future on an equity index. However, as the volatility index is a forward (hence linear) payout of the square root of variance, the payoff is different from a variance swap (whose payout is on variance itself). The price of a forward on a volatility index lies between the fair value of a forward volatility swap and the square root of the fair value of a forward variance swap.

⇨ $\sigma_{\text{Forward volatility swap}} \leq \text{Future on volatility index} \leq \sigma_{\text{Forward variance swap}}$

FUTURES ON VOLATILITY INDICES ARE SHORT VOL OF VOL

A variance swap can be hedged by delta hedging a portfolio of options[21] with the same maturity as the variance swap (portfolio is known as a log contract, where the weight of each option is $1/K^2$ where K is the strike). As the portfolio of options does not change, the only hedging costs are the costs associated with delta hedging. A volatility swap has to be hedged through buying and selling variance swaps (or a log contract of options); hence, it needs to have a volatility of volatility (vol of vol) model. A variance swap is more convex than a volatility swap (as a variance swap payout is on volatility squared); thus, a volatility swap is short convexity compared to a variance swap. A volatility swap is therefore short vol of vol as a variance swap has no vol of vol risk. As the price of a future on a volatility index is linear in volatility, a future on a volatility index is short vol of vol (like volatility swaps).

As vol of vol is underpriced, futures on volatility indices are overpriced

Short volatility future and long forward variance is a popular trade

While the price of volatility futures should be below that of forward variance swaps, retail demand and potential lack of knowledge of the client base means that they have traded at similar levels. This overpricing of volatility futures means that volatility of volatility is underpriced in these products. Being short volatility futures and long forward variance is a popular trade to arbitrage this mispricing.

EUREX, NOT CBOE, WAS THE FIRST EXCHANGE TO LIST VOL FUTURES

While futures on the VIX (launched by the CBOE in March 2004) are the oldest currently traded, the DTB (now Eurex) was the first exchange to list volatility futures, in January 1998. These VOLAX futures were based on 3-month ATM implieds, but they ceased trading in December of the same year.

STRUCTURED PRODUCTS ON VOL FUTURES IMPROVED LIQUIDITY

Initially VIX and vStoxx futures had limited liquidity. The creation of structured products has improved liquidity of these products. Similarly, the introduction of options on these futures has increased the need to delta hedge using these futures, also increasing liquidity. In the US, the size of structured products on VIX futures is so large at times it appears to have moved the underlying market.

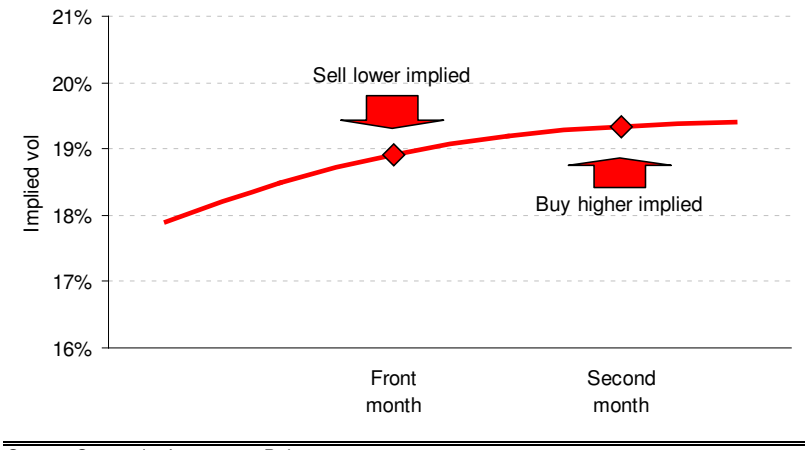
Open-ended volatility products on volatility indices steepen term structure

Structured products often sell near-dated expiries and buy longer-dated expiries, lifting term structure

While futures on a volatility index have the advantage of being a listed instrument, they have the disadvantage of having an expiry and, therefore, a longer-term position needs to be rolled. In response to investor demand, many investment banks sold products based on having a fixed maturity exposure on an underlying volatility index. As time passes, these banks hedge their exposure by selling a near-dated expiry and buying a far-dated expiry. The weighted average maturity is therefore kept constant, but the flow puts upward pressure on the term structure. For products of sufficient size, the impact of structured products on the market ensures the market moves against them. As term structure is lifted, there is a high carry cost (as a low near-dated volatility future is sold and a high far-dated volatility future is bought).

[21] As the payout of a delta hedged option is based on variance, not volatility (ie, delta hedging an option whose underlying moves 2%, earns four times as much as if the underlying moved 1%).

Figure 58. Volatility Future Term Structure



Source: Santander Investment Bolsa.

(3) FORWARD STARTING VAR SWAPS

In the section *Measuring Historical Volatility* in the *Appendix* we show that variance is additive (variance to time T_2 = variance to time T_1 + forward variance T_1 to T_2). This allows the payout of a forward starting variance swap between T_1 and T_2 to be replicated via a long variance swap to T_2, and short variance swap to T_1. We define N_1 and N_2 to be the notionals of the variance swaps to T_1 and T_2, respectively. It is important to note that N_1 and N_2 are the notionals of the variance swap, not the vega (N = vega ÷ 2 σ). As the variance swap payout of the two variance swaps must cancel up to T_1, the following relationship is true (we are looking at the floating leg of the variance swaps, and ignore constants that cancel such as the annualisation factor):

Payout long variance to T_2 = Payout long variance to T_1 + Payout long variance T_1 to T_2

$$\Rightarrow \quad N_2 \frac{\sum_{i=1}^{T_2}[\text{Ln}(\text{return}_i)]^2}{T_2} = N_2 \frac{\sum_{i=1}^{T_1}[\text{Ln}(\text{return}_i)]^2}{T_2} + N_2 \frac{\sum_{i=T_1+1}^{T_2}[\text{Ln}(\text{return}_i)]^2}{T_2}$$

$$\Rightarrow \quad \text{Payout short variance} = N_2 \frac{\sum_{i=1}^{T_1}[\text{Ln}(\text{return}_i)]^2}{T_2} \quad (= \text{Payout long variance to } T_1)$$

$$\Rightarrow \quad N_1 \frac{\sum_{i=1}^{T_1}[\text{Ln}(\text{return}_i)]^2}{T_1} = N_2 \frac{\sum_{i=1}^{T_1}[\text{Ln}(\text{return}_i)]^2}{T_2}$$

$$\Rightarrow \quad \frac{N_1}{T_1} = \frac{N_2}{T_2}$$

$$\Rightarrow \quad N_1 = \frac{T_1}{T_2} N_2 = (\text{Notional of near dated variance as factor of far dated variance notional})$$

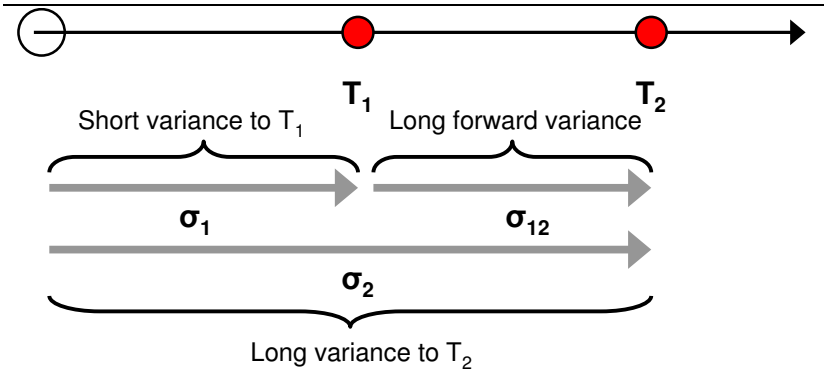
$$\Rightarrow \quad 0 < N_1 < N_2 \ (\text{as } T_1 > T_2)$$

Notional of the near-dated variance is smaller than notional of far-dated

The above proof shows that in order to construct a forward starting variance swap from two vanilla variance swaps, the near-dated variance should have a notional of T_1/T_2 (which is less than 1) of the notional of the far-dated variance. Intuitively, this makes sense as the near-dated variance swap to T_1 only needs to cancel the overlapping period of the longer-dated variance swap to T_2. The notional of the near-dated variance swap to T_1 therefore has to be scaled down, depending on its relative maturity to T_2. For example, if T_1 is 0, then there is no need to short any near-dated variance and N_1 is similarly zero. In addition, if $T_1 = T_2$, then the two legs must cancel, which occurs as $N_1 = N_2$.

The notional N_{12} must be equal to the difference of the notionals of the two vanilla variance swaps that hedge it (ie, $N_{12} = N_2 - N_1$) by considering the floating legs and having constant realised volatility ($N_2\sigma_2^2 = N_1\sigma_1^2 + N_{12}\sigma_{12}^2$, hence $N_2 = N_1 + N_{12}$ if volatility σ^2 is constant).

Figure 59. Constructing Forward Variance from Vanilla Variance Swaps



Source: Santander Investment Bolsa.

CALCULATING FORWARD VARIANCE

The additive variance rule allows the level of forward variance to be calculated (as variance time weighted is additive).

$$\sigma_2^2 T_2 = \sigma_1^2 T + \sigma_{12}(T_2 - T_1)$$

$$\Rightarrow \quad \sigma_{12}^2 = \frac{\sigma_2^2 T_2 - \sigma_1^2 T_1}{T_2 - T_1} = \text{forward volatility } T_1 \text{ to } T_2$$

SKEW CAUSES NEGATIVE SHADOW DELTA

The presence of skew causes a correlation between volatility and spot. This correlation produces a negative shadow delta for all forward starting products. The rationale is similar to the argument that variance swaps have negative shadow delta.

Fixed dividends also causes shadow delta

Forward starting products also have a shadow delta caused by fixed (or discrete) dividends[22] (calls have positive shadow delta, while puts have negative shadow delta).

[22] Fixed dividends also lifts value of calls, puts and variance swaps

BARRIER OPTIONS

Barrier options are the most popular type of light exotic product, as they are used within structured products or to provide cheap protection. The payout of a barrier option knocks in or out depending on whether a barrier is hit. There are eight types of barrier option, but only four are commonly traded, as the remaining four have a similar price to vanilla options. Barrier puts are more popular than calls (due to structured product and protection flow), and investors like to sell visually expensive knock-in options and buy visually cheap knock-out options. Barrier options (like all light exotics) are always European (if they were American, the price would be virtually the same as a vanilla option, as the options could be exercised just before the barrier was hit).

BARRIER OPTIONS CAN HAVE DELTA OF MORE THAN ±100%

Extra hedging risk of barriers widens the bid-offer spread

The hedging of a barrier option is more involved than for vanilla options, as the delta near the barrier can be significantly more than ±100% near expiry. The extra hedging risk of barriers widens the bid-offer spread in comparison with vanilla options. Barrier options are always European and are traded OTC.

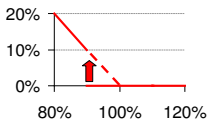
THERE ARE THREE KEY VARIABLES FOR BARRIER OPTIONS

There are three key variables to a barrier option, each of which has two possibilities. These combinations give eight types of barrier option (8=2×2×2).

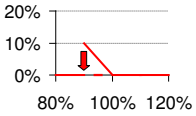
- **Down/up.** The direction of the barrier in relation to spot. Almost all put barriers are down barriers and, similarly, almost all call barriers are up barriers.
- **Knock in/out.** Knock-out options have a low premium and give the impression of being cheap; hence, they are usually bought by investors. Conversely, knock-in options are visually expensive (as knock-in options are a similar price to a vanilla) and are usually sold by investors (through structured products). For puts, a knock-in is the most popular barrier (structured product selling of down and knock-in puts). However, for calls this is reversed and knock-outs are the most popular. Recent volatility has made knock-out products less popular than they once were, as many hit their barrier and became worthless.
- **Put/call.** The type of payout of the option. Put barriers are three to four times more popular than call barriers, due to the combination of selling from structured products (down and knock-in puts) and cheap protection buying (down and knock-out puts).

ONLY FOUR OF THE EIGHT TYPES OF BARRIER ARE USUALLY TRADED

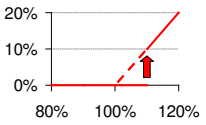
The difference in price between a vanilla option and barrier option is only significant if the barrier occurs when the option has intrinsic value. If the only value of the option when the barrier knocks in/out is time value, then the pricing for the barrier option will be roughly equal to the vanilla option. Because of this, the naming convention for barrier options can be shortened to knock in (or out) followed by call/put (as puts normally have a down barrier, and calls an up barrier). The four main types of barrier option and their uses are shown below.



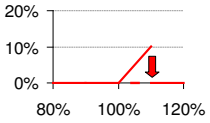
Knock-in put (down and knock-in put). Knock-in puts are the most popular type of barrier option, as autocallables are normally hedged by selling a down and knock-in put to fund the high coupon. They have a barrier which is below both spot and strike and give an identical payoff to a put only once spot has gone below the down barrier. Until spot reaches the down barrier there is no payout. However, as this area has the least intrinsic value, the theoretical price is similar to a vanilla and therefore visually expensive.



Knock-out put (down and knock-out put). Knock-out puts are the second most popular barrier option after knock-in puts (although knock-in puts are three times as popular as knock-out puts due to structured product flow). Knock-out puts give an identical payout to a put, until spot declines through the down barrier (which is below both spot and strike), in which case the knock-out option becomes worthless. As the maximum payout for a put lies below the knock-out barrier, knock-out puts are relatively cheap and are often thought of as a cheap method of gaining protection.



Knock-in call (up and knock-in call). Knock-in calls give an identical payout to a call, but only when spot trades above the up barrier, which lies above spot and the strike. They are the least popular barrier option, as their high price is similar to the price of a call and structured product flow is typically less keen on selling upside than downside.

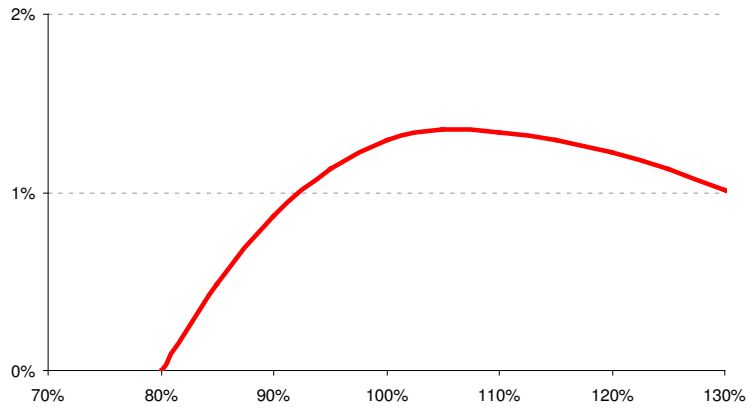


Knock-out call (up and knock-out call). Knock-out calls are the most popular barrier option for calls, but their popularity still lags behind both knock-in and knock-out puts. As they give the same upside participation as a vanilla call until the up barrier (which is above spot and strike) is reached, they can be thought of as a useful way of gaining cheap upside.

KNOCK-OUT OPTIONS DECREASE IN VALUE AS STRIKE APPROACHES

While vanilla options (and knock-in options) will increase in value as spot moves further in the money, this is not the case for knock-out options, where spot is near the strike. This effect is caused by the payout equalling zero at the barrier, which can cause delta to be of opposite sign to the vanilla option. This effect is shown below for a one-year ATM put with 80% knock-out. The peak value of the option is at c105%; hence, for values lower than that value the delta is positive not negative. This is a significant downside to using knock-out puts for protection, as their mark to market can increase (not decrease) equity sensitivity to the downside.

Figure 60. Price of One-Year ATM Put with 80% Knock-Out

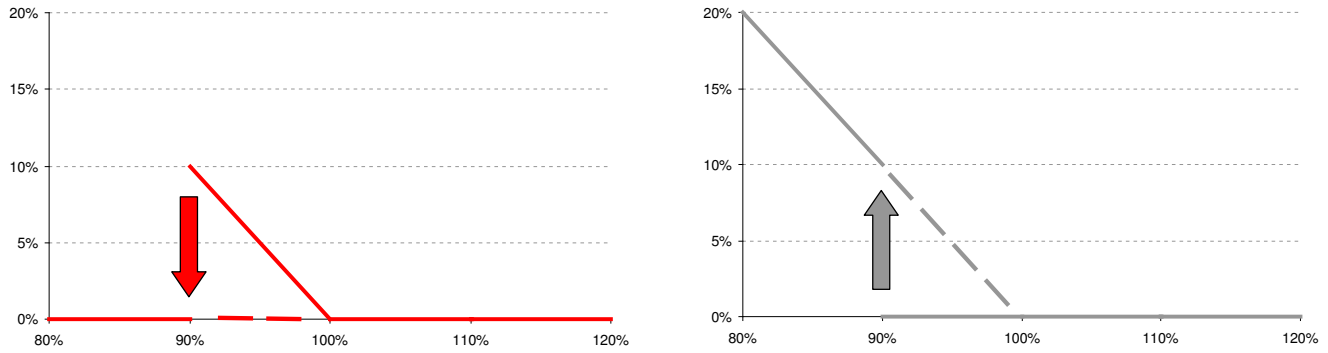


Source: Santander Investment Bolsa.

KNOCK-OUT PUT + KNOCK-IN PUT = (VANILLA) PUT

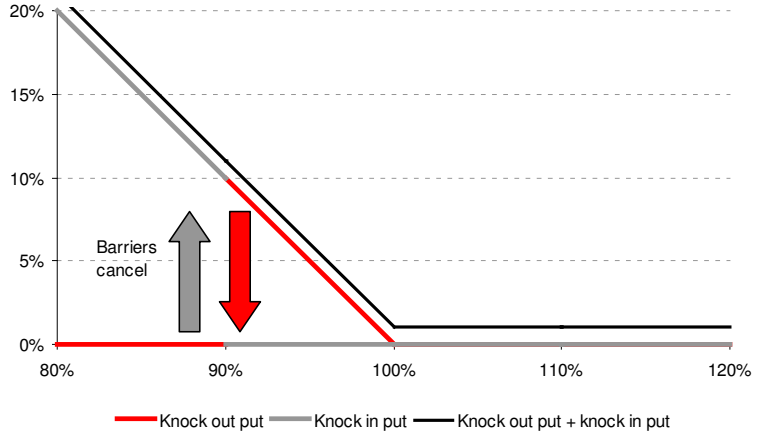
If a knock-out put and a knock-in put have the same strike and barrier, then together the combined position is equal to a long vanilla put ($P_{KO} + P_{KI} = P$). This is shown in the charts below. The same argument can apply to calls ($P_{KO} + P_{KI} = P$). This relationship allows us to see mathematically that if knock-out options are seen as visually cheap, then knock-in options must be visually expensive (as a knock-in option must be equal to the price of a vanilla less the value of a visually cheap knock-out option).

Figure 61. Knock-out Put **Knock-in Put**



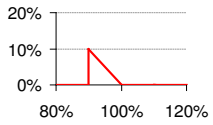
Source: Santander Investment Bolsa.

Figure 62. Knock-out Put + Knock-in Put = Put



Source: Santander Investment Bolsa.

KNOCK-OUTS COST C15%-25% OF PUT SPREAD COST



The payout of a knock-out put is equal to a 'shark fin' (see figure on the left) until the barrier is reached. A 'shark fin' is equal to a short digital position (at the barrier) plus a put spread (long put at strike of knock-out put, short put at barrier of knock-out put). The price of a knock-out put can therefore be considered to be the cost of a put spread, less a digital and less the value of the knock-out. As pricing digitals and barriers is not trivial, comparing the price of a knock-out put as a percentage of the appropriate put spread can be a quick way to evaluate value (the knock-out will have a lower value as it offers less payout to the downside). For reasonable barriers between 10% and 30% below the strike, the price of the knock-out option should be between c15% and c25% of the cost of the put spread.

CONTINUOUS BARRIERS ARE CHEAPER THAN DISCRETE

Continuous barriers are more popular as they are cheaper

There are two types of barriers, continuous and discrete. A continuous barrier is triggered if the price hits the barrier intraday, whereas a discrete barrier is only triggered if the closing price passes through the barrier. Discrete knock-out barriers are more expensive than continuous barriers, while the reverse holds for knock-in barriers (especially during periods of high volatility). There are also additional hedging costs to discrete barriers, as it is possible for spot to move through the barrier intraday without the discrete barrier being triggered (ie, if the close is the correct side of the discrete barrier). As these costs are passed on to the investor, discrete barriers are far less popular than continuous barriers for single stocks (c10%-20% of the market), although they do make up almost half the market for indices.

Jumps in stock prices between close and open is a problem for all barriers

While the hedge for a continuous barrier should, in theory, be able to be executed at a level close to the barrier, this is not the case should the underlying jump between close and open. In this case, the hedging of a continuous barrier suffers a similar problem to the hedging of a discrete barrier (delta hedge executed at a significantly different level to the barrier).

DOUBLE BARRIERS ARE POSSIBLE, BUT RARE

Double barrier options have both an up barrier and a down barrier. As only one of the barriers is significant for pricing, they are not common (as their pricing is similar to an ordinary single-barrier option). They make up less than 5% of the light exotic market.

REBATES CAN COMPENSATE INVESTORS IF BARRIER TRIGGERED

The main disadvantage of knock-out barrier options is that the investor receives nothing for purchasing the option if they are correct about the direction of the underlying (option is ITM) but incorrect about the magnitude (underlying passes through barrier). In order to provide compensation, some barrier options give the long investors a rebate if the barrier is triggered: for example, an ATM call with 120% knock-out that gives a 5% rebate if the barrier is touched. Rebates comprise approximately 20% of the index barrier market but are very rare for single-stock barrier options.

WORST-OF/BEST-OF OPTIONS

Worst-of (or best-of) options give payouts based on the worst (or best) performing asset. They are the second most popular light exotic due to structured product flow. Correlation is a key factor in pricing these options, and investor flow typically buys correlation (making uncorrelated assets with low correlation the most popular underlyings). The underlyings can be chosen from different asset classes (due to low correlation), and the number of underlyings is typically between three and 20. They are always European, and normally ATM options.

MATURITY IS NORMALLY ONE YEAR AND CAN BE CALLS OR PUTS

Worst-of/best-of options can be any maturity. Although the most popular is one-year maturity, up to three years can trade. As an option can be a call or a put, and either 'worst-of' or 'best-of'; there are four types of option to choose from. However, the most commonly traded are worst-of options (call or put). The payouts of the four types are given below:

Worst-of call payout = Max (Min (r_1, r_2, ... , r_N) , 0) where r_i is the return of N assets

Worst-of put payout = Max (-Min (r_1, r_2, ... , r_N) , 0) where r_i is the return of N assets

Best-of call payout = Max (Max (r_1, r_2, ... , r_N) , 0) where r_i is the return of N assets

Best-of put payout = Max (-Max (r_1, r_2, ... , r_N) , 0) where r_i is the return of N assets

WORST-OF CALLS POPULAR TO BUY (AS CHEAPER THAN ANY CALL)

The payout of a worst-of call option will be equal to the lowest payout of individual call options on each of the underlyings. As it is therefore very cheap, they are popular to buy. If all the assets are 100% correlated, then the value of the worst-of call is equal to the value of calls on all the underlyings (hence, in the normal case of correlation less than 100%, a worst-of call will be cheaper than any call on the underlying). If we lower the correlation, the price of the worst-of call also decreases (eg, the price of a worst-of call on two assets with -100% correlation is zero, as one asset moves in the opposite direction to the other). A worst-of call option is therefore long correlation. As worst-of calls are cheap, investors like to buy them and, therefore, provide buying pressure to implied correlation.

Rumour of QE2 lifted demand for worst-of calls on cross assets

Cross asset worst-of/best-of options are popular when macro risks dominate

Before QE2 (second round of quantitative easing) was announced, there was significant buying flow for worst-of calls on cross assets. The assets chosen were all assets that were likely to be correlated should QE2 occur but that would normally not necessarily be correlated (giving attractive pricing). QE2 was expected to cause USD weakening (in favour of other G10 currencies like the JPY, CHF and EUR), in addition to lifting 'risk-on' assets, like equities and commodities. The buying of worst-of calls on these three assets would therefore be a cheap way to gain exposure to the expected movements of markets if quantitative easing was extended (which it was).

WORST-OF PUTS ARE EXPENSIVE AND USUALLY SOLD

Flow from worst-of/best-of lifts implied correlation

A worst-of put will have a greater value than any of the puts on the underlying assets and is therefore very expensive to own. However, as correlation increases towards 100%, the value of the worst-of put will decrease towards the value of the most valuable put on either of the underlyings. A worst-of put is therefore short correlation. As selling (expensive) worst-of puts is popular, this flow puts buying pressure on implied correlation (the same effect as the flow for worst-of calls).

BEST-OF CALLS AND BEST-OF PUTS ARE RELATIVELY RARE

While worst-of options are popular, there is relatively little demand for best-of options. There are some buyers of best-of puts (which again supports correlation); however, best-of calls are very rare. Figure 63 below summarises the popularity and direction of investor flows (normally from structured products) and the effect on implied correlation. A useful rule of thumb for worst-of/best-of options is that they are short correlation if the price of the option is expensive (worst-of put and best-of call) and the reverse if the price of the option is cheap. This is why the buying of cheap and selling of expensive worst-of/best-of options results in buying flow to correlation.

Figure 63. Best-of/Worst-of Options

Option	Correlation	Flow	Cost	Notes
Worst-of put	Short	Sellers	Expensive	Popular structure to sell as cost is greater than that of most expensive put
Worst-of call	Long	Buyers	Cheap	Popular way to buy upside as low cost is less than cheapest call on any of the assets
Best-of put	Long	Some buyers	Cheap	Some buyers as cost is lower than cheapest put
Best-of call	Short	Rare	Expensive	Benefits from correlation falling as markets rise

Source: Santander Investment Bolsa.

LIGHT EXOTIC OPTIONS FLOW LIFTS IMPLIED CORRELATION

As the flow from worst-of/best-of products tends to support the levels of implied correlation, implied correlation typically trades above fair value. While other light exotic flow might not support correlation (eg, outperformance options, which are described below), worst-of/best-of options are the most popular light exotic, whose pricing depends on correlation and are therefore the primary driver for this market. We would point out that the most popular light exotics – barrier options – have no impact on correlation markets. In addition, worst-of/best-of flow is concentrated in uncorrelated assets, whereas outperformance options are usually on correlated assets.

OUTPERFORMANCE OPTIONS

Outperformance options are an option on the difference between returns on two different underlyings. They are a popular method of implementing relative value trades, as their cost is usually cheaper than an option on either underlying. The key unknown parameter for pricing outperformance options is implied correlation, as outperformance options are short correlation. The primary investor base for outperformance options is hedge funds, which are usually buyers of outperformance options on two correlated assets (to cheapen the price). Outperformance options are European and can always be priced as a call. Unless they are struck with a hurdle, they are an ATM option.

OUTPERFORMANCE OPTIONS ARE USUALLY SHORT-DATED CALLS

Outperformance options give a payout based on the difference between the *returns* of two underlyings. While any maturity can be used, they tend to be for maturities up to a year (maturities less than three months are rare). The payout formula for an outperformance option is below – by convention always quoted as a call of 'r_A over r_B,' (as a put of 'r_A over r_B' can be structured as a call on 'r_B over r_A'). Outperformance options are always European (like all light exotics) and are traded OTC.

Payout = Max ($r_A - r_B$, 0) where r_A and r_B are the returns of assets A and B, respectively

OPTIONS USUALLY ATM, CAN HAVE HURDLE AND ALLOWABLE LOSS

While outperformance options are normally structured ATM, they can be cheapened by making it OTM through a hurdle or by allowing an allowable loss at maturity (which simply defers the initial premium to maturity). While outperformance options can be structured ITM by having a negative hurdle, as this makes the option more expensive, this is rare. The formula for outperformance option payout with these features is:

Payout = Max ($r_A - r_B$ – hurdle, – allowable loss)

OUTPERFORMANCE OPTIONS ARE SHORT CORRELATION

The pricing of outperformance options depends on both the volatility of the two underlyings and the correlation between them. As there tends to be a more liquid and visible market for implied volatility than correlation, it is the implied correlation that is the key factor in determining pricing. Outperformance options are short correlation, which can be intuitively seen as: the price of an outperformance option must decline to zero if one assumes correlation rises towards 100% (two identical returns give a zero payout for the outperformance option).

As flow is to the buy side, some hedge funds outperformance call overwrite

Outperformance options are ideal for relative value trades

Outperformance options are ideal for implementing relative value trades, as they benefit from the upside, but the downside is floored to the initial premium paid. The primary investor base for outperformance options are hedge funds. While flow is normally to the buy side, the overpricing of outperformance options due to this imbalance has led some hedge funds to call overwrite their relative value position with an outperformance option.

MARGRABE'S FORMULA CAN BE USED FOR PRICING

An outperformance option volatility σ_{A-B} can be priced using Margrabe's formula given the inputs of the volatilities σ_A and σ_B of assets A and B, respectively, and their correlation ρ. This formula is shown below.

$$\sigma_{A-B} = \sqrt{\sigma_A{}^2 + \sigma_B{}^2 - 2\rho\sigma_A\sigma_B}$$

TEND TO BE USED FOR CORRELATED ASSETS

The formula above confirms mathematically that outperformance options are short correlation (due to the negative sign of the final term with correlation ρ). From an investor perspective, it therefore makes sense to sell correlation at high levels; hence, outperformance options tend to be used for correlated assets (so cross-asset outperformance options are very rare). This is why outperformance options tend to be traded on indices with a 60%-90% correlation and on single stocks that are 30%-80% correlated. The pricing of an outperformance option offer tends to have an implied correlation 5% below realised for correlations of c80%, and 10% below realised for correlations of c50% (outperformance option offer is a bid for implied correlation).

Best pricing is with assets of similar volatility

The price of an outperformance is minimised if volatilities σ_A and σ_B of assets A and B are equal (assuming the average of the two volatilities is kept constant). Having two assets of equal volatility increases the value of the final term $2\rho\sigma_A\sigma_B$ (reducing the outperformance volatility σ_{A-B}).

LOWER FORWARD FLATTERS OUTPERFORMANCE PRICING

ATMf strikes should be used when comparing relative costs

Assuming that the two assets have a similar interest rate and dividends, the forwards of the two assets approximately cancel each other out, and an ATM outperformance option is also ATMf (ATM forward or At The Money Forward). When comparing relative costs of outperformance options with call options on the individual underlyings, ATMf strikes must be used. If ATM strikes are used for the individual underlyings, the strikes will usually be lower than ATMf strikes and the call option will appear to be relatively more expensive compared to the ATMf (= ATM) outperformance option.

Pricing of ATM outperformance options is usually less than ATMf on either underlying

If two assets have the same volatility ($\sigma_A = \sigma_B$) and are 50% correlated ($\rho = 50\%$), then the input for outperformance option pricing σ_{A-B} is equal to the volatilities of the two underlyings ($\sigma_{A-B} = \sigma_A = \sigma_B$). Hence, ATMf (ATM forward) options on either underlying will be the same as an ATMf ($\approx$ATM) outperformance option. As outperformance options tend to be used on assets with higher than 50% correlation and whose volatilities are similar, outperformance options are usually cheaper than similar options on either underlying.

LOOK-BACK OPTIONS

There are two types of look-back options, strike look-back and payout look-back, and both are usually multi-year options. Strike reset (or look-back) options have their strike set to the highest, or lowest, value within an initial look-back period (of up to three months). These options are normally structured so the strike moves against the investor in order to cheapen the cost. Payout look-back options conversely tend to be more attractive and expensive than vanilla options, as the value for the underlying used is the best historical value. As with all light exotics, these options are European and OTC.

STRIKE OF RESET OPTIONS MOVES AGAINST INVESTOR

Strike look-back, and payout look-back are the two types of look-back options

There are two main strike reset options, and both have an initial look-back period of typically one to three months, where the strike is set to be the highest (for a call) or lowest (for a put) traded value. While the look-back optionality moves against the investor, as the expiry of these options is multi-year (typically three), there is sufficient time for spot to move back in the investor's favour, and the strike reset cheapens the option premium. While having a strike reset that moves the strike to be the most optimal for the investor is possible, the high price means they are unpopular and rarely trade. While the cheaper form of strike reset options does attract some flow due to structured products, they are not particularly popular.

Strike reset options perform best when there is an initial period of range trading

There are three possible outcomes to purchasing a strike reset option. Strike reset options can be considered a cheaper alternative to buying an ATM option at the end of the strike reset period, as the strike is roughly identical for two of the three possible outcomes (but at a lower price).

- **Spot moves in direction of option payout.** If spot moves in a direction that would make the option ITM, the strike is reset to be equal to spot as it moves in a favourable direction, and the investor is left with a roughly ATM option.
- **Range-trading markets.** Should markets range trade, the investor will similarly receive a virtually ATM option at the end of the strike reset period.
- **Spot moves in opposite direction to option payout.** If spot initially moves in the opposite direction to the option payout (down for calls, up for puts), then the option strike is identical to an option that was initially ATM (as the key value of the underlying for the strike reset is the initial value) and, hence, OTM at the end of the strike rest period. The downside of this outcome is why strike reset options can be purchased for a lower cost than an ATM option.

Strike reset options are therefore most suitable for investors who believe there will be an initial period of range trading, before the underlying moves in a favourable direction.

PAYOUT LOOK-BACK OPTIONS

Having a look-back option that selects the best value of the underlying (highest for calls, lowest for puts) increases the payout of an option – and cost. These options typically have a five-year maturity and typically use end-of-month or end-of-year values for the selection of the optimal payout.

CONTINGENT PREMIUM OPTIONS

Contingent premium options are initially zero-premium and only require a premium to be paid if the option becomes ATM on the close. The contingent premium to be paid is, however, larger than the initial premium would be, compensating for the fact that it might never have to be paid. Puts are the most popular, giving protection with zero initial premium. These typically one-year put options are OTM (or the contingent premium would almost certainly have to be paid immediately) and European.

CONTINGENT PREMIUM OPTIONS ALLOW ZERO UPFRONT COST

Contingent premium allows protection to be bought at zero cost

While contingent premium calls are possible, the most popular form is for a contingent premium put to allow protection to be bought with no initial cost. The cost of the premium to be paid is roughly equal to the initial premium of the vanilla option, divided by the probability of spot trading through the strike at some point during the life of the option (eg, an 80% put whose contingent premium has to be paid if the underlying goes below 80%). Using contingent premium options for protection has the benefit that no cost is suffered if the protection is not needed, but if spot dips below the strike/barrier, then the large premium has to be paid (which is likely to be more than the put payout unless there was a large decline). These can be thought of as a form of 'crash put'.

Having a conditional premium on a level other than strike is possible, but rare

The usual structure for contingent premium options is to have the level at which the premium is paid equal to the strike. The logic is that although investors have to pay a large premium, they do have the benefit of holding an option that is slightly ITM. Having the conditional premium at a level other than strike is possible, but rare (eg, an 80% put whose contingent premium has to be paid if the underlying reaches 110%).

COMPOSITE AND QUANTO OPTIONS

There are two types of option involving different currencies. The simplest is a composite option, where the strike (or payoff) currency is in a different currency to the underlying. A slightly more complicated option is a quanto option, which is similar to a composite option, but the exchange rate of the conversion is fixed.

COMPOSITE OPTIONS USE DIFFERENT VOLATILITY INPUT

Composite options are usually more expensive than vanilla options

A composite option is a cash or physical option on a security whose currency is different from the strike or payoff currency (eg, Euro strike option on Apple). If an underlying is in a foreign currency, then its price in the payout (or strike) currency will usually be more volatile (and hence more expensive) due to the additional volatility associated with currency fluctuations. Only for significantly negative correlations will a composite option be less expensive than the vanilla option (if there is zero correlation the effect of FX still lifts valuations). The value of a composite option can be calculated using Black-Scholes as usual, by substituting the volatility of the asset with the volatility of the asset in payout currency terms. The payout (or strike) currency risk-free rate should be used rather than the (foreign) security currency risk-free rate. The dividend yield assumption is unchanged (as it has no currency) between a composite option and a vanilla option.

$$\sigma_{Payout} = \sqrt{\sigma_{Security}{}^2 + \sigma_{FX}{}^2 + 2\rho\sigma_{Security}\sigma_{FX}}$$

where

σ_{Payout} = volatility of asset in payout (strike) currency

$\sigma_{Security}$ = volatility of asset in (foreign) security currency

σ_{FX} = volatility of FX rate (between payout currency and security currency)

ρ = correlation of FX rate (security currency in payoff currency terms) and security price

Composite options are long correlation (if FX is foreign currency in domestic terms)

The formula to calculate the volatility of the underlying is given above. As the payoff increases with a positive correlation between FX and the underlying, a composite option is long correlation (the positive payout will be higher due to FX, while FX moving against the investor is irrelevant when the payout is zero). Note that care has to be taken when considering the definition of the FX rate; it should be the (foreign) security currency given in (domestic) payoff currency terms.

For example, if we are pricing a euro option on a dollar-based security and assume an extreme case of $\rho = 100\%$, the volatility of the USD underlying in EUR will be the sum of the volatility of the underlying and the volatility of USD.

QUANTO OPTIONS USE DIFFERENT DIVIDEND INPUT

Quanto options cost a similar amount to vanilla options if FX correlation is small

Quanto options are similar to a composite option, except the payout is always cash settled and a fixed FX rate is used to determine the payout. Quanto options can be modelled using Black-Scholes. As the FX rate for the payout is fixed, quanto options are modelled using the normal volatility of the underlying (as FX volatility has no effect). The payout is simply the fixed FX rate multiplied by the price of a vanilla option with the same volatility, but a different carry. The carry (risk-free rate - dividend) to be used is shown below (the risk-free rate for quanto options is assumed to be the risk-free rate of the security currency, ie, it is not the same as for composite options).

$$c_{Quanto} = rfr_{Security} - d - \rho\sigma_{Security}\sigma_{FX}$$

$$\Rightarrow \quad d_{Quanto} = d + \rho\sigma_{Security}\sigma_{FX} \text{ as } d_{quanto} = rfr_{Security} - c_{Quanto}$$

where

c_{Quanto} = carry for quanto pricing

d_{Quanto} = dividend for quanto pricing

d = dividend yield

$rfr_{Security}$ = risk free rate of security currency

rfr_{Payout} = risk free rate of payout currency

Quanto options are either long or short correlation depending on the sign of the delta

The correlation between the FX and the security has an effect on quanto pricing, the direction (and magnitude) of which depends on the delta of the option. This is because the dividend risk of an option is equal to its delta, and the dividend used in quanto pricing increases as correlation increases.

Quanto option calls are short correlation (if FX is foreign currency in domestic terms)

As a call option is short dividends (call is an option on the price of underlying, not the total return of the underlying), a quanto call option is short correlation. A quanto put option is therefore slightly long correlation. In both cases, we assume the FX rate is the foreign security currency measured in domestic payout terms.

Intuitively, we can see a quanto call option is short correlation by assuming the dividend yield and both currency risk-free rates are all zero and comparing its value to a vanilla call option priced in the (foreign) security currency. If correlation is high, the vanilla call option is worth more than the quanto call option (as FX moves in favour of the investor if the price of the security rises). The reverse is also true (negative correlation causes a vanilla call option to be worth less than a quanto call option). As the price of a vanilla (single currency) call does not change due to the correlation of the underlying with the FX rate, this shows a quanto call option is short correlation.

This page left blank intentionally.

ADVANCED VOLATILITY TRADING

RELATIVE VALUE TRADING

Relative value is the name given to a variety of trades that attempt to profit from the mean reversion of two related assets that have diverged. The relationship between the two securities chosen can be fundamental (different share types of same company or significant cross-holding) or statistical (two stocks in same sector). Relative value can be carried out via cash (or delta-1), options or outperformance options.

TRADES ARE USUALLY CHOSEN ON CORRELATED ASSETS

Relative value trades attempt to profit from mean reversion

The payout of a relative value trade on two uncorrelated securities is completely random, and the investor on average gains no benefit. However, if two securities have a strong fundamental or statistical reason to be correlated, they can be thought of as trading in a similar direction with a random noise component. Assuming the correlation between the securities is sufficiently strong, the noise component should mean revert. Relative value trades attempt to profit from this mean reversion. There are five main types of relative value trades.

- **Dual listing.** If a share trades on different exchanges (eg, an ADR), the two prices should be equal. This is not always the case due to execution risk (different trading times) and perhaps due to indexation flow. Non-fungible shares or those with shorting restrictions are most likely to show the largest divergence in price. Of all relative value trades, dual-listing ones are likely to show the strongest correlation.

- **Share class.** If there is more than one type of share, perhaps with voting or ownership restrictions, then the price of these shares can diverge from one another. For example, preference shares typically have a higher dividend to compensate for lack of voting rights, but suffer from less liquidity and (normally) exclusion from equity indices. During special situations, for example, during the Porsche/VW saga, the difference in price between the two shares can diverge dramatically.

- **Cross-holding.** If one company (potentially a holding company) owns a significant amount of another company, the prices of the two companies will be linked. Sometimes putting on a cross-holding trade is difficult in practice due to the high borrow cost of the smaller company. This trade is also known as a stub trade when the investor wants pure exposure to the larger company, and hedges out the unwanted exposure to the equity holdings of the larger company. Potentially, these trades can occur when a larger company spins off a subsidiary but keeps a substantial stake post spin-off.

- **Event-driven.** In the event of a takeover that is estimated to have a significant chance of succeeding, the share prices of the acquiring and target company should be correlated. The target will usually trade at a discount to the bid price, to account for the probability the deals falls through (although if the offer is expected to be improved, or beaten by another bidder, the target could trade above the offer price).

- **Long-short.** If a long and short position is initiated in two securities that do not have one of the above four reasons to be correlated, it is a long-short trade. The correlation between the two securities of a long-short trade is likely to be lower than for other relative values trades. Because of this, often two stocks within a sector are chosen, as they should have a very high correlation and the noise component is likely to be bounded (assuming market share and profitability is unlikely to change substantially over the period of the relative value trade).

Long-short can focus returns on stock picking ability (which is c10% of equity return)

Only c20% of equity returns are due to stock picking (c10% is sector selection and c70% is the broad equity market)

General market performance is typically responsible for c70% of equity returns, while c10% is due to sector selection and the remaining c20% due to stock picking. If an investor wishes to focus returns on the proportion due to sector or stock picking, they can enter into a long position in that security and a short position in the appropriate market index (or vice versa). This will focus returns on the c30% due to sector and stock selection. Typically, relatively large stocks are selected, as their systematic risk (which should cancel out in a relative value trade) is usually large compared to specific risk. Alternatively, if a single stock in the same sector (or sector index) is used instead of the market index, then returns should be focused on the c20% due to stock picking within a sector.

SIZE OF POSITIONS SHOULD BE WEIGHTED BY BETA

If the size of the long-short legs are chosen to have equal notional (share price × number of shares × FX), then the trade will break even if both stock prices go to zero. However, the legs of the trade are normally weighted by beta to ensure the position is market neutral for more modest moves in the equity market. The volatility (historical or implied) of the stock divided by the average volatility of the market can be used as an estimate of the beta.

DELTA-1, OPTIONS AND OUTPERFORMANCE OPTIONS

Relative value trades can be implemented via cash/delta-1, vanilla options or outperformance options. They have very different trade-offs between liquidity and risk. No one method is superior to others, and the choice of which instrument to use depends on the types of liquidity and risk the investor is comfortable with.

Figure 64. Different Methods of Relative Value Trading

Asset Class	Position	Benefits	Disadvantages
Cash/delta-1	Long A, short B using stock/CFD, future, forwards, total return swap or ETF	High liquidity (volatility products might not be available)	Unlimited risk
Options	Long call on A, short call on B (or put/call spread/put spread)	Limited downside on long leg and convex payoff	Unlimited risk on short side (unless call spreads/put spreads)
Outperformance option	Long outperformance option on A vs B	Limited downside and convex payoff	Poor liquidity/wide bid-offer spreads

Source: Santander Investment Bolsa estimates.

(1) CASH/DELTA-1: BEST LIQUIDITY, BUT UNLIMITED RISK

The deepest and most liquid market for relative value trades is the cash (or delta-1) market. While there are limited restrictions in the size or stocks available, the trade can suffer potentially unlimited downside. While there are many similarities between cash or delta-1 instruments, there are also important differences.

Benefits of more beneficial taxation can be shared

For many delta-1 products, the presence of investors with more beneficial taxation can result in more competitive pricing. Products that have to be based in one location, such as ETFs, suffer from being unable to benefit from the different taxation of other investors.

Figure 65. Delta-1 Product Summary

	Listed / OTC	Counter party risk	Leverage	Maintenance	Benefit from borrow	Dividend risk	Cost	Notes	Overall
ETFs	Listed	Low	No (cash)	Low	No	No	Medium	Dividend taxation	
Futures	Listed	Low	Yes	Medium	Yes	Yes	Medium	Tracking error	
Forwards	OTC	Medium	Yes	Low	Yes	Yes	Low	Rare	
Total return swaps	OTC	Medium	Yes	Low	Yes	No	Low	Best OTC	
CFDs	OTC	Medium	Yes	Medium	No	No	Low	Similar to TRS	
Cash	Listed	Low	No (cash)	High	No	No	Low	Keep voting rights	
Certificates / p-notes	Listed	High (issuer)	No (cash)	Low	No	No	High	Flexible underlying	

Source: Santander Investment Bolsa.

CFDs have many advantages over stock

A relative value trade in the cash (stock) market can be initiated by a long stock position combined with a short stock position. The short stock position needs a functioning stock borrow market, as the stock needs to be borrowed before it can be sold short. Using stock can tie up a lot of capital, as the long position needs to be funded, as does the short position. Normally, a prime broker can help fund the position; however, for simplicity, CFDs are often used instead of stock.

As CFDs remove the overhead of corporate actions such as dividends, they are very popular with hedge funds that wish to quickly initiate long/short positions. As CFDs avoid paying stamp duty in certain countries, there can be yield enhancement benefits from using CFDs. While in theory a CFD has counterparty risk, a CFD is often created to be a stock equivalent with daily resets (exchange of cash flows) limiting this disadvantage. The main disadvantage of using CFDs is the loss of voting rights; however, relative value investors are not usually interested in voting. While stock and CFDs can be used to trade indices, this is rare as it usually requires more maintenance than other delta-1 products.

Total return swaps and forwards are best for indices

Total return swap is the best instrument for trading an index

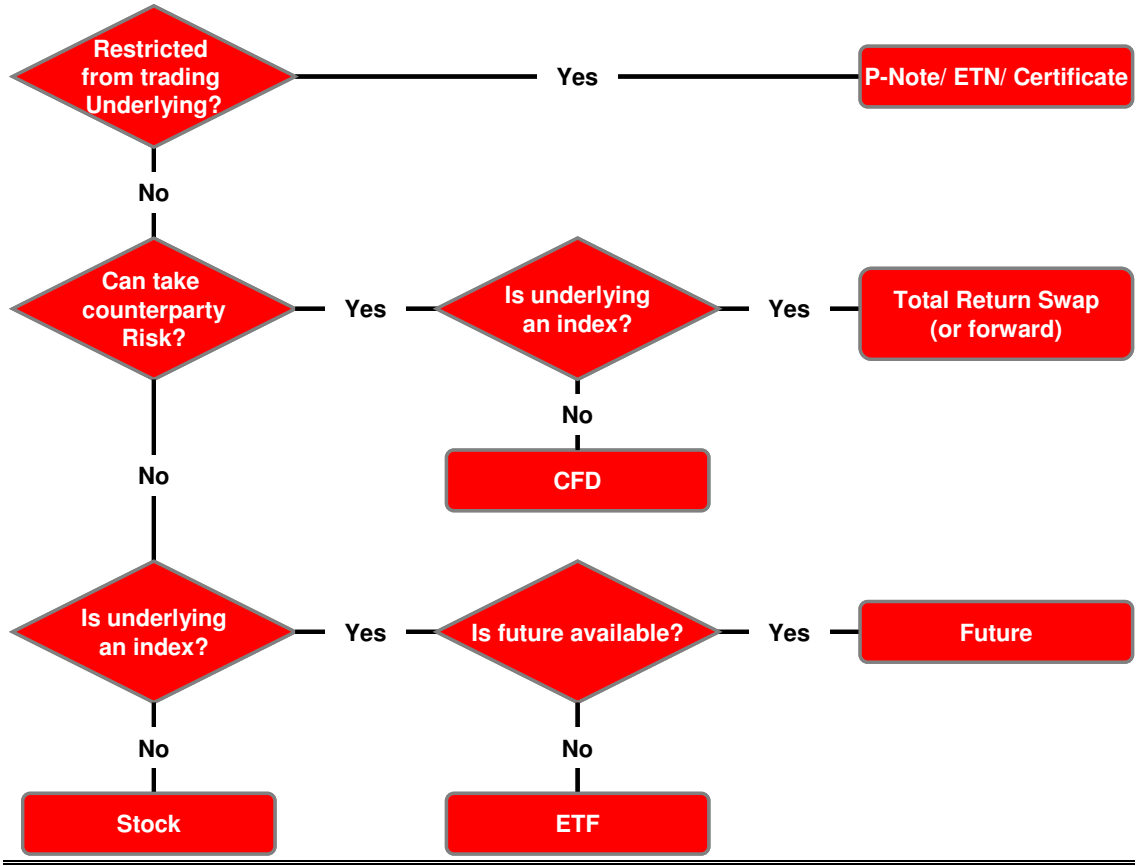
The index equivalent of a CFD is a total return swap. Potentially, a portfolio swap could be used instead of a total return swap to manage the long and short legs in one trade. A forward is essentially a price return equivalent of a total return swap and is normally only used instead of a TRS (total return swap) for internal reasons (for example, if IT systems can only handle forwards). Both futures and forwards benefit from the more optimal taxation treatment of other investors, allowing yield enhancement.

Futures and ETFs provide a listed method of trading indices with no counterparty risk

Future is the best listed (no counterparty risk) instrument for trading an index

A future is a listed equivalent of a forward (hence, it suffers from exchange fees) and benefits from the same yield enhancement factors. Futures often require slightly larger margins to take into consideration there is no counterparty risk. For investors requiring a listed instrument with no counterparty risk, futures are the best instrument for trading indices. There is, however, a significant maintenance cost to futures, as liquidity is concentrated on the front months and therefore requires rolling. While futures and forwards can be used to trade single stocks, they are usually used for indices (although in Europe futures can be crossed, allowing them to be used instead of stock). An ETF can be thought of as the index equivalent of a stock, being a fully funded listed instrument with no potential for yield enhancement. As there are more ETFs than futures, they can be used for a wider variety of underlyings.

Figure 66. Delta-1 Product Decision Tree



Source: Santander Investment Bolsa.

Certificates, ETNs and p-notes are effectively the same

Certificates/ETNs and participation notes (p-notes) are traded on an exchange and can give exposure to strategies, markets and currencies that an investor might normally be unable to invest in. For example, if an investor is prohibited from investing in volatility then a certificate or ETN that wraps a volatility strategy (call overwriting, selling one-month variance swaps, long VIX/vStoxx futures, etc) can be bought instead. Access to Chinese and Indian markets is not trivial, but can be traded via p-notes (as can trading in markets with restricted currencies, as the product can be redenominated in USD or another currency). As 100% upfront payment is required, certificate/ETNs and p-notes can be considered a fully funded equity swap (total or price return) with a listed price. Despite being listed, the investor is 100% exposed to the credit risk of the counterparty.

(2) OPTIONS: CONVEX PAYOFF AND CAN LIMIT DOWNSIDE ON LONG LEG

Long leg of relative value trade is often rotated into options when long leg becomes profitable

Options can be used in place of stock or delta-1 for either the long or short leg, or potentially both. Options offer convexity, allowing a position to profit from the expected move while protecting against the potentially unlimited downside. Often a relative value trade will be put on in the cash/delta-1 market, and the long leg rotated into a call once the long leg is profitable (in order to protect profits). While volatility is a factor in determining the attractiveness of using options, the need for safety or convexity is normally the primary driver for using options (as relative value traders do not delta hedge, the change in implied volatility is less of a factor in profitability than the delta/change in equity market). Investors who are concerned about the cost of options can cheapen the trade by using call spreads or put spreads in place of vanilla calls or puts.

Weighting options by volatility is similar to weighting by beta and roughly zero cost

The most appropriate weighting for two relative value legs is beta weighting the size of the delta hedge of the option (ie, same beta × number of options × delta × FX), rather than having identical notional (share price × number of options × FX). Beta weighting ensures the position is market neutral. Volatility weighting can be used as a substitute for beta weighting, as volatility divided by average volatility of the market is a reasonable estimate for beta. Volatility weighting ATM (or ATMf) options is roughly zero cost, as the premium of ATM options is approximately linear in volatility.

Choosing strike and maturity of option is not trivial

One disadvantage of using options in place of equity is the need to choose a maturity, although some investors see this as an advantage as it forces a view to be taken on the duration or exit point of the trade at inception. If the position has to be closed or rolled before expiry, there are potentially mark-to-market risks. Similarly, the strike of the option needs to be chosen, which can be ATM (at the money), ATMf (ATM forward), same percentage of spot/forward or same delta. Choosing the same delta of an OTM option means trading a strike further away from spot/forward for the more volatile asset (as delta increases as volatility increases). We note that trading the same delta option is not the same as volatility weighting the options traded as delta is not linear in volatility. Delta also does not take into account the beta of the underlyings.

(3) OUTPERFORMANCE OPTIONS: LIMITED DOWNSIDE BUT LOW LIQUIDITY

Outperformance options are ideally suited to relative value trades, as the maximum loss is the premium paid and the upside is potentially unlimited. However, outperformance options suffer from being relatively illiquid. While pricing is normally cheaper than vanilla options (for normal levels of correlation), it might not be particularly appealing depending on the correlation between the two assets. As there are usually more buyers than sellers of outperformance options, some hedge funds use outperformance options to overwrite their relative value trades.

RELATIVE VALUE VOLATILITY TRADING

Relative value volatility trades can impact the volatility market

Volatility investors can trade volatility pairs, in the same way as trading equity pairs. For indices, this can be done via options, variance swaps or futures on a volatility index (such as the VIX or vStoxx). The wider bid-offer spreads for single stocks reduces the attractiveness of single-stock relative value volatility trading. For indices that are popular volatility trading pairs, if they have significantly different skews this can impact the volatility market.

Implied volatility spread between pairs is kept stable

While skew of different indices is dependent on correlation, traders tend to keep the spreads of volatility between indices similar. This is why if equity markets move down, the implied volatility of the S&P500 or FTSE (as they are large diversified indices that hence have high skew) tends to come under pressure, while the implied volatility of country indices with fewer members, such as the DAX, are likely to be supported. The SX5E tends to lie somewhere in between, as it has fewer members than the S&P500 or FTSE but is more diverse than other European country indices. Should markets rise, the reverse tends to occur (high skew indices implieds lifted, low skew implieds are weighed on).

TRADING EARNINGS ANNOUNCEMENTS/JUMPS

From the implied volatilities of near dated options, it is possible to calculate the implied jump on key dates. Trading these options in order to take a view on the likelihood of unanticipated (low or high) volatility on reporting dates is a very common strategy. We examine the different methods of calculating the implied jump, and show how the jump calculation should normalise for index term structure.

TOTAL VOLATILITY = DIFFUSIVE VOLATILITY + JUMP VOLATILITY

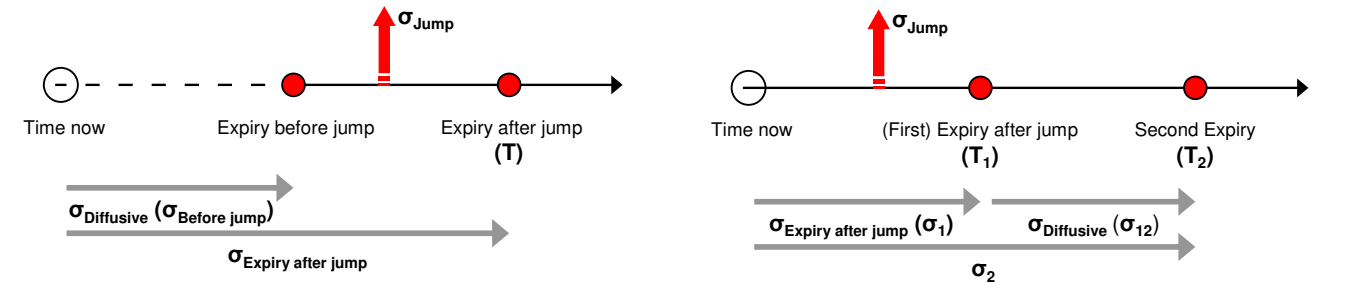
Expiry to trade jump is the expiry just after the event (can hedge by shorting expiry before event if available)

While stock prices under Black-Scholes are modelled as having a GBM (Geometric Brownian Motion) with constant volatility, in reality there are certain dates where there is likely to be more volatility than average. These key dates are usually reporting dates, but could also coincide with conference dates or investor days (in fact, any day where material non-public information is released to the public). The implied volatility of an option whose expiry is after a key date can be considered to be the sum of the normal diffusive volatility (normal volatility for the stock in the absence of any significantly material information being released) and the volatility due to the anticipated jump on the key date. While options of any expiry after the key date could be used, we shall assume the expiry chosen is the expiry just after the key date (to ensure the greatest percentage of the options' time value is associated with the jump). This position can be hedged by shorting the expiry before the key date, if one exists.

ESTIMATING DIFFUSIVE VOLATILITY IS NOT TRIVIAL

In order to calculate the implied jump due to a key date, the diffusive (normal) volatility of the stock needs to be estimated. While the diffusive volatility could be estimated by looking at historical volatility, it is usual to look at implied volatility (as there are several measures of historical volatility, but only one implied volatility). If there is an option that expires just before the key date, then the implied volatility of this option can be used. If not, the forward volatility after the key date is used as the estimate for the normal volatility of the security.

Figure 67. Diffusive Assumption Using Implied Vol **Diffusive Assumption Using Forward Vol**



Source: Santander Investment Bolsa estimates.

Implied jumps normally calculated for near-dated events

Implied jumps are normally only calculated for near-dated events, as the effect of the jump tends to be too diluted for far dated expiries (and hence would be untradeable taking bid-offer spreads into account). Forward starting options could be used to trade jumps after the first expiry, but the wider bid-offer spread could be greater than potential profits.

Forward volatility can be calculated with implied of two options

The calculation for forward volatility is derived from the fact variance (time weighted) is additive. The formula is given below (σ_x is the implied volatility for options of maturity T_x).

$$\sigma_{12} = \sqrt{\frac{\sigma_2{}^2 T_2 - \sigma_1{}^2 T_1}{T_2 - T_1}} = \text{forward volatility } T_1 \text{ to } T_2$$

JUMP VOLATILITY CAN BE CALCULATED FROM DIFFUSIVE VOLATILITY

Volatility just after reporting tends to be ¾ of the volatility just before reporting

As variance is additive, the volatility due to the jump can be calculated from the total volatility and the diffusive volatility. We note this assumes the normal diffusive volatility is constant, whereas volatility just after a reporting date is, in fact, typically ¾ of the volatility just before a reporting date (as the previously uncertain earnings are now known).

$$\sigma_{\text{Expiry after jump}}{}^2 T = \sigma_{\text{Jump}}{}^2 + \sigma_{\text{Diffusive}}{}^2 (T-1)$$

$$\Rightarrow \quad \sigma_{\text{Jump}} = \sqrt{(\sigma_{\text{Expiry after jump}}{}^2 T - \sigma_{\text{Diffusive}}{}^2 (T-1)}$$

where

$\sigma_{\text{Expiry after jump}}$= implied volatility of option whose expiry is after the jump

T = time to the expiry after jump (= T_1)

$\sigma_{\text{Diffusive}}$ = diffusive volatility ($\sigma_{\text{Before jump}}$ if there is an expiry before the jump, if not it is σ_{12})

σ_{Jump}= implied volatility due to the jump

IMPLIED JUMP CALCULATED FROM JUMP VOLATILITY

From the above implied volatility due to jump (σ_{Jump}) it is possible to calculate the implied daily return on the day of the jump (which is a combination of the normal daily move and the effect of the jump).

$$\text{Expected daily return} = e^{\frac{\sigma_{\text{Jump}}{}^2}{2}} \left[(2 \times N(\sigma_{\text{Jump}}) - 1) \right]$$

A proof of the above formula is below.

$$\Delta S = S_1 - S_0 = S_0 \left(\frac{S_1}{S_0} - 1 \right) = S_0 \left(e^r - 1 \right) \text{ where } r = Ln \left(\frac{S_1}{S_0} \right)$$

$$\Rightarrow \quad \text{E} \left(\frac{|\Delta S|}{S_0} \right) = \text{E} \left(\left| e^r - 1 \right| \right)$$

$$\Rightarrow \quad \text{Expected daily return} = \text{E} \left(\left| e^r - 1 \right| \right)$$

as r is normally distributed

$$\Rightarrow \text{Expected daily return} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \left(e^r - 1\right) e^{\frac{-r^2}{2\sigma^2}} dr$$

$$\Rightarrow \text{Expected daily return} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{\infty} \left(e^r - 1\right) e^{\frac{-r^2}{2\sigma^2}} dr + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{0} \left(e^r - 1\right) e^{\frac{-r^2}{2\sigma^2}} dr$$

$$\Rightarrow \text{Expected daily return} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{\infty} \left(e^r - e^{-r}\right) e^{\frac{-r^2}{2\sigma^2}} dr$$

if we define x such that r = xσ

$$\Rightarrow \text{Expected daily return} = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \left(e^{x\sigma} - e^{-x\sigma}\right) e^{\frac{-x^2}{2}} dx$$

$$\Rightarrow \text{Expected daily return} = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{x\sigma - \frac{x^2}{2}} - e^{-x\sigma - \frac{x^2}{2}} dx$$

$$\Rightarrow \text{Expected daily return} = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{x\sigma - \frac{x^2}{2}} dx - \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-x\sigma - \frac{x^2}{2}} dx$$

$$\Rightarrow \text{Expected daily return} = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{\frac{-(x-\sigma)^2}{2} + \frac{\sigma^2}{2}} dx - \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{\frac{-(x+\sigma)^2}{2} + \frac{\sigma^2}{2}} dx$$

$$\Rightarrow \text{Expected daily return} = e^{\frac{\sigma^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{\frac{-(x-\sigma)^2}{2}} dx - e^{\frac{\sigma^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{\frac{-(x+\sigma)^2}{2}} dx$$

$$\Rightarrow \text{Expected daily return} = e^{\frac{\sigma^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{\frac{-(x-\sigma)^2}{2}} dx - e^{\frac{\sigma^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{\frac{-(x+\sigma)^2}{2}} dx$$

$$\Rightarrow \text{Expected daily return} = e^{\frac{\sigma^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sigma} e^{\frac{-x^2}{2}} dx - e^{\frac{\sigma^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{\sigma}^{\infty} e^{\frac{-x^2}{2}} dx$$

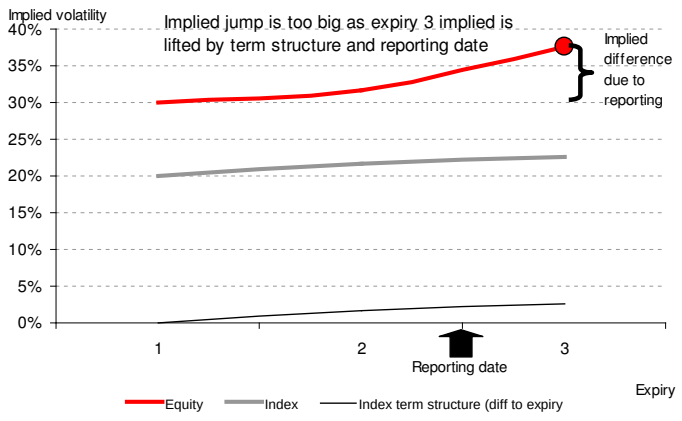
$$\Rightarrow \text{Expected daily return} = e^{\frac{\sigma^2}{2}} N(\sigma) - e^{\frac{\sigma^2}{2}} \left[1 - N(\sigma)\right]$$

$$\Rightarrow \text{Expected daily return} = e^{\frac{\sigma^2}{2}} \left[2 \times N(\sigma) - 1\right]$$

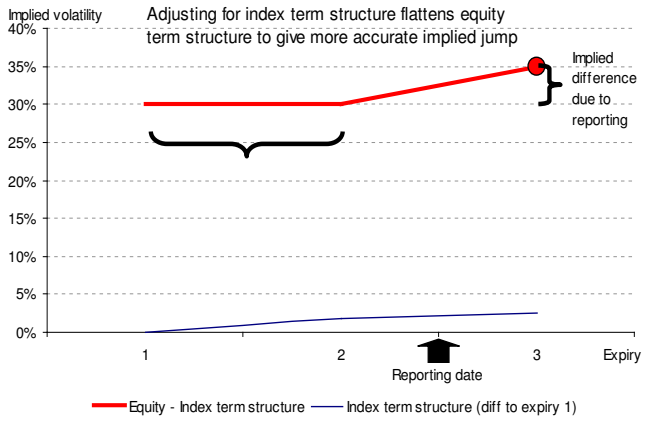
EQUITY TERM STRUCTURE CAN BE ADJUSTED BY INDEX TERM STRUCTURE

The methodology for extracting jumps from the difference between the front month implieds is simply a case of mathematics, assuming the volatility of a stock is equal to a 'normal' volatility on every day plus an 'abnormal' jump on reporting. In order to calculate the 'abnormal' jump, we need to estimate the 'normal' volatility, and this estimate usually requires a flat term structure to be assumed. If the index term structure is used to adjust the single-stock term structure, then a more accurate implied jump can be calculated [23](assuming the single-stock term structure would be identical to index term structure without the effect of a reporting date). For simplicity, the diagrams below assume reporting is between expiry 2 and 3, but the effect will be similar if earnings is between expiry 1 and 2.

Figure 68. Equity and Index Term Structure



Equity Term Structure Adjusted by Index Term Structure



Source: Santander Investment Bolsa estimates.

[23] This assumes a flat implied correlation term structure, which is a reasonable assumption for the very near-dated expiries.

STRETCHING BLACK-SCHOLES ASSUMPTIONS

The Black-Scholes model assumes an investor knows the future volatility of a stock, in addition to being able to continuously delta hedge. In order to discover what the likely profit (or loss) will be in reality, we stretch these assumptions. If the future volatility is unknown, the amount of profit (or loss) will vary depending on the path, but buying cheap volatility will always show some profit. However, if the position is delta-hedged discretely, the purchase of cheap volatility may reveal a loss. The variance of discretely delta-hedged profits can be halved by hedging four times as frequently. We also show why traders should hedge with a delta calculated from expected – not implied – volatility, especially when long volatility.

BLACK-SCHOLES ASSUMES KNOWN VOL AND CONTINUOUS HEDGING

While there are a number of assumptions behind Black-Scholes, the two which are the least realistic are: (1) a continuous and known future realised volatility; and (2) an ability to delta hedge continuously. There are, therefore, four different scenarios to investigate. We assume that options are European (can only be exercised at maturity), although most single-stock options are American (can be exercised at any time).

- **Continuous delta hedging with known volatility.** In this scenario, the profit (or loss) from volatility trading is fixed. If the known volatility is constant, then the assumptions are identical to Black-Scholes. Interestingly, the results are the same if volatility is allowed not to be constant (while still being known).

- **Continuous delta hedging with unknown volatility.** If volatility is unknown, then typically traders hedge with the delta calculated using implied volatility. However, as implied volatility is not a perfect predictor of future realised volatility, this causes some variation in the profit (or loss) of the position. However, with these assumptions, if realised volatility is above the implied volatility price paid, it is impossible to suffer a loss.

- **Discrete delta hedging with known volatility.** As markets are not open 24/7, continuous delta hedging is arguably an unreasonable assumption. The path dependency of discrete delta hedging adds a certain amount of variation in profits (or losses), which can cause the purchase of cheap volatility (implied less than realised) to suffer a loss. The variance of the payout is inversely proportional to the frequency of the delta hedging. For example, the payout from hedging four times a day has a variance that is a quarter of the variance that results if the position is hedged only once a day. The standard deviation is therefore halved if the frequency of hedging is quadrupled (as standard deviation squared = variance).

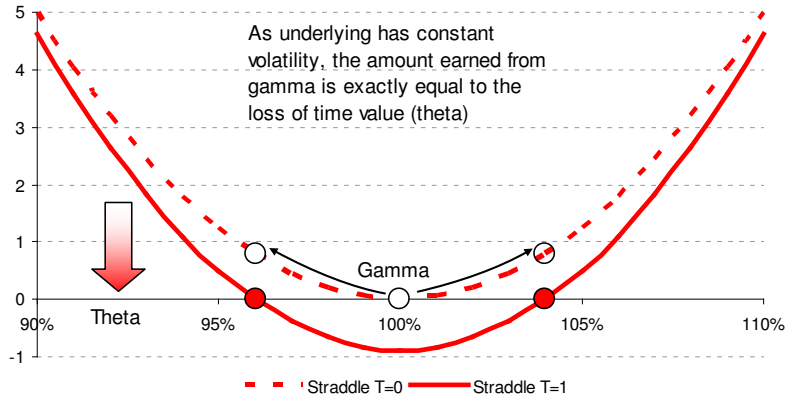
- **Discrete delta hedging with unknown volatility.** The most realistic assumption is to hedge discretely with unknown volatility. In this case, the payout of volatility trading is equal to the sum of the variance due to hedging with unknown volatility plus the variance due to discretely delta hedging.

CONTINUOUS DELTA HEDGING WITH KNOWN VOLATILITY

With continuous and known volatility, the correct delta can be calculated

In a Black-Scholes world, the volatility of a stock is constant and known. While a trader is also able to continuously delta hedge, Figure 69 below will assume we hedge discretely but in an infinitesimally small amount of time. In each unit of time, the stock can either go up or down. As the position is initially delta-neutral (ie, delta is zero), the gamma (or convexity) of the position gives it a profit for both downward and upward movements. While this effect is always profitable, the position does lose time value (due to theta). If an option is priced using the actual fixed constant volatility of the stock, the two effects cancel each other and the position does not earn an abnormal profit or loss as the return is equal to the risk-free rate. There is a very strong relationship between gamma and theta (theta pays for gamma)[24].

Figure 69. Constant and Known Realised Volatility to Calculate Delta



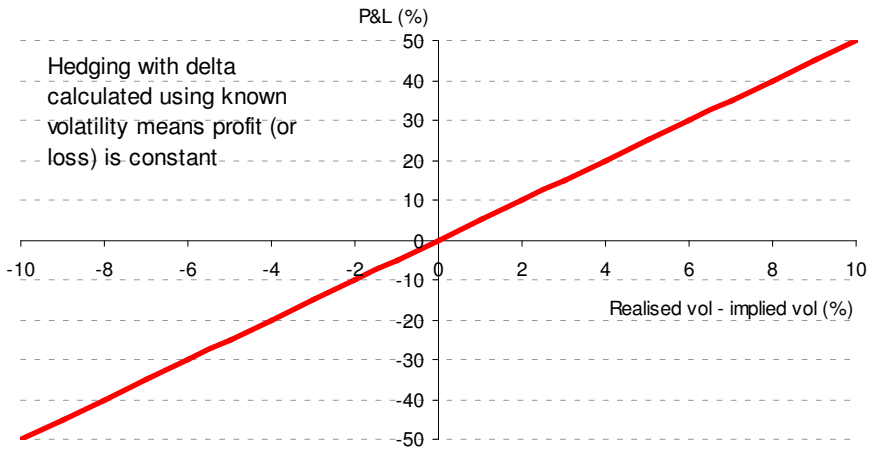
Source: Santander Investment Bolsa.

Profit from delta hedging is equal to the difference between price and theoretical price

The theoretical price of an option, using the known volatility, can be extracted by delta hedging. Should an option be bought at an implied volatility less than realised volatility, the difference between the theoretical price and the actual price will equal the profit of the trade. Figure 70 shows the profit vs the difference in implied and realised volatility. As there is no path dependency, the profit (or loss) of the trade is fixed and cannot vary.

[24] They are not perfectly correlated, due to the interest paid on borrowing the shares (which varies with spot).

Figure 70. Profit (or Loss) from Continuously Delta-Hedging Known Volatility



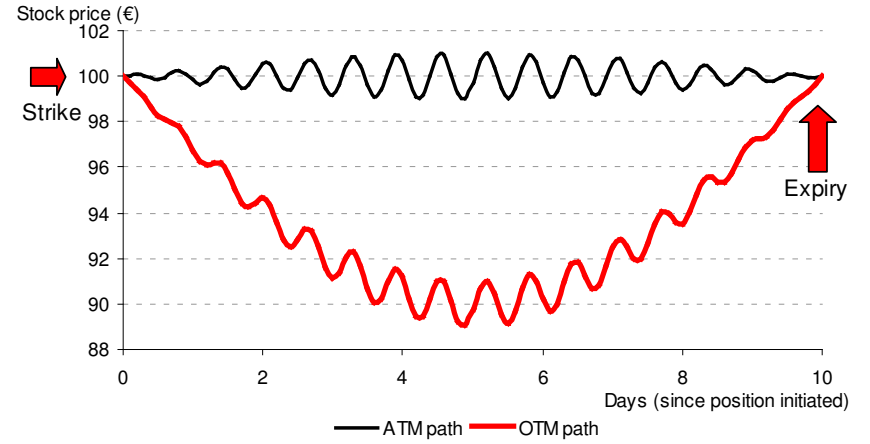
Source: Santander Investment Bolsa.

As theta and gamma are either both high or both low, profits are not path dependent

Delta hedging: an option that remains ATM and earns more gamma but pays more theta

If a position is continuously delta hedged with the correct delta (calculated from the known future volatility over the life of the option), then the payout is not path dependent. Figure 71 below shows two paths with equal volatility and the same start and end point. Even though one path is always ATM while the other has most volatility OTM, delta hedging gives the same profit for both. The cause of this relationship is the fact that, while the ATM option earns more due to delta hedging, the total theta cost is also higher (and exactly cancels the delta hedging profit).

Figure 71. Two Security Paths with Identical Volatility, Start and End Points



Source: Santander Investment Bolsa.

Profits are path independent, even if volatility is not constant (but still known)

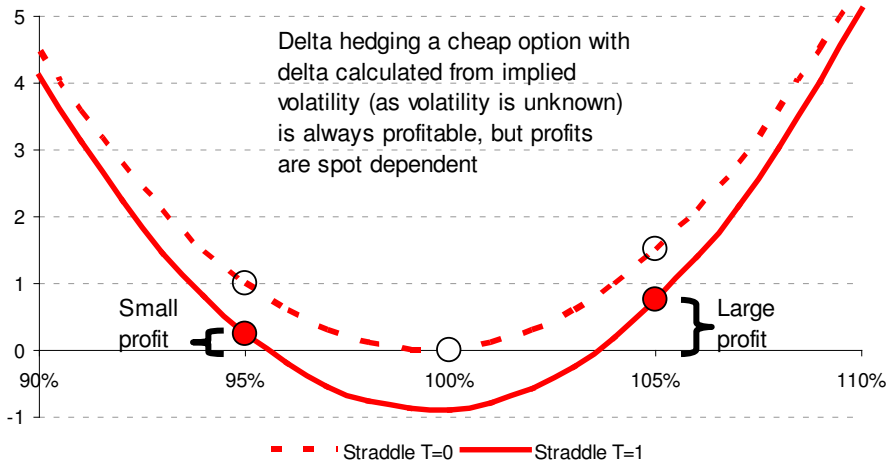
While Black-Scholes assumes a constant known volatility, there are similar results for non-constant known volatility. This result is due to the fact that a European option payout depends only on the stock price at expiry. Therefore, the volatility over the life of the option is the only input to pricing. The timing of this volatility is irrelevant.

CONTINUOUS DELTA HEDGING WITH UNKNOWN VOLATILITY

If volatility is unknown, the correct delta cannot be calculated

As it is impossible to know in advance what the future volatility of a security will be, the implied volatility is often used to calculate deltas. Delta hedging using this estimate causes the position to have equity market risk and, hence, it becomes path dependent (although the average or expected profit remains unchanged). Figure 72 shows that the profits from delta hedging are no longer independent of the direction in which the underlying moves. The fact that there is a difference between the correct delta (calculated using the remaining volatility to be realised over the life of the option) and the delta calculated using the implied volatility means returns are dependent on the direction of equity markets.

Figure 72. Profit from Cheap Options Is Not Constant if Volatility Is Not Known

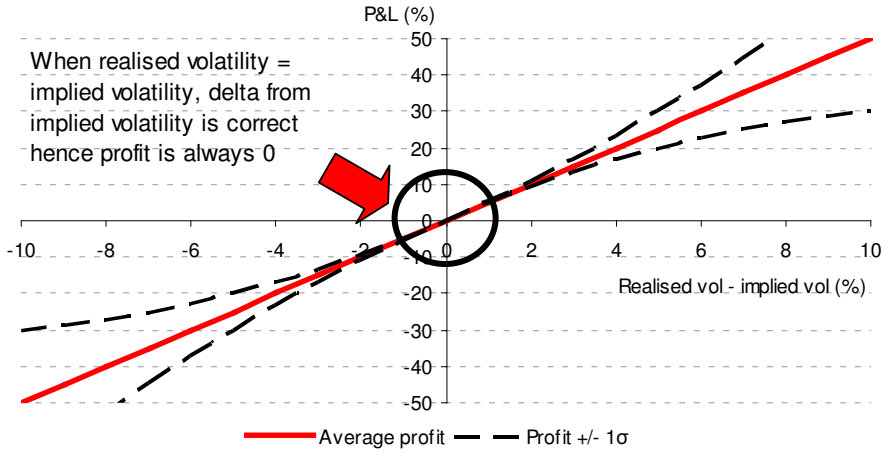


Source: Santander Investment Bolsa.

If implied volatility = realised volatility, profits are path independent

If the implied volatility is equal to the realised volatility, then the estimated delta calculated from the implied will be equal to the actual delta (calculated from the realised). In this case, profits from hedging will exactly match the theta cost for all paths, so it is path independent.

Figure 73. Profit (or Loss) from Continuously Delta Hedging Unknown Volatility



Source: Santander Investment Bolsa.

With continuous hedging, buying a cheap option is always profitable

If there is a difference between the actual delta and estimated delta, there is market risk but not enough to make a cheap option unprofitable (or an expensive option profitable). This is because in each infinitesimally small amount of time a cheap option will always reveal a profit from delta hedging (net of theta), although the magnitude of this profit is uncertain. The greater the difference between implied and realised, the greater the market risk and the larger the potential variation in profit.

DISCRETELY DELTA HEDGING WITH KNOWN VOLATILITY

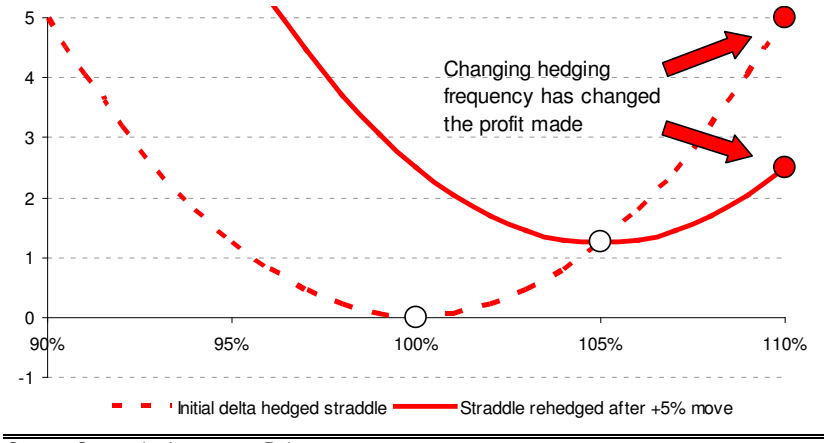
While assuming continuous delta hedging is mathematically convenient, it is impossible in practice. Issues such as the cost of trading and minimum trading size (even if this is one share) make continuous trading impossible, as do fundamental reasons, such as trading hours (if you cannot trade 24 hours then it is impossible to trade overnight and prices can jump between the close of one day and start of another) and weekends.

Discrete hedging errors can be reduced by increasing the frequency of hedging

Discrete delta hedging adds noise to returns

The more frequent the discrete hedging, the less variation in the returns. If 24-hour trading were possible, then with an infinite frequency of hedging with known volatility the returns converge to the same case as continuous hedging with known volatility (ie, Black-Scholes). In order to show how the frequency of hedging can affect the payout of delta hedging, we shall examine hedging for every 5% and 10% move in spot.

Figure 74. Profit from Discrete Delta Hedging with Different Frequencies



Source: Santander Investment Bolsa.

Hedging every 5% move in spot

If an investor delta hedges every 5% move in spot, then an identical profit is earned if the underlying rises 10% as if the underlying rises 5% and then returns to its starting point. This shows that the hedging frequency should ideally be frequent enough to capture the major turning points of an underlying.

Hedging every 10% move in spot

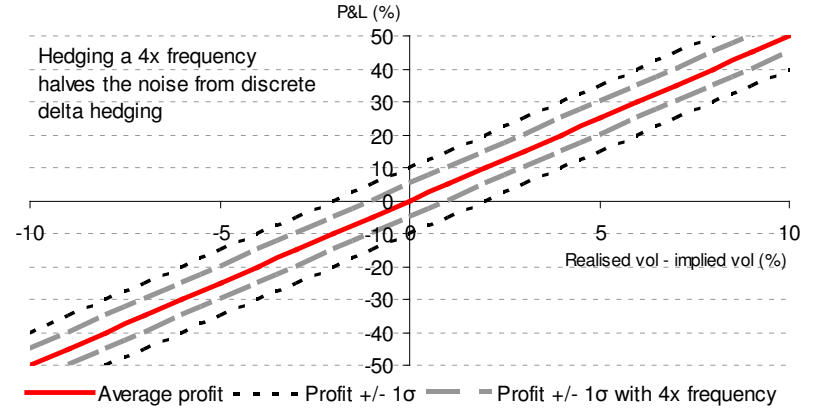
If the investor is hedged for every 10% move in the underlying, then no profit will be earned if the underlying rises 5% and then returns to its starting point. However, if the underlying rises 10%, a far larger profit will be earned than if the position was hedged every 5%. This shows that in trending markets it is more profitable to let positions run than to re-hedge them frequently.

Hedging error is independent of average profitability of trade

Noise from discrete delta hedging is independent of how cheap the option is

As the volatility of the underlying is known, there is no error due to the calculation of delta. As the only variation introduced is essentially 'noise', the size of this noise, or variation, is independent from the average profitability (or difference between realised vol and implied vol) of the trade.

Figure 75. Profit (or Loss) from Discrete Delta Hedging Known Volatility



Source: Santander Investment Bolsa.

With discrete hedging, cheap options can lose money

With continuous delta hedging (with known or unknown volatility) it is impossible to lose money on a cheap option (an option whose implied volatility is less than the realised volatility over its life). However, as the error from discrete hedging is independent from the profitability of the trade, it is possible to lose money on a cheap option (and make money on an expensive option).

Hedging error is halved if frequency of hedging increased by factor of four

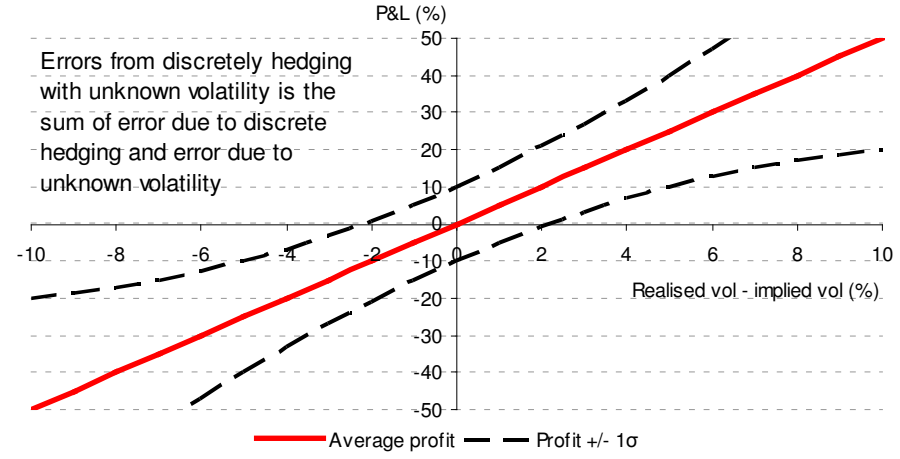
The size of the hedging error can be reduced by increasing the frequency of hedging. An approximation (shown below) is that if the frequency of hedging is increased by a factor of four, the hedging error term halves. This rule of thumb breaks down for very high-frequency hedging, as no frequency of hedging can eliminate the noise from non-24x7 trading (it will always have noise, due to the movement in share prices from one day's close to the next day's open).

$$\sigma_{P\&L} \approx \sigma \times vega \times \sqrt{\frac{\pi}{4N}}$$ where N is the number of times position is hedged in a year

DISCRETE DELTA HEDGING WITH UNKNOWN VOLATILITY

The most realistic assumption for profitability comes from the combination of discrete delta hedging and unknown volatility. Trading hours and trading costs are likely to limit the frequency at which a trader can delta hedge. Equally, the volatility of a stock is unknown, so implied volatility is likely to be used to calculate the delta. The variation in the profit (or loss) is caused by the variation due to discrete hedging and the inaccuracy of the delta (as volatility is unknown). Figure 76 shows this combined variation in profit (or loss) and, as for discrete hedging with known volatility, it is possible for a delta hedged cheap option to reveal a loss.

Figure 76. Profit (or Loss) from Discrete Delta Hedging with Unknown Volatility



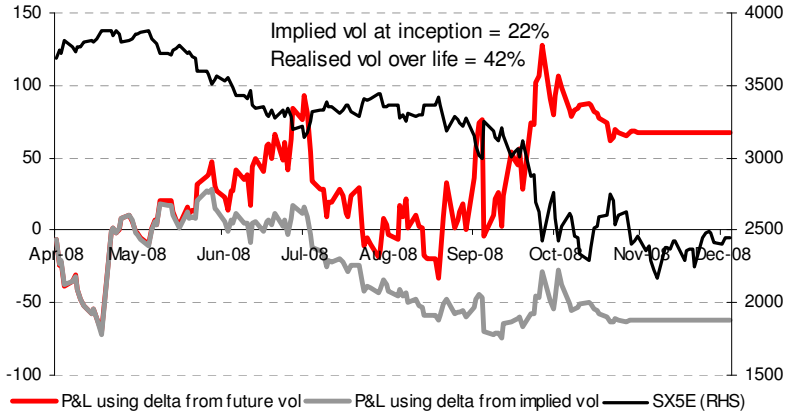
Source: Santander Investment Bolsa.

USE EXPECTED, NOT IMPLIED VOL FOR DELTA CALCULATIONS

As an example, let us assume a Dec08 SX5E ATM straddle was purchased in April 2008. In theory, it should be very profitable as the realised volatility of 39% was more than 50% above the 25% implied. However, most of the volatility came after the Lehman bankruptcy, which occurred towards the end of the option's life. If implied volatility was used to calculate the delta, then the time value would be assumed to be near zero. As equity markets had declined since April, the strike of the straddle would be above spot, hence we would have a delta ≈ -100% (the call would be OTM with a delta ≈ 0, while the put would be ITM with a delta ≈ 100%). To be delta-hedged, the investor would then buy 100% of the underlying per straddle. If the delta was calculated using the actual volatility (which was much higher), then the time value would be higher and the delta greater than -100% (eg, -85%). As the delta-hedged investor would have bought less than 100% of the underlying per straddle, this position outperformed hedging with implied volatility when the market fell after Lehman collapsed (as delta was lower, so less of the underlying was bought).

These results can be seen in Figure 77, which gives a clear example of why traders should hedge with the delta calculated from expected volatility rather than implied volatility. Because of the extreme volatility at the end of 2008, the two deltas differed at times by 24% (60% vs 84%).

Figure 77. Payout from Delta Hedging with Implied vs Realised Volatility



Source: Santander Investment Bolsa.

Hedging with delta using implied volatility is bad for long volatility strategies

Using expected, not implied, volatility to calculate delta is most important for long volatility strategies

Typically, when volatility rises there is often a decline in the markets. The strikes of the option are therefore likely to be above spot when actual volatility is above implied. This reduces the profits of the delta-hedged position as the position is actually long delta when it appears to be delta flat. Alternatively, the fact that the position hedged with the realised volatility over the life of the option is profitable can be thought of as due to the fact it is properly gamma hedged, as it has more time value than is being priced into the market. Hence, if a trader buys an option when the implied looks 5pts too cheap, then the hedge using delta should be calculated from a volatility 5pts above current implied volatility. Using the proper volatility means the profit is approximately the difference between the theoretical value of the option at inception (ie, using actual realised volatility in pricing) and the price of the option (ie, using implied volatility in pricing).

This page left blank intentionally.

SKEW AND TERM STRUCTURE TRADING

SKEW AND TERM STRUCTURE ARE LINKED

When there is an equity market decline, there is normally a larger increase in ATM implied volatility at the near end of volatility surfaces than the far end. Assuming sticky strike, this causes near-dated skew to be larger than far-dated skew. The greater the term structure change for a given change in spot, the higher skew is. Skew is also positively correlated to term structure (this relationship can break down in panicked markets). For an index, skew (and potentially term structure) is also lifted by the implied correlation surface. Diverse indices tend to have higher skew for this reason, as the ATM correlation is lower (and low strike correlation tends to 100% for all indices).

SKEW AND TERM STRUCTURE CUT SURFACE IN DIFFERENT DIMENSIONS

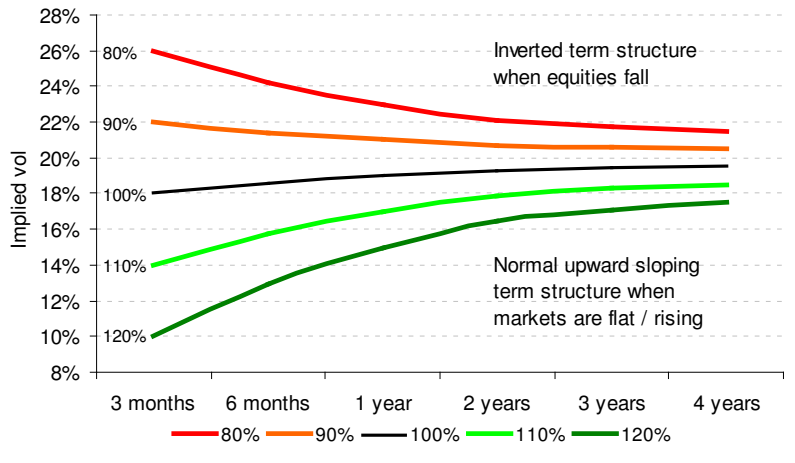
A volatility surface has three dimensions (strike, expiry and implied volatility), which is difficult to show on a two dimensional page. For simplicity, a volatility surface is often plotted as two separate two dimensional graphs. The first plots implied volatility vs expiry (similar to the way in which a yield curve plots credit spread against expiry) in order to show term structure (the difference in implied volatility for options with different maturities and the same strike). The second plots implied volatility vs strike to show skew (the difference in implied volatility for options with different strikes and the same maturity). We examine a volatility surface in both these ways (ie, term structure and skew) and show how they are related.

TERM STRUCTURE IS NORMALLY UPWARD SLOPING

Near-dated implieds are more volatile than far-dated implieds

When there is a spike in realised volatility, near-dated implied volatility tends to spike in a similar way (unless the spike is due to a specific event such as earnings). This is because the high realised volatility is expected to continue in the short term. Realised volatility can be expected to mean revert over a c8-month period, on average. Hence far-dated implied volatilities tend to rise by a smaller amount than near-dated implied volatilities (as the increased volatility of the underlying will only last a fraction of the life of a far-dated option). Near-dated implieds are therefore more volatile than far-dated implieds. The theoretical term structure for different strikes is shown in Figure 78, which demonstrates that near-dated implieds are more volatile. We have shown ATM (100%) term structure as upward sloping as this is how it trades on average (for the same reasons credit spread term structure is normally upward sloping, ie, risk aversion and supply-demand imbalances for long maturities).

Figure 78. Term Structure for Options of Different Strikes



Source: Santander Investment Bolsa.

If equity markets decline, term structure becomes inverted

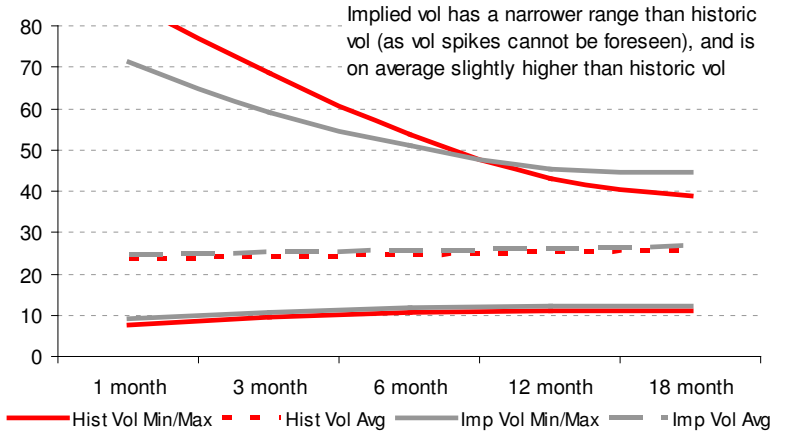
ATM term structure is typically positive in stable or rising markets, but negative in declining markets

Typically, an increase in volatility tends to be accompanied by a decline in equity markets, while a decline in volatility tends to occur in periods of calm or rising markets. If volatility surfaces are assumed not to move as spot moves (ie, sticky strike), then this explains why the term structure of low strike implied volatility is normally downward sloping (as the 80% strike term structure will be the ATM term structure when equities fall 20%). Similarly, this explains why the term structure of high strike implieds is normally upward sloping (as the 120% strike term structure will become the ATM term structure when equities rise 20%).

Slope of rising term structure is shallower than slope of inverted term structure

While Figure 78 shows the term structure of a theoretical volatility surface, in practice the slope of rising term structure is shallower than the slope of inverted term structure. This can be seen by looking at a volatility cone (Figure 79). Despite the fact that the inverted term structure is steeper, the more frequent case of upward sloping term structure means the average term structure is slightly upward sloping[25].

Figure 79. Implied and Historic Volatility Cone (SX5E since 2006)



Source: Santander Investment Bolsa.

Implied volatility is usually greater than realised volatility and less volatile

While historic and realised volatility are linked, there are important differences which can be seen when looking at empirical volatility cones. Average implied volatility lies slightly above average realised volatility, as implieds are on average slightly expensive. Implied volatility is also less volatile (it has a smaller min-max range) than realised volatility for near-dated maturities. This is because implieds are forward looking (ie, similar to an average of possible outcomes) and there is never 100% probability of the maximum or minimum possible realised. This effect fades for longer maturities, potentially due to the additional volatility caused by supply-demand imbalances (eg, varying demand for put protection). This causes inverted implied volatility term structure to be less steep than realised volatility term structure.

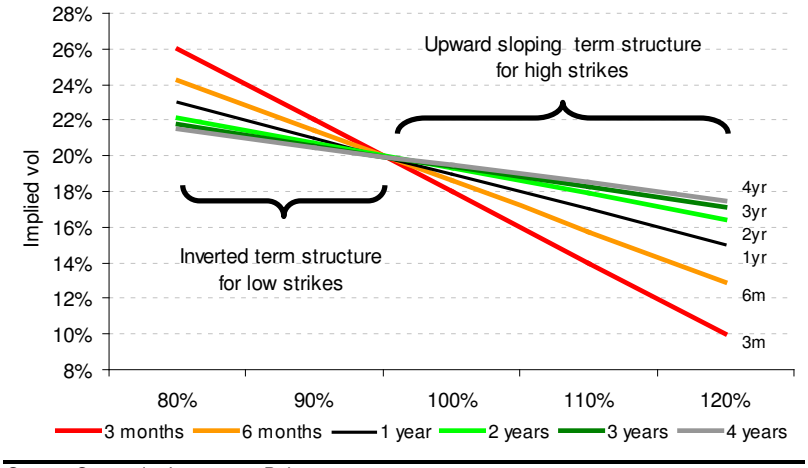
[25] Positive implied correlation term structure will also lift index term structure relative to single stock.

SKEW IS INVERTED AND IS HIGHER FOR NEAR-DATED EXPIRIES

Near-term skew is steeper than far-dated skew as near-dated ATM is more volatile

Assuming volatility surfaces stay constant (ie, sticky strike), the effect of near-dated ATM implieds moving further than far-dated implieds for a given change in spot is priced into volatility surfaces by having a larger near-dated skew. The example data given in Figure 78 is plotted in Figure 80 with a change of axes to show skew for options of different maturity. This graph shows that near-dated implieds have higher skew than far-dated implieds. The more term structure changes for a given change in spot, the steeper skew is. As near-dated ATM volatility is more volatile than far-dated ATM volatility, near-dated implied volatility has higher skew.

Figure 80. Skew for Options of Different Maturity



Source: Santander Investment Bolsa.

Skew for equities is normally inverted

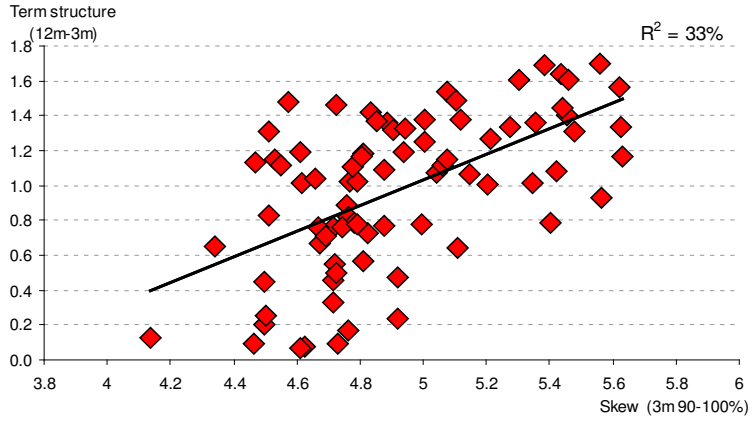
Unless there is a high likelihood of a significant jump upwards (eg, if there were a potential takeover event), equities normally have negative skew (low strike implied greater than high strike implied). There are many possible explanations for this, some of which are listed below.

- **Big jumps in spot tend to be down, rather than up.** If there is a jump in the stock price, this is normally downwards as it is more common for an unexpectedly bad event to occur (bankruptcy, tsunami, terrorist attack, accident, loss or death of key personnel, etc) than an unexpectedly good event to occur (positive drivers are normally planned for).
- **Volatility is a measure of risk and leverage (hence risk) increases as equities decline.** If we assume no change in the number of shares in issue or amount of debt, then as a company's stock price declines its leverage (debt/equity) increases. Both leverage and volatility are a measure of risk and, hence, they are correlated, with volatility rising as equities fall.
- **Demand for protection and call overwriting.** Typically, investors are interested in buying puts for protection, rather than selling them. This lifts low strike implieds. Additionally, some investors like to call overwrite their positions, which weighs on higher strike implieds.

REASONS WHY SKEW AND TERM STRUCTURE ARE CORRELATED

The correlation between skew and term structure is shown below. The diagram only shows data for positive term structure, as the relationship tends to break down during a crisis.

Figure 81. SX5E Skew and Positive Term Structure (2007-10)[26]



Source: Santander Investment Bolsa.

There are three reasons why skew and term structure are correlated

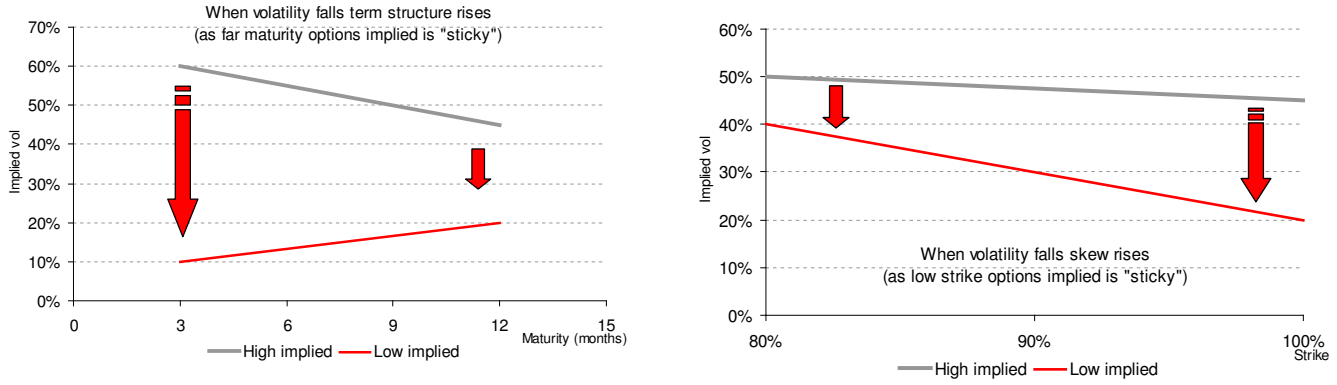
- Credit events, such as bankruptcy, lift both skew and term structure
- Implied volatility is 'sticky' for low strikes and long maturities
- Implied correlation is 'sticky' for low strikes and long maturities (only applies to index)

(1) BANKRUPTCY LIFTS BOTH SKEW AND TERM STRUCTURE

There are various models that show the effect of bankruptcy (or credit risk) lifting both skew and term structure. As implieds with lower strikes have a greater sensitivity to credit risk (as most of the value of low strike puts is due to the risk of bankruptcy), their implieds rise more, which causes higher skew. Similarly, options with longer maturity are more sensitive to credit risk (causing higher term structure, as far-dated implieds rise more). Longer-dated options have a higher sensitivity to credit risk as the probability of entering default increases with time (hence a greater proportion of an option's value will be associated with credit events as maturity increases). More detail on the link between volatility and credit can be seen in section *Capital Structure Arbitrage* in the *Appendix*.

[26] Excludes data from April 2010 onwards, as the change in US regulation regarding prop desks and moving equity derivatives onto exchanges (hence increased margin requirements) caused a spike in skew for major indices.

Figure 82. Term Structure Rising with Falling Volatility **Skew Rising with Falling Volatility**



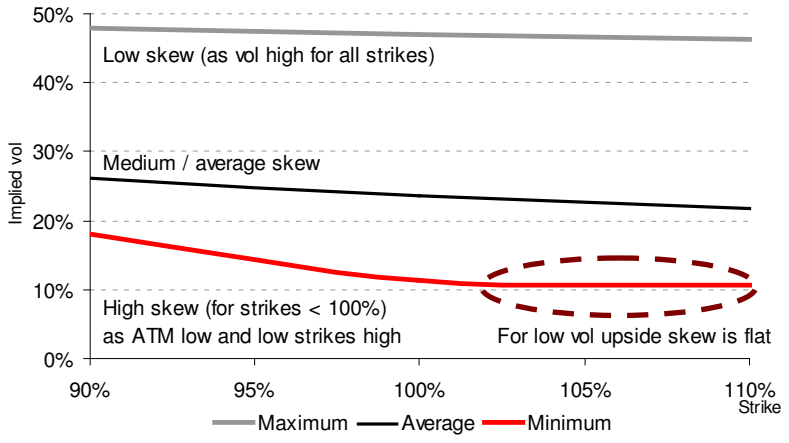
Source: Santander Investment Bolsa.

(2) IMPLIED VOL IS 'STICKY' FOR LOW STRIKES AND LONG MATURITIES

If there is a sudden decline in equity markets, it is reasonable to assume realised volatility will jump to a level in line with the peak of realised volatility. Therefore, low-strike, near-dated implieds should be relatively constant (as they should trade near the all-time highs of realised volatility). If a low-strike implied is constant, the difference between a low-strike implied and ATM implied increases as ATM implieds falls. This means near-dated skew should rise if near-dated ATM implieds decline (see figure above on the right). For this reason, we do not view skew as a reliable risk indicator, as it can be inversely correlated to ATM volatility[27]. The effect of falling implieds causing an increase in 90%-100% skew is shown with empirical data in Figure 83 below (we prefer to use 90%-100% skew rather than 90%-110%, as upside 100%-110% skew flattens as implieds reach a bottom).

Similarly, term structure should also rise if near-dated ATM implieds fall, as far-dated ATM implieds are relatively constant (as they tend to include complete economic cycles). This is shown in Figure 82 above on the left. Hence skew and term structure should be correlated as a fall in near-dated ATM implied lifts both of them.

Figure 83. SX5E 1 Year Max, Min and Average Implied Vol Since 2006



Source: Santander Investment Bolsa.

[27] ATM volatility is a risk measure; hence, a measure often inversely correlated to ATM volatility, such as skew, is not a reliable risk measure.

(3) CORRELATION SURFACE CAUSES INDEX SKEW AND TERM STRUCTURE TO BE CORRELATED

In the same way implied volatility is 'sticky' for low strikes and long maturities, so is implied correlation. This can be an additional reason why index skew and index term structure are correlated.

CORRELATION LIFTS INDEX SKEW ABOVE SINGLE-STOCK SKEW

Index skew is caused by both single-stock skew, and by the skew of the correlation surface

An approximation for implied correlation is the index volatility squared divided by the average single-stock volatility squared [$\rho = \sigma_{Index}^2 \div \text{average}(\sigma_{Single\ stock})^2$]. Implied correlation is assumed to tend towards 100% for low strikes, as all stocks can be expected to decline in a crisis. This causes index skew to be greater than single stock skew. Index skew can be thought of as being caused by both the skew of the single stock implied volatility surface, and the skew of the implied correlation surface.

Example of how index skew can be positive with flat single-stock skew

We shall assume all single stocks in an index have the same (flat) implied volatility and single-stock skew is flat. Low strike index volatility will be roughly equal to the constant single-stock volatility (as implied correlation is close to 100%), but ATM index volatility will be less than this figure due to diversity (as implied correlation ρ for ATM strikes is less than 100% and $\sigma_{Index}^2 = \rho \times \text{average}(\sigma_{Single\ stock})^2$. Despite single stocks having no skew, the index has a skew (as low strike index implieds > ATM index implieds) due to the change in correlation. For this reason, index skew is always greater than the average single-stock skew.

Implied correlation is likely to be sticky for low strikes and long maturities

A correlation surface can be constructed for options of all strikes and expiries, and this surface is likely to be close to 100% for very low strikes. The surface is likely to be relatively constant for far maturities; hence, implied correlation term structure and skew will be correlated (as both rise when near-dated ATM implied correlation falls, similar to volatility surfaces). This also causes skew and term structure to be correlated for indices.

DIVERSE INDICES HAVE HIGHER SKEW THAN LESS DIVERSE INDICES

As index skew is caused by both single-stock skew and implied correlation skew, a more diverse index should have a higher skew than a less diverse index (assuming there is no significant difference in the skew of the single-stock members). This is due to the fact that diverse indices have a lower ATM implied, but low strike implieds are in line with (higher) average single-stock implieds for both diverse and non-diverse indices.

SQUARE ROOT OF TIME RULE CAN COMPARE DIFFERENT TERM STRUCTURES AND SKEWS

When implied volatility changes, the change in ATM volatility multiplied by the square root of time is generally constant. This means that different (T_2-T_1) term structures can be compared when multiplied by $\sqrt{(T_2T_1)}/(\sqrt{T_2}-\sqrt{T_1})$, as this normalises against 1Y-3M term structure. Skew weighted by the square root of time should also be constant. Looking at the different term structures and skews, when normalised by the appropriate weighting, can allow us to identify calendar and skew trades in addition to highlighting which strike and expiry is the most attractive to buy (or sell).

REALISED VOLATILITY MEAN REVERTS AFTER EIGHT MONTHS

When there is a spike in realised volatility, it takes on average eight months for three-month realised volatility to settle back down to levels seen before the spike. The time taken for volatility to normalise is generally longer if the volatility is caused by a negative return, than if it is caused by a positive return (as a negative return is more likely to be associated with an event that increases uncertainty than a positive return). This mean reversion is often modelled via the square root of time rule.

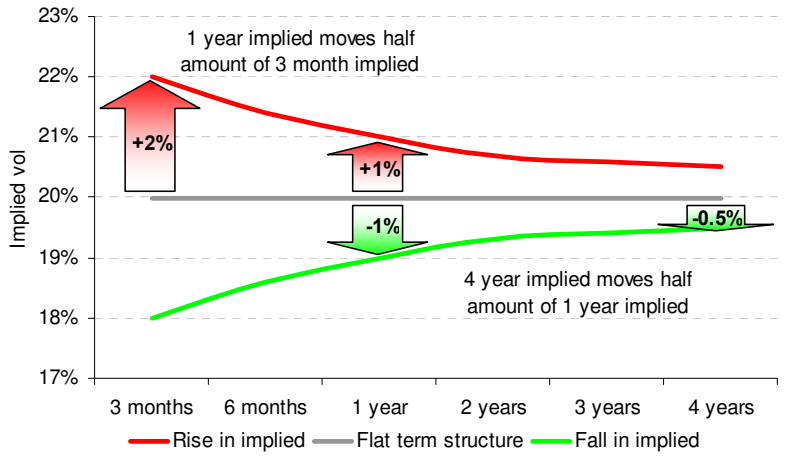
VOLATILITY MOVE MULTIPLIED BY √TIME IS USUALLY CONSTANT

Square root of time rule is a quick way to sensibly adjust an entire volatility surface with just one parameter

The near-dated end of volatility surfaces is highly correlated to realised volatility, as hedge funds and prop desks typically initiate long/short gamma positions should there be a significant divergence. As volatility mean reverts, the far-dated end of volatility surfaces is more stable (as investors know that any spike in volatility will be short-lived and not last for the full length of a far-dated option). A common way to model the movement of volatility surfaces, is to define the movement of one-year implied and then adjust the rest of the curve by that move divided by time (in years) to the power of p. Only two parameters (the one-year move and p) are needed to adjust the whole surface. Fixing the power (or p) at 0.5 is the most common and is known as the square root of time rule (which only has one parameter, the one-year change).

$$\text{Implied volatility move for maturity T years} = \frac{\text{One year implied volatility move}}{T^{p}}$$

Figure 84. ATM Implied Volatility Moving in a Square Root of Time Manner



Source: Santander Investment Bolsa.

Square root of time rule has power 0.5, parallel moves are power 0

Historically, the power is 0.44, not 0.5 (but very close to square root of time)

While the above method is usually used with power 0.5 (square root of time rule), any power can be used. If there is a parallel movement in volatility surfaces (all maturities move the same amount), then a power of 0 should be used. In practice, implied volatility tends to move with power 0.44, suggesting that surfaces move primarily in a square root of time manner but at times also in parallel. If implieds rise (or decline) in a square root of time manner when equities decline (or rise), then this causes skew to decay by the square root of time as well (assuming sticky strike). This means that the skews of different maturities can be compared with each other by simply multiplying the skew by the square root of the maturity (see Figure 85 below).

Figure 85. Skew by Maturity (with same skew when multiplied by square root of time)

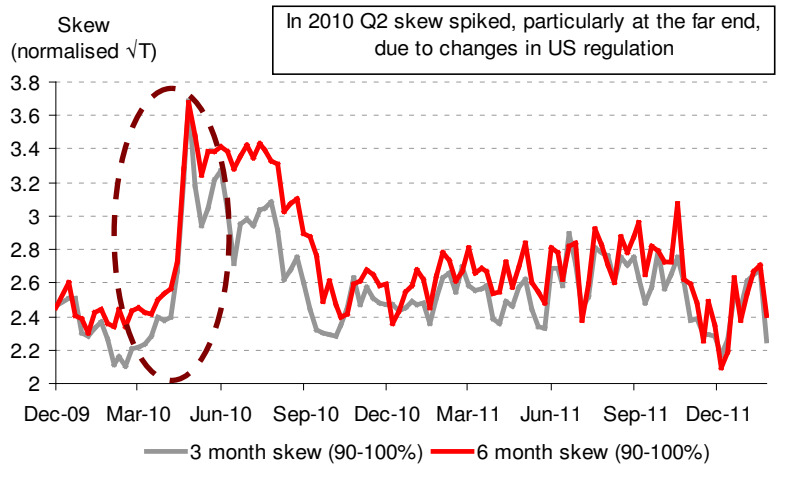
Maturity	3 Months	6 Months	1 Year	2 Years	3 Years	4 Years
Time (years)	0.25	0.5	1	2	3	4
Square root of time	**0.5**	**0.71**	**1**	**1.41**	**1.73**	**2**
90% implied	22.0%	21.4%	21.0%	20.7%	20.6%	20.5%
100% implied	18.0%	18.6%	19.0%	19.3%	19.4%	19.5%
Skew (per 10% move spot)	**4.0%**	**2.8%**	**2.0%**	**1.4%**	**1.2%**	**1.0%**
Skew × square root of time	**2.0%**	**2.0%**	**2.0%**	**2.0%**	**2.0%**	**2.0%**

Source: Santander Investment Bolsa.

POSITIVE PUT/CALL SPREADS IMPLY √TIME RULE FOR SKEW

Structures such as put spreads or call spreads, which can only have a positive payout, must have a cost associated with them, or investors would simply purchase an infinite amount of them for zero cost (or small profit) and enter a position which could never suffer a loss. This means that when the strike of a put is increased, its premium must rise too (intuitively correct, as the strike is the amount of money you receive when you 'put' the stock, so the higher the strike the better). Conversely, as the strike of a call increases, its premium must decrease. It can be shown that enforcing positive put/call spreads puts a cap/floor on skew, which decays by the square root of time. This provides mathematical support for the empirical evidence, suggesting implied volatility should normally move in a power weighted (by square root of time) manner[28]. For more details, see the section *Modelling Volatility Surfaces* in the *Appendix*.

Figure 86. SX5E Skew Multiplied by the Square Root of Time (R^2=83%)



Source: Santander Investment Bolsa.

[28] Looking at ratio put spreads, it can be shown that for long maturities (five years) skew should decay by time, ie, 1/T or power=1 (rather than √T or power 0.5).

√TIME RULE CAN COMPARE DIFFERENT TERM STRUCTURES

ATM term structure can be modelled as flat volatility, with a square root of time adjustment on top. With this model, flat volatility is equal to the volatility for an option of infinite maturity. There are, therefore, two parameters to this model, the volatility at infinity (V_∞) and the scale of the square root of time adjustment ,which we define to be z (for one-year implied).

$$\text{Volatility} = V_\infty - z/\sqrt{T}$$

where

z = scale of the square root of time adjustment (which we define as normalised term structure)

We have a negative sign in front of z, so that a positive z implies an upward sloping term structure and a negative z is a downward sloping term structure.

Different term structures are normalised by multiplying by $\sqrt{(T_2T_1)}/(\sqrt{T_2}-\sqrt{T_1})$

Using the above definition, we can calculate the normalised term structure z from two volatility points V_1 and V_2 (whose maturity is T_1 and T_2).

$$V_1 = V_\infty - z/\sqrt{T_1}$$

$$V_2 = V_\infty - z/\sqrt{T_2}$$

$$\Rightarrow \quad V_1 + z/\sqrt{T_1} = V_2 + z/\sqrt{T_2} = V_\infty$$

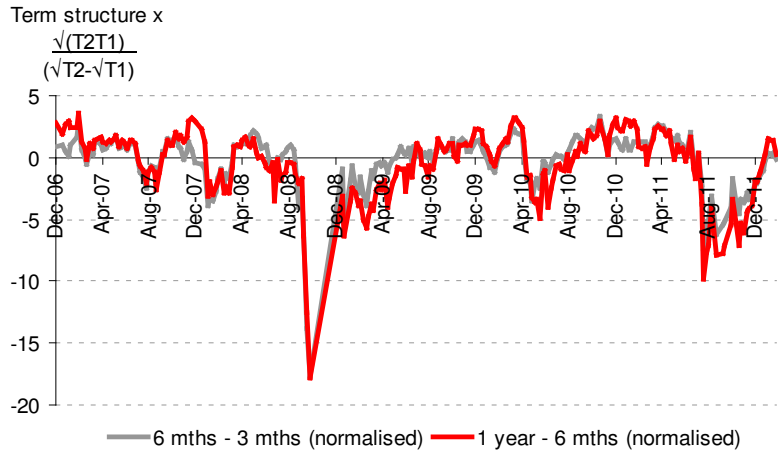
$$\Rightarrow \quad z\,(1/\sqrt{T_1} - 1/\sqrt{T_2}) = V_2 - V_1$$

$$\Rightarrow \quad z = (V_2 - V_1) \times \frac{\sqrt{T_2T_1}}{\sqrt{T_2} - \sqrt{T_1}}$$

Normalising term structure by $\sqrt{(T_2T_1)}/(\sqrt{T_2}-\sqrt{T_1})$ puts it in the same 'units' as 1Y-3M term structure

$V_2 - V_1$ is the normal definition for term structure. Hence, term structure can be normalised by multiplying by $\sqrt{(T_2T_1)}/(\sqrt{T_2}-\sqrt{T_1})$. We note that the normalisation factor for 1Y-3M term structure is 1. Therefore, normalising allows all term structure to be compared to 1Y-3M term structure.

Figure 87. SX5E Normalised Term Structure (R2=80%)

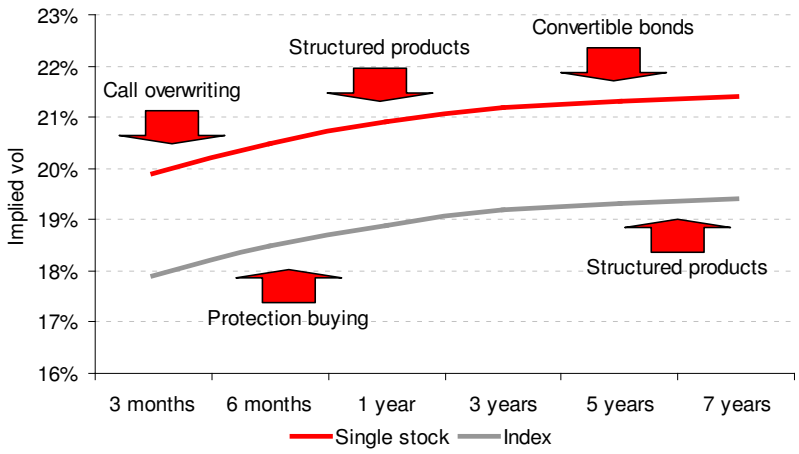


Source: Santander Investment Bolsa.

TERM STRUCTURE TRADES CAN PROFIT FROM IMBALANCES

The supply-demand imbalances of different products on implied volatility surfaces can create opportunity for other investors. The degree of the imbalance depends on the popularity of the product at the time. Investors who are willing to take the other side of the trade should be able to profit from the imbalance, and the risk taken can be hedged with other maturities or related securities.

Figure 88. Implied Volatility Imbalances by Maturity



Source: Santander Investment Bolsa.

CALENDARS REMAIN CONSTANT IF SURFACES FOLLOW √TIME RULE

Calendars can be used to trade term structure imbalances

Given that the square root of time appears in the Black-Scholes formula for premium, the price of a 1x1 calendar (long one far-dated option, short one near-dated option) remains approximately constant if implied volatility surfaces move in a square root of time manner. Calendars can therefore be used to trade term structure imbalances as the trade is indifferent to the level of volatility as long as volatility moves in a power weighted manner.

IDENTIFYING WHEN TO GO LONG, OR SHORT, CALENDARS

When examining term structure trades, the power of the movement in volatility surfaces can be compared to the expected 0.5 power of the square root of time rule. If the movement has a power significantly different from 0.5, then a long (or short) calendar position could be initiated to profit from the anticipated correction. This method assumes calendars were previously fairly priced (otherwise the move could simply be a mean reversion to the norm).

If volatility rises with power less than 0.5, investors should short calendars

If surfaces rise with a power less than 0.5 (ie, a more parallel move) then near-dated implieds have not risen as much as expected and a short calendar (long near-dated, short far-dated) position should be initiated. This position will profit from the anticipated correction. Should surfaces fall with a power less than 0.5, a long calendar (short near-dated, long far-dated) would profit from the anticipated further decline of near-dated implieds.

If volatility rises with power more than 0.5, investors should go long calendars

Conversely, if surfaces rise with a power greater than 0.5, near-dated implieds have risen too far and a long calendar position should be initiated. On the other hand, if surfaces fall with a power greater than 0.5, a short calendar position should be initiated (as near-dated implieds have fallen too far).

POWER VEGA IS VEGA DIVIDED BY THE SQUARE ROOT OF TIME

As volatility surfaces tend to move in a square root of time manner, many systems report power vega (vega divided by square root of time). Power vega takes into account the fact that the implied volatility of near-dated options is more volatile than far-dated options.

VARIANCE TERM STRUCTURE CAN INDENTIFY TRADES

Variance term structure is similar to ATM term structure

To determine if a term structure trade is needed, we could look at variance term structure rather than implied volatility term structure. Using variance term structure eliminates the need to choose a strike (an ATM term structure will not be ATM as soon as the spot moves, so it is effectively strike dependent, but simply delegates the choice of strike to the equity market). Variance term structure is similar to ATM term structure, despite variance being long skew and skew being greater for near-dated implieds. This is because the time value of an OTM option increases with maturity. Hence, the increased weight associated with OTM options cancels the effect of smaller skew for longer maturities.

Forward starting variance swaps (or options) can be used to trade term structure

Trading term structure via a long and short variance swap is identical to a position in a forward starting variance swap (assuming the weights of the long and short variance swap are correct; if not, there will be a residual variance swap position left over). The correct weighting for long and short variance swaps to be identical to a forward starting variance swap is detailed in the section *Forward Starting Products*. If an investor wants to trade term structure, but does not want to have exposure to current volatility (ie, wants to have zero theta and gamma), then forward starting products (variance swaps or options) can be used. Note that while forward starting products have no exposure to current realised volatility, they do have exposure to future expectations of volatility (ie, implied volatility hence has positive vega).

HOW TO MEASURE SKEW AND SMILE

The implied volatilities for options of the same maturity, but of different strike, are different from each other for two reasons. Firstly, there is skew, which causes low strike implieds to be greater than high strike implieds due to the increased leverage and risk of bankruptcy. Secondly, there is smile (or convexity/kurtosis), when OTM options have a higher implied than ATM options. Together, skew and smile create the 'smirk' of volatility surfaces. We look at how skew and smile change by maturity in order to explain the shape of volatility surfaces both intuitively and mathematically. We also examine which measures of skew are best, and why.

MOMENTS DESCRIBE THE PROBABILITY DISTRIBUTION

Moments 1-4 describe forward, variance, skew and kurtosis

In order to explain skew and smile, we shall break down the probability distribution of log returns into moments. Moments can describe the probability distribution[29]. From the formula below we can see that the zero-th moment is 1 (as the sum of a probability distribution is 1, as the probability of all outcomes is 100%). The first moment is the expected value (ie, mean or forward) of the variable. The second, third and fourth moments are variance, skew and kurtosis, respectively (see table on the left below). For moments of two or greater it is usual to look at central moments, or moments about the mean (we cannot for the first moment as the first central moment would be 0). We shall normalise the central moment by dividing it by σ^n in order to get a dimensionless measure. The higher the moment, the greater the number of data points that are needed in order to get a reasonable estimate.

$$\text{Raw moment} = \mathrm{E}(X^k) = \int_{-\infty}^{\infty} x^k f(x)$$

$$\text{Normalised central moment} = \mathrm{E}((X-\mu)^k)/\sigma^k = \int_{-\infty}^{\infty} (x-\mu)^k f(x)/\sigma^k$$

where

$f(x)$ is the probability distribution function

Figure 89. Moments

Moment	Name	Decay/Movement over Time	Maturity Where Dominates Surface	Related Option Position: Related (Long) Position	Related Option Position: Key Greek for Position	Related Option Position: P&L Driver for Position
1†	Forward (expected price)	Random walk	NA	Stock/futures	Delta	Price
2	Variance (volatility²)	Mean reverts	Far-dated/all maturities	ATM options	Vega	Implied volatility
3	Skew	Decay square root of time	Medium-dated	Risk reversal (long low strikes & short high strikes)	Vanna	Skew (90-110%)
4	Kurtosis	Decay by time	Near-dated	Butterfly (long wings, short body)	Volga (gamma of volatility)	Vol of vol

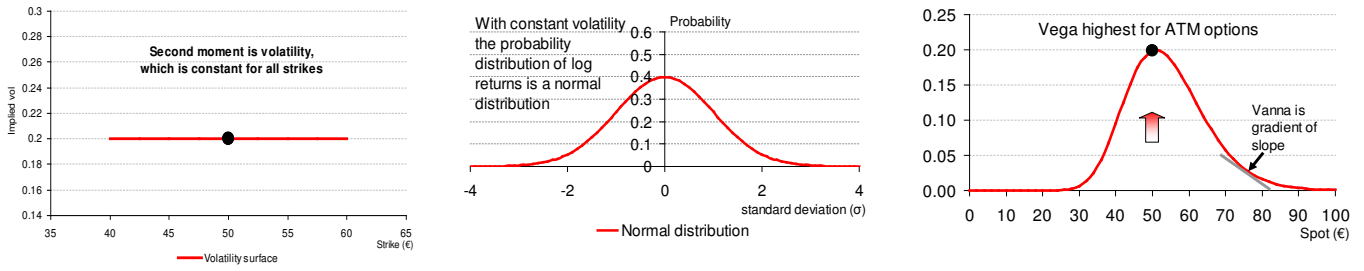
(†) First raw moment (other moments are normalised central moments).
Source: Santander Investment Bolsa.

[29] The combination of all moments can perfectly explain any distribution as long as the distribution has a positive radius of convergence or is bounded (eg, a sine wave is not bounded; hence, it cannot be explained by moments alone).

VEGA MEASURES SIZE OF VOLATILITY POSITION

Vega measures the change in price of an option for a given (normally 1%) move in implied volatility. Implied volatility for far-dated options is relatively flat compared to near-dated, as both skew and kurtosis decay with maturity. Vega is highest for ATM options, as can be seen in the right hand chart in Figure 90 below.

Figure 90. Moment 2 = Variance | **Distribution for Constant Volatility** | **Vega is Size of Volatility Position**

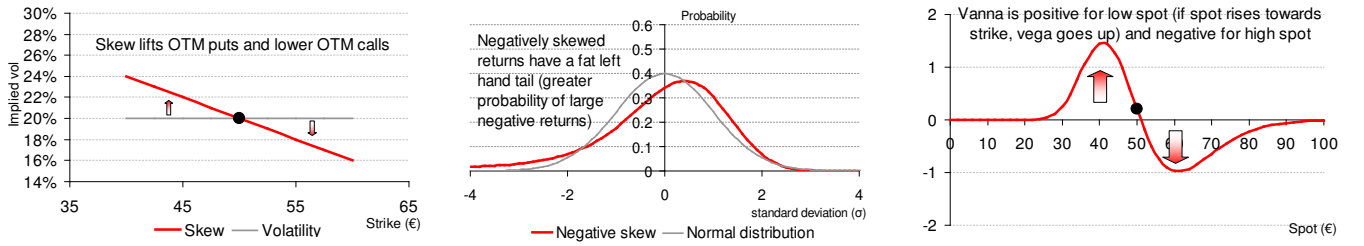


Source: Santander Investment Bolsa estimates.

VANNA MEASURES SIZE OF SKEW POSITION

Vanna (dVega/dSpot which is equal to dDelta/dVol) measures the size of a skew position[30], and is shown on the right side of Figure 91 below. Vanna is the slope of vega plotted against spot (see graph on right above).

Figure 91. Moment 3 = Skew[31] | **Distribution with Skew** | **Vanna is Size of Skew Position**

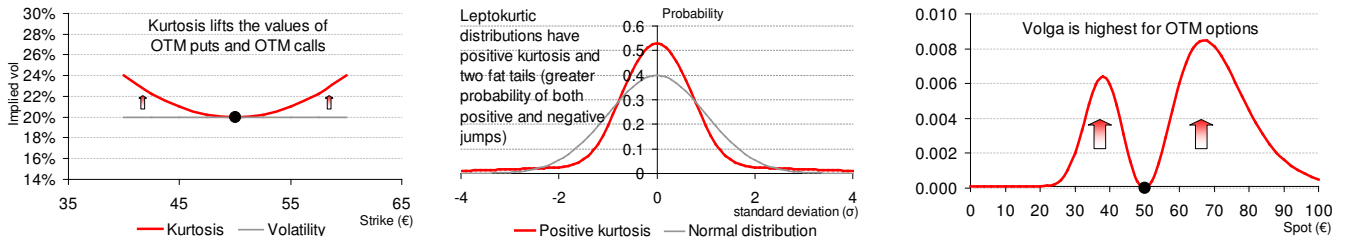


Source: Santander Investment Bolsa estimates.

VOLGA MEASURES SIZE OF GAMMA OF VOLATILITY

The gamma of volatility is measured by Volga (dVega/dVolatility), which is also known as volatility gamma or vega convexity. Volga is always positive (similar to option gamma always being positive) and peaks for c10-15 delta options (like Vanna).

Figure 92. Moment 4 = Kurtosis[31] | **Distribution with Kurtosis** | **Volga = Gamma of Vol**



Source: Santander Investment Bolsa estimates.

[30] For details, see the next section *Skew Trading*.
[31] This is an approximation as the effect of moments on slope and convexity are intertwined.

Options with high volga benefit from volatility of volatility

Vol of vol gives volatility surfaces a 'smile' profile

Just as an option with high gamma benefits from high stock price volatility, an option with high volga benefits from volatility of volatility. The level of volatility of volatility can be calculated in a similar way to how volatility is calculated from stock prices (taking log returns is recommended for volatility as well). The more OTM an option is, the greater the volatility of volatility exposure. This is because the more implied volatility can change, the greater the chance of it rising and allowing an OTM option to become ITM. This gives the appearance of a 'smile', as the OTM option's implied volatility is lifted while the ATM implied volatility remains constant.

Stock returns have positive excess kurtosis and are leptokurtic

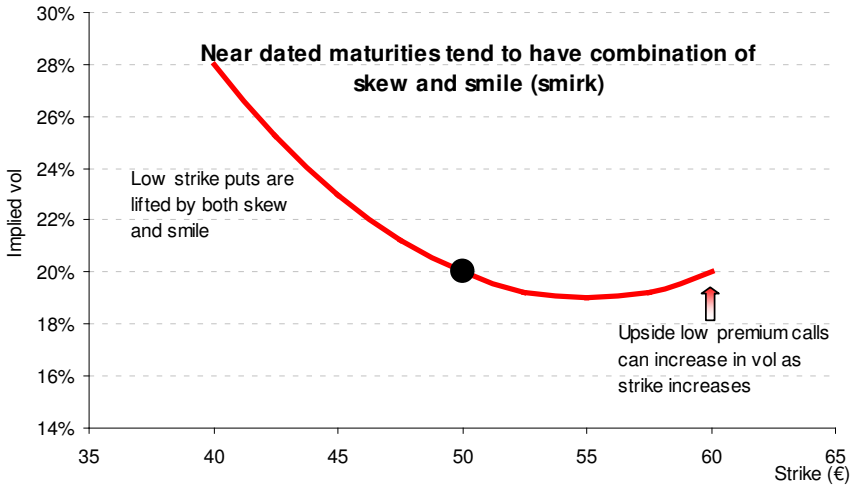
Kurtosis is always positive[32]. Hence, excess kurtosis (kurtosis -3) is usually used. The kurtosis (or normalised fourth moment) of the normal distribution is three; hence, normal distributions have zero excess kurtosis (and are known as mesokurtic). High kurtosis distributions (eg, stock price log returns) are known as leptokurtic, whereas low kurtosis distributions (pegged currencies that change infrequently by medium-sized adjustments) are known as platykurtic.

IMPLIED VOLATILITY SMIRK IS A COMBINATION OF SKEW AND SMILE

Kurtosis decays by 1/time, hence smile is most important for near-dated expiries

The final 'smirk' for options of the same maturity is the combination of skew (3rd moment) and smile (4th moment). The exact smirk depends on maturity. Kurtosis (or smile) can be assumed to decay with maturity by dividing by time[33] and, hence, is most important for near-dated expiries. For medium- (and long-) dated expiries, the skew effect will dominate kurtosis, as skew usually decays by the reciprocal of the square root of time (for more details, see the section *Modelling Volatility Surfaces* in the *Appendix*). Skew for equities is normally negative and therefore have mean < median < mode (max) and a greater probability of large negative returns (the reverse is true for positively skewed distributions). For far-dated maturities, the effect of both skew and kurtosis fades; hence, implieds converge to a flat line for all strikes. Skew can be thought of as the effect of changing volatility as spot moves, while smile can be thought of as the effect of jumps (up or down).

Figure 93. Near-Dated Implied Volatilities with Smirk (Skew and Smile)



Source: Santander Investment Bolsa.

[32] Kurtosis is only zero for a point distribution.
[33] Assuming stock price is led by Lévy processes (eg, accumulation of independent identical shocks).

THERE ARE THREE WAYS TO MEASURE SKEW

There are three main ways skew can be measured. While the first is the most mathematical, in practice the other two are more popular with market participants.

- Third moment
- Strike skew (eg, 90%-110%)
- Delta skew (eg, [25 delta put – 25 delta call] / 50 delta)

(1) THIRD MOMENT IS DEFINITION OF CBOE SKEW INDEX

CBOE have created a skew index on the S&P500. This index is based on the normalised third central moment; hence, it is strike independent. The formula for the index is given below. For normal negative skew, if the size of skew increases, so does the index (as negative skew is multiplied by -10).

SKEW = $100 - 10 \times 3^{rd}$ moment

(2) STRIKE SKEW SHOULD NOT BE DIVIDED BY VOLATILITY

The most common method of measuring skew is to look at the difference in implied volatility between two strikes, for example 90%-110% skew or 90%-ATM skew. It is a common mistake to believe that strike skew should be divided by ATM volatility in order to take into account the fact that a 5pt difference is more significant for a stock with 20% volatility than 40% volatility. This ignores the fact that the strikes chosen (say 90%-110% for 20% volatility stocks) should also be wider for high volatility stocks (say 80%-120%, or two times wider, for 40% volatility stocks as the volatility is 2×20%). The difference in implied volatility should be taken between two strikes whose width between the strikes is proportional to the volatility (similar to taking the implied volatility of a fixed delta, eg, 25% delta). An approximation to this is to take the fixed strike skew, and multiply by volatility, as shown below. As the two effects cancel each other out, we can simply take a fixed strike skew without dividing by volatility.

Difference in vol between 2 strikes = 90-110%

⇨ Difference in vol between 2 strikes whose width increases with vol = 90-110% × ATM

$$\text{Skew} = \frac{\text{Difference in vol between 2 strikes whose width increases with vol}}{\text{ATM}}$$

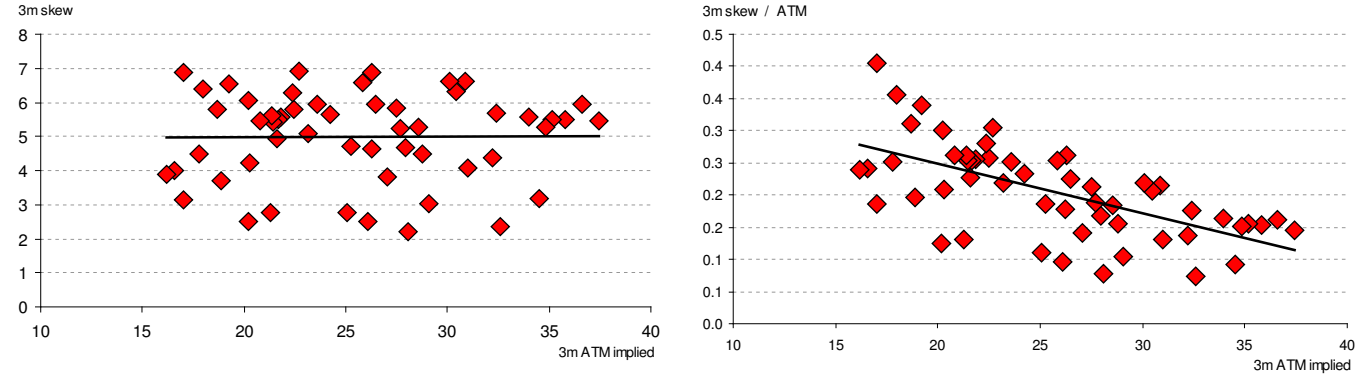
⇨ $$\text{Skew} = \frac{\text{90-110\% } \times \text{ ATM}}{\text{ATM}}$$

⇨ Skew = 90-110%

Empirically, 90%-100% (or 90%-110%) skew is correct measure for fixed strike skew

The best measure of skew is one that is independent of the level of volatility. If this were not the case, then the measure would be partly based on volatility and partly on skew, which would make it more difficult to determine if skew was cheap or expensive. We have shown mathematically that an absolute difference (eg, 90%-110% or 90%-100%) is the correct measure of skew, but we can also show it empirically. The left-hand chart in Figure 94 below shows that there is no correlation between volatility and skew (90%-110%) for any European stocks that have liquid equity derivatives. If skew is divided by volatility, there is unsurprisingly a negative correlation between this measure and volatility (see right-hand chart below).

Figure 94. Strike Skew (90%-110%) Plotted vs Volatility **Strike Skew/Volatility Plotted vs Volatility**

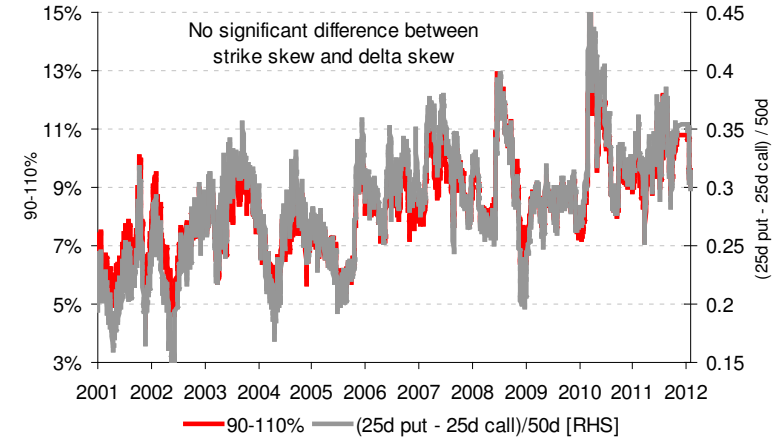


Source: Santander Investment Bolsa.

(3) DELTA SKEW IS VIRTUALLY IDENTICAL TO STRIKE SKEW

Arguably the best measure of skew is delta skew, where the difference between constant delta puts and calls is divided by 50 delta implied. An example of skew measured by delta is [25 delta put – 25 delta call] / 50 delta. As this measure widens the strikes examined as vol rises, in addition to normalising (ie, dividing) by the level of volatility, it is a 'pure' measure of skew (ie, not correlated to the level of volatility). While delta skew is theoretically the best measure, in practice it is virtually identical to strike skew. As there is a R^2 of 93% between delta skew and strike skew, we believe both are viable measures of skew (although strike skew is arguably more practical as it represents a more intuitive measure).

Figure 95. Strike Skew vs Delta Skew



Source: Santander Investment Bolsa.

SKEW TRADING

The profitability of skew trades is determined by the dynamics of a volatility surface. We examine sticky delta (or 'moneyness'), sticky strike, sticky local volatility and jumpy volatility regimes. Long skew suffers a loss in both a sticky delta and sticky strike regimes due to the carry cost of skew. Long skew is only profitable with jumpy volatility. We also show how the best strikes for skew trading can be chosen.

FOUR IDEALISED REGIMES DESCRIBE MOVEMENT OF VOL SURFACE

There are four idealised regimes for a volatility surface. While sticky delta, sticky strike and (sticky) local volatility are well known and widely accepted names, we have added 'jumpy volatility' to define volatility with a high negative correlation with spot. These regimes are summarised below, and more details are given on pages 159-163 of this section.

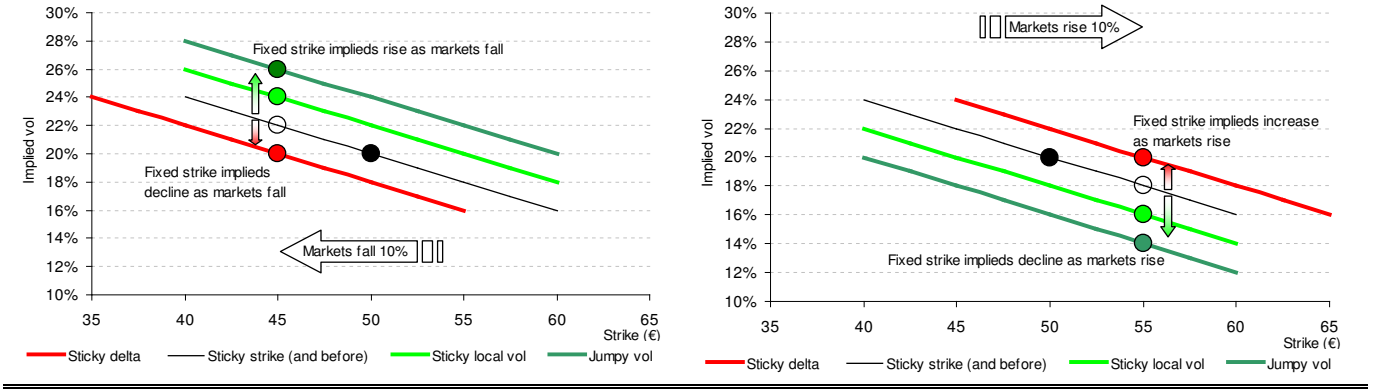
(1) **Sticky delta (or sticky moneyness).** Sticky delta assumes a constant volatility for options of the same strike as a percentage of spot. For example, ATM or 100% strike volatility has constant volatility. As this model implies there is a positive correlation between volatility and spot, the opposite of what is usually seen in the market, it is not a particularly realistic model (except over a very long time horizon).

(2) **Sticky strike.** A sticky strike volatility surface has a constant volatility for options with the same fixed currency strike. Sticky strike is usually thought of as a stable (or unmoving) volatility surface as real-life options (with a fixed currency strike) do not change their implied volatility.

(3) **Sticky local volatility.** Local volatility is the instantaneous volatility of stock at a certain stock price. When local volatility is static, implied volatility rises when markets fall (ie, there is a negative correlation between stock prices and volatility). Of all the four volatility regimes, it is arguably the most realistic and fairly prices skew.

(4) **Jumpy volatility.** We define a jumpy volatility regime as one in which there is an excessive jump in implied volatility for a given movement in spot. There is a very high negative correlation between spot and volatility. This regime usually occurs over a very short time horizon in panicked markets (or a crash).

Figure 96. Volatility Surfaces When Equities Fall 10% **Volatility Surfaces When Equities Rise 10%**



Source: Santander Investment Bolsa estimates.

PUTTING ON LONG SKEW TRADES HAS A COST (SKEW THETA)

Skew theta is the difference between the cost of gamma of an OTM option compared to an ATM option

If an investor initiates a long skew position by buying an OTM put and selling an OTM call, the implied volatility of the put purchased has a higher implied volatility than the implied volatility sold through the call. The long skew position therefore has a cost associated with it, which we shall define as 'skew theta'. Skew theta is the difference between the cost of gamma (theta per unit of dollar gamma) of an OTM option compared to an ATM option. If skew is flat, then all strikes have an identical cost of gamma, but as OTM puts have a higher implied volatility than ATM ones they pay more per unit of gamma. Skew theta is explained in greater depth at the end of this section. If the long skew position does not give the investor enough additional profit to compensate for the skew theta paid, then skew can be sold at a profit.

Skew trades profit from negative spot volatility correlation

Skew trades profit from negative spot volatility correlation

If there is a negative correlation between the movement of a volatility surface and spot (as is usually seen in practice), then this movement will give a long skew position a profit when the volatility surface is re-marked. For example, let us assume an investor is long skew via a risk reversal (long an OTM put and short an OTM call). If equity markets decline, the put becomes ATM and is the primary driver of value for the position (as the OTM call becomes further OTM it is far less significant). The rise in the volatility surface (due to negative correlation between spot and volatility) boosts the value of the (now ATM) put and, hence, the value of the risk reversal.

SKEW TRADES BREAK EVEN IF LOCAL VOL SURFACE IS CONSTANT

If the local volatility surface stays constant, the amount volatility surfaces move for a change in spot is equal to the skew (ie, ATM volatility moves by twice the skew, once for moving up the skew and another by the movement of the volatility surface itself). This movement is exactly the correct amount for the profit (or loss) on a volatility surface re-mark to compensate for the cost (or benefit) of skew theta[34]. The profit (or loss) caused by skew trades given the four volatility regimes are shown below.

Figure 97. Different Volatility Regimes and Breakdown of P&L for Skew Trades

	Fixed strike implied volatility change		P&L breakdown for long skew (e.g. long put, short call)		
Volatility regime	Equity decline	Equity rise	Remark	Skew theta	Total
Sticky delta	Falls	Rises		+ 	= 
Sticky strike	-	-		+ 	= 
Sticky local volatility	Rises	Falls		+ 	= 
Jumpy volatility	Rises significantly	Falls significantly		+ 	= 

Source: Santander Investment Bolsa.

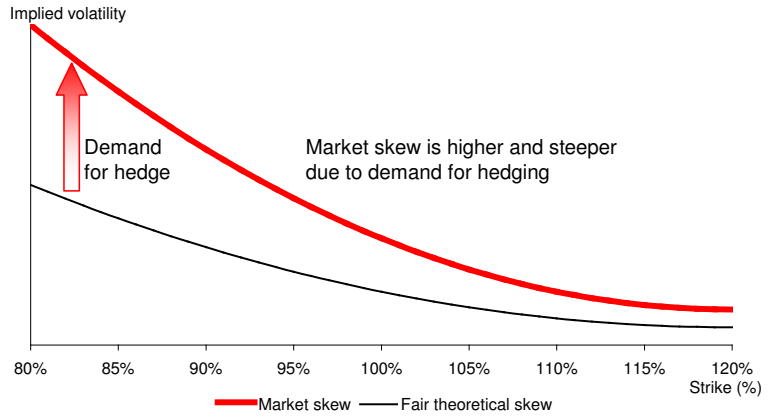
[34] More details on local volatility can be found in the section *Local Volatility* in the *Appendix.*

SKEW IS USUALLY OVERPRICED DUE TO HEDGING

Demand for hedges causes skew to be overpriced

As volatility markets tend to trade between a static strike and static local volatility regime, long skew trades are usually unprofitable (usually there is negative spot volatility correlation, but not enough to compensate for the skew theta). As long skew trades break even during static local volatility regimes, they are only profitable in periods of jumpy volatility. This overpricing of skew can be considered to be a result of excessive demand for downside put options, potentially caused by hedging. Another reason for the overpricing of skew could be the popularity of short volatility long (downside) skew trades (traders often hedge a short volatility position with a long skew (OTM put) position, in order to protect themselves should markets suddenly decline). The profits from shorting expensive volatility are likely to more than compensate for paying an excessive amount for the long skew position (OTM put).

Figure 98. Market and Theoretical Skew



Source: Santander Investment Bolsa.

VOL REGIME DETERMINED BY TIME AND SENTIMENT

Implied volatility can be thought of as the equity derivative market's estimate of future volatility[35]. It is therefore investor sentiment that determines which implied volatility regime the market trades in, and this choice is largely determined by how much profit (or loss) a long skew position is expected to reveal over a certain time period (ie, investor sentiment). The choice of regime is also determined by the time horizon chosen.

Figure 99. Characteristics of Different Volatility Regimes

Characteristic	Sticky Delta		Sticky Strike		Sticky Local Vol		Jumpy Vol
Sentiment	Calm/trending		Normal				Panicked
Time horizon	Long term		Medium term				Short term
Spot vol correlation	Positive		Zero		Negative		Very negative
Call delta	δ_{call}	>	$\delta_{Black\text{-}Scholes}$	>	δ_{call}	>	δ_{call}
Put delta	δ_{put}	>	$\delta_{Black\text{-}Scholes}$	>	δ_{put}	>	δ_{put}
Abs(Put delta)	$Abs(\delta_{put})$	<	$Abs(\delta_{Black\text{-}Scholes})$	<	$Abs(\delta_{put})$	<	$Abs(\delta_{put})$

Source: Santander Investment Bolsa estimates.

[35] In the absence of any supply-demand imbalance in the market.

Sticky delta regimes occur over long time horizon or trending markets

A sticky delta regime is typically one in which markets are trending in a stable manner (either up or down, with ATM volatility staying approximately constant) or over a very long time horizon of months or years (as over the long term the implied volatility mean reverts as it cannot go below zero or rise to infinity).

Jumpy volatility regimes occur over very short time horizons and panicked markets

It is rare to find a jumpy volatility regime that occurs over a long time horizon, as they tend to last for periods of only a few days or weeks. Markets tend to react in a jumpy volatility manner after a sudden and unexpected drop in equity markets (large increase in implied volatility given a decline in spot) or after a correction from such a decline (a bounce in the markets causing implied volatility to collapse).

Markets tend to trade between a sticky strike and sticky local volatility regime

On average, markets tend to trade between sticky strike and sticky local volatility

Sticky delta and jumpy volatility are the two extremes of volatility regimes. Sticky strike and sticky local volatility are far more common volatility regimes. Sticky strike is normally associated with calmer markets than sticky local volatility (as it is closer to a sticky delta model than jumpy volatility).

DELTA OF OPTION DEPENDS ON VOLATILITY REGIME

How a volatility surface reacts to a change in spot changes the value of the delta of the option. For sticky strike, as implied volatilities do not change, the delta is equal to the Black-Scholes delta.

Sticky delta regime implies a higher delta for options

However, if we assume a sticky delta volatility regime if an investor is long a call option, then the implied volatility of that option will decline if there is a fall in the market. The value of the call is therefore lower than expected for falls in the market. The reverse is also true as implied volatility increases if equities rise. As the value of the call is lower for declines and higher for rises (as volatility is positively correlated to spot), the delta is higher than that calculated by Black-Scholes (which is equal to the sticky strike delta).

A similar argument can be made for sticky local volatility (as volatility is negatively correlated to spot, the delta is less than that calculated by Black-Scholes). Figure 99 summarises the differences in delta for the different volatility regimes.

VOL CAN BE EXAMINED IN RELATIVE OR ABSOLUTE DIMENSIONS

To evaluate the profit – or loss – from a skew trade, assumptions have to be made regarding the movement of volatility surfaces over time (as we assume a skew trader always delta hedges, we are not concerned with the change in premium only the change in volatility). Typically, traders use two main ways to examine implied volatility surfaces. Absolute dimensions tend to be used when examining individual options, a snapshot of volatilities, or plotting implied volatilities over a relatively short period of time. Relative dimensions tend to be used when examining implied volatilities over relatively long periods of time[36].

- **Absolute dimensions.** In absolute dimensions, implied volatility surfaces are examined in terms of fixed maturity (eg, Dec14 expiry) and fixed strike (eg, €4,000). This surface is a useful way of examining how the implied volatility of actual traded options changes.
- **Relative dimensions.** An implied volatility surface is examined in terms of relative dimensions when it is given in terms of relative maturity (eg, three months or one year) and relative strike (eg, ATM, 90% or 110%). Volatility surfaces tend to move in relative dimensions over a very long period of time, whereas absolute dimensions are more suitable for shorter periods of time.

Care must be taken when examining implieds in relative dimensions

Relative dimensions should only be used for analysing long-term trends

As the options (and variance swaps) investors buy or sell are in fixed dimensions with fixed expiries and strikes, the change in implied volatility in absolute dimensions is the key driver of volatility profits (or losses). However, investors often use ATM volatility to determine when to enter (or exit) volatility positions, which can be misleading. For example, if there is a skew (downside implieds higher than ATM) and equity markets decline, ATM implieds will rise even though volatility surfaces remain stable. A plot of ATM implieds will imply buying volatility was profitable over the decline in equity markets; however, in practice this is not the case.

Absolute implied volatility is the key driver for equity derivative profits

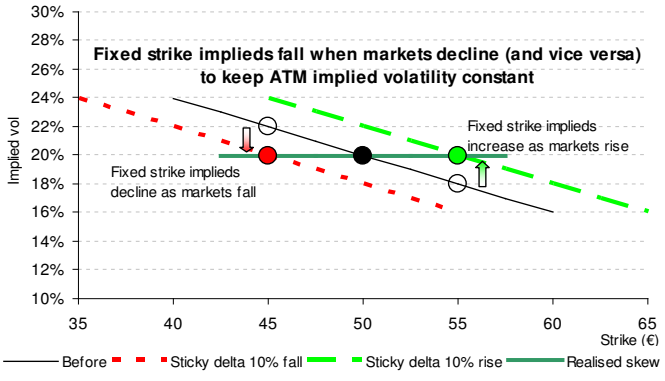
As options that are traded have a fixed strike and expiry, it is absolute implied volatility that is the driver for equity derivative profits and skew trades. However, we accept that relative implied volatility is useful when looking at long-term trends. For the volatility regimes (1) sticky delta and (2) sticky strike, we shall plot implieds using both absolute and relative dimensions in order to explain the difference. For the remaining two volatility regimes (sticky local volatility and jumpy volatility), we shall only plot implied volatility using absolute dimensions (as that is the driver of profits for traded options and variance swaps).

[36] This is usually for liquidity reasons, as options tend to be less liquid for maturities greater than two years (making implied volatility plots of more than two years problematic in absolute dimensions).

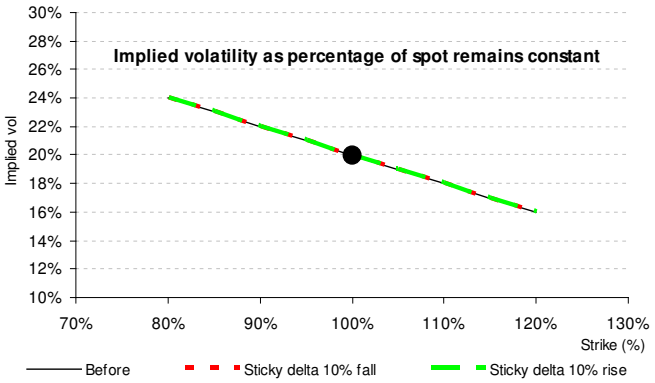
1. STICKY DELTA ASSUMES ATM VOL NEVER CHANGES

A sticky delta model assumes a constant implied volatility for strikes as a percentage of spot (eg, ATM stays constant). How a volatility surface moves with a change in spot is shown below for both absolute/fixed strike (Figure 100 on the left) and relative strike (Figure 100 on the right).

Figure 100. Sticky Delta Absolute/Fixed Strike



Sticky Delta Relative Strike (as Percentage of Spot)



Source: Santander Investment Bolsa estimates.

RANGE-BOUND VOLATILITIES SUPPORT A STICKY DELTA MODEL

As implied volatility cannot be negative, it is therefore usually floored close to the lowest levels of realised volatility. Although an infinite volatility is theoretically possible, in practice implied volatility is typically capped close to the all-time highs of realised volatility. Over a long period of time, ATM implied volatility can be thought of as being range bound and likely to trend towards an average value (although this average value will change over time as the macro environment varies). As the trend towards this average value is independent of spot, the implied volatility surface in absolute dimensions (fixed currency strike) has to move to keep implied volatility surface in relative dimensions (strike as percentage of spot) constant. Thinking of implied volatility in this way is a sticky delta (or sticky moneyness) implied volatility surface model.

Sticky delta most appropriate over long term (many months or years)

Sticky delta most appropriate for time horizon of one year or more

While over the long term implied volatility tends to return to an average value, in the short term volatility can trade away from this value for a significant period of time. Typically, when there is a spike in volatility it takes a few months for volatility to revert back to more normal levels. This suggests a sticky delta model is most appropriate for examining implied volatilities for periods of time of a year or more. As a sticky delta model implies a positive correlation between (fixed strike) implied volatility and spot, the opposite of what is normally seen, it is not usually a realistic model for short periods of time. Trending markets (calmly rising or declining) are usually the only situation when a sticky delta model is appropriate for short periods of time. In this case, the volatility surface tends to reset to keep ATM volatility constant, as this implied volatility level is in line with the realised volatility of the market.

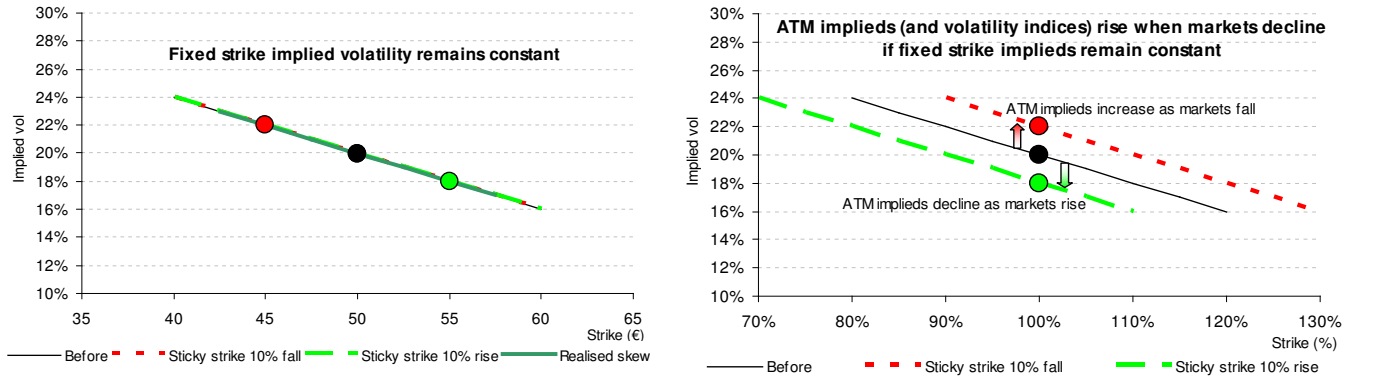
LONG SKEW IS UNPROFITABLE IN STICKY DELTA VOLATILITY REGIME

In a sticky delta volatility regime the fixed strike implied volatility (and, therefore, the implied volatility of traded options) has to be re-marked when spot moves. The direction of this re-mark for long skew positions causes a loss, as skew should be flat if ATM volatility is going to remain unchanged as markets move (we assume the investor has bought skew at a worse level than flat). Additionally, the long skew position carries the additional cost of skew theta, the combination of which causes long skew positions to be very unprofitable.

2. STICKY STRIKE HAS ZERO SPOT VOL CORRELATION

A sticky strike model assumes that options of a fixed currency strike are fixed (absolute dimensions). The diagrams below show how a volatility surface moves in both absolute/fixed strike and relative strike due to a change in spot.

Figure 101. Sticky Strike Absolute/Fixed Strike **Sticky Strike Relative Strike (as Percentage of Spot)**



Source: Santander Investment Bolsa estimates.

TRADER'S SYSTEMS CAN GIVE ILLUSION OF STICKY STRIKE

While Figure 101 above describes which volatility regime normally applies in any given environment, there are many exceptions. A particular exception is that for very small time horizons volatility surfaces can seem to trade in a sticky strike regime. We believe this is due to many trading systems assuming a static strike volatility surface, which then has to be re-marked by traders (especially for less liquid instruments, as risk managers are likely to insist on volatilities being marked to their last known traded implied volatility)[37]. As the effect of these trading systems on pricing is either an illusion (as traders will re-mark their surface when asked to provide a firm quote) or well within the bid-offer arbitrage channel, we believe this effect should be ignored.

LONG SKEW UNPROFITABLE WITH STICKY STRIKE

Sticky strike can be thought of as a Black-Scholes model

While there is no profit or loss from re-marking a surface in a sticky strike model, a long skew position still has to pay skew theta. Overall, a long skew position is still unprofitable in sticky strike regimes, but it is less unprofitable than for a sticky delta regime.

[37] Anchor delta measures the effect of re-marking a volatility surface and is described in the section *Advanced (Practical or Shadow) Greeks* in the *Appendix*.

3. STICKY LOCAL VOLATILITY PRICES SKEW FAIRLY

As a sticky local volatility causes a negative correlation between spot and Black-Scholes volatility (shown below), this re-mark is profitable for long skew positions. As the value of this re-mark is exactly equal to the cost of skew theta, skew trades break even in a sticky local volatility regime. If volatility surfaces move as predicted by sticky local volatility, then skew is priced fairly (as skew trades do not make a loss or profit).

BLACK-SCHOLES VOL IS AVERAGE OF INSTANTANEOUS LOCAL VOL

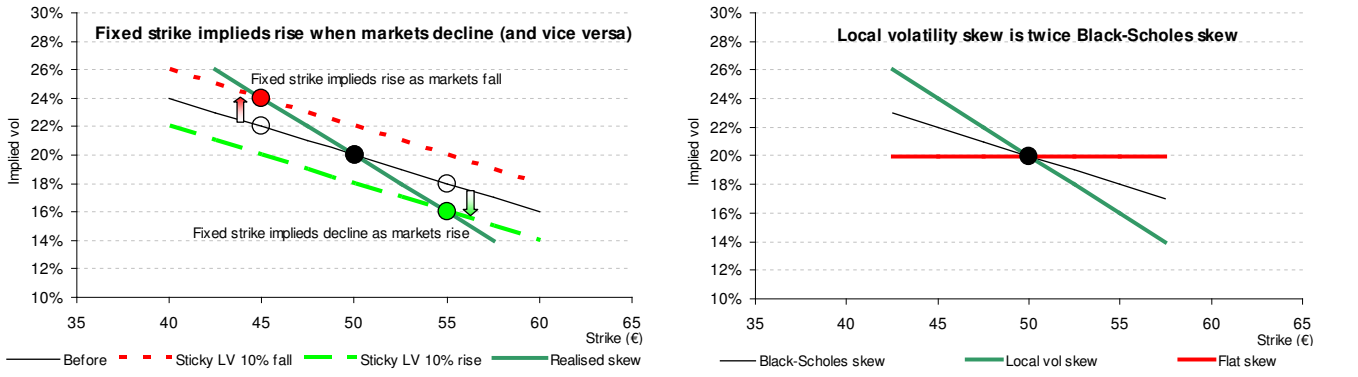
Local volatility is the name given for the instantaneous volatility of an underlying (ie, the exact volatility it has at a certain point). The Black-Scholes volatility of an option with strike K is equal to the average local (or instantaneous) volatility of all possible paths of the underlying from spot to strike K. This can be approximated by the average of the local volatility at spot and the local volatility at strike K. This approximation gives two results[38]:

- The ATM Black-Scholes volatility is equal to the ATM local volatility.
- Black-Scholes skew is half the local volatility skew (due to averaging).

Example of local volatility skew = 2x Black-Scholes skew

The second point can be seen if we assume the local volatility for the 90% strike is 22% and the ATM local volatility is 20%. The 90%-100% local volatility skew is therefore 2%. As the Black-Scholes 90% strike option will have an implied volatility of 21% (the average of 22% and 20%), it has a 90%-100% skew of 1% (as the ATM Black-Scholes volatility is equal to the 20% ATM local volatility).

Figure 102. Sticky Local Volatility Absolute/Fixed Strike Black-Scholes and Local Volatility Skew



Source: Santander Investment Bolsa estimates.

STICKY LOCAL VOL IMPLIES NEGATIVE SPOT VOL CORRELATION

As local volatility skew is twice the Black-Scholes skew, and ATM volatilities are the same, a sticky local volatility surface implies a negative correlation between spot and implied volatility. This can be seen by the ATM Black-Scholes volatility resetting higher if spot declines and is shown in the diagrams above.

[38] As this is an approximation, there is a slight difference which we shall ignore.

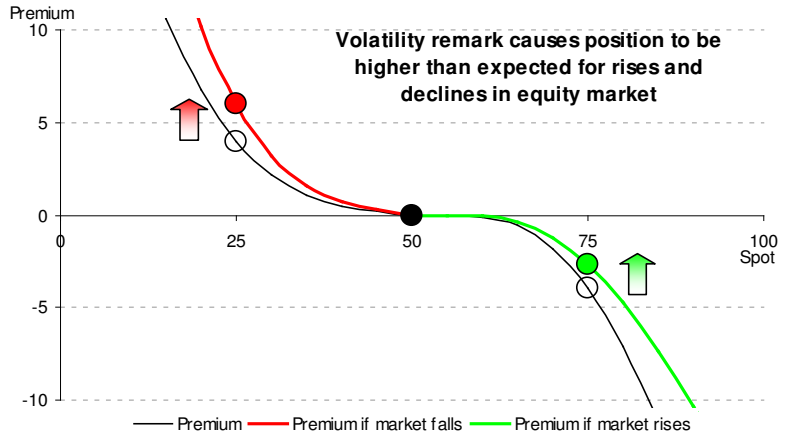
Example of negative correlation between spot and Black-Scholes volatility

We shall use the values from the previous example, with the local volatility for the 90% strike = 22%, Black-Scholes of the 90% strike = 21% and the ATM volatility for both (local and Black-Scholes) = 20%. If markets decline 10%, then the 90% strike option Black-Scholes volatility will rise 1% from 21% to 22% (as ATM for both local and Black-Scholes volatility must be equal). This 1% move will occur in parallel over the entire surface (as the Black-Scholes skew has not changed). Similarly, should markets rise 10%, the Black-Scholes volatility surface will fall 1% (assuming constant skew).

LONG SKEW PROFITS FROM VOLATILITY SURFACE RE-MARK

In order to demonstrate how the negative correlation between spot and (Black-Scholes) implied volatility causes long skew positions to profit from volatility surfaces re-mark, we shall assume an investor is long a risk reversal (long OTM put, short OTM call). This position is shown in Figure 103 below. When markets fall, the primary driver of the risk reversal's value is the put (which is now more ATM than the call), and the put value will increase due to the rise in implied volatility (due to negative correlation with spot). Similarly, the theoretical value of the risk reversal will rise (as the call is now more ATM – and therefore the primary driver of value – and, as implied volatilities decline as markets rise, the value of the short call will rise as well). The long skew position therefore profits from both a movement up or down in equity markets, as can be seen in the diagram below as both the long call and short put position increase in value.

Figure 103. Premium of Long Put, Short Call (long skew) Risk Reversal



Source: Santander Investment Bolsa.

VOLATILITY RE-MARK WITH STICKY LOCAL VOL = SKEW THETA

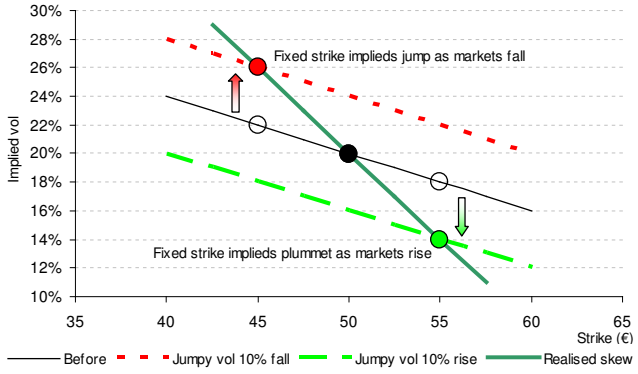
Skew trades break even with a static local volatility model

While a sticky local volatility regime causes long skew positions to profit from (Black-Scholes) implied volatility changes, the position still suffers from skew theta. The combination of these two cancel exactly, causing a long (or short) skew trade to break even. As skew trades break even under a static local volatility model, and as there is a negative spot vol correlation, it is arguably the most realistic volatility model.

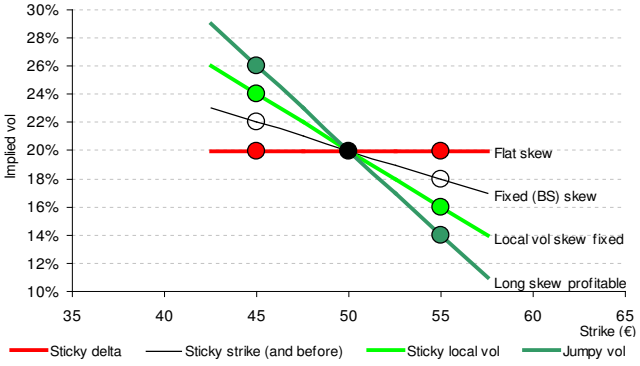
4. JUMPY VOLATILITY IS ONLY REGIME WHERE LONG SKEW IS PROFITABLE

During very panicked markets, or immediately after a crash, there is typically a very high correlation between spot and volatility. During this volatility regime (which we define as jumpy volatility) volatility surfaces move in excess of that implied by sticky local volatility. As the implied volatility surface re-mark for a long skew position is in excess of skew theta, long skew positions are profitable. A jumpy volatility regime tends to last for a relatively short period of time.

Figure 104. Jumpy Volatility Absolute/Fixed Strike



Realised Skew of Sticky Delta/Strike/Local/Jumpy Vol



Source: Santander Investment Bolsa estimates.

Example of volatility regimes and skew trading

If one-year 90%-100% skew is 25bp per 1% (ie, 2.5% for 90-100%) and markets fall 1%, volatility surfaces have to rise by 25bp for the profit from realised skew to compensate for the cost of skew theta. If surfaces move by more than 25bp, surfaces are moving in a jumpy volatility way and skew trades are profitable. If surfaces move by less than 25bp then skew trades suffer a loss.

Figure 105. Breakdown of P&L for Skew Trades

	P&L breakdown for long skew (e.g. long put, short call)				
Volatility regime	Remark		Skew theta		Total
Sticky delta		+		=	
Sticky strike		+		=	
Sticky local volatility		+		=	
Jumpy volatility		+		=	

Source: Santander Investment Bolsa.

SKEW ONLY FAIRLY PRICED IF ATM MOVES BY TWICE THE SKEW

We define realised skew to be the profit due to re-marking the surface

For a given movement in spot from S_0 to S_1, we shall define the movement of the (Black-Scholes) implied volatility surface divided by the skew (implied volatility of strike S_1 – implied volatility of strike S_0) to be the realised skew. The realised skew can be thought of as the profit due to re-marking the volatility surface. Defining realised skew to be the movement in the volatility surface is similar to the definition of realised volatility, which is the movement in spot.

realised skew = movement of surface/skew

where:

movement of surface = movement of surface when spot moves from S_0 to S_1

skew = difference in implied volatility between S_1 and S_0

The ATM volatility can then be determined by the below equation:

$$ATM_{time\ 1} = ATM_{time\ 0} + skew + movement\ of\ surface$$

$$\Rightarrow ATM_{time\ 1} = ATM_{time\ 0} + skew + (skew \times realised\ skew)$$

$$\Rightarrow ATM_{time\ 1} = ATM_{time\ 0} + skew \times (1 + realised\ skew)$$

The realised skew for sticky delta is therefore -1 in order to keep ATM constant (and hence skew flat) for all movements in spot. A sticky strike regime has a realised skew of 0, as there is no movement of the volatility surface and skew is fixed. A local volatility model has a realised skew of 1, which causes ATM to move by twice the value implied by a fixed skew. As local volatility prices skew fairly, skew is only fairly priced if ATM moves by twice the skew. We shall assume the volatility surface for jumpy volatility moves more than it does for sticky local volatility, hence has a realised skew of more than 1.

Skew profit is proportional to realised skew – 1 (due to skew theta)

In order to calculate the relative profit (or loss) of trading skew, the value of skew theta needs to be taken away, and this value can be thought of as -1. Skew profit is then given by the formula below:

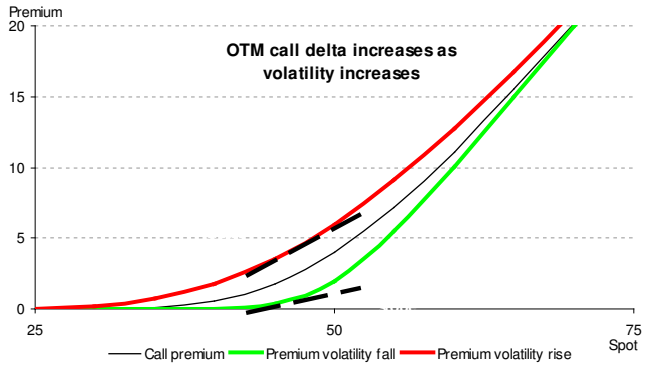
Skew profit α realised skew - 1

SKEW TRADING IS EQUIVALENT TO TRADING 2ND ORDER GAMMA

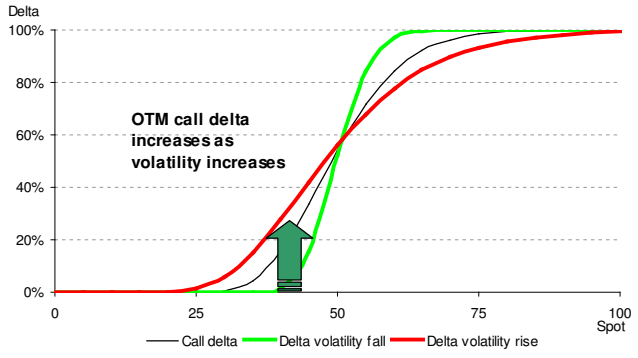
Comparing vanna to skew theta can identify trading opportunities

Determining the current volatility regime helps a trader decide if skew trades are likely to be profitable. In order to determine the strikes used to initiate long or short skew positions, a trader needs to evaluate the richness or cheapness of skew across different strikes. It is possible to show intuitively, and mathematically, that skew trading is very similar to delta hedging gamma. Given this relationship, comparing vanna (dVega/dSpot), weighted by the square root of time, to skew theta can be a useful rule of thumb to identify potential trading opportunities.

Figure 106. Call Option with 50 Strike



Delta of Call Option with Rise and Fall in Implied Volatility

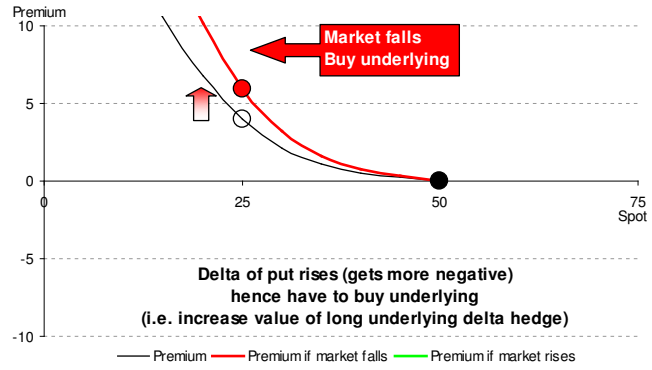


Source: Santander Investment Bolsa estimates.

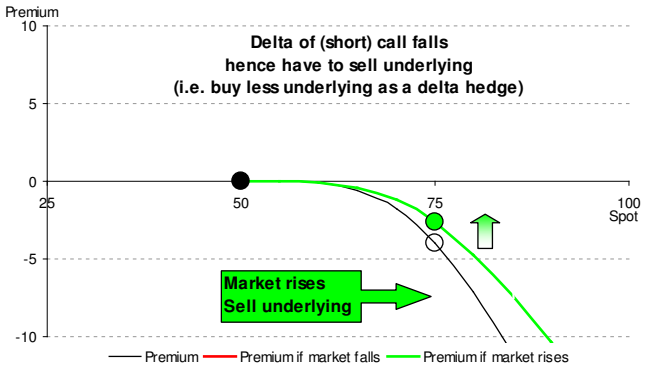
MOVEMENT OF IMPLIED VOL SURFACE CHANGES DELTA OF OPTIONS

We shall assume we are in a sticky local vol (or jumpy vol) market, ie, volatility rises if markets fall, and a trader is trading skew using a long OTM put and short OTM call (ie, a risk reversal). As the delta of OTM options increases in value if implied volatility increases, and vice versa, the delta hedging of the long skew position is impacted by the movement in volatility surfaces.

Figure 107. Delta of Long Put if Market Falls (Local Vol)



Delta of Short Call if Market Rises (Local Vol)



Source: Santander Investment Bolsa estimates.

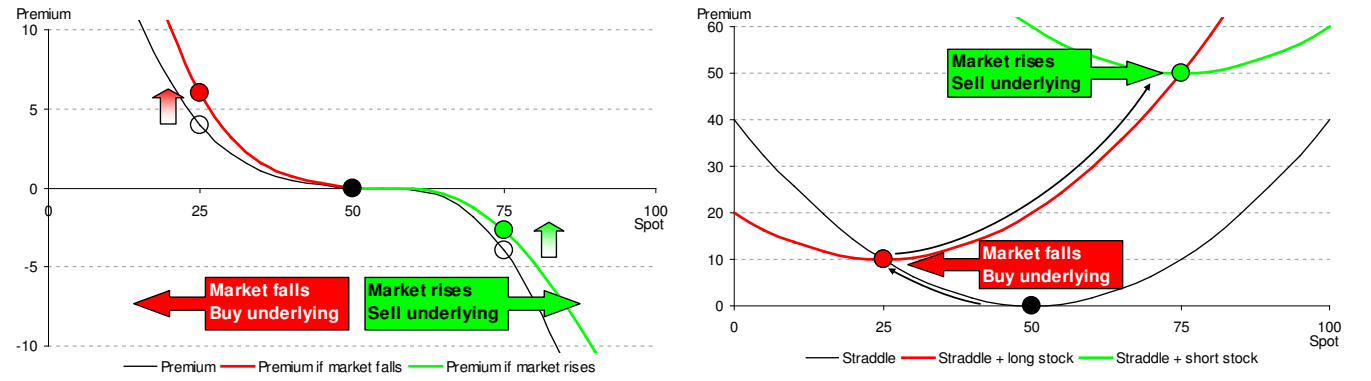
Trader needs to buy stock (or futures) if market declines

If there is a decline in spot, the volatility of the long put (which is now more ATM and the primary driver of value) increases. This causes the delta of the position to decrease (absolute delta of put increases and, as delta of put is negative, the delta decreases). A trader has to buy more stock (or futures) than expected in order to compensate for this change, as shown on the left of Figure 107 above.

Conversely, trader needs to sell stock (or futures) if the market rises

The opposite trade occurs if markets rise as, for an increase in spot, the volatility of the short call (which is now more ATM and the primary driver of value) decreases. This causes the delta of the position to increase (delta of call decreases as delta of short call increases). Traders have to sell more stock (or futures) than they expect to compensate for this change (as shown on the right of Figure 107 above), which is the reverse trade of that which occurs for a decline in the market.

Figure 108. Delta Hedging Due to Trading Skew **Delta Hedging Due to Trading Gamma**



Source: Santander Investment Bolsa estimates.

DELTA HEDGING SKEW IS SIMILAR TO DELTA HEDGING GAMMA

Trading skew involves delta hedging the same way as trading gamma

Let us assume a negative correlation between spot and volatility (ie, for sticky local volatility or jumpy volatility) and that a trader is initially delta hedged[39] and intends to remain so. The movement of the volatility surface means the trader has to buy more stock (or futures) than he expects if markets fall and sell more stock (or futures) if markets rise. Buying low and selling high locks in the profit from the long skew position. This trade is identical to delta hedging a long gamma position, which can be seen in Figure 108 above.

If there is a positive relationship between spot and volatility (ie, a sticky delta volatility regime), then the reverse trade occurs with stock (or futures) being sold if markets decline and bought if markets rise. For sticky delta regimes, a long skew position is similar to being short gamma (and hence very unprofitable, given skew theta has to be paid as well).

[39] A long put and short call risk reversal would be delta hedged with a long stock (or futures) position.

MATHEMATICALLY, SKEW TRADING IS SIMILAR TO GAMMA TRADING

It is possible to show mathematically the relationship between skew trading and gamma trading if one assumes a correlation between spot and volatility. Vanna, the rate of change in vega for a change in spot (dVega/dSpot) measures the size of a skew position. This can be seen intuitively from the arguments above; as markets decline, the OTM put becomes more ATM and hence the primary driver of value. It is this change in vega (long put dominating the short call) for a change in spot, that causes volatility surface re-marks to be profitable for skew trading. Vanna is not only equal to dVega/dSpot, but is also equal to dDelta/dVol[40]. The equations below show that this relationship, when combined with spot being correlated to volatility, links skew and gamma trading.

Vanna = dDelta/dVol (and = dVega/dSpot)

As Vol α Spot

⇨ Vanna α dDelta/dSpot

As Gamma = dDelta/dSpot

⇨ Vanna α Gamma

Therefore, gamma can be considered to be second order gamma due to the negative correlation between volatility and spot.

[40] The proof of this relationship is outside of the scope of this publication

SKEW THETA PAYS FOR SKEW, GAMMA THETA PAYS FOR GAMMA

In order to break down an option's profit into volatility and skew, the total theta paid needs to be separated into gamma theta and skew theta. Gamma theta pays for gamma (or volatility) while skew theta pays for skew. We note that skew across the term structure can be compared with each other if weighted by the square root of time. As skew is measured by vanna, skew theta should therefore be compared to power vanna (vanna weighted by the square root of time) to identify skew trading opportunities. This is equivalent to comparing gamma to gamma theta. The method for calculating skew theta is given below.

Total theta = gamma theta + skew theta (all measured in theta per units of cash gamma)

Cash (or dollar) gamma = $\gamma \times S^2 / 100$ = notional cash value bought (or sold) per 1% spot move

GAMMA THETA IDENTICAL FOR ALL OPTIONS IF IMPLIEDS IDENTICAL

Gamma theta is the cost (or income) from a long (or short) gamma position. To calculate the cost of gamma, we shall assume an index has a volatility of 20% for all strikes and maturities. We shall ignore interest rates, dividends and borrowing costs and assume spot is currently at 3000pts.

Figure 109. Theta (per Year) of Index with 20% Implied

Strike	3 Months	1 Year	4 Years
80%	-0.070	-0.227	-0.178
90%	-0.517	-0.390	-0.213
100%	-0.949	-0.473	-0.233
110%	-0.632	-0.442	-0.237
120%	-0.197	-0.342	-0.230

Source: Santander Investment Bolsa estimates.

Cash Gamma per 1% Move of Index with 20% Implied

Strike	3 Months	1 Year	4 Years
80%	9	29	22
90%	65	49	27
100%	120	60	29
110%	80	56	30
120%	25	43	29

Both gamma and theta are high for short-dated ATM options

As can be seen in Figure 109, both cash gamma and theta are highest for near-dated and ATM options. When the cost per unit of cash gamma is calculated, it is identical for all strikes and expiries as the implied volatility is 20% for them all. We shall define the ATM theta cost per unit of cash gamma to be gamma theta (in units of 1 million cash gamma to have a reasonably sized number). This is, essentially, the values on the left in Figure 109 divided by the values on the right in Figure 109.

Figure 110. Theta per 1 million Cash Gamma (Gamma Theta) of Index with 20% Implied

Strike	3 Months	1 Year	4 Years
80%	-7,937	-7,937	-7,937
90%	-7,937	-7,937	-7,937
100% = Gamma Theta	-7,937	-7,937	-7,937
110%	-7,937	-7,937	-7,937
120%	-7,937	-7,937	-7,937

Source: Santander Investment Bolsa.

TERM STRUCTURE CHANGES GAMMA THETA BY MATURITY

In order to have a more realistic volatility surface we shall introduce positive sloping term structure, while keeping the implied volatility of one-year maturity options identical. As there is no skew in the surface, all the theta is solely due to the cost of gamma or gamma theta. The gamma theta is now lower for near-dated maturities, which is intuitively correct as near-dated implieds are now lower than the far-dated implieds.

Figure 111. Volatility Surface with Term Structure

Strike	3 Months	1 Year	4 Years
80%	19%	20%	21%
90%	19%	20%	21%
100%	19%	20%	21%
110%	19%	20%	21%
120%	19%	20%	21%

Source: Santander Investment Bolsa estimates.

Theta per 1mn Cash Gamma (Gamma Theta)

Strike	3 Months	1 Year	4 Years
80%	-7,163	-7,937	-8,338
90%	-7,163	-7,937	-8,338
100%= Gamma Theta	-7,163	-7,937	-8,338
110%	-7,163	-7,937	-8,338
120%	-7,163	-7,937	-8,338

SKEW MAKES IT MORE EXPENSIVE TO OWN PUTS THAN CALLS

If we introduce skew to the volatility surface we increase the cost of gamma for puts and decrease it for calls. This can be seen on the right of Figure 112; the ATM options have the same cost of gamma as before but the wings now have a different value.

Figure 112. Vol Surface with Skew and Term Structure

Strike	3 Months	1 Year	4 Years
80%	27%	24%	23%
90%	23%	22%	22%
100%	19%	20%	21%
110%	15%	18%	20%
120%	11%	16%	19%

Source: Santander Investment Bolsa estimates.

Theta per 1mn Cash Gamma

Strike	3 Months	1 Year	4 Years
80%	-14,464	-11,429	-10,045
90%	-10,496	-9,603	-9,172
100%= Gamma Theta	-7,163	-7,937	-8,338
110%	-4,464	-6,429	-7,545
120%	-2,401	-5,079	-6,791

SKEW THETA IS THE COST OF GOING LONG SKEW

As we have defined the theta paid for ATM option gamma (or gamma theta) as the fair price for gamma, the difference between this value and other options' cost of gamma is the cost of skew (or skew theta). Skew theta is therefore calculated by subtracting the cost of ATM gamma from all other options (and hence skew theta is zero for ATM options by definition).

Figure 113. Theta per 1mn Cash Gamma

Strike	3 Months	1 Year	4 Years
80%	-14,464	-11,429	-10,045
90%	**-10,496**	-9,603	-9,172
100%= Gamma Theta	**-7,163**	-7,937	-8,338
110%	-4,464	-6,429	-7,545
120%	-2,401	-5,079	-6,791

Source: Santander Investment Bolsa estimates.

Skew Theta

Strike	3 Months	1 Year	4 Years
80%	-7,302	-3,492	-1,706
90%	**-3,333**	-1,667	-833
100%	0	0	0
110%	2,698	1,508	794
120%	4,762	2,857	1,548

Example of skew theta calculation

The annual cost for a million units of cash gamma for three-month 90% strike options is €10,496, whereas ATM options only have to pay €7,163. The additional cost of being long 90% options (rather than ATM) is therefore €10,496 - €7,163 = €3,333. This additional €3,333 cost is the cost of being long skew, or skew theta.

Strikes lower than ATM suffer from skew theta

For low strike options there is a cost (negative sign) to owning the option and hence being long skew. High strike options benefit from an income of skew theta (which causes the lower cost of gamma) to compensate for being short skew (hence they have a positive sign).

VOLATILITY SLIDE THETA HAS A MINOR EFFECT ON SKEW TRADING

If one assumes volatility surfaces have relative time (one-year skew stays the same) rather than absolute time (ie, Dec14 skew stays the same) then one needs to take into account volatility slide theta (to factor in the increase in skew as the maturity of the option decreases). Volatility slide theta partly compensates for the cost of skew theta. For more details, see the section *Advanced (Practical or Shadow) Greeks* in the *Appendix*.

APPENDIX

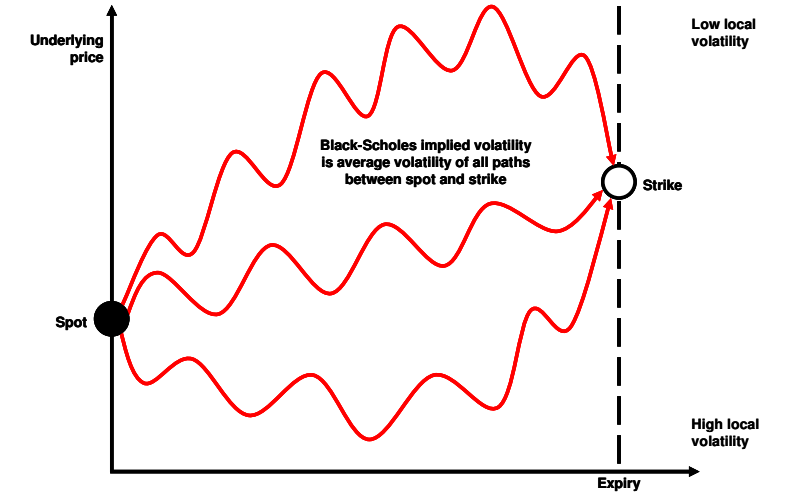
LOCAL VOLATILITY

While Black-Scholes is the most popular method for pricing vanilla equity derivatives, exotic equity derivatives (and ITM American options) usually require a more sophisticated model. The most popular model after Black-Scholes is a local volatility model as it is the only completely consistent volatility model[41]. A local volatility model describes the instantaneous volatility of a stock, whereas Black-Scholes is the average of the instantaneous volatilities between spot and strike.

LOCAL VOLATILITY IS INSTANTANEOUS VOLATILITY OF UNDERLYING

Instantaneous volatility is the volatility of an underlying at any given local point, which we shall call the local volatility. We shall assume the local volatility is fixed and has a normal negative skew (higher volatility for lower spot prices). There are many paths from spot to strike and, depending on which path is taken, they will determine how volatile the underlying is during the life of the option (see Figure 114).

Figure 114. Different Paths between Spot and Strike



Source: Santander Investment Bolsa.

BLACK-SCHOLES VOLATILITY IS AVERAGE OF LOCAL VOLATILITIES

It is possible to calculate the local (or instantaneous) volatility surface from the Black-Scholes implied volatility surface. This is possible as the Black-Scholes implied volatility of an option is the average of all the paths between spot (ie, zero maturity ATM strike) and the maturity and strike of the option. A reasonable approximation is the average of all local volatilities on a direct straight-line path between spot and strike. For a normal relatively flat skew, this is simply the average of two values, the ATM local volatility and the strike local volatility.

[41] Strictly speaking, this is true only for deterministic models. However, as the expected volatility of non-deterministic models has to give identical results to a local volatility model to be completely consistent, they can be considered to be a 'noisy' version of a local volatility model.

Black-Scholes skew is half local volatility skew as it is the average

ATM implieds are identical for local vol and Black-Scholes, but local vol skew is twice Black-Scholes

If the local volatility surface has a 22% implied at the 90% strike, and 20% implied at the ATM strike, then the Black-Scholes implied volatility for the 90% strike is 21% (average of 22% and 20%). As ATM implieds are identical for both local and Black-Scholes implied volatility, this means that 90%-100% skew is 2% for local volatility but 1% for Black-Scholes. Local volatility skew is therefore twice the Black-Scholes skew.

ATM volatility is the same for both Black-Scholes and local volatility

For ATM implieds, the local volatility at the strike is equal to ATM, hence the average of the two identical numbers is simply equal to the ATM implied. For this reason, Black-Scholes ATM implied is equal to local volatility ATM implied.

LOCAL VOL IS THE ONLY COMPLETE CONSISTENT VOL MODEL

A local volatility model is complete (it allows hedging based only on the underlying asset) and consistent (does not contain a contradiction). It is often used to calculate exotic option implied volatilities to ensure the prices for these exotics are consistent with the values of observed vanilla options and hence prevent arbitrage. A local volatility model is the only complete consistent volatility model; a constant Black-Scholes volatility model (constant implied volatility for all strikes and expiries) can be considered to be a special case of a static local volatility model (where the local volatilities are fixed and constant for all strikes and expiries).

MEASURING HISTORICAL VOLATILITY

The implied volatility for a certain strike and expiry has a fixed value. There is, however, no single calculation for historical volatility. The number of historical days for the historical volatility calculation changes the calculation, in addition to the estimate of the drift (or average amount stocks are assumed to rise). There should, however, be no difference between the average daily or weekly historical volatility. We also examine different methods of historical volatility calculation, including close-to-close volatility and exponentially weighted volatility, in addition to advanced volatility measures such as Parkinson, Garman-Klass (including Yang-Zhang extension), Rogers and Satchell and Yang-Zhang. We also show that it is best to assume a zero drift assumption for close-to-close volatility, and that under this condition variance is additive.

CLOSE-TO-CLOSE HISTORICAL VOLATILITY IS THE MOST COMMON

Volatility is defined as the annualised standard deviation of log returns. For historical volatility the usual measure is close-to-close volatility, which is shown below.

Log return = $x_i = \text{Ln}\left(\frac{c_i + d_i}{c_{i-1}}\right)$ where d_i = ordinary dividend and c_i is close price

Volatility[42] (not annualised) = $\sigma_x \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$

where $\bar{x}$ = drift = Average (x_i)

BEST TO ASSUME ZERO DRIFT FOR VOLATILITY CALCULATION

For relatively short time periods (daily, weekly), the drift should be close to zero and can be ignored

The calculation for standard deviation calculates the deviation from the average log return (or drift). This average log return has to be estimated from the sample, which can cause problems if the return over the period sampled is very high or negative. As over the long term very high or negative returns are not realistic, the calculation of volatility can be corrupted by using the sample log return as the expected future return. For example, if an underlying rises 10% a day for ten days, the volatility of the stock is zero (as there is zero deviation from the 10% average return). This is why volatility calculations are normally more reliable if a zero return is assumed. In theory, the expected average value of an underlying at a future date should be the value of the forward at that date. As for all normal interest rates (and dividends, borrow cost) the forward return should be close to 100% (for any reasonable sampling frequency, ie, daily/weekly/monthly). Hence, for simplicity reasons it is easier to assume a zero log return as Ln(100%) = 0.

[42] We take the definition of volatility of John Hull in *Options, Futures and Other Derivatives* in which n day volatility uses n returns and n+1 prices. We note Bloomberg uses n prices and n-1 returns.

VARIANCE IS ADDITIVE IF ZERO MEAN IS ASSUMED

Frequency of returns in a year = F (eg, 252 for daily returns)

$$\sigma_{\text{Annualised}} = \sqrt{\text{F}} \times \sigma_{\text{x}} = \sqrt{\text{F}} \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

As $\bar{x} \approx 0$ if we assume zero average returns

$$\sigma_{\text{Annualised}} = \sqrt{\text{F}} \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^{\,2}}$$

$$\sigma_{\text{Annualised}}^{\,2} = \frac{\text{F}}{N}\sum_{i=1}^{N} x_i^{\,2}$$

Now if we assume that the total sample N can be divided up into period 1 and period 2 where period 1 is the first M returns then:

$$\sigma_{\text{Total}}^{\,2} = \frac{\text{F}}{N_{\text{Total}}} \times \sum_{i=1}^{N} x_i^{\,2}$$

$$\sigma_{\text{Period 1}}^{\,2} = \frac{\text{F}}{N_{\text{Period 1}}} \times \sum_{i=1}^{M} x_i^{\,2} \text{ (where } \text{N}_{\text{Period 1}} = \text{M)}$$

$$\sigma_{\text{Period 2}}^{\,2} = \frac{\text{F}}{N_{\text{Period 2}}} \times \sum_{i=M+1}^{N} x_i^{\,2} \text{ (where } \text{N}_{\text{Period 2}} = \text{N} - \text{M)}$$

then

$$\sigma_{\text{Total}}^{\,2} = \frac{\text{F}}{N_{\text{Total}}} \times \sum_{i=1}^{N} x_i^{\,2} = \frac{\text{F}}{N_{\text{Total}}} \left(\sum_{i=1}^{M} x_i^{\,2} + \sum_{i=M+1}^{N} x_i^{\,2} \right)$$

$$\sigma_{\text{Total}}^{\,2} = \frac{\text{F}}{N_{\text{Total}}} \times \sum_{i=1}^{M} x_i^{\,2} + \frac{\text{F}}{N_{\text{Total}}} \times \sum_{i=M+1}^{N} x_i^{\,2}$$

$$\sigma_{\text{Total}}^{\,2} = \frac{N_{\text{Period 1}}}{N_{\text{Total}}} \left(\frac{\text{F}}{N_{\text{Period 1}}} \times \sum_{i=1}^{M} x_i^{\,2} \right) + \frac{N_{\text{Period 2}}}{N_{\text{Total}}} \left(\frac{\text{F}}{N_{\text{Period 2}}} \times \sum_{i=M+1}^{N} x_i^{\,2} \right)$$

$$\sigma_{\text{Total}}^{\,2} = \frac{N_{\text{Period 1}}}{N_{\text{Total}}} \sigma_{\text{Period 1}}^{\,2} + \frac{N_{\text{Period 2}}}{N_{\text{Total}}} \sigma_{\text{Period 2}}^{\,2}$$

Hence variance is additive (when weighted by the time in each period / total time)

LOG RETURNS CAN BE APPROXIMATED BY PERCENTAGE RETURNS

As returns are normally close to 1 (=100%) the log of returns is very similar to return – 1 (which is the percentage change of the price). If the return over the period is assumed to be the same for all periods, and if the mean return is assumed to be zero (it is normally very close to zero), the standard deviation of the percentage change is simply the absolute value of the percentage return. Hence an underlying that moves 1% has a volatility of 1% for that period. As volatility is usually quoted on an annualised basis, this volatility has be multiplied by the square root of the number of samples in a year (ie, $\sqrt{252}$ for daily returns,$\sqrt{52}$ for weekly returns and $\sqrt{12}$ for monthly returns).

Number of trading days in year	= 252	=> Multiply daily returns by $\sqrt{252}$	≈ 16
Number of weeks in year	= 52	=> Multiply weekly returns by $\sqrt{52}$	≈ 7
Number of months in year	= 12	=> Multiply monthly returns by $\sqrt{12}$	≈ 3.5

WHICH HISTORICAL VOLATILITY SHOULD I USE?

When examining how attractive the implied volatility of an option is, investors will often compare it to historical volatility. However, historical volatility needs two parameters.

- Length of time (eg, number of days/weeks/months)
- Frequency of measurement (eg, daily/weekly)

LENGTH OF TIME FOR HISTORICAL VOLATILITY

Historical volatility should be a multiple of 3 months to have a constant number of quarterly reporting periods

Choosing the historical volatility number of days is not a trivial choice. Some investors believe the best number of days of historical volatility to look at is the same as the implied volatility of interest. For example, one-month implied should be compared to 21 trading day historical volatility (and three-month implied should be compared to 63-day historical volatility, etc). While an identical duration historical volatility is useful to arrive at a realistic minimum and maximum value over a long period of time, it is not always the best period of time to determine the fair level of long-dated implieds. This is because volatility mean reverts over a period of c8 months. Using historical volatility for periods longer than c8 months is not likely to be the best estimate of future volatility (as it could include volatility caused by earlier events, whose effect on the market has passed). Arguably a multiple of three months should be used to ensure that there is always the same number of quarterly reporting dates in the historical volatility measure. Additionally, if there has been a recent jump in the share price that is not expected to reoccur, the period of time chosen should try to exclude that jump.

The best historical volatility period does not have to be the most recent

If there has been a rare event which caused a volatility spike, the best estimate of future volatility is not necessary the current historical volatility. A better estimate could be the past historical volatility when an event that caused a similar volatility spike occurred. For example, the volatility post credit crunch could be compared to the volatility spike after the Great Depression or during the bursting of the tech bubble.

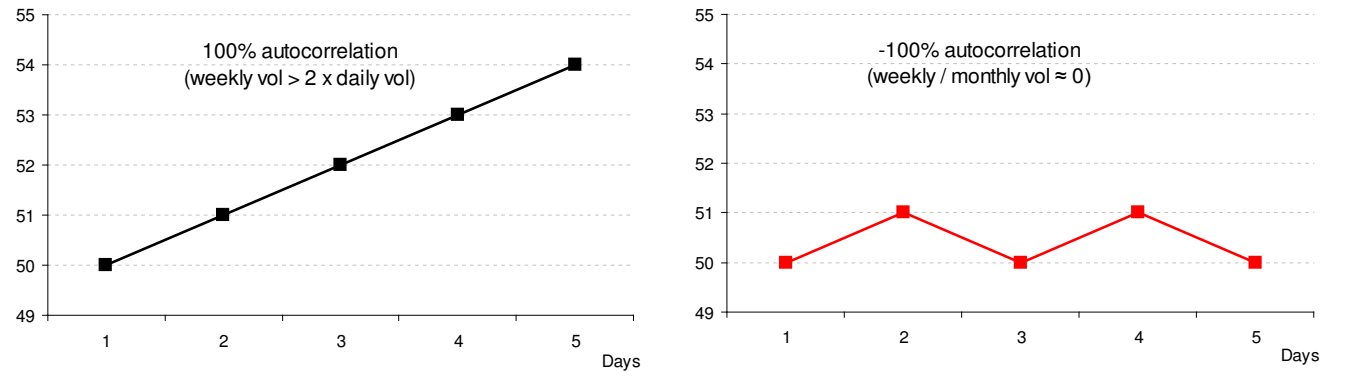
FREQUENCY OF HISTORICAL VOLATILITY

While historical volatility can be measured monthly, quarterly or yearly, it is usually measured daily or weekly. Normally, daily volatility is preferable to weekly volatility as five times as many data points are available. However, if volatility over a long period of time is being examined between two different markets, weekly volatility could be the best measure to reduce the influence of different public holidays (and trading hours[43]). If stock price returns are independent, then the daily and weekly historical volatility should on average be the same. If stock price returns are not independent, there could be a difference. Autocorrelation is the correlation between two different returns so independent returns have an autocorrelation of 0%.

Trending markets imply weekly volatility is greater than daily volatility

With 100% autocorrelation, returns are perfectly correlated (ie, trending markets). Should autocorrelation be -100% correlated, then a positive return is followed by a negative return (mean reverting or range trading markets). If we assume markets are 100% daily correlated with a 1% daily return, this means the weekly return is 5%. The daily volatility is therefore c16% (1% × √252), while the weekly volatility of c35% (5% × √52) is more than twice as large.

Figure 115. Stock Price with 100% Daily Autocorrelation **Stock Price with -100% Daily Autocorrelation**



Source: Santander Investment Bolsa estimates.

High market share of high frequency trading should prevent autocorrelation

Historically (decades ago), there could have been positive autocorrelation due to momentum buying, but once this became understood this effect is likely to have faded. Given the current high market share of high frequency trading (accounting for up to three-quarters of US equity trading volume), it appears unlikely that a simple trading strategy such as 'buy if security goes up, sell if it goes down' will provide above-average returns over a significant period of time[44].

Panicked markets could cause temporary negative autocorrelation

While positive autocorrelation is likely to be arbitraged out of the market, there is evidence that markets can overreact at times of stress as market panic (rare statistical events can occur under the weak form of efficient market hypotheses). During these events human traders and some automated trading systems are likely to stop trading (as the event is rare, the correct response is unknown), or potentially exaggerate the trend (as positions get 'stopped out' or to follow the momentum of the move). A strategy that is long daily variance and short weekly variance will therefore usually give relatively flat returns, but occasionally give a positive return.

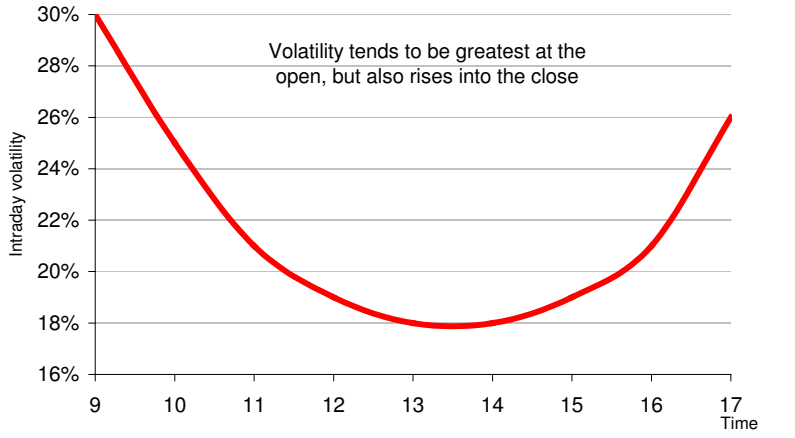
[43] Advanced volatility measures could be used to remove part of the effect of different trading hours.
[44] Assuming there are no short selling restrictions.

INTRADAY VOLATILITY IS NOT CONSTANT

Advanced volatility measures should be used by traders wishing to take into account intraday prices

For most markets, intraday volatility is greatest just after the open (as results are often announced around the open) and just before the close (performance is often based upon closing prices). Intraday volatility tends to sag in the middle of the day due to the combination of a lack of announcements and reduced volumes/liquidity owing to lunch breaks. For this reason, using an estimate of volatility more frequent than daily tends to be very noisy. Traders who wish to take into account intraday prices should instead use an advanced volatility measure.

Figure 116. Intraday Volatility



Source: Santander Investment Bolsa.

EXPONENTIALLY WEIGHTED VOLATILITIES ARE RARELY USED

Exponentially weighted moving average can be used to reduce effect of spikes in volatility disappearing

An alternate measure could be to use an exponentially weighted moving average model, which is shown below. The parameter λ is between zero (effectively one-day volatility) and one (ignore current vol and keep vol constant). Normally, values of c0.9 are used. Exponentially weighted volatilities are rarely used, partly due to the fact they do not handle regular volatility-driving events such as earnings very well. Previous earnings jumps will have least weight just before an earnings date (when future volatility is most likely to be high) and most weight just after earnings (when future volatility is most likely to be low). It could, however, be of some use for indices.

$$\sigma_i^2 = \lambda\sigma_{i-1}^2 + (1-\lambda)x_i^2$$

Exponentially weighted volatility avoids volatility collapse of historic volatility

Exponential volatility has the advantage over standard historical volatility in that the effect of a spike in volatility gradually fades (as opposed to suddenly disappearing causing a collapse in historic volatility). For example, if we are looking at the historical volatility over the past month and a spike in realised volatility suddenly occurs the historical volatility will be high for a month, then collapse. Exponentially weighted volatility will rise at the same time as historical volatility and then gradually decline to lower levels (arguably in a similar way to how implied volatility spikes, then mean reverts).

ADVANCED VOLATILITY MEASURES

Volatility measures can use open, high and low prices in addition to closing price

Close-to-close volatility is usually used as it has the benefit of using the closing auction prices only. Should other prices be used, then they could be vulnerable to manipulation or a 'fat fingered' trade. However, a large number of samples need to be used to get a good estimate of historical volatility, and using a large number of closing values can obscure short-term changes in volatility. There are, however, different methods of calculating volatility using some or all of the open (O), high (H), low (L) and close (C). The methods are listed in order of their maximum efficiency (close-to-close variance divided by alternative measure variance).

- **Close to close (C).** The most common type of calculation that benefits from only using reliable prices from closing auctions. By definition its efficiency is one at all times.

- **Parkinson (HL).** As this estimate only uses the high and low price for an underlying, it is less sensitive to differences in trading hours. For example, as the time of the EU and US closes are approximately half a trading day apart, they can give very different returns. Using the high and low means the trading over the whole day is examined, and the days overlap. As it does not handle jumps, on average it underestimates the volatility, as it does not take into account highs and lows when trading does not occur (weekends, between close and open). Although it does not handle drift, this is usually small. The Parkinson estimate is up to 5.2 times more efficient than the close-to-close estimate. While other measures are more efficient based on simulated data, some studies have shown it to be the best measure for actual empirical data.

- **Garman-Klass (OHLC).** This estimate is the most powerful for stocks with Brownian motion, zero drift and no opening jumps (ie, opening price is equal to closing price of previous period). While it is up to 7.4 times as efficient as the close to close estimate, it also underestimates the volatility (as like Parkinson it assumes no jumps).

- **Rogers-Satchell (OHLC).** The efficiency of the Rogers-Satchell estimate is similar to that for Garman-Klass; however, it benefits from being able to handle non-zero drift. Opening jumps are not handled well though, which means it underestimates the volatility.

- **Garman-Klass Yang-Zhang extension (OHLC).** Yang-Zhang extended the Garman-Klass method that allows for opening jumps hence it is a fair estimate, but does assume zero drift. It has an efficiency of eight times the close-to-close estimate.

- **Yang-Zhang (OHLC).** The most powerful volatility estimator which has minimum estimation error. It is a weighted average of Rogers-Satchell, the close-open volatility and the open-close volatility. It is up to a maximum of 14 times as efficient (for two days of data) as the close-to-close estimate.

Figure 117. Summary of Advanced Volatility Estimates

Estimate	Prices Taken	Handle Drift?	Handle Overnight Jumps?	Efficiency (max)
Close to close	C	No	No	**1**
Parkinson	HL	No	No	**5.2**
Garman-Klass	OHLC	No	No	**7.4**
Rogers-Satchell	OHLC	Yes	No	**8**
Garman-Klass Yang-Zhang ext.	OHLC	No	Yes	**8**
Yang-Zhang	OHLC	Yes	Yes	**14**

Source: Santander Investment Bolsa.

EFFICIENCY AND BIAS DETERMINE BEST VOLATILITY MEASURE

Standard close-to-close is best for large samples, Yang-Zhang is best for small samples

There are two measures that can be used to determine the quality of a volatility measure: efficiency and bias. Generally, for small sample sizes the Yang-Zhang measure is best overall, and for large sample sizes the standard close to close measure is best.

- **Efficiency.** Efficiency $(\sigma_x^2) = \frac{\sigma_{cc}^2}{\sigma_x^2}$ where σ_x is the volatility of the estimate and σ_{cc} is the volatility of the standard close to close estimate.

- **Bias.** Difference between the estimated variance and the average (ie, integrated) volatility.

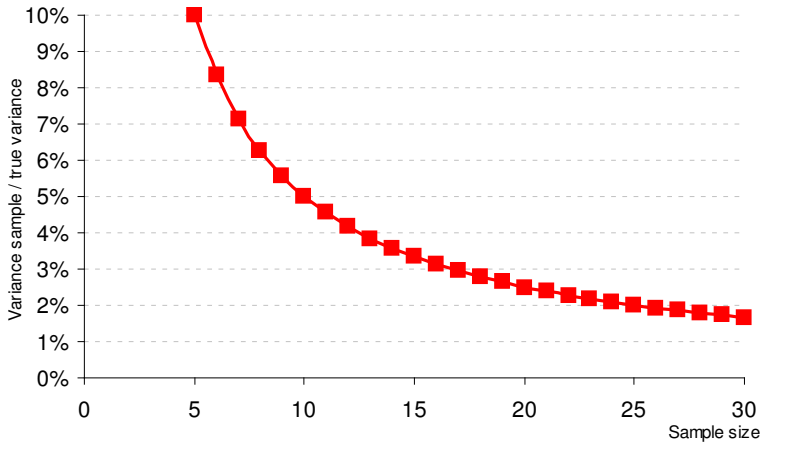
Efficiency measures the volatility of the estimate

The efficiency describes the variance, or volatility of the estimate. The efficiency is dependent on the number of samples, with efficiency decreasing the more samples there are (as close-to-close will converge and become less volatile with more samples). The efficiency is the theoretical maximum performance against an idealised distribution, and with real empirical data a far smaller benefit is usually seen (especially for long time series). For example, while the Yang-Zhang based estimators deal with overnight jumps if the jumps are large compared to the daily volatility the estimate will converge with the close-to-close volatility and have an efficiency close to one.

Close-to-close volatility should use at least five samples (and ideally 20 or more)

The variance of the close-to-close volatility can be estimated as a percentage of the actual variance by the formula 1/(2N) where N is the number of samples. This is shown in Figure 118 below and demonstrates that at least five samples are needed (or the estimate has a variance of over 10%) and that only marginal extra accuracy is gained for each additional sample above 20.

Figure 118. Variance of Close-To-Close Volatility/Actual Variance



Source: Santander Investment Bolsa.

Bias depends on the type of distribution of the underlying

Bias can be positive or negative depending on the distribution

While efficiency (how volatile the measure is) is important, so too is bias (whether the measure is, on average, too high or low). Bias depends on the sample size, and the type of distribution the underlying security has. Generally, the close-to-close volatility estimator is too big[45] (as it does not model overnight jumps), while alternative estimators are too small (as they assume continuous trading, and discrete trading will have a smaller difference between the maximum and minimum). The key variables that determine the bias are:

- **Sample size.** As the standard close-to-close volatility measure suffers with small sample sizes, this is where alternative measures perform best (the highest efficiency is reached for only two days of data).

- **Volatility of volatility.** While the close-to-close volatility estimate is relatively insensitive to a changing volatility (vol of vol), the alternative estimates are far more sensitive. This bias increases the more vol of vol increases (ie, more vol of vol means a greater underestimate of volatility).

Approximately 1/6 of total volatility occurs overnight

- **Overnight jumps between close and open.** Approximately one-sixth of equity volatility occurs outside the trading day (and approximately twice that amount for ADRs). Overnight jumps cause the standard close-to-close estimate to overestimate the volatility, as jumps are not modelled. Alternative estimates that do not model jumps (Parkinson, Garman Klass and Rogers-Satchell) underestimate the volatility. Yang-Zhang estimates (both Yang-Zhang extension of Garman Klass and the Yang-Zhang measure itself) will converge with standard close-to-close volatility if the jumps are large compared to the overnight volatility.

- **Drift of underlying.** If the drift of the underlying is ignored as it is for Parkinson and Garman Klass (and the Yang Zhang extension of Garman Glass), then the measure will overestimate the volatility. This effect is small for any reasonable drifts (ie, if we are looking at daily, weekly or monthly data).

- **Correlation daily volatility and overnight volatility.** While Yang-Zhang measures deal with overnight volatility, there is the assumption that overnight volatility and daily volatility are uncorrelated. Yang-Zhang measures will underestimate volatility when there is a correlation between daily return and overnight return (and vice versa), but this effect is small.

Variance, volatility and gamma swaps should look at standard volatility (or variance)

As the payout of variance, volatility and gamma swaps are based on close-to-close prices, the standard close-to-close volatility (or variance) should be used for comparing their price against realised. Additionally, if a trader only hedges at the close (potentially for liquidity reasons) then again the standard close-to-close volatility measure should be used.

[45] Compared to integrated volatility.

CLOSE-TO-CLOSE

As the average is taken from the sample, close-to-close volatility has N-1 degrees of freedom

The simplest volatility measure is the standard close-to-close volatility. We note that the volatility should be the standard deviation multiplied by √N/(N-1) to take into account the fact we are sampling the population (or take standard deviation of the sample)[46]. We ignored this in the earlier definition as for reasonably large n it √N/(N-1) is roughly equal to one.

$$\text{Standard dev of x} = s_x = \sqrt{\frac{F}{N}}\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$\text{As } \sigma = \sqrt{\sigma^2} = \sqrt{E(s^2)} < E(\sqrt{s^2}) = E(s) \text{ by Jensens´s inequality}$$

$$\text{Volatility} = \sigma_x = s_x \times \sqrt{\frac{N}{N-1}}$$

$$\text{Volatility}_{\text{close to close}} = \sigma_{cc} = \sqrt{\frac{F}{N-1}}\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \sqrt{\frac{F}{N-1}}\sqrt{\sum_{i=1}^{N} Ln(\frac{c_i}{c_{i-1}})} \text{ assuming zero drift}$$

PARKINSON

The first advanced volatility estimator was created by Parkinson in 1980, and instead of using closing prices it uses the high and low price. One drawback of this estimator is that it assumes continuous trading, hence it underestimates the volatility as potential movements when the market is shut are ignored.

$$\text{Volatility}_{\text{Parkinson}} = \sigma_P = \sqrt{\frac{F}{N}}\sqrt{\frac{1}{4\,Ln(2)}\sum_{i=1}^{N}\left(Ln(\frac{h_i}{l_i})\right)^2}$$

GARMAN-KLASS

Later in 1980 the Garman-Klass volatility estimator was created. It is an extension of Parkinson which includes opening and closing prices (if opening prices are not available the close from the previous day can be used instead). As overnight jumps are ignored the measure underestimates the volatility.

$$\text{Volatility}_{\text{Garman-Klass}} = \sigma_{GK} = \sqrt{\frac{F}{N}}\sqrt{\sum_{i=1}^{N}\frac{1}{2}\left(Ln(\frac{h_i}{l_i})\right)^2 - (2Ln(2)-1)\left(Ln(\frac{c_i}{o_i})\right)^2}$$

[46] As the formula for standard deviation has N-1 degrees of freedom (as we subtract the sample average from each value of x)

ROGERS-SATCHELL

All of the previous advanced volatility measures assume the average return (or drift) is zero. Securities that have a drift, or non-zero mean, require a more sophisticated measure of volatility. The Rogers-Satchell volatility created in the early 1990s is able to properly measure the volatility for securities with non-zero mean. It does not, however, handle jumps; hence, it underestimates the volatility.

$$\text{Volatility}_{\text{Rogers-Satchell}} = \sigma_{RS} = \sqrt{\frac{\text{F}}{N}}\sqrt{\sum_{i=1}^{N} Ln(\frac{h_i}{c_i})Ln(\frac{h_i}{o_i}) + Ln(\frac{l_i}{c_i})Ln(\frac{l_i}{o_i})}$$

GARMAN-KLASS YANG-ZHANG EXTENSION

Yang-Zhang modified the Garman-Klass volatility measure in order to let it handle jumps. The measure does assume a zero drift; hence, it will overestimate the volatility if a security has a non-zero mean return. As the effect of drift is small, the fact continuous prices are not available usually means it underestimates the volatility (but by a smaller amount than the previous alternative measures).

$$\text{Volatility}_{GKYZ} = \sigma_{GKYZ} = \sqrt{\frac{\text{F}}{N}}\sqrt{\sum_{i=1}^{N}\left(Ln(\frac{o_i}{c_{i-1}})\right)^2 + \frac{1}{2}\left(Ln(\frac{h_i}{l_i})\right)^2 - (2Ln(2)-1)\left(Ln(\frac{c_i}{o_i})\right)^2}$$

YANG-ZHANG

Yang-Zhang is the sum of overnight volatility, and a weighted average of Rogers-Satchell and open-to-close volatility

In 2000 Yang-Zhang created a volatility measure that handles both opening jumps and drift. It is the sum of the overnight volatility (close-to-open volatility) and a weighted average of the Rogers-Satchell volatility and the open-to-close volatility. The assumption of continuous prices does mean the measure tends to slightly underestimate the volatility.

$$\text{Volatility}_{\text{Yang-Zhang}} = \sigma_{YZ} = \sqrt{\text{F}}\sqrt{\sigma_{\text{overnight volatility}}^2 + k\,\sigma_{\text{open to close volatility}}^2 + (1-k)\sigma_{RS}^2}$$

$$\text{where k} = \frac{0.34}{1.34 + \frac{N+1}{N-1}}$$

$$\sigma_{\text{overnight volatility}}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[Ln(\frac{o_i}{c_{i-1}}) - \overline{Ln(\frac{o_i}{c_{i-1}})}\right]^2$$

$$\sigma_{\text{open to close volatility}}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[Ln(\frac{c_i}{o_i}) - \overline{Ln(\frac{c_i}{o_i})}\right]^2$$

PROOF VARIANCE SWAPS CAN BE HEDGED BY LOG CONTRACT (=1/K^2)

A log contract is a portfolio of options of all strikes (K) weighted by 1/K^2. When this portfolio of options is delta hedged on the close, the payoff is identical to the payoff of a variance swap. We prove this relationship and hence show that the volatility of a variance swap can be hedged with a static position in a log contract.

PORTFOLIO OF OPTIONS WITH CONSTANT VEGA WEIGHTED 1/K^2

In order to prove that a portfolio of options with flat vega has to be weighted 1/K^2, we will define the variable x to be K/S (strike K divided by spot S). With this definition and assuming zero interest rates, the standard Black-Scholes formula for vega of an option simplifies to:

Vega of option = $\tau \times S \times f(x, v)$

where

x = K / S (strike a ratio of spot)

τ = time to maturity

$v = \sigma^2 \tau$ (total variance)

$$f(x, v) = \frac{1}{\sqrt{2\pi}} \times e^{\frac{-d_1^2}{2}}$$

$$d_1 = \frac{\mathrm{Ln}(\frac{1}{x}) + \frac{v}{2}}{\sqrt{v}}$$

If we have a portfolio of options where the weight of each option is w(K), then the vega of the portfolio of options V(S) is:

$$V(S) = \tau \int_{K=0}^{\infty} w(K) \times S \times f(x,v) dK$$

As K = xS this means dK / dx = S, hence dK = S × dx and we can change variable K for x.

$$V(S) = \tau \int_{x=0}^{\infty} w(xS) \times S^2 \times f(x,v) dx$$

In order for the portfolio of options to have a constant vega – no matter what the level of spot – dV(S)/dS has to be equal to zero.

$$\frac{dV}{dS} = \tau \int_{x=0}^{\infty} \frac{d}{dS}\left[S^2 w(xS)\right] \times f(x,v) dx = 0$$

And by the chain rule:

$$\Rightarrow \quad \tau \int_{x=0}^{\infty} \left[2Sw(xS) + S^2 \frac{d}{dS} w(xS) \right] \times f(x,v)dx = 0$$

$$\Rightarrow \quad \tau \int_{x=0}^{\infty} S \left[2w(xS) + S \frac{d}{dS} w(xS) \right] \times f(x,v)dx = 0$$

As d/dS = (d/dK) × (dK/dS), and dK/dS = x

$$\Rightarrow \quad \tau \int_{x=0}^{\infty} S \left[2w(xS) + xS \frac{d}{dK} w(xS) \right] \times f(x,v)dx = 0$$

As K = xS

$$\Rightarrow \quad \tau \int_{x=0}^{\infty} S \left[2w(xS) + K \frac{d}{dK} w(K) \right] \times f(x,v)dx = 0$$

$$\Rightarrow \quad 2w + K \frac{d}{dK} w(K) = 0 \text{ for all values of S}$$

$$\Rightarrow \quad w(K) = \frac{\text{constant}}{K^2}$$

MODELLING VOLATILITY SURFACES

There are a variety of constraints on the edges of a volatility surface, and this section details some of the most important constraints from both a practical and theoretical point of view. We examine the considerations for very short-dated options (a few days or weeks), options at the wings of a volatility surface and very long-dated options.

IMPLIED VOLATILITY IS LESS USEFUL FOR NEAR-DATED OPTIONS

Options that only have a few days or a few weeks to expiry have a very small premium. For these low-value options, a relatively small change in price will equate to a relatively large change in implied volatility. This means the implied volatility bid-offer arbitrage channel is wider, and hence less useful. The bid-offer spread is more stable in cash terms for options of different maturity, so shorter-dated options should be priced more by premium rather than implied volatility.

Need to price short-dated options with a premium after a large collapse in the market

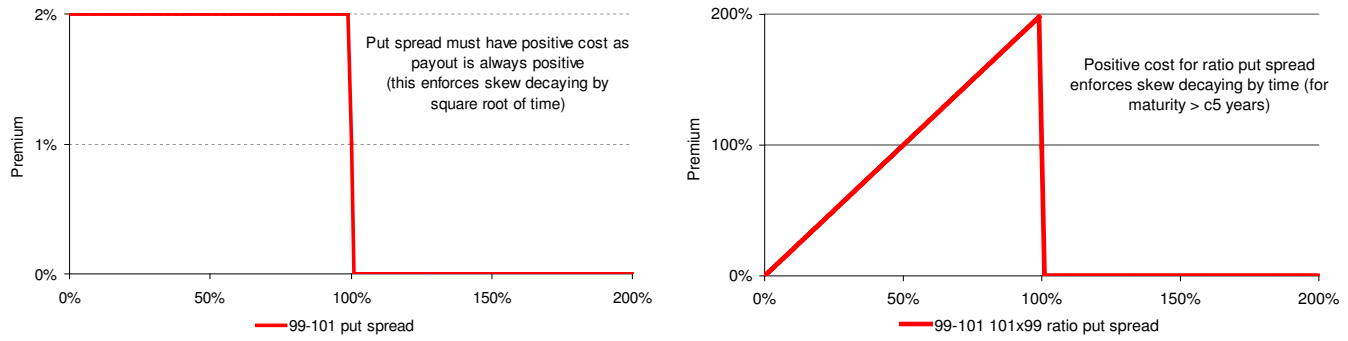
Near-dated options need to take premium, as well as implied volatility, into account

If there has been a recent dip in the market, there is a higher than average probability that the markets could bounce back to their earlier levels. The offer of short-dated ATM options should not be priced at a lower level than the size of the decline. For example, if markets have dropped 5%, then a one-week ATM call option should not be offered for less than c5% due to the risk of a bounce-back.

SKEW SHOULD DECAY BY SQUARE ROOT OF TIME

The payout of a put spread (and call spread) is always positive; hence, it should always have a positive cost. If it was possible to enter into a long put (or call) spread position for no cost (or potentially earning a small premium), any rational investor would go long as large a position as possible and earn risk-free profits (as the position cannot suffer a loss). A put spread will have a negative cost if the premium earned by selling the lower strike put is more than the premium of the higher strike put bought. This condition puts a cap on how negative skew can be: for high (negative) skew, the implied of the low strike put could be so large the premium is too high (ie, more than the premium of higher strike puts). The same logic applies for call spreads, except this puts a cap on positive skew (ie, floor on negative skew). As skew is normally negative, the condition on put spreads (see figure below on the left) is usually the most important. As time increases, it can be shown that the cap and floor for skew (defined as the gradient of first derivative of volatility with respect to strike, which is proportional to 90%-100% skew) decays by roughly the square root of time. This gives a mathematical basis for the 'square root of time rule' used by traders.

Figure 119. Put Spread **Ratio put spread**



Source: Santander Investment Bolsa estimates.

Far-dated skew should decay by time for long maturities (c5 years)

In theory skew should decay by time (not square root of time)

It is possible to arrive at a stronger limit to the decay of skew by considering leveraged ratio put spreads (see chart above on the right). For any two strikes A and B (assume A<B), then the payout of going long A× puts with strike B, and going short B× puts with strike A creates a ratio put spread whose value cannot be less than zero. This is because the maximum payouts of both the long and short legs (puts have maximum payout with spot at zero) is A×B. This can be seen in the figure above on the right (showing a 99-101 101x99 ratio put spread). Looking at such leveraged ratio put spreads enforces skew decaying by time, not by the square root of time. However, for reasonable values of skew this condition only applies for long maturities (c5 years).

PROOF SKEW IS CAPPED AND FLOORED BY SQUARE ROOT OF TIME

Enforcing positive values for put and call spreads is the same as the below two conditions:

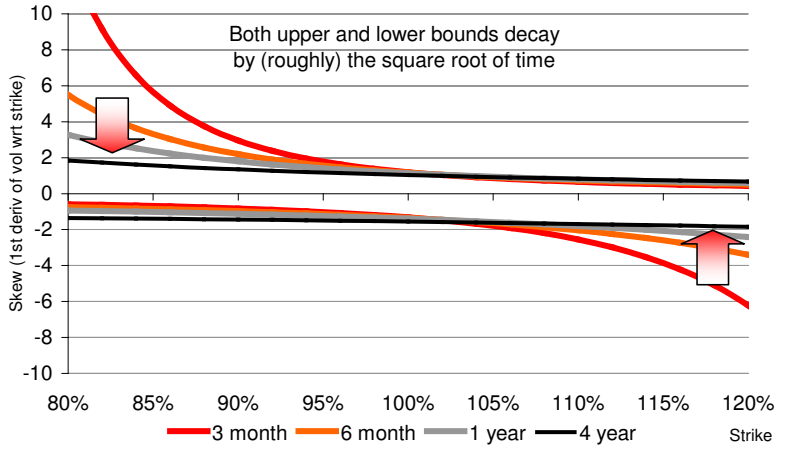
- **Change in price of a call when strike increases has to be negative** (intuitively makes sense, as you have to pay more to exercise the higher strike call).
- **Change in price of a put when strike increases has to be positive** (intuitively makes sense, as you receive more value if the put is exercised against you).

These conditions are the same as saying the gradient of x (=Strike/Forward) is bound by:

$$\text{Lower bound} = -\sqrt{2\pi}.e^{\frac{d_1^2}{2}}\left[1-N(d_2)\right] \le x \le \sqrt{2\pi}.e^{\frac{d_1^2}{2}}.N(d_2) = \text{upper bound}$$

It can be shown that these bounds decay by (roughly) the square root of time. This is plotted below.

Figure 120. Upper and Lower Bound for Skew (given 25% volatility)



Source: Santander Investment Bolsa.

Proof of theoretical cap for skew works in practice

In the above example, for a volatility of 25% the mathematical lower bound for one-year skew (gradient of volatility with respect to strike) is -1.39. This is the same as saying that the maximum difference between 99% and 100% strike implied is 1.39% (ie, 90%-100% or 95%-105% skew is capped at 13.9%). This theoretical result can be checked by pricing one-year put options with Black-Scholes.

- Price 100% put with 25% implied = 9.95%
- Price 99% put with 26.39% implied = 9.95% (difference of implied of 1.39%)

In practice, skew is likely to be bounded well before mathematical limits

While a 90%-100% one-year skew of 13.9% is very high for skew, we note buying cheap put spreads will appear to be attractive long before the price is negative. Hence, in practice, traders are likely to sell skew long before it hits the mathematical bounds for arbitrage (as a put spread's price tends to zero as skew approaches the mathematical bound). However, as the mathematical bound decays by the square root of time, so too should the 'market bound'.

OTM IMPLIEDS AT THE WINGS HAVE TO BE FLAT IN LOG SPACE

While it is popular to plot implieds vs delta, it can be shown for many models[47] that implied volatility must be linear in log strike (ie, Ln[K/F]) as log strike goes to infinity. Hence a parameterisation of a volatility surface should, in theory, be parameterised in terms of log strike, not delta. In practice, however, as the time value of options for a very high strike is very small, modelling implieds against delta can be used as the bid-offer should eliminate any potential arbitrage.

[47] Eg, stochastic volatility plus jump models.

BLACK-SCHOLES FORMULA

The most popular method of valuing options is the Black-Scholes-Merton model. We show the key formulas involved in this calculation. The assumptions behind the model are also discussed.

BLACK-SCHOLES MAKES A NUMBER OF ASSUMPTIONS

It is often joked that Black-Scholes is the wrong model with the wrong assumptions that gets the right price. The simplicity of the model has ensured that it is still used despite the competition from other, more complicated models. The assumptions are below:

- Constant (known) volatility
- Constant interest rates
- No dividends (a constant dividend yield can, however, be incorporated into the interest rate)
- Zero borrow cost, zero trading cost and zero taxes
- Constant trading
- Stock price return is log normally distributed
- Can trade infinitely divisible amounts of securities
- No arbitrage

BLACK-SCHOLES PRICE OF CALL AND PUT OPTIONS

Call option price = $S \times N(d_1) - Ke^{-rT} N(d_2)$

Put option price = $-S \times N(-d_1) + Ke^{-rT} N(d_2)$

where

$$d_1 = \frac{Ln\left(\frac{S}{K}\right) + (r + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T} = \frac{Ln\left(\frac{S}{K}\right) + (r - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}$$

S = Spot

K = Strike

r = risk free rate (– dividend yield)

σ = volatility

T = time (years)

GREEKS AND THEIR MEANING

Greeks is the name given to the (usually) Greek letters used to measure risk. We give the Black-Scholes formula for the key Greeks and describe which risk they measure.

VEGA IS NOT A GREEK LETTER

Although Vega is a Greek, it is not a Greek letter. It is instead the brightest star in the constellation Lyra. The main greeks and their definition are in the table below.

Figure 121. Greeks and their definition

Greek	Symbol	Measures	Definition
Delta	δ or Δ	Equity exposure	Change in option price due to spot
Gamma	γ or Γ	Convexity of payout	Change in delta due to spot
Theta	θ or Θ	Time decay	Change in option price due to time passing
Vega	ν	Volatility exposure	Change in option price due to volatility
Rho	ω or Ω	Interest rate exposure	Change in option price due to interest rates
Volga	λ or Λ	Vol of vol exposure	Change in vega due to volatility
Vanna	ψ or Ψ	Skew	Change in vega due to spot OR change in delta due to volatility

Source: Santander Investment Bolsa estimates.

The variables for the below formulae are identical to the earlier definitions in the previous section *Black-Scholes Formula*. In addition:

N'(z) is the normal density function, $\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$

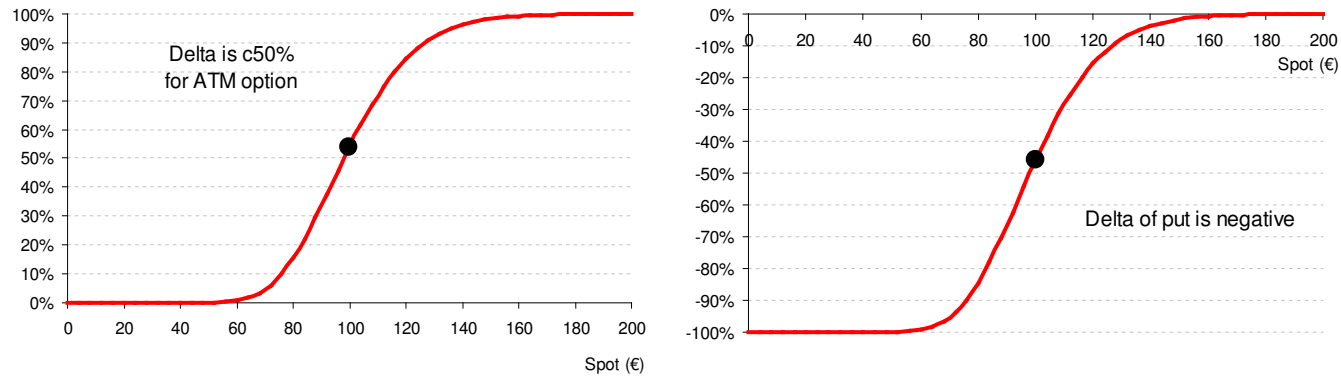
N(z) is the cumulative normal distribution, ie, N(0) = 0.5.

DELTA MEASURES EQUITY EXPOSURE

Call delta:
$N(d_1)$
Put delta:
$-N(-d_1)$
$= N(d_1)-1$

The most commonly examined Greek is delta, as it gives the equity sensitivity of the option (change of option price due to change in underlying price). Delta is normally quoted in percent. For calls it lies between 0% (no equity sensitivity) and 100% (trades like a stock). The delta of puts lies between -100% (trades like short stock) and 0%. If a call option has a delta of 50% and the underlying rises €1, the call option increases in value €0.50 (= €1 * 50%). Note the values of the call and put delta in the formula below give the equity sensitivity of a forward of the same maturity as the option expiry. The equity sensitivity to spot is slightly different. Please note that there is a (small) difference between the probability that an option expires ITM and delta.

Figure 122. Delta (for Call) **Delta (for Put)**



Source: Santander Investment Bolsa estimates.

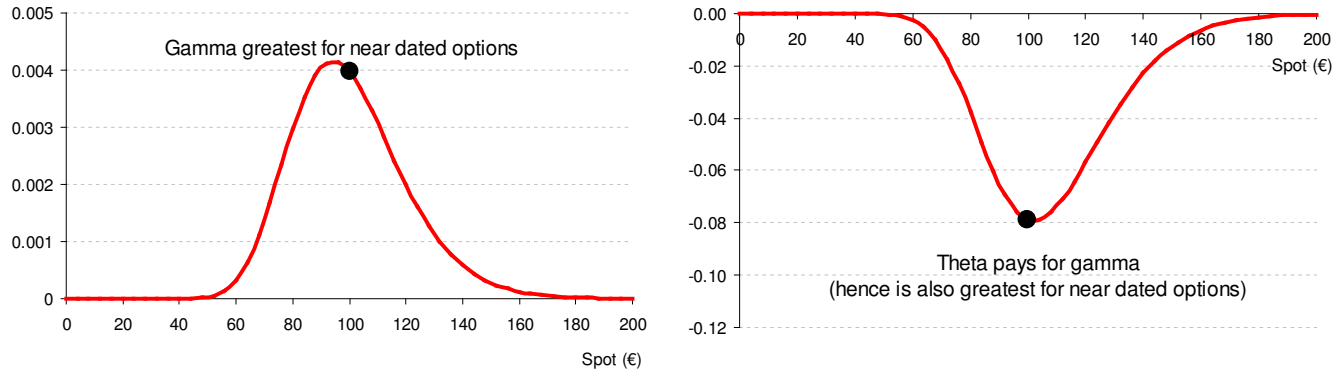
GAMMA MEASURES CONVEXITY (AMOUNT EARNED DELTA HEDGING)

Gamma:

$$-\frac{N'(d_1)}{S\sqrt{T}\sigma}$$

Gamma measures the change in delta due to the change in underlying price. The higher the gamma, the more convex is the theoretical payout. Gamma should not be considered a measure of value (low or high gamma does not mean the option is expensive or cheap); implied volatility is the measure of an option's value. Options are most convex, and hence have the highest gamma, when they are ATM and also about to expire. This can be seen intuitively as the delta of an option on the day of expiry will change from c0% if spot is just below expiry to c100% if spot is just above expiry (a small change in spot causes a large change in delta; hence, the gamma is very high).

Figure 123. Gamma **Theta**



Source: Santander Investment Bolsa estimates.

Call theta:

$$-\frac{S\sigma \times N'(d_1)}{2\sqrt{T}}$$

$$-rKe^{-rT}N(d_2)$$

Put theta:

$$-\frac{S\sigma \times N'(d_1)}{2\sqrt{T}}$$

$$+rKe^{-rT}N(-d_2)$$

THETA MEASURES TIME DECAY (COST OF BEING LONG GAMMA)

Theta is the change in the price of an option for a change in time to maturity; hence, it measures time decay. In order to find the daily impact of the passage of time, its value is normally divided by 252 (trading days in the year). If the second term in the formula below is ignored, the theta for calls and puts are identical and proportional to gamma. Theta can therefore be considered the cost of being long gamma.

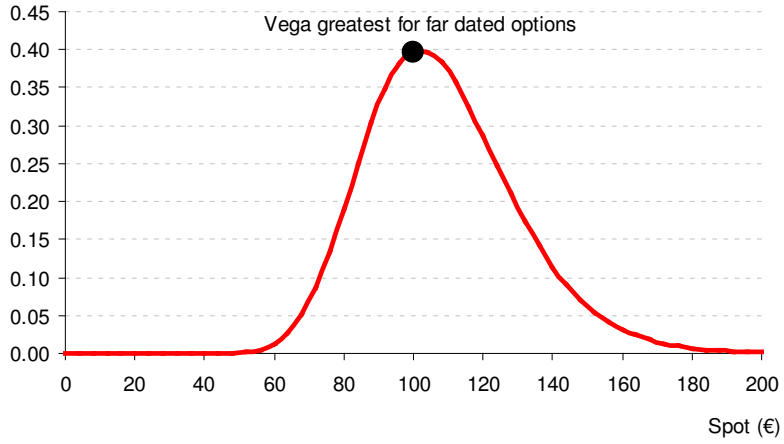
VEGA MEASURES VOLATILITY EXPOSURE (AVERAGE OF GAMMAS)

Vega:

$S\sqrt{T} \times N'(d_1)$

Vega gives the sensitivity to volatility of the option price. Vega is normally divided by 100 to give the price change for a 1 volatility point (ie, 1%) move in implied volatility. Vega can be considered to be the average gamma (or non-linearity) over the life of the option. As vega has a √T in the formula power vega (vega divided by square root of time) is often used as a risk measure (to compensate for the fact that near dated implieds move more than far-dated implieds).

Figure 124. Vega



Source: Santander Investment Bolsa.

RHO MEASURES INTEREST RATE RISK (RELATIVELY SMALL)

Call rho:

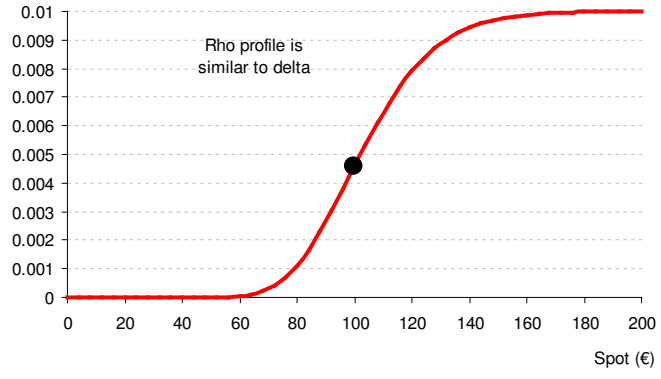
$KTe^{-rT}N(d_2)$

Put rho:

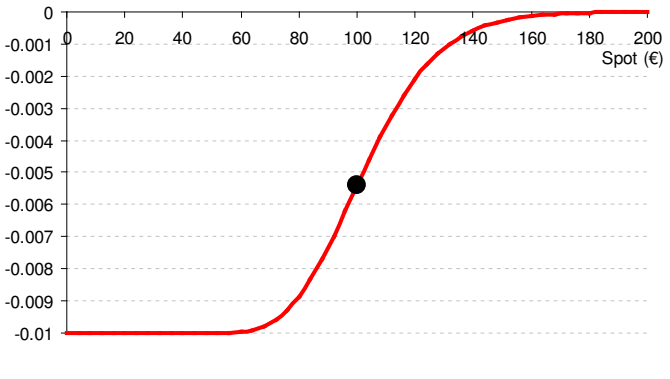
$-KTe^{-rT}N(-d_2)$

Rho measures the change in the value of the option due to a move in the risk-free rate. The profile of rho vs spot is similar to delta, as the risk-free rate is more important for more equity-sensitive options (as these are the options where there is the most benefit in selling stock and replacing it with an option and putting the difference in value on deposit). Rho is normally divided by 10,000 to give the change in price for a 1bp move.

Figure 125. Rho (for call)



Rho (for put)



Source: Santander Investment Bolsa estimates.

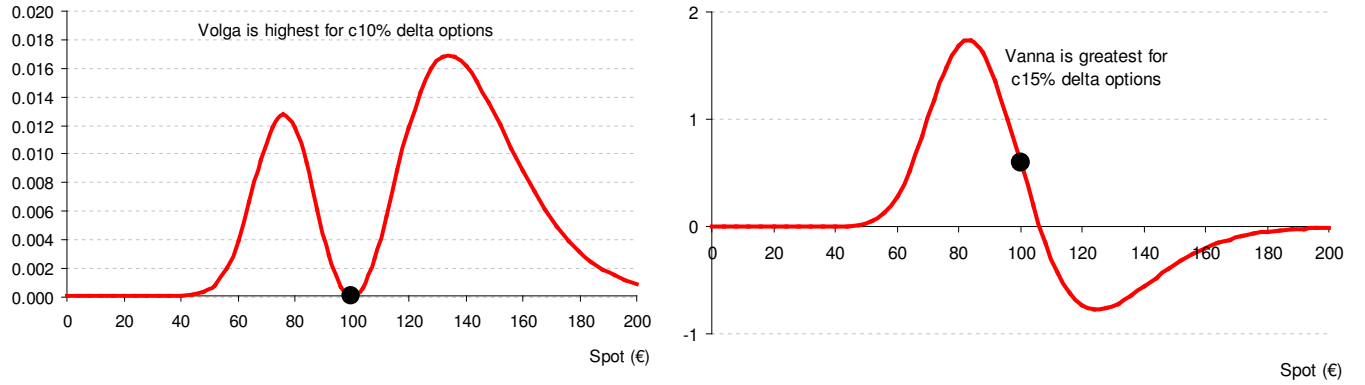
VOLGA MEASURES VOLATILITY OF VOLATILITY EXPOSURE

Volga:

$$\frac{S\sqrt{T}d_1d_2N'(d_1)}{\sigma}$$

Volga is short for VOLatility GAmma, and is the rate of change of vega due to a change in volatility. Volga (or Vomma/vega convexity) is highest for OTM options (approximately 10% delta), as these are the options where the probability of moving from OTM to ITM has the greatest effect on its value. For more detail on Volga, see the section *How to Measure Skew and Smile*.

Figure 126. Volga **Vanna**



Source: Santander Investment Bolsa estimates.

VANNA MEASURES SKEW EXPOSURE

Vanna:

$$\frac{-d_2N'(d_1)}{\sigma}$$

Vanna has two definitions as it measures the change in vega given a change in spot and the change in delta due to a change in volatility. The change in vega for a change in spot can be considered to measure the skew position, as this will lead to profits on a long skew trade if there is an increase in volatility as spot declines. The extreme values for vanna occur for c15 delta options, similar to volga's c10 delta peaks. For more detail on vanna, see the section *How to Measure Skew and Smile*.

ADVANCED (PRACTICAL OR SHADOW) GREEKS

How a volatility surface changes over time can impact the profitability of a position. While the most important aspects have already been covered (and are relatively well understood by the market) there are 'second order' Greeks that are less well understood. Two of the most important are the impact of the passage of time on skew (volatility slide theta), and the impact of a movement in spot on OTM options (anchor delta). These Greeks are not mathematical Greeks, but are practical or 'shadow' Greeks.

INCREASE IN SKEW AS TIME PASSES CAUSES 'VOL SLIDE THETA'

As an option approaches expiry, its maturity decreases. As near-dated skew is larger than far-dated skew, the skew of a fixed maturity option will increase as time passes. This can be seen by assuming that skew by maturity (eg, three-month or one-year) is constant (ie, relative time, the maturity equivalent of sticky moneyness or sticky delta). We also assume that three-month skew is larger than the value of one year skew. If we buy a low strike one year option (ie, we are long skew) then, assuming spot and ATM volatility stay constant, when the option becomes a three-month option its implied will have risen (as three-month skew is larger than one-year skew and ATM volatility has not changed). We define 'volatility slide theta' as the change in price of an option due to skew increasing with the passage of time[48].

VOLATILITY SLIDE THETA IS MOST IMPORTANT FOR NEAR EXPIRIES

Volatility slide theta measures the increase of skew as expiry approaches

Given that skew increases as maturity decreases, this change in skew will increase the value of long skew positions (as in the example) and decrease the value of short skew positions. The effect of 'volatility slide theta' is negligible for medium- to far-dated maturities, but increases in importance as options approach expiry. If a volatility surface model does not take into account 'volatility slide theta', then its impact will be seen when a trader re-marks the volatility surface.

VOL SLIDE THETA MEASURES IMPACT OF CONSTANT SMILE RULE

The constant smile rule (CSR) details how forward starting options should be priced. The impact of this rule on valuations is given by the 'volatility slide theta' as they both assume a fixed maturity smile is constant. The impact of this assumption is more important for forward starting options than for vanilla options.

[48] While we concentrate on Black-Scholes implied volatilities, volatility slide theta also affects local volatility surfaces.

WHEN TRADERS CHANGE THEIR 'ANCHOR' THIS INTRODUCES A SECOND ORDER DELTA ('ANCHOR DELTA')

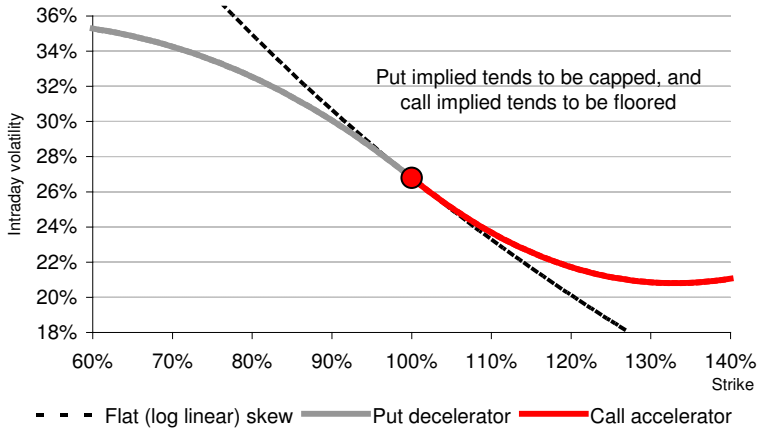
Volatility surfaces are normally modelled via a parameterisation. One of the more popular parameterisations is to set the ATMf volatility from a certain level of spot, or 'anchor', and then define the skew (slope). While this builds a reasonable volatility surface for near ATM options, the wings will normally need to be slightly adjusted. Normally a fixed skew for both downside puts and upside calls will cause upside calls to be too cheap (as volatility will be floored) and downside puts to be too expensive (as volatility should be capped at some level, even for very low strikes). As the 'anchor' is raised, the implied volatility of OTM options declines (assuming the wing parameters for the volatility surface stay the same). We call this effect 'anchor delta'.

Implied volatility has to be floored, and capped, for values to be realistic

When traders change their 'anchor' this introduces a second order 'anchor delta'

There are many different ways a volatility surface parameterisation can let traders correct the wings, but the effect is usually similar. We shall simply assume that the very OTM call implied volatility is lifted by a call accelerator, and very OTM put implied volatility is lowered by a put decelerator. This is necessary to prevent call implieds going too low (ie, below minimum realised volatility), or put implieds going too high (ie, above maximum realised volatility). The effect of these wing parameters is shown in Figure 127 below.

Figure 127. Skew with Put Decelerators and Call Accelerators

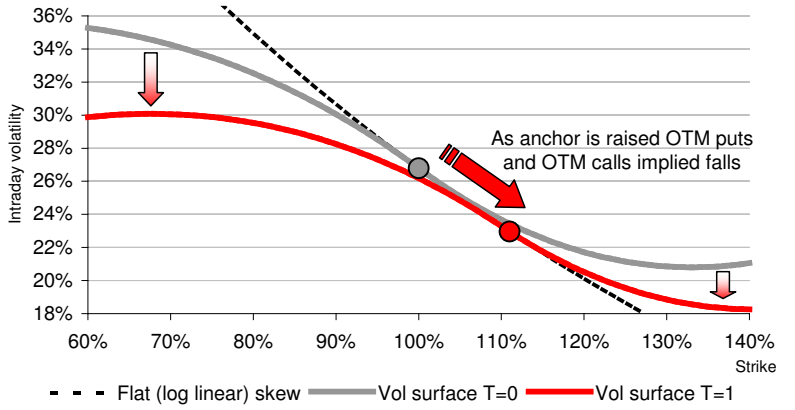


Source: Santander Investment Bolsa.

Traders tend to refresh a surface by only changing the key parameters

For liquid underlyings such as indices, a volatility surface is likely to be updated several times a day (especially if markets are moving significantly). Usually only the key parameters will be changed, and the less key parameters such as the wing parameters are changed less frequently. We shall assume that there will be many occasions where there is a movement in spot along the skew (ie, static strike for near ATM strikes). In these cases a trader is likely to change the anchor (and volatility at the anchor, which has moved along the skew), but leave the remaining skew and wing parameters (which are defined relative to the anchor) unchanged. In order to have the same implied volatility for OTM options after changing the anchor, the call accelerator should be increased and the put decelerator decreased. In practice this does not always happen, as wing parameters are typically changed less frequently. The effect of an increase in anchor along the (static strike) skew while leaving the wing parameters unchanged is shown below.

Figure 128. Moving Anchor 10% Higher Along the Skew



Source: Santander Investment Bolsa.

OTM options have a second order 'anchor delta'

To simplify the example we shall assume the call wing parameter increases the implied volatility for strikes 110% and more, and the put wing parameter decreases the implied volatility for strikes 90% or lower. If spot rises 10%, the 120% call implied volatility will suffer when the anchor is re-marked 10% higher, because the call implied volatility is initially lifted by the call wing parameter (which no longer has an effect). OTM calls therefore have a negative 'anchor delta' as they lose value as anchor rises. Similarly, as anchor rises the effect of the put wing will increase, lowering the implied volatility of puts of strike 90% or less as anchor rises. So, under this scenario all options that are OTM have a negative 'anchor delta' that needs to be hedged.

SHORTING STOCK BY BORROWING SHARES

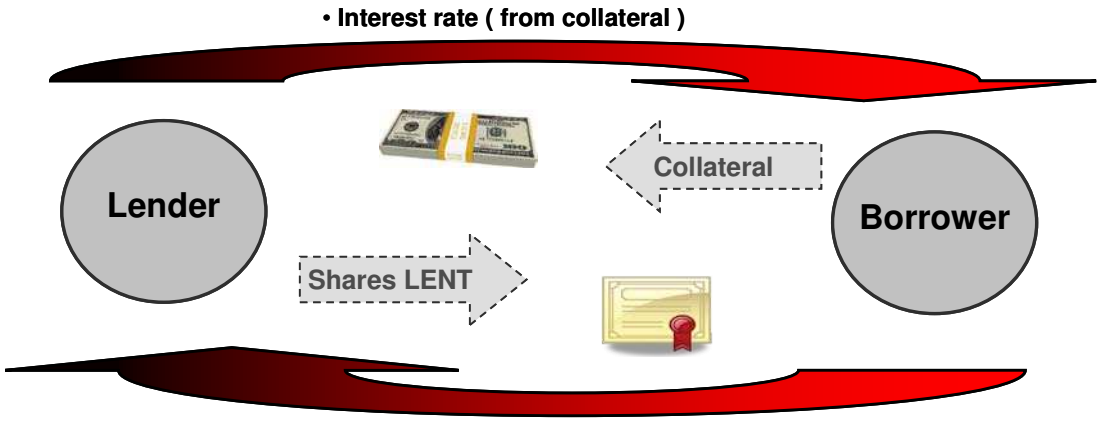
The hedging of equity derivatives assumes you can short shares by borrowing them. We show the processes involved in this operation. The disadvantages, and benefits, to an investor who lends out shares are also explained.

THERE IS NO COUNTERPARTY RISK WHEN YOU LEND SHARES

Voting rights lost when shares are lent

To short shares initially, the shares must first be borrowed. In order to remove counterparty risk, when an investor lends out shares he/she receives collateral (cash, stock, bonds, etc) for the same value. Both sides retain the beneficial ownership of both the lent security and the collateral, so any dividends, coupons, rights issues are passed between the two parties. If cash is used as collateral, the interest on the cash is returned. Should a decision have to be made, ie, to receive a dividend in cash or stock, the decision is made by the original owner of the security. The only exception is that the voting rights are lost, which is why lent securities are often called back before votes. To ensure there is no counterparty risk during the time the security is lent out, the collateral and lent security is marked to market and the difference settled for cash (while a wide range of securities can be used as initial collateral, only cash can be used for the change in value of the lent security).

Figure 129. Borrowing Shares



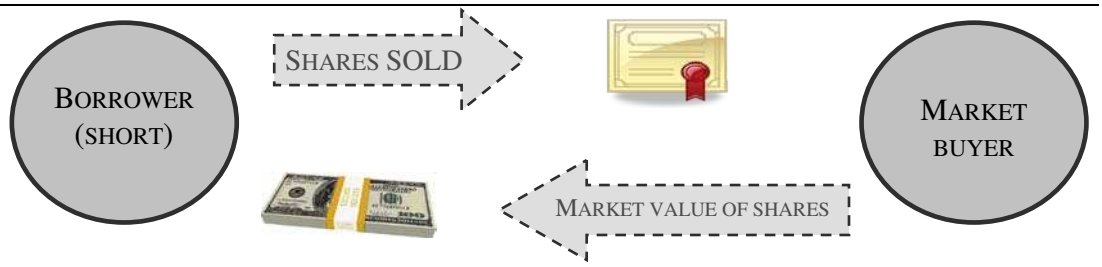
LENDER:	BORROWER:
Retains	**Retains**
• Full exposure to movement share price	• Benefits of collateral
•All non-voting benefit of shares (dividends, rights)	**Gains**
Gains	• Voting rights
• Borrow costs	• Right to sell shares (go short)
Loses	**Loses**
• Voting rights of shares	• Borrow cost

Source: Santander Investment Bolsa.

SELLING THE STOCK YOU HAVE BORROWED GIVES A SHORT POSITION

Once an investor has borrowed shares, these shares can be sold in the market. The proceeds from this sale can be used as the collateral given to the lender. Selling borrowed shares gives a short position, as profits are earned if the stock falls (as it can be bought back at a lower price than it was sold for, and then returned to the original owner).

Figure 130. Borrowing Shares – can you resize so double line is closer



BORROWER:

Retains

- Benefits of collateral
- Responsibility to return shares to lender

Gains

- Profits from share price declines
- Cash value of shares (can be used as collateral)

Loses

- Borrow cost

Source: Santander Investment Bolsa.

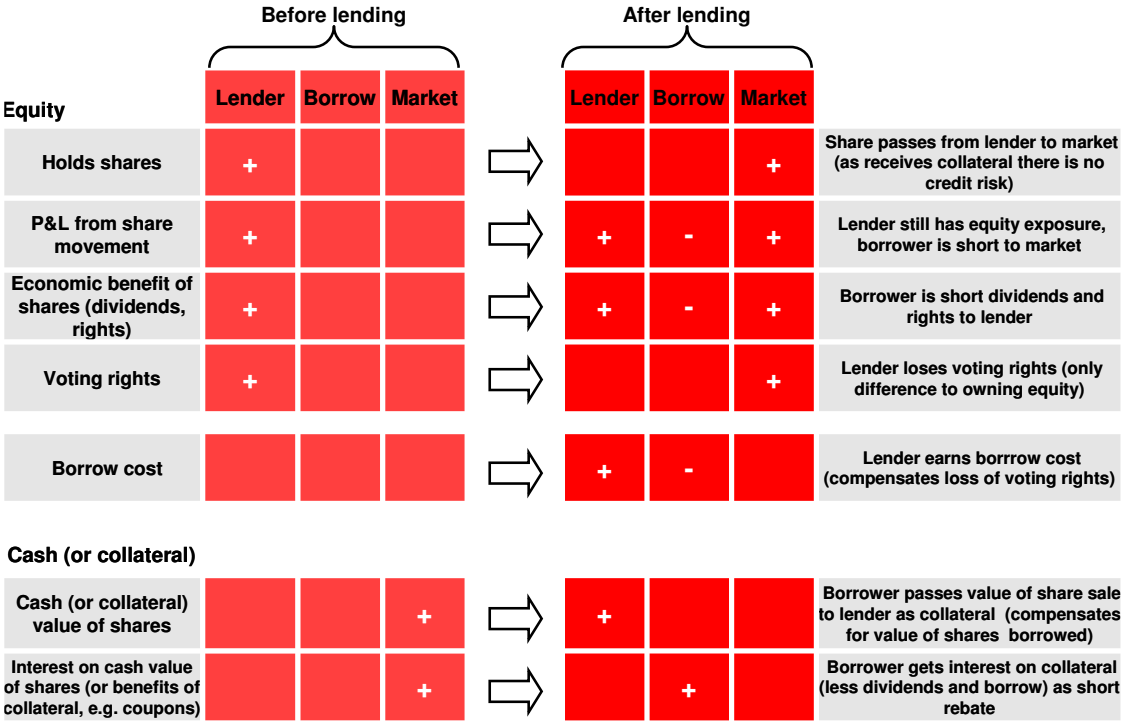
EARN SHORT REBATE WHEN YOU SHORT

Short seller receives interest rate on collateral (less dividends and borrow) as short rebate

The investor who has shorted the shares receives interest on the collateral, but has to pass dividends and borrow cost to the original owner. The net of these cash flows is called the short rebate, as it is the profit (or loss for high dividend paying stocks) that occurs if there is no change in the price of the shorted security. Shorting shares is therefore still profitable if shares rise by less than the short rebate.

Short rebate = interest rate (normally central bank risk free rate) – dividends – borrow cost

Figure 131. Initial and Final Position of Lender, Borrower and Market Following Shorting of Shares



Source: Santander Investment Bolsa.

SORTINO RATIO

If an underlying is distributed normally, standard deviation is the perfect measure of risk. For returns with a skewed distribution, such as with option trading or call overwriting, there is no one perfect measure of risk; hence, several measures of risk should be used. The Sortino is one of the most popular measures of skewed risk, as it only takes into account downside risk.

SORTINO RATIO IS MODIFICATION OF SHARPE RATIO

Sortino ratio is Sharpe ratio where excess return is divided by downside risk (not total risk)

The Sharpe ratio measures the excess return, or amount of return (R) that is greater than the target rate of return (T). Often zero or risk-free rate is used as the target return. To take volatility of returns into account, this excess return is divided by the standard deviation. However, this takes into account both upside and downside risks. As investors are typically more focused on downside risks, the Sortino ratio modifies the Sharpe ratio calculation to only divide by the downside risk (DR). The downside risk is the square root of the target semivariance, which can be thought of as the amount of standard deviation due to returns less than the target return. The Sortino ratio therefore only penalises large downside moves, and is often thought of as a better measure of risk for skewed returns.

$$Sortino = \frac{R - T}{DR}$$

where

$$DR = \left[\int_{\infty}^{T} (T - x) f(x) dx \right]^{0.5}$$

R = Return

T = Target return

CAPITAL STRUCTURE ARBITRAGE

When Credit Default Swaps were created in the late 1990's, they traded independently of the equity derivative market. The high levels of volatility and credit spreads during the bursting of the TMT bubble demonstrated the link between credit spreads, equity, and implied volatility. We examine four models that demonstrate this link (Merton model, jump diffusion, put vs CDS, and implied no-default volatility).

NORMALLY TRADE CREDIT VS EQUITY, NOT VOLATILITY

Capital structure arbitrage models can link the price of equity, credit and implied volatility. However, the relatively wide bid-offer spreads of equity derivatives mean trades are normally carried out between credit and equity (or between different subordinations of credit and preferred shares vs ordinary shares). The typical trade is for an investor to go long the security that is highest in the capital structure, for example, a corporate bond (or potentially a convertible bond), and short a security that is lower in the capital structure, for example, equity. Reverse trades are possible, for example, owning a subordinated higher yielding bond and shorting a senior lower yielding bond (and earning the positive carry as long as bankruptcy does not occur). Only for very wide credit spreads and high implied volatility is there a sufficiently attractive opportunity to carry about an arbitrage between credit and implied volatility. We shall concentrate on trading credit vs equity, as this is the most common type of trade.

Credit spread is only partly due to default risk

The OAS (Option Adjusted Spread) of a bond over the risk-free rate can be divided into three categories. There is the expected loss from default; however, there is also a portion due to general market risk premium and additionally a liquidity cost. Tax effects can also have an effect on the corporate bond market. Unless a capital structure arbitrage model takes into account the fact that not all of a bond's credit spread is due to the risk of default, the model is likely to fail. The fact that credit spreads are higher than they should be if bankruptcy risk was the sole risk of a bond was often a reason why long credit short equity trades have historically been more popular than the reverse (in addition to the preference to being long the security that is highest in the capital structure in order to reduce losses in bankruptcy).

CDS usually better than bonds for credit leg, as they are unfunded and easier to short

Using CDS is likely to be superior to corporate bonds for funding and technical reasons

Using CDS rather than corporate bonds can reduce many of the discrepancies in spread that a corporate bond suffers and narrow the difference between the estimated credit spread and the actual credit spread. We note that CDS are an unfunded trade (ie, leveraged), whereas corporate bonds are a funded trade (have to fund the purchase of the bond) that has many advantages when there is a funding squeeze (as occurred during the credit crunch). CDS also allow a short position to be easily taken, as borrow for corporate bonds is not always available, is usually expensive and can be recalled at any time. While borrow for bonds was c50bp before the credit crunch it soared to c5% following the crisis.

Credit derivatives do not have established rules for equity events

While credit derivatives have significant language against credit events, they have no language for equity events, such as special dividends or rights issues. Even for events such as takeovers and mergers, where there might be relevant documentation, credit derivatives are likely to behave differently than equity (and equity derivatives).

CREDIT MARKET CAN LEAD EQUITY MARKET AND VICE VERSA

While the bond market lags the CDS market, on average there is no difference between CDS and equity

We note that there are occasions when corporate bond prices lag a movement in equity prices, simply as traders have not always updated levels (but this price would be updated should an investor request a firm price). CDS prices suffer less from this effect, and we note for many large companies the corporate bond market is driven by the CDS market and not vice versa (the tail wags the dog). Although intuitively the equity market should be more likely to lead the CDS market than the reverse (due to high frequency traders and the greater number of market participants), when the CDS market is compared to the equity market on average neither consistently leads the other. Even if the CDS and equity on average react equally as quickly to new news, there are still occasions when credit leads equity and vice versa. Capital structure arbitrage could therefore be used on those occasions when one market has a delayed reaction to new news compared to the other.

GREATEST OPPORTUNITY ON BBB OR BB RATED COMPANIES

In order for capital structure arbitrage to work, there needs to be a strong correlation between credit and equity. This is normally found in companies that are rated BBB or BB. The credit spread for companies with ratings of A or above is normally more correlated to the general credit supply and interest rates than the equity price. For very speculative companies (rated B or below), the performance of their debt and equity is normally very name-specific, and often determined by the probability of takeover or default.

Capital structure arbitrage works best when companies don't default

Capital structure arbitrage is a bet on the convergence of equity and credit markets. It has the best result when a company in financial distress recovers, and the different securities it has issued converge. Should the company enter bankruptcy, the returns are less impressive. The risk to the trade is that the company becomes more distressed, and as the likelihood of bankruptcy increases the equity and credit markets cease to function properly. This could result in a further divergence or perhaps closure of one of the markets, potentially forcing a liquidation of the convergence strategy.

FUNDAMENTAL FACTORS CAN DWARF STATISTICAL RELATIONSHIPS

Capital structure arbitrage assumes equity and credit markets move in parallel. However, there are many events that are bullish for one class of investors and bearish for another. This normally happens when the leverage of a company changes suddenly. Takeovers and rights issues are the two main events that can quickly change leverage. Special dividends, share buybacks and a general reduction of leverage normally have a smaller, more gradual effect.

Rights issue. A rights issue will always reduce leverage, and is effectively a transfer of value from equity holders to debt holders (as the company is less risky, and earnings are now divided amongst a larger number of shares).

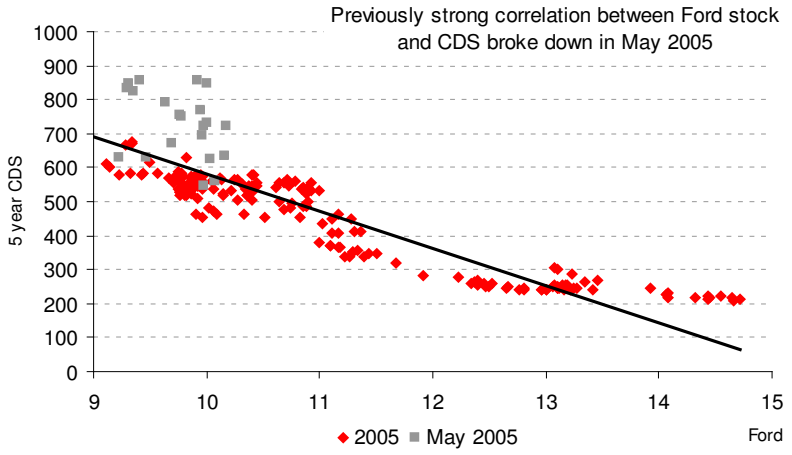
Takeover bid (which increases leverage). When a company is taken over, unless the acquisition is solely for equity, a portion of the acquisition will have to be financed with cash or debt (particularly during a leveraged buyout). In this case, the leverage of the acquiring company will increase, causing an increase in credit spreads and a reduction in the value of debt. Conversely, the equity price of the acquiring company is more stable. For the acquired company, the equity price should jump close to the level of the bid and, depending on the structure of the offer, the debt could fall (we note that if the acquired company is already in distress the value of debt can rise; for example, when Household was acquired by HSBC).

GM EQUITY SOARED A DAY BEFORE CREDIT SANK, CAUSING LOSSES

GM was downgraded by S&P a day after Kirk Kerkorian bought a 5% stake

On May 4, 2005, Kirk Kerkorian announced the intention to increase his (previously unknown) stake in GM, causing the troubled company's share price to soar 18% intraday (7.3% close to close). The following day, S&P downgraded GM and Ford to 'junk', causing a collapse in the credit market and a 122bp CDS rise in two days. As many capital structure arbitrage investors had a long credit short equity position, both legs were loss making and large losses were suffered.

Figure 132. Equity vs Credit Spread (5-Year CDS) for Ford

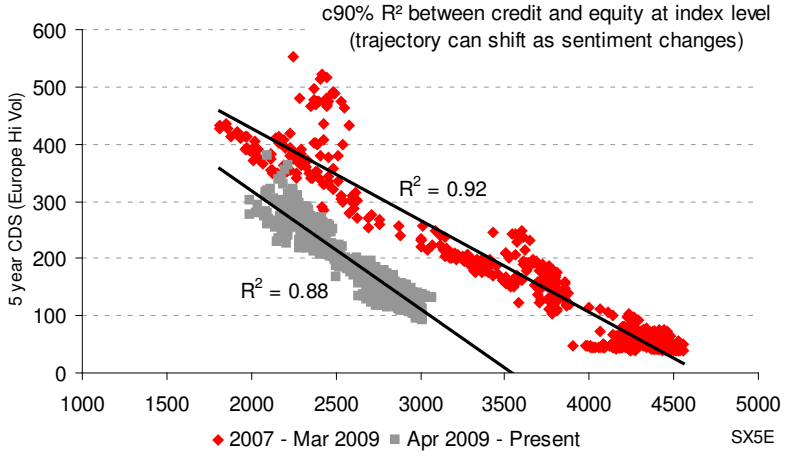


Source: Santander Investment Bolsa.

CORRELATION BETWEEN CREDIT AND EQUITY LOW AT STOCK LEVEL

For many companies the correlation between equity and credit is not particularly strong, with a typical correlation between 5% and 15%. Hence it is necessary for a capital structure arbitrage investor to have many different trades on simultaneously. The correlation of a portfolio of bonds and equities is far higher (c90%).

Figure 133. SX5E vs 5-Year CDS (European HiVol)



Source: Santander Investment Bolsa.

MODELLING THE LINK BETWEEN CREDIT SPREADS AND IMPLIED VOL

While there are many models that show the link between the equity, equity volatility and debt of a company, we shall restrict ourselves to four of the most popular.

- **Merton model.** The Merton model uses the same model as Black-Scholes, but applies it to a firm. If a firm is assumed to have only one maturity of debt, then the equity of the company can be considered to be a European call option on the value of the enterprise (value of enterprise = value of debt + value of equity) whose strike is the face value of debt. This model shows how the volatility of equity rises as leverage rises. The Merton model also shows that an increase in volatility of the enterprise increases the value of equity (as equity is effectively long a call on the value of the enterprise), and decreases the value of the debt (as debt is effectively short a put on the enterprise, as they suffer the downside should the firm enter bankruptcy but the upside is capped).

- **Jump diffusion.** A jump diffusion model assumes there are two parts to the volatility of a stock. There is the diffusive (no-default) volatility, which is the volatility of the equity without any bankruptcy risk, and a separate volatility due to the risk of a jump to bankruptcy. The total volatility is the sum of these two parts. While the diffusive volatility is constant, the effect on volatility due to the jump to bankruptcy is greater for options of low strike than high strike causing 'credit induced skew'. This means that as the credit spread of a company rises, this increases the likelihood of a jump to bankruptcy and increases the skew. A jump diffusion model therefore shows a link between credit spread and implied volatility.

- **Put vs CDS.** As the share price of a company in default tends to trade close to zero, a put can be assumed to pay out its strike in the even of default. This payout can be compared to the values of a company's CDS, or its debt market (as the probability of a default can be estimated from both). As a put can also have a positive value even if a company does not default, the value of a CDS gives a floor to the value of puts. As 1xN put spreads can be constructed to never have a negative payout, various caps to the value of puts can be calculated by ensuring they have a cost. The combination of the CDS price floor, and put price cap, gives a channel for implieds to trade without any arbitrage between CDS and put options.

- **No-default implied volatility.** Using the above put vs CDS methodology, the value of a put price due to the payout in default can be estimated. If this value is taken away from the put price, the remaining price can be used to calculate a no-default implied volatility (or implied diffusive volatility). The skew and term structure of implied no-default implied volatilities are flatter than Black-Scholes implied volatility, which allows an easier comparison and potential for identifying opportunities.

(1) MERTON MODEL

Value of enterprise is equal to sum of equity and debt

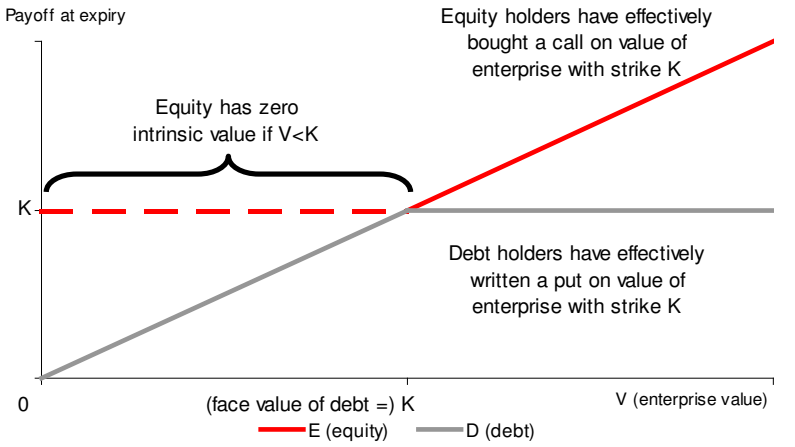
The Merton model assumes that a company has an enterprise value (V) whose debt (D) consists of only one zero coupon bond whose value at maturity is K. These assumptions are made in order to avoid the possibility of a default before maturity (which would be possible if there was more than one maturity of debt, or a coupon had to be paid. The company has one class of equity (E) that does not pay a dividend. The value of equity (E) and debt (D) at maturity is given below.

Enterprise value = V = E + D

Equity = Max(V – K, 0) = call on V with strike K

Debt = Min(V, K) = K – Max(K – V, 0) = Face value of debt K – put on V with strike K

Figure 134. Graph of Value of Enterprise, Equity and Debt



Source: Santander Investment Bolsa.

Enterprise value of a firm at maturity has to be at least K or it will enter bankruptcy

Before the maturity of the debt, the enterprise has obligations to both the equity and debt holders. At the maturity of the debt, if the value of the enterprise is equal to or above K, the enterprise will pay off the debt K and the remaining value of the firm is solely owned by the equity holders. If the value of the enterprise is below K then the firm enters bankruptcy. In the event of bankruptcy, the equity holders get nothing and the debt holders get the whole value of the enterprise V (which is less than K).

Equity is long a call on the value of a firm

Equity holders are long a call, put holders are short a put

If the value of the enterprise V is below the face value of debt K at maturity the equity holders receive nothing. However, if V is greater than K, the equity holders receive V - K. The equity holders therefore receive a payout equal to a call option on V of strike K.

Debt is short a put on value of firm

The maximum payout for owners of debt is the face value of debt. This maximum payout is reduced by the amount the value of the enterprise is below the face value of debt at maturity. Debt is therefore equal to the face value of debt less the value of a put on V of strike K.

DEBT HAS A DELTA THAT CAN BE USED TO ARBITRAGE VS EQUITY

As the value of the short put has a delta, debt has a delta. It is therefore possible to go long debt and short equity (at the calculated delta using the Merton model) as part of a capital structure arbitrage trade.

If enterprise value is unchanged, then if value of equity rises, value of credit falls

Equity holders are long vol, debt holders are short vol

As enterprise value is equal to the sum of equity and debt, if enterprise value is kept constant then for equity to rise the value of debt must fall. An example would be if a company attempts to move into higher-risk activity, lifting its volatility. As equity holders are long a call on the value of the company they benefit from the additional time value. However, as debt holders are short a put they suffer should a firm move into higher-risk activities.

Merton model assumes too high a recovery rate

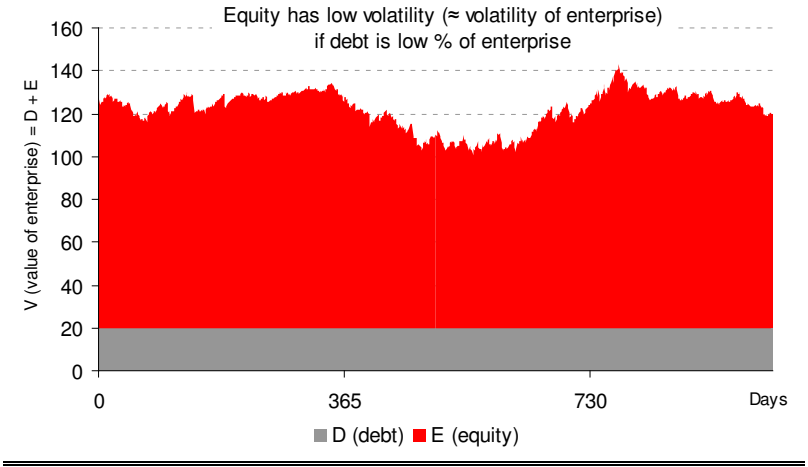
Using the vanilla Merton model gives unrealistic results with credit spreads that are too tight. This is because the recovery rate (of V/K) is too high. However, using more advanced models (eg, stochastic barrier to take into account the default point is unknown), the model can be calibrated to market data.

MERTON MODEL EXPLAINS EQUITY SKEW

The volatility of an enterprise should be based on the markets in which it operates, interest rates and other macro risks. It should, however, be independent of how it is funded. The proportion of debt to equity therefore should not change the volatility of the enterprise V; however, it does change the volatility of the equity E. It can be shown that the volatility of equity is approximately equal to the volatility of the enterprise multiplied by the leverage (V/E). Should the value of equity fall, the leverage will rise, lifting the implied volatility. This explains skew: the fact that options of lower strike have an implied volatility greater than options of high strike.

$\sigma_E \approx \sigma_V \times V / E \ (= \sigma_V \times \text{leverage})$

Figure 135. Value of Enterprise and Equity with Low Debt



Source: Santander Investment Bolsa.

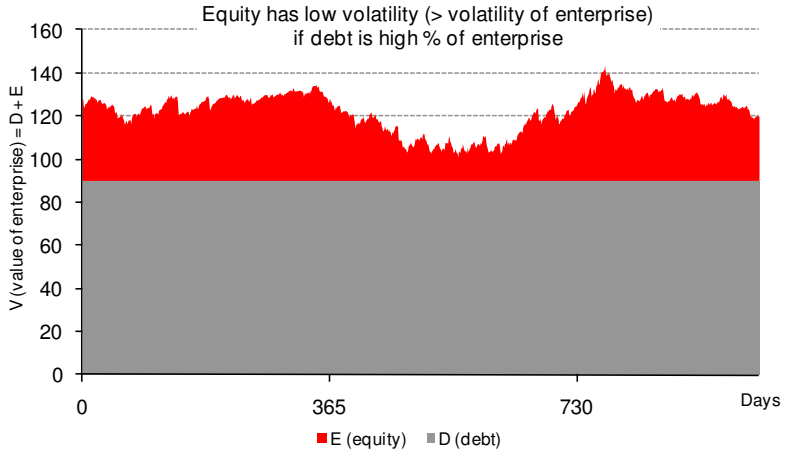
Firms with a small amount of debt have equity volatility roughly equal to firm volatility

If a firm has a very small (or zero) amount of debt, then the value of equity and the enterprise are very similar. In this case, the volatility of the equity and enterprise should be very similar (see Figure 135 above).

Firms with high value of debt to equity have very high equity volatility

For enterprises with very high levels of debt, a relatively small percentage change in the value of the enterprise V represents a relatively large percentage change in the value of equity. In these cases equity volatility will be substantially higher than the enterprise volatility (see Figure 136 below).

Figure 136. Value of Enterprise and Equity with High Debt



Source: Santander Investment Bolsa.

Proof equity volatility is proportional to leverage

The mathematical relationship between the volatility of the enterprise and volatility of equity is given below. The $N(d_1)$ term adjusts for the delta of the equity.

$$\sigma_E = N(d_1) \times \sigma_V \times V / E$$

Equity volatility is proportional to enterprise volatility multiplied by leverage

If we assume the enterprise is not distressed and the equity is ITM, then $N(d_1)$ or delta of the equity should be very close to 1 (it is usually c90%). Therefore, the equation can be simplified so the volatility of equity is proportional to leverage (V / E).

$$\sigma_E \approx \sigma_V \times \text{leverage}$$

(2) JUMP DIFFUSION

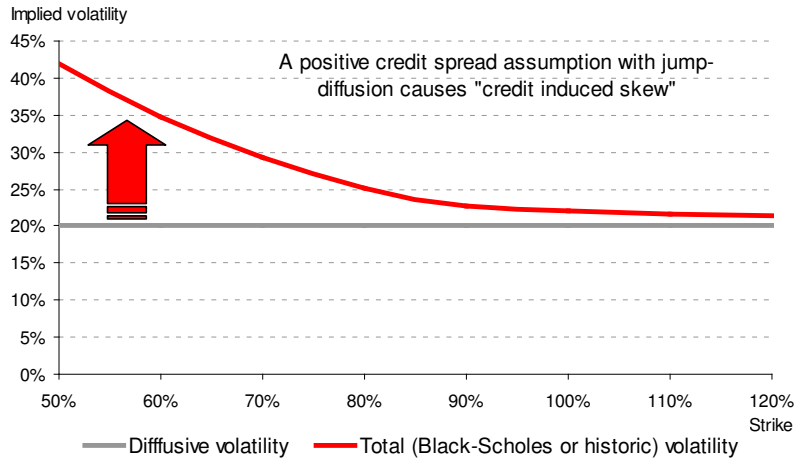
A jump diffusion model separates the movement of equities into two components. There is the diffusive volatility, which is due to random log-normally distributed returns occurring continuously over time. In addition, there are discrete jumps the likelihood of which is given by a credit spread. The total of the two processes is the total volatility of the underlying. It is this total volatility that should be compared to historic volatility or Black-Scholes volatility.

Default risk explained by credit spread

Zero credit spread implies a company can never default

For simplicity, we shall assume that in a jump diffusion model the jumps are to a zero stock price as the firm enters bankruptcy, but results are similar for other assumptions. The credit spread determines the risk of entering bankruptcy. If a zero credit spread is used, the company will never default. The probability of default increases as the credit spread increases (approximately linearly).

Figure 137. Credit-Induced Skew (with 100bp credit spread)



Source: Santander Investment Bolsa.

JUMP DIFFUSION CAUSES CREDIT-INDUCED SKEW

To show how credit spread (or bankruptcy) causes credit-induced skew, we shall price options of different strike with jump diffusion, keeping the diffusive volatility and credit spread constant. Using the price of the option, we shall then calculate the Black-Scholes implied volatility. The Black-Scholes implied volatility is higher for lower strikes than higher strikes, causing skew.

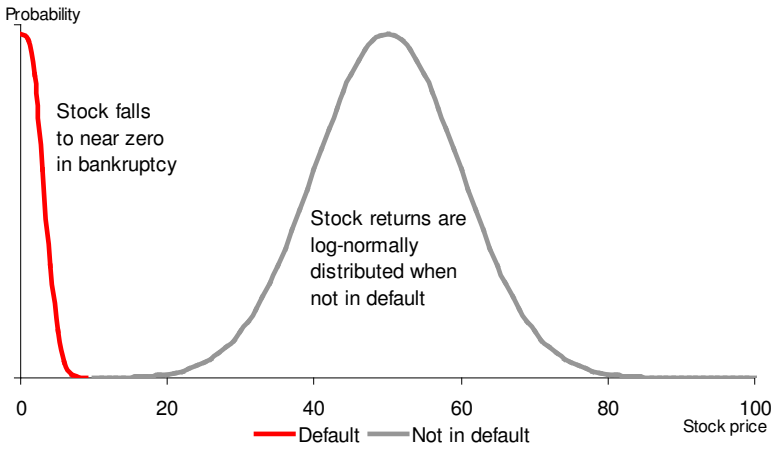
Credit-induced skew is caused by 'option on bankruptcy'

The time value of an option will be divided between the time value due to diffusive volatility and the time value due to the jump to zero in bankruptcy. High strike options will be relatively unaffected by the jump to bankruptcy, and the Black-Scholes implied volatility will roughly be equal to the diffusive volatility. However, the value of a jump to a zero stock price will be relatively large for low strike put options (which, due to put call parity, is the implied for all options). The difference between the Black-Scholes implied and diffusive volatility could be considered to be the value due to the 'option on bankruptcy'.

(3) PUT VS CDS

The probability distribution of a stock price can be decomposed into the probability of a jump close to zero due to credit events or bankruptcy, and the log-normal probability distribution that occurs when a company is not in default. While the value of a put option will be based on the whole probability distribution, the value of a CDS will be driven solely by the probability distribution due to default. The (bi-modal) probability distribution of a stock price due to default, and when not in default, is shown below.

Figure 138. Stock Price Probability Distribution



Source: Santander Investment Bolsa.

Puts can be used instead of CDS (as puts pay out strike price in event of bankruptcy)

Stock price jumps to near zero in event of default

When a stock defaults, the share price tends to fall to near zero. The recovery rate of equities can only be above zero if debt recovers 100% of face value, and most investors price in a c40% recovery rate for debt. A put can therefore be assumed to pay out the maximum level (ie, the strike) in the event of default. Puts can therefore be used as a substitute for a CDS. The number of puts needed is shown below.

Value of puts in default = Strike × Number of Puts

Value of CDS in default = (100% – Recovery Rate) × Notional

In order to substitute value of puts in default has to equal value of CDS in default.

- ⇨ Strike × Number of Puts = (100% – Recovery Rate) × Notional
- ⇨ Number of Puts = (100% – Recovery Rate) × Notional / Strike

CDS PRICES PROVIDE FLOOR FOR PUTS

Long put vs short CDS was popular in 2000-03 bear market

As a put can have a positive value even if a stock is not in default, a CDS must be cheaper than the equivalent number of puts (equivalent number of puts chosen to have same payout in event of default, ie, using the formula above). If a put is cheaper than a CDS, an investor can initiate a long put-short CDS position and profit from the difference. This was a popular capital structure arbitrage trade in the 2000-03 bear market, as not all volatility traders were as focused on the CDS market as they are now, and arbitrage was possible.

CDS in default must have greater return than put in default (without arbitrage)

As a CDS has a lower price for an identical payout in default, a CDS must have a higher return in default than a put. Given this relationship, it is possible to find the floor for the value of a put. This assumes the price of a CDS is 'up front' ie, full cost paid at inception of the contract rather than quarterly.

Puts return in default = Strike / Put Price

CDS return in default = (100% – Recovery Rate) / CDS Price

As CDS return in default must be greater than or equal to put return in default.

- ⇨ (100% – Recovery Rate) / CDS Price ≥ Strike / Put Price
- ⇨ Put Price ≥ Strike × CDS Price / (100% – Recovery Rate)

PUT VS CDS IS A POPULAR CAPITAL STRUCTURE ARBITRAGE TRADE

As the prices of the put and CDS are known, the implied recovery rate can be backed out using the below formula. If an investor's estimate of recovery value differs significantly from this level, a put vs CDS trade can be initiated. For a low (or zero) recovery rate, the CDS price is too high and a short CDS long put position should be initiated. Conversely, if the recovery rate is too high, a CDS price is too cheap and the reverse (long CDS, short put) trade should be initiated.

Put Price = Strike × CDS Price / (100% – Implied Recovery Rate)

RATIO PUT SPREADS CAP VALUE OF PUTS

CDSs provide a floor to the price of a put. It is also possible to cap the price of a put by considering ratio put spreads. For example, if we have the price for the ATM put, this means we know that the value of a 50% strike put cannot be greater than half the ATM put price. If not, we could purchase an ATM-50% 1×2 put spread (whose payout is always positive) and earn a premium for free. This argument can be used for all strikes K and all 1xN put spreads, and is shown below:

$$N \times \text{put of strike } \frac{K}{N} \leq \text{put of strike K}$$

ARBITRAGE MOST LIKELY WITH LOW STRIKE AND LONG MATURITY

Implied volatility is floored by CDSs, and capped by put ratios

The combination of CDS prices providing a floor, and put prices of higher strikes providing a cap, gives a corridor for the values of puts. The width of this corridor is narrowest for low strike long maturity options, as these options have the greatest percentage of their value associated with default risk. As for all capital structure arbitrage strategies, companies with high credit spreads are more likely to have attractive opportunities and arbitrage is potentially possible for near-dated options.

(4) NO-DEFAULT IMPLIED VOLATILITY

The volatility of a stock price can be decomposed into the volatility due to credit events or bankruptcy and the volatility that occurs when a company is not in default. This is similar to the volatility due to jumps and the diffusive volatility of a jump diffusion model. As the value of a put option due to the probability of default can be calculated from the CDS or credit market, if this value was taken away from put prices this would be the 'no-default put price' (ie, the value the put would have if a company had no credit risk). The implied volatility calculated using this 'no-default put price' would be the 'no-default implied volatility'. No-default implied volatilities are less than the vanilla implied volatility, as vanilla implied volatilities include credit risk).

No-default implied volatilities have lower skew and term structure

While we derive the no-default implied volatility from put options, due to put call parity the implied volatility of calls and puts is identical for European options. As the value of a put associated with a jump to default is highest for low-strike and/or long-dated options, no-default implied volatilities should have a lower skew and term structure than vanilla Black-Scholes implied volatilities. A no-default implied volatility surface should therefore be flatter than the standard implied volatility surface and, hence, could be used to identify potential trading opportunities.

The products and strategies addressed in this report are complex, typically involve a high degree of risk and are intended for sale only to sophisticated investors who are capable of understanding and assuming the risks involved. The market value of any structured security may be affected by changes in economic, financial and political factors (including, but not limited to, spot and forward interest and exchange rates), time to maturity, market conditions and volatility, and the credit quality of any issuer or reference issuer. Any investor interested in purchasing a structured product should conduct their own investigation and analysis of the product and consult with their own professional advisers as to the risks involved in making such a purchase.

The opinions and recommendations included in this report are not necessarily those of the Equity Research Department of Santander Investment Bolsa or of its affiliates. A “Trading Places” rating on a specific company equating to that associated with a conventional “Buy, Hold or Underweight” recommendation should not be construed as a fundamental or official rating of a Santander Investment Bolsa analyst. Furthermore, the opinions and strategies contained in this report are completely independent of those that the Equity Research and Sales/Trading Departments of Santander Investment Bolsa may have from time to time.

Some investments discussed in this report may have a high level of volatility. High volatility investments may experience sudden and large falls in their value causing losses when that investment is realised. Those losses may equal your original investment. Indeed, in the case of some investments the potential losses may exceed the amount of initial investment and, in such circumstances, you may be required to pay more money to support those losses. Income yields from investments may fluctuate and, in consequence, initial capital paid to make the investment may be used as part of that income yield. Some investments may not be readily realisable and it may be difficult to sell or realise those investments, similarly it may prove difficult for you to obtain reliable information about the value, or risks, to which such an investment is exposed.

Past performance should not be taken as an indication or guarantee of future performance, and no representation or warranty, express or implied, is made regarding future performance. Information, opinions and estimates contained in this report reflect a judgement at its original date of publication by Grupo Santander and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. The value of securities and financial instruments is subject to exchange rate fluctuation that may have a positive or adverse effect on the price or income of such securities or financial instruments.

Information and opinions presented in this report have been obtained or derived from sources believed by Grupo Santander to be reliable, but Grupo Santander makes no representation as to their accuracy or completeness. Grupo Santander accepts no liability for loss arising from the use of the material presented in this report, except that this exclusion of liability does not apply to the extent that such liability arises under specific statutes or regulations applicable to Grupo Santander. This report is not to be relied upon in substitution for the exercise of independent judgment. Grupo Santander may have issued, and may in the future issue, other reports that are inconsistent with, and reach different conclusions from, the information presented in this report. Those reports reflect the different assumptions, views and analytical methods of the analysts who prepared them and Grupo Santander is under no obligation to ensure that such other reports are brought to the attention of any recipient of this report.

See back cover of this report for further disclaimer disclosures.

Important Disclosures

This report has been prepared by Santander Investment Bolsa, Sociedad de Valores, S.A. ("Santander Investment Bolsa") and is provided for information purposes only. This document must not be considered as an offer to sell or a solicitation of an offer to buy. Any decision by the recipient to buy should be based on publicly available information on the related security and, where appropriate, should take into account the content of the related prospectus filed with the CNMV (Spanish National Securities Market Commission) and available from the CNMV, the company governing the related market (Sociedad Rectora de la Bolsa) and the company issuing the security. This report is issued in the United States by Santander Investment Securities, Inc. ("SIS"), in Spain by Santander Investment Bolsa and in the United Kingdom by Banco Santander, S.A., London Branch ("Santander London"). Santander London is authorised by the Bank of Spain. SIS, Santander Investment Bolsa and Santander London are members of Grupo Santander. This report is not being issued to private customers.

The information contained herein has been compiled from sources believed to be reliable, but while all reasonable care has been taken to ensure that the information contained herein is not untrue or misleading at the time of publication, we make no representation that it is accurate or complete and it should not be relied upon as such. All opinions and estimates included herein constitute our judgement as at the date of this report and are subject to change without notice. Santander Investment Bolsa may change the recommendation it has on a stock at any given time. It may also cease to cover a stock or initiate coverage of new stocks. There is no specific calendar for any such actions. From time to time, Grupo Santander and/or any of its officers or directors may have a position, or otherwise be interested in, transactions in securities which are directly or indirectly the subject of this report. Santander Investment Bolsa has internal rules of conduct that contain, among other things, procedures to prevent conflicts of interest with respect to recommendations, including: the consideration of its Research Department as a separate area, Chinese Walls, and the possibility of establishing specific restrictions on research activity where appropriate. Santander Investment Bolsa's research reports contain a certification stating that they reflect the authors' own opinions.

Grupo Santander may from time to time perform services for or solicit business from any company mentioned in this report. Neither Grupo Santander nor any other person accepts any liability whatsoever for any direct or consequential loss arising from any use of this report or its contents. This report may not be reproduced, distributed or published by any recipient for any purpose. Santander Investment Bolsa is under the supervision of the CNMV.

Any US recipient of this report (other than a registered broker-dealer or a bank acting in a broker-dealer capacity) that would like to effect any transaction in any security discussed herein should contact and place orders in the United States with the company distributing the research, SIS at (212) 692-2550, which, without in any way limiting the foregoing, accepts responsibility (solely for purposes of and within the meaning of Rule 15a-6 under the US Securities Exchange Act of 1934) under this report and its dissemination in the United States. US recipients of this report should be advised that this research has been produced by a non-member affiliate of SIS and, therefore, by rule, not all disclosures required under NASD Rule 2711 apply.

LOCAL OFFICES

Madrid
Tel: 34-91-257-2309
Fax: 34-91-257-1811

Lisbon
Tel: 351-21-381-6580
Fax: 351-21-389-3680

London
Tel: 44-207-756-6605
Fax: 44-207-756-5033

New York
Tel: 212-692-2550
Fax: 212-407-4540

Bogotá
Tel: 571-644-8006
Fax: 571-592-0638

Buenos Aires
Tel: 54114-341-1052
Fax: 54114-341-1226

Caracas
Tel: 582-401-4306
Fax: 582-401-4219

Lima
Tel: 511-215-8100
Fax: 511-215-8185

Mexico City
Tel: 5255-5629-5040
Fax: 5255-5629-5846

Santiago
Tel: 562-336-3300
Fax: 562-697-3869

São Paulo
Tel: 5511-5538-8226
Fax: 5511-5538-8407

Tokyo
Tel: 813-3211-0356
Fax: 813-3211-0362