GLOBAL VOLATILITY SUMMIT 2013

April 2013 Newsletter

2014 Event Details

Date. Plans are already underway to secure a convenient date for the 2014 event!

Please continue to check the website for registration, updates and tentative agenda (www.globalvolatilitysummit.com).

2013 Event Recap

Keynote speaker. Sal Khan, founder of The Khan Academy and author of *The One World Schoolhouse* gave an insightful presentation on using technology to innovate the way education is provided across the globe.

Special Guest Speaker. Mike Edleson followed up to his 2012 GVS talk about the decision to implement a tail hedge, with an informative discussion on implementation of a tail hedge and how to identify the right managers for your mandate. Mr. Edleson's presentation is available on the GVS website.

Managers. The following managers participated:

Blue Mountain Capital
Capstone Investment Advisors
Fortress Investment Group
Forty4 Fund
Ionic Capital Management
JD Capital Management
Parallax Fund
PIMCO
Pine River Capital Management
Saiers Capital

Questions?
Please contact info@globalvolatilitysummit.com

2013 Event Summary and April research piece

The fourth annual Global Volatility Summit ("GVS") was a success. The event took place on February 25th in New York City, and ten volatility and tail hedge managers hosted an audience of over 350 people. The event featured a thought provoking key note speech by Sal Khan regarding the transformation of the educational process to a web based mode of communication, a presentation by Mike Edleson from The University of Chicago on tail hedging implementation, and four panels including a pension and consultant panel.

The primary goal of the GVS is to educate the investment community about volatility and how it can help investors attain their growth targets. The GVS is an evolving community of managers, investors, and industry experts. We rely on the feedback and guidance of our investors to shape the event and line-up of speakers each year. Following the summit in February, a number of you requested more fundamental knowledge on volatility trading strategies. As a result, we are sharing a comprehensive piece on volatility trading strategies co-authored by Colin Bennett and Miguel Gil of Santander. We thank them for sharing this piece, which we believe you will find to be informative.

If you have any topics you would like to see the managers address in future newsletters please send us an email.

Cheers,
Global Volatility Summit

Equity Derivatives

Europe

Madrid, November 5, 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

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 seven 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 later chapters are aimed at hedge funds and proprietary trading desks.

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. We also show that while the profit from delta hedging is based on percentage move squared (ie, variance or volatility^2), it is the absolute difference between realised and implied that determines carry.

- **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 labour-intensive and expensive.

- **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 ¾ 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 VOLATILITY INDICES

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 volatility futures such as VIX and vStoxx futures has become increasingly popular. However, as is the case with forward starting options, there are modelling issues.

- **Forward starting products.** As the exposure is to forward volatility rather than volatility, more sophisticated models need to be used to price forward starting products than ordinary options. Forward starting options will usually have wider bid-offer spreads than vanilla options, as their pricing and hedging is more complex. Forward starting variance swaps are easier to price as the price is determined by two variance swaps.

- **Volatility indices.** While volatility indices were historically based on ATM implied, most providers have swapped to a variance swap based calculation. The price of a volatility index will, however, typically be 0.2-0.7pts below the price of a variance swap of the same maturity, as the calculation of the volatility index typically chops the tails to remove illiquid prices. Each volatility index provider has to use a different method of chopping the tails in order to avoid infringing the copyright of other providers.

- **Futures on volatility indices.** While futures on volatility indices were first launched on the VIX in March 2004, it has only been since the more recent launch of structured products and options on volatility futures that liquidity has improved enough to be a viable method of trading volatility. As a volatility future payout is based on the square root of variance, the payout is linear in volatility not variance. The fair price of a future on a volatility index is between the forward volatility swap, and the square root of the forward variance swap. Volatility futures are, therefore, short vol of vol, just like volatility swaps. It is therefore possible to get the implied vol of vol from the listed price of volatility futures.

- **Volatility future ETN/ETF.** Structured products based on constant maturity volatility futures have become increasingly popular, and in the US have at times had a greater size than the underlying volatility futures market. As a constant maturity volatility product needs to sell near-dated expiries and buy far-dated expiries, this flow supports term structure for volatility futures and the underlying options on the index itself. The success of VIX-based products has led to their size being approximately two-thirds of the vega of the relevant VIX futures market (which is a similar size to the net listed S&P500 market) and, hence, appears to be artificially lifting near-dated term structure. The size of vStoxx products is not yet sufficient to significantly impact the market, hence they are a more viable method of trading volatility in our view. We recommend shorting VIX-based structured products to profit from this imbalance, potentially against long vStoxx-based products as a hedge. Investors who wish to be long VIX futures should consider the front-month and fourth-month maturities, as their values are likely to be depressed from structured flow.

- **Options on volatility futures.** The arrival of options on volatility futures has encouraged trading on the underlying futures. It is important to note that an option on a volatility future is an option on future implied volatility, whereas an option on a variance swap is an option on realised volatility. As implieds always trade at a lower level to peak realised (as you never know when peak realised will occur) the volatility of implied is lower than the volatility of realised, hence options on volatility futures should trade at a lower implied than options on var. Both have significantly downward sloping term structure and positive skew. We note that the implied for options on volatility futures should not be compared to the realised of volatility indices.

LIGHT EXOTICS

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

- **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. Conversely, payout look-back options 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 than 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.

- **Relative value volatility trading.** 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). For indices that are popular volatility trading pairs, if they have significantly different skews this can impact the volatility market. Single-stock relative value volatility trading is possible, but less attractive due to the wider bid-offer spreads.

- **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-T_1) term structures can be compared when multiplied by $\sqrt{}(T_2 T_1)/(\sqrt{T_2}-\sqrt{T_1})$, 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).

- **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.
- **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.
- **Proof variance swaps can be notional = vega/σ).** The payout of a volatility swap can be approximated by a variance swap. We show how the difference in their notionals should be weighted by 2σ.
- **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, the Sortino ratio is more appropriate.
- **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).

This page left blank intentionally.

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

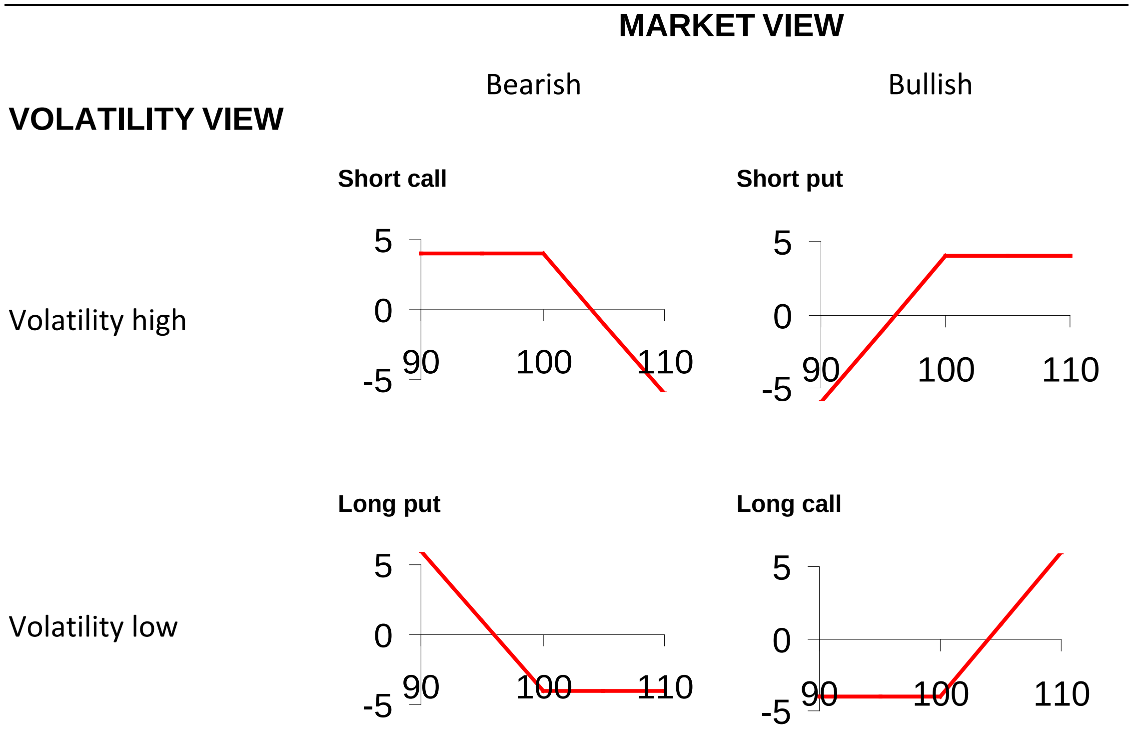
Options increase in value as volatility rises

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



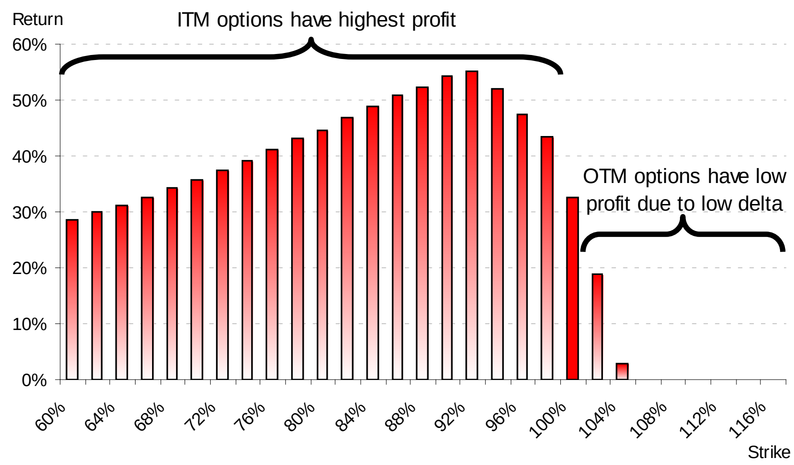
Source: Santander Investment Bolsa.

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

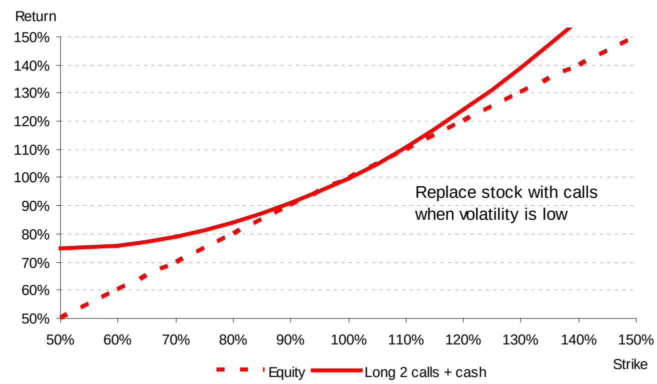
As d_2 is less than d_1 (see above) and N(z) is a monotonically increasing function, this means that $N(d_2)$ is less than $N(d_1)$. Hence, the probability of a call being in the money = $N(d_2)$ is less than the delta = $N(d_1)$. As the delta of a put = delta of call – 1, and the sum of call and put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

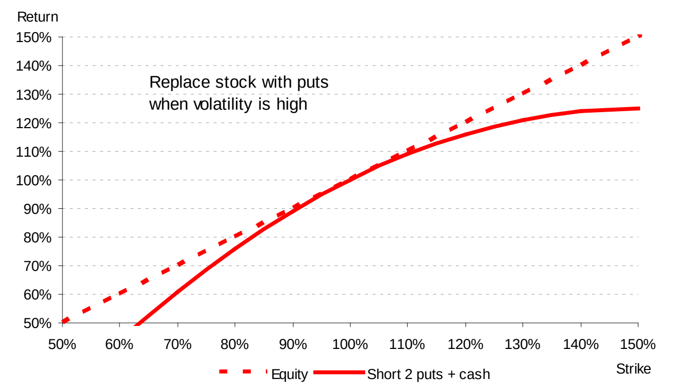
STOCK REPLACING WITH LONG CALL OR SHORT PUT

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.

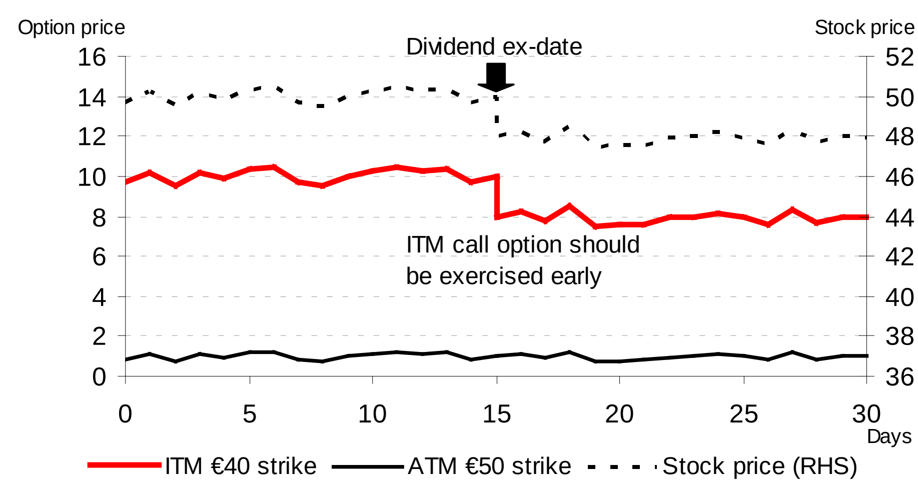
CONVERTING OPTIONS EARLY IS RARE, BUT SOMETIMES NECESSARY

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 ($=e^{rf r\times T}-1$) 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:

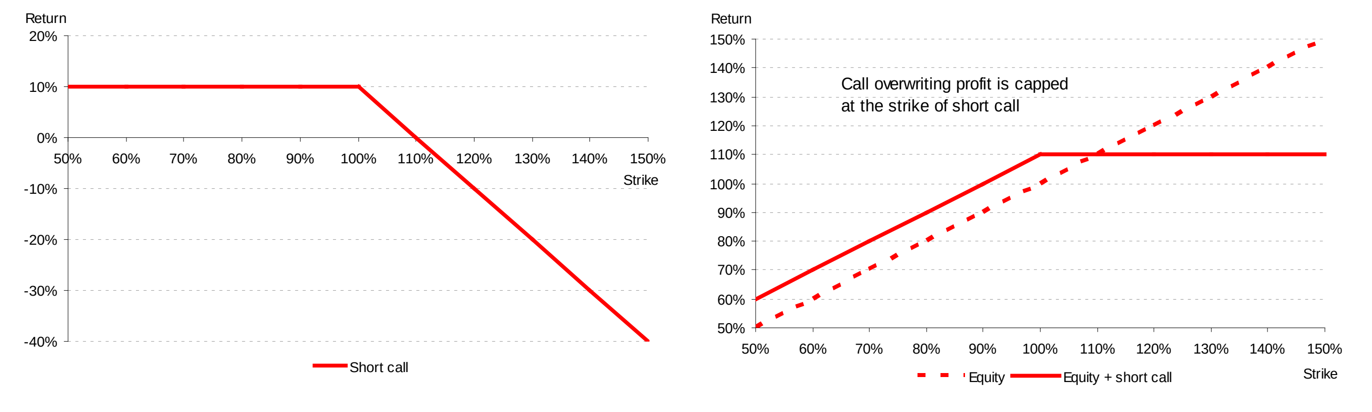
- **Option buying for protection**.
- **Unwillingness to sell low premium options causes market makers to raise their prices** (selling low premium options, like selling lottery tickets, has to be done on a large scale to be attractive).
- **High gamma of near-dated options has a gap risk premium** (risk 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

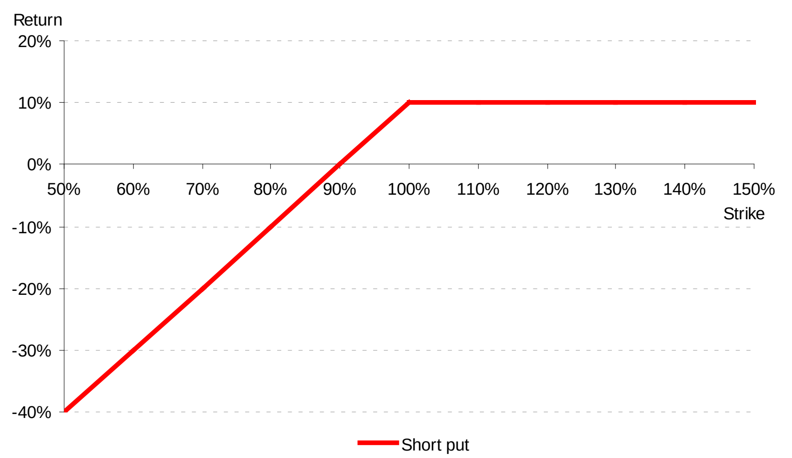
As there are typically more Buy ideas than Sell ideas, call overwriting is useful to enforce an exit strategy

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

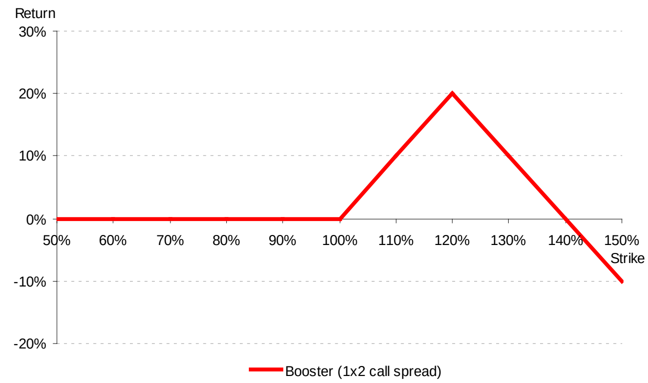


Source: Santander Investment Bolsa.

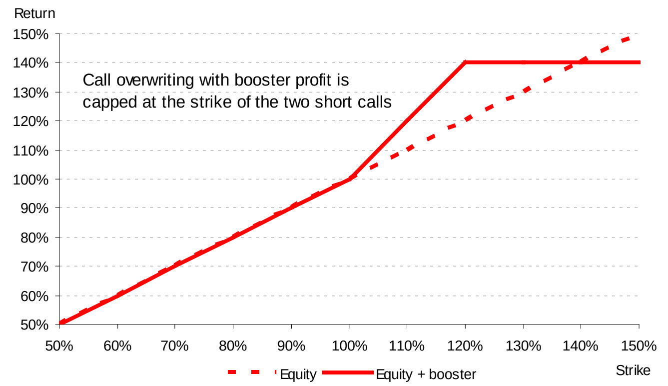
Boosters (1×2 call spreads) are useful when a bounce-back is expected

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



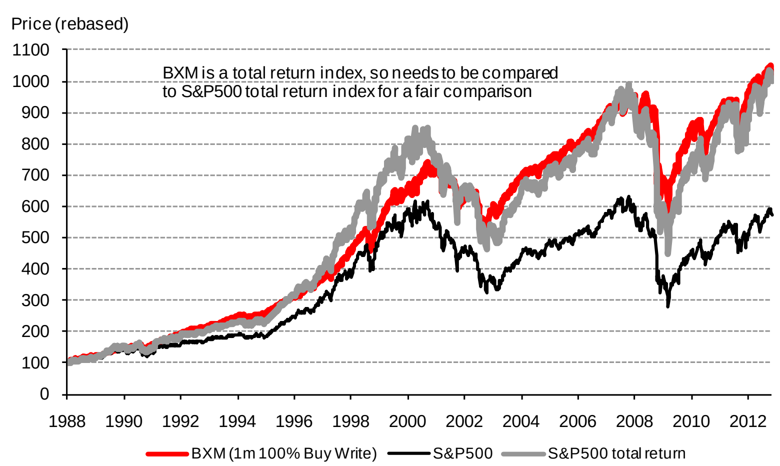
Source: Santander Investment Bolsa estimates.

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)

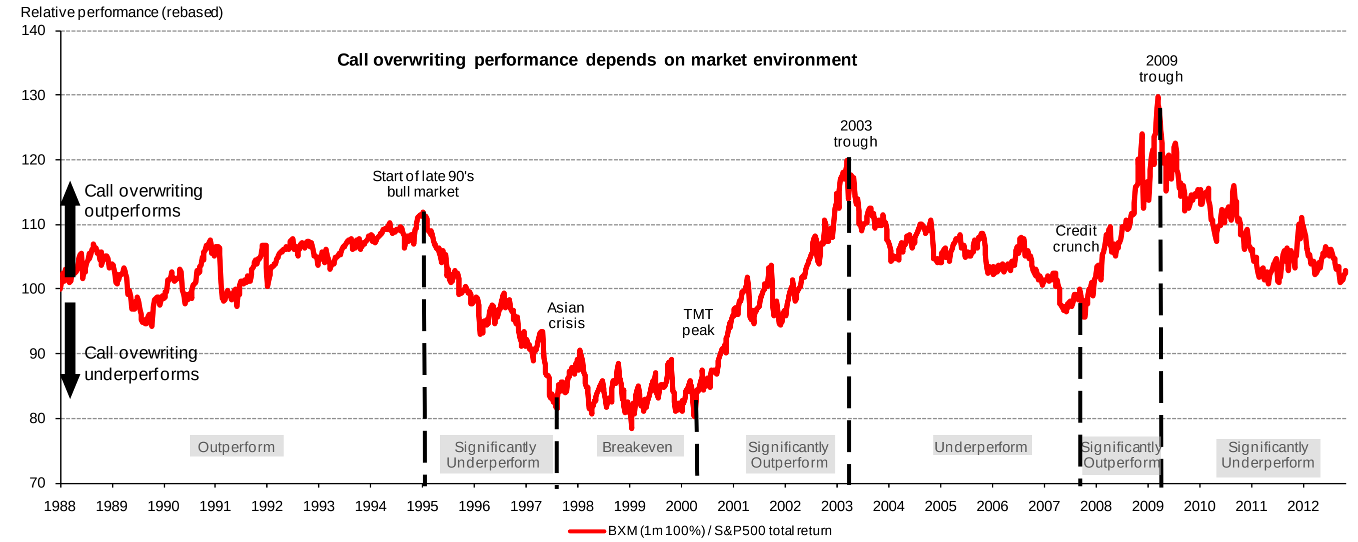


Source: Santander Investment Bolsa.

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

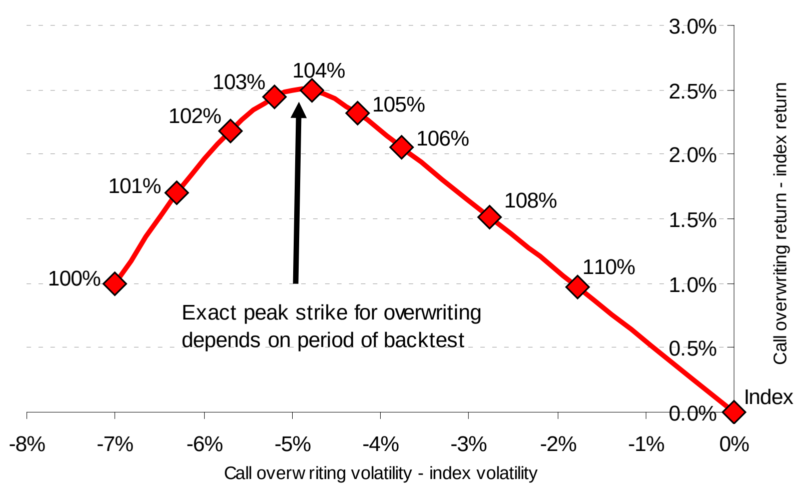
Overwriting with near-dated options has highest return, and highest risk

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

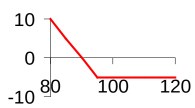
Should a systematic way to enhance call overwriting be viable, this method could be applied to volatility trading without needing an existing long position in the underlying. Given the presence of statistical arbitrage funds and high frequency traders, we believe it is unlikely that a simple automated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

PROTECTION STRATEGIES USING OPTIONS

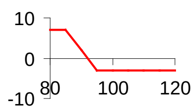
For both economic and regulatory reasons, one of the most popular uses of options is to provide protection against a long position in the underlying. The cost of buying protection through a put is lowest in calm, low-volatility markets, but in more turbulent markets the cost can be too high. In order 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

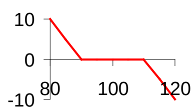
Buying a put against a long position gives complete and total protection for underlying moves below the strike (as the investor can simply put the long position back for the strike price following severe declines). The disadvantage of a put is the relatively high cost, as an investor is typically unwilling 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.



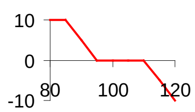
- **Puts give complete protection without capping performance.** As puts give such good protection, their cost is usually prohibitive unless the strike is low. For this reason, put protection is normally bought for strikes around 90%. Given that this protection will not kick in until there is a decline 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.

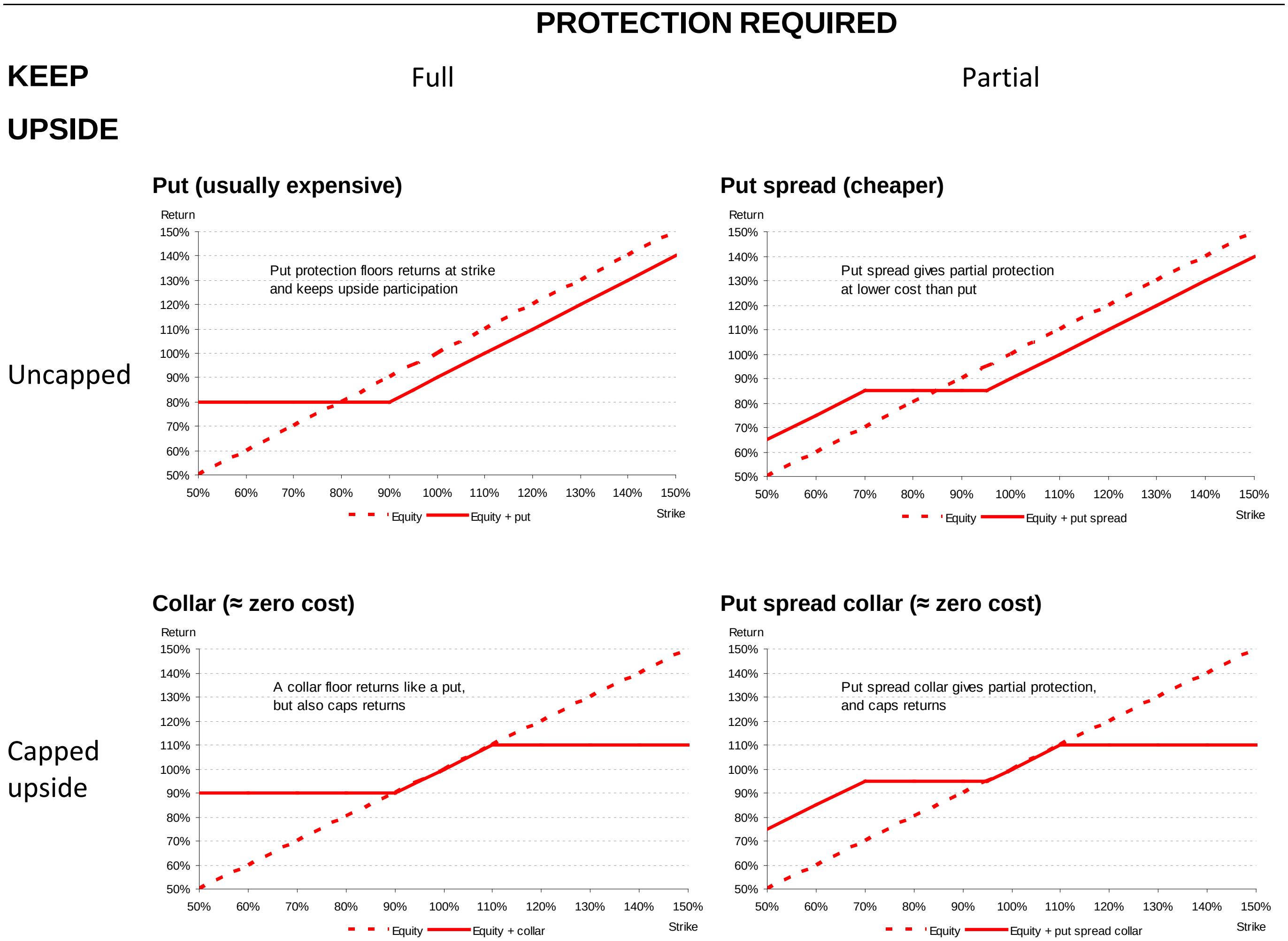


- **Put spread collars best when volatility is high, as two OTM options are sold.** Selling both an OTM put and OTM call against a long put (a put spread collar) is typically attractive when volatility is high, as this lifts the value of the two OTM options sold more than the long put bought. If equity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

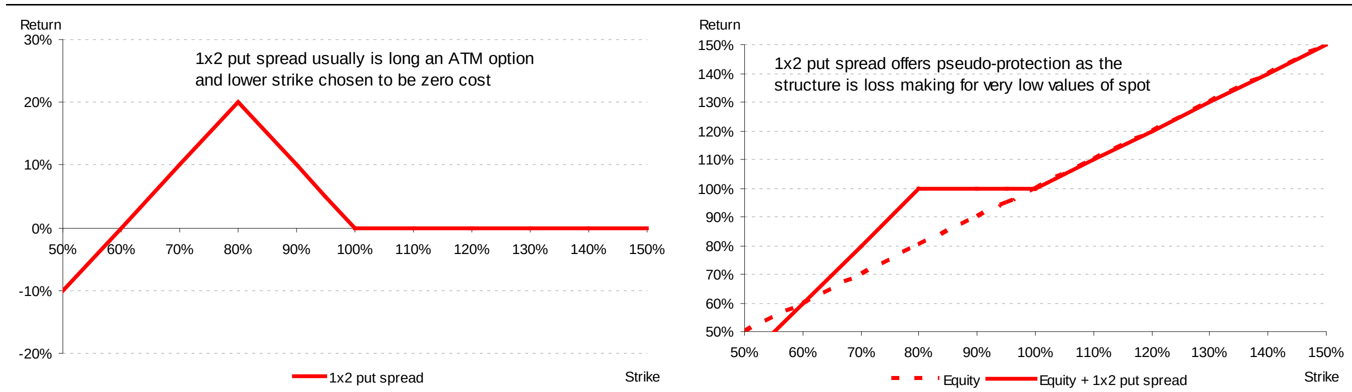
Capping performance should only be used when a long-lasting rally is unlikely

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).

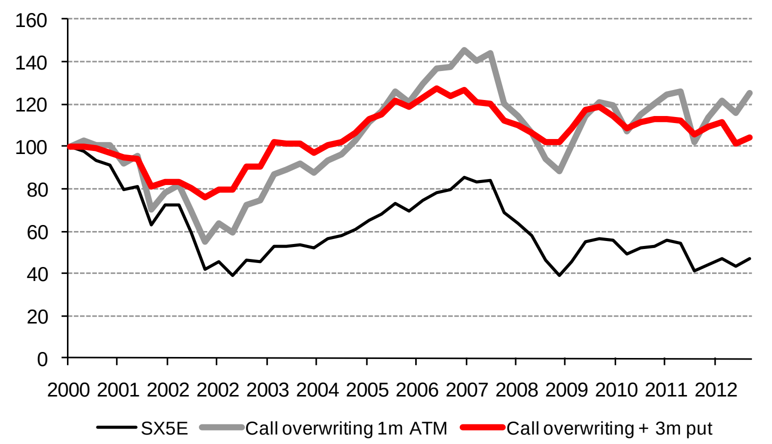
MULTIPLE EXPIRY PROTECTION STRATEGIES

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.

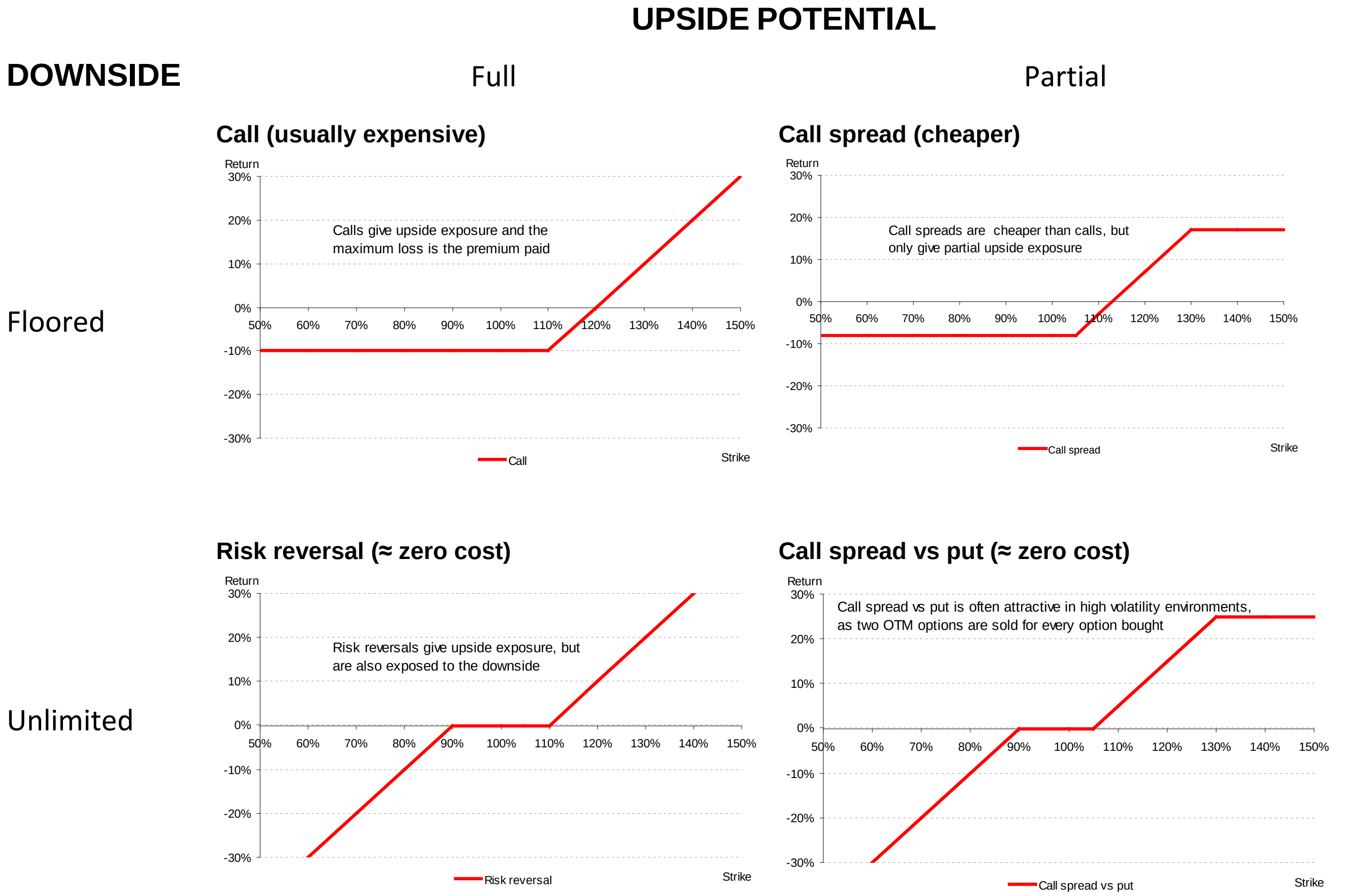
BULLISH COMBOS ARE MIRROR IMAGE OF PROTECTION STRATEGIES

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

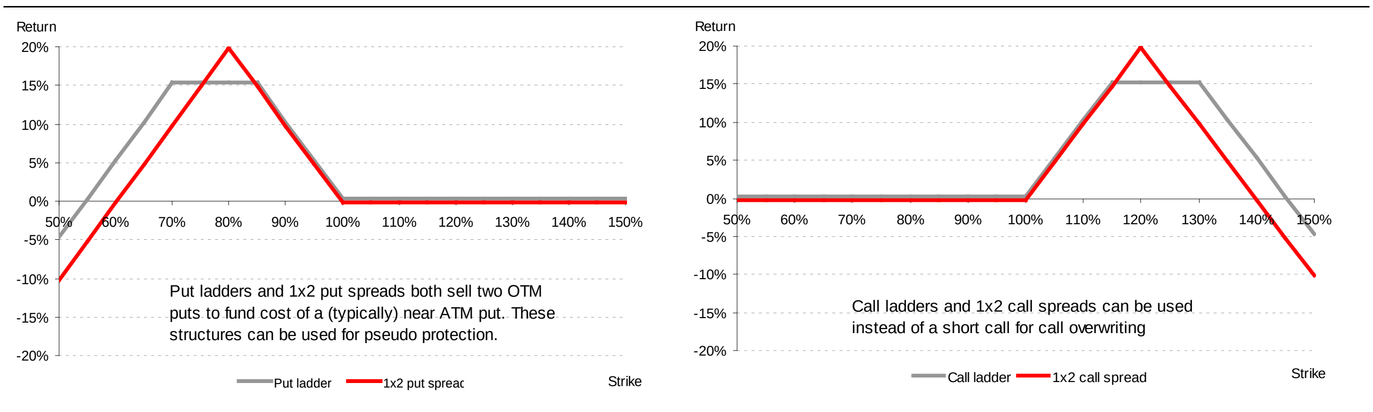


Source: Santander Investment Bolsa estimates.

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

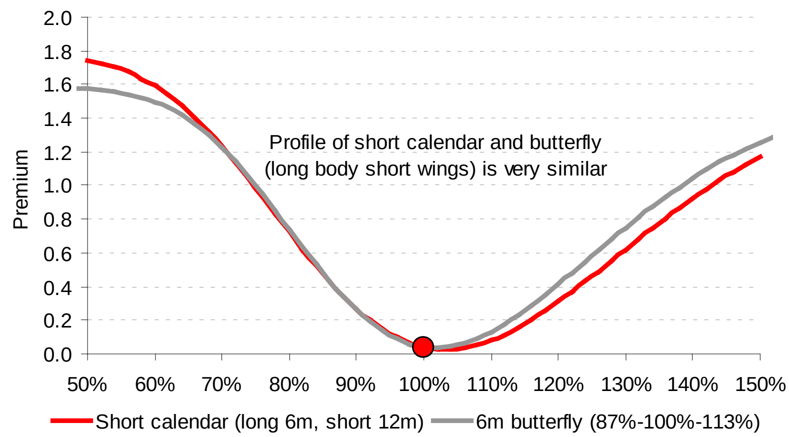
Using option structures allows a straddle (long call and put of same strike) or strangle (long call and put of different strikes) to be traded. These structures are long volatility, but do not have any exposure to the direction of the equity market. For more details, see next section *Volatility Trading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

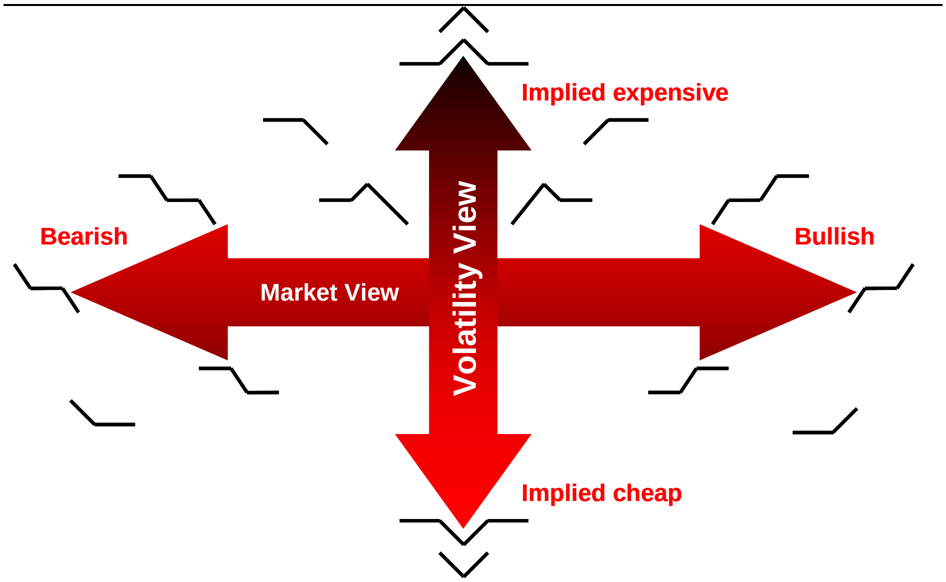


Source: Santander Investment Bolsa.

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

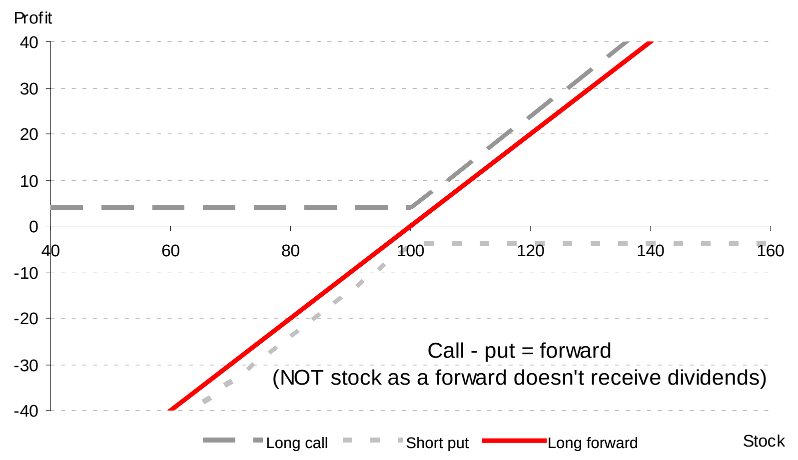
VOLATILITY TRADING USING OPTIONS

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. We also show that while the profit from delta hedging is based on percentage move squared (ie, variance or volatility^2), it is the absolute difference between realised and implied that determines carry (not the difference between realised^2 and implied^2).

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)



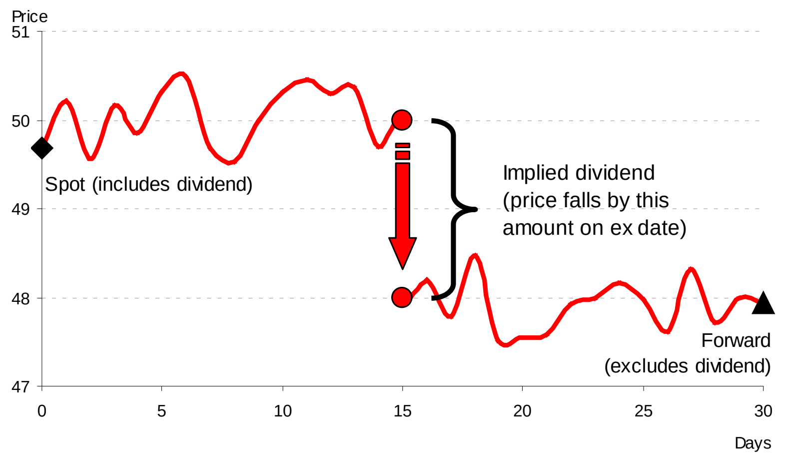
Source: Santander Investment Bolsa.

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

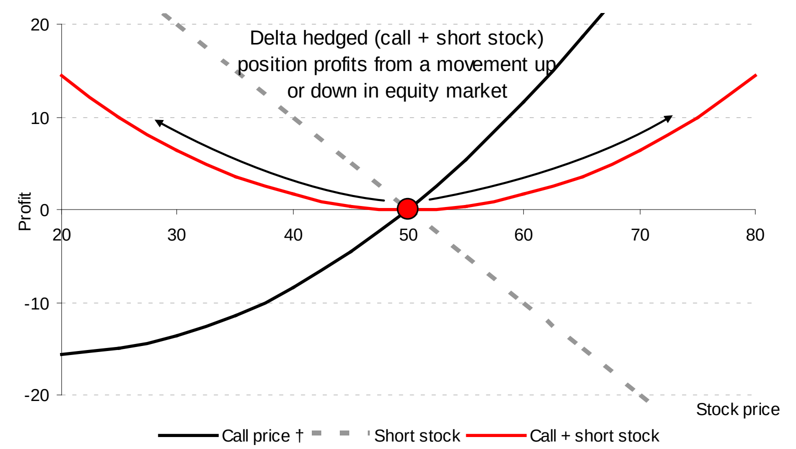
Delta-hedged option gives a position whose profitability is determined by volatility

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.

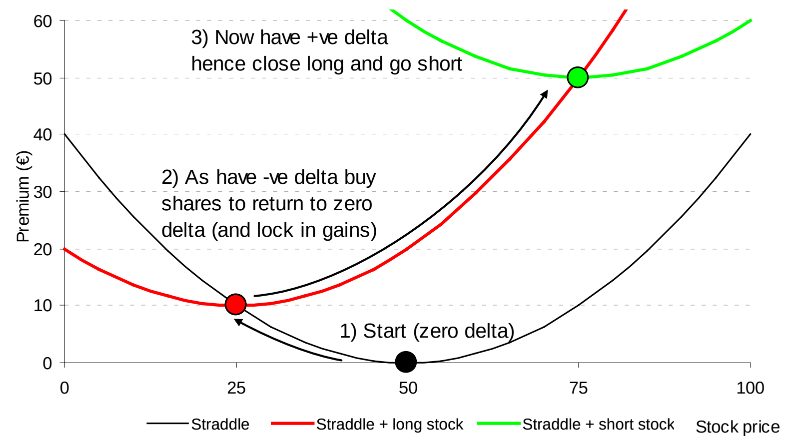
Constant delta hedging is called gamma scalping

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

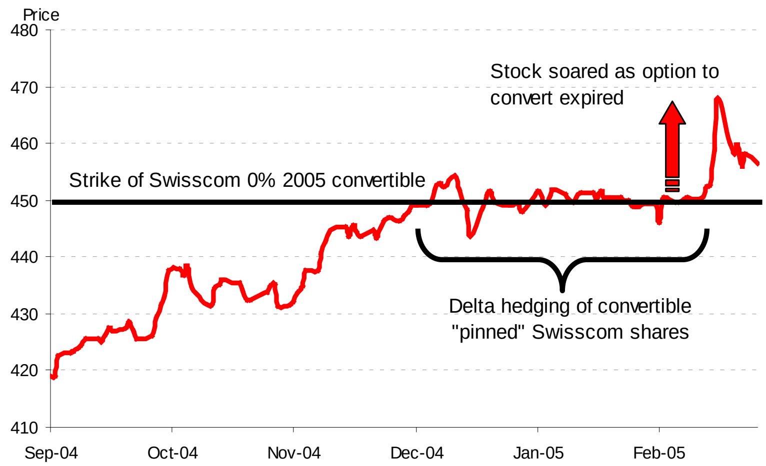
Single stocks (but not indices) can be pinned approaching expiry

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.

- **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).

- **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.

- **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.

- **Difference between implied and realised determines carry.** While variance is the driver of profits if the implied volatility of an option is constant, the carry is determined by the absolute difference between realised and implied (ie, the same carry is earned by going long a 20% implied option that realises 21% as by going long a 40% implied option that realises 41%.

ANNUALISED VOLATILITY IS 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 0.4 × VOLATILITY × √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 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

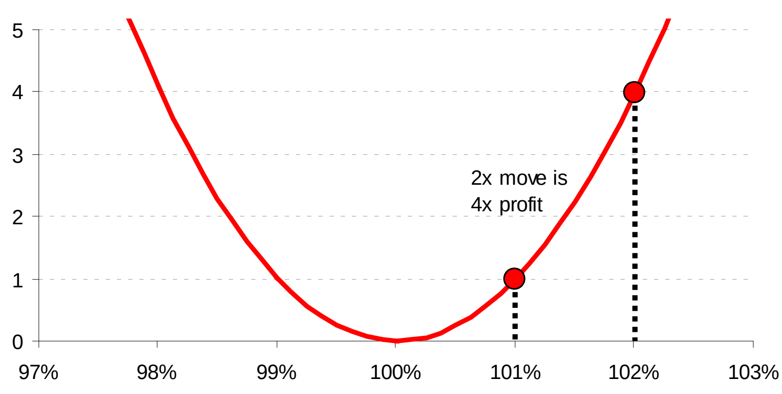
θ = theta

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

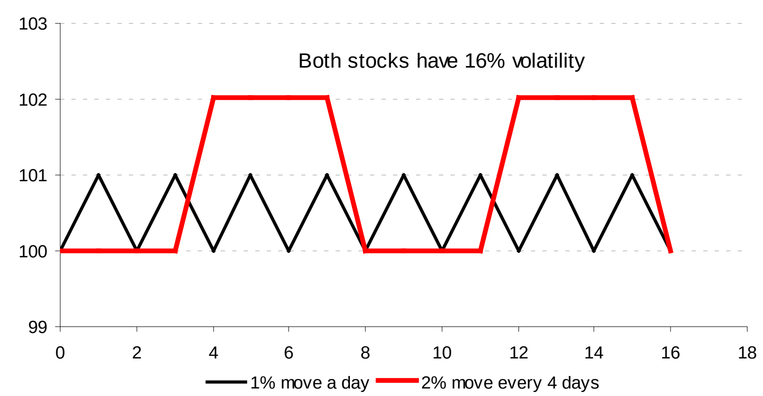


Source: Santander Investment Bolsa.

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.

DIFFERENCE BETWEEN IMPLIED AND REALISED DETERMINES CARRY

Carry is proportional to difference between implied and realised volatility

Assuming all other inputs are constant, the payout of a delta-hedged option is based on the variance (return squared). However, when examining how much carry is earned, or lost, when delta hedging an option, it is the difference between realised and implied (not realised^2 - implied^2) that matters. This is because the gamma of an option is proportional to 1/σ; hence, if volatility doubles the gamma halves. Thus, the same carry (profit from gamma less cost of theta) is earned by going long a 20% implied option that realises 21% as by going long a 40% implied option that realises 41%. The proof of this is below.

$$\text{Delta hedged P\&L from option} = \text{Dollar gamma} \times (\text{return}^2 - \sigma^2 dt)$$

where:

σ = implied volatility

$\gamma = - N'(d_1) / (S \times \sigma \times \sqrt{T})$

Dollar gamma $= 0.5 \times \gamma \times S^2 \approx \text{constant} / \sigma$ for constant spot S and time T

⇨ Daily P&L from option $\approx \text{constant} \times (\text{return}^2 - \sigma^2 dt) / \sigma$

If we define return to be similar to volatility, then return = (σ + x)dt where x is small

⇨ Daily P&L from option $\approx \text{constant} \times dt \times ((\sigma + x)^2 - \sigma^2) / \sigma$

⇨ Daily P&L from option $\approx \text{constant} \times dt \times ((\sigma^2 + 2\sigma x + x^2) - \sigma^2) / \sigma$

⇨ Daily P&L from option $\approx \text{constant} \times dt \times (2x + x^2/\sigma)$

⇨ Daily P&L from option $\approx \text{constant} \times dt \times 2x$ as x is small

⇨ Daily P&L from option proportional to x, where x = realised volatility - σ

Hence, it is the difference between realised and implied volatility that is the key to daily P&L (or carry).

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

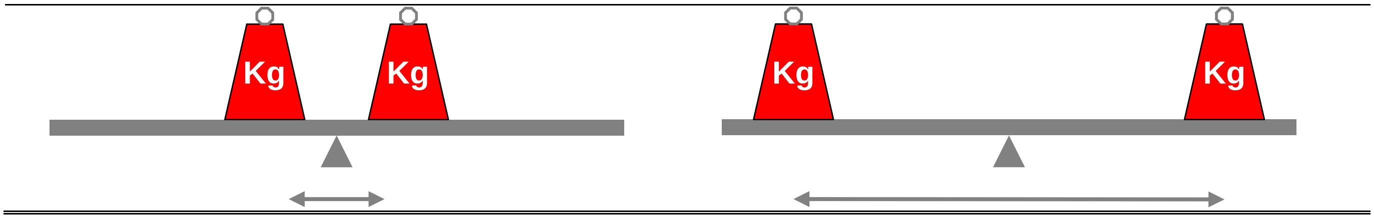
When looking at how rich or cheap options with the same maturity are, rather than looking at the implied volatility for a certain strike (ie, ATM or another suitable strike) it is better to look at the implied variance as it takes into account the implied volatility of all strikes. For example, if an 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 $\text{price}_n/\text{price}_0$.

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

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

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

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

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.

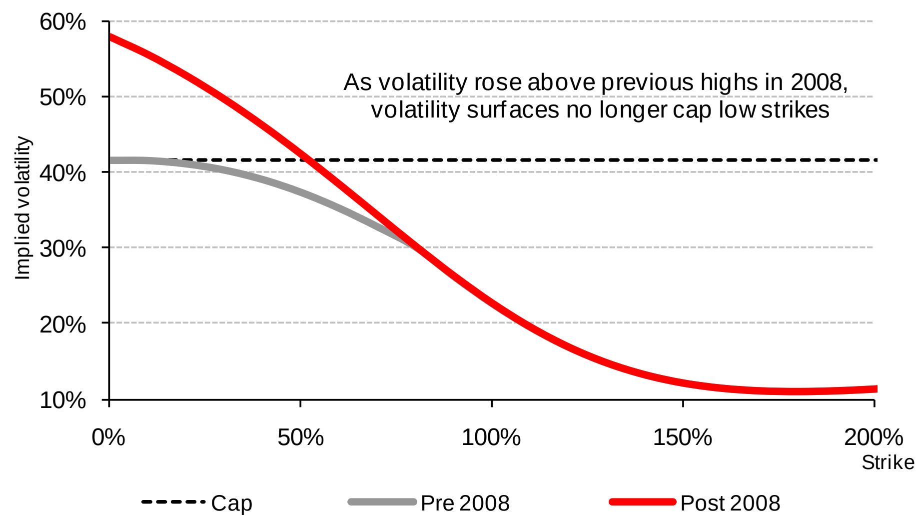
Basis risk between cash and futures can cause traders problems

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.

VARIANCE PRICING CHANGED POST THE 2008 SPIKE IN VOLATILITY

The turmoil seen in 2008 caused 3-month realised volatility to spike above 70%. This was higher than the mid-60's high reached during the Great Depression. Before the Lehman bankruptcy, volatility traders used to cap implied volatility surfaces at a level similar to the all-time highs of realised volatility. The realisation that there could be an event that occurs in the future that has not occurred in the past, a so called 'black swan', has removed this cap (as it is now understood that volatility can spike above historical highs in a severe crisis).

Figure 30. Volatility Surfaces Pre- and Post-2008



Source: Santander Investment Bolsa.

Removal of the implied volatility cap has lifted variance swap levels

The removal of the cap on implied volatility has caused low strike puts to be priced with a far higher implied volatility[6]. While the effect on premium for vanilla options (where the time value of very low strike puts is small) is small, for variance swaps the effect is very large. As variance swaps are more sensitive to low strike implied volatility (shown below), the removal of the cap lifted levels of variance swaps from c2pts above ATMf to c7pts above.

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.

[6] Note the slope of Ln(strike) cannot become steeper as spot declines without arbitrage occurring

Figure 31. Vega of Options of Different Strikes

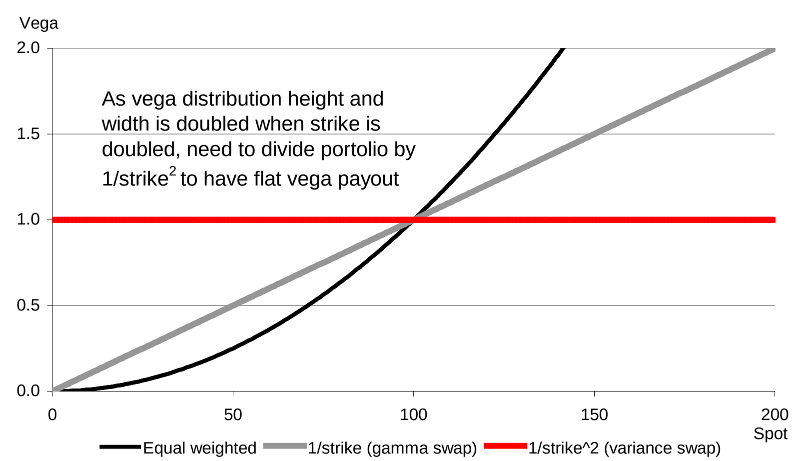
Source: Santander Investment Bolsa.

Variance swaps are hedged with portfolio weighted 1/K²

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 ×σ)2 The vega profile of a volatility swap is of course a flat line (as vega is equal to the volatility notional).

Figure 32. Vega of Portfolio of Options of All Strikes



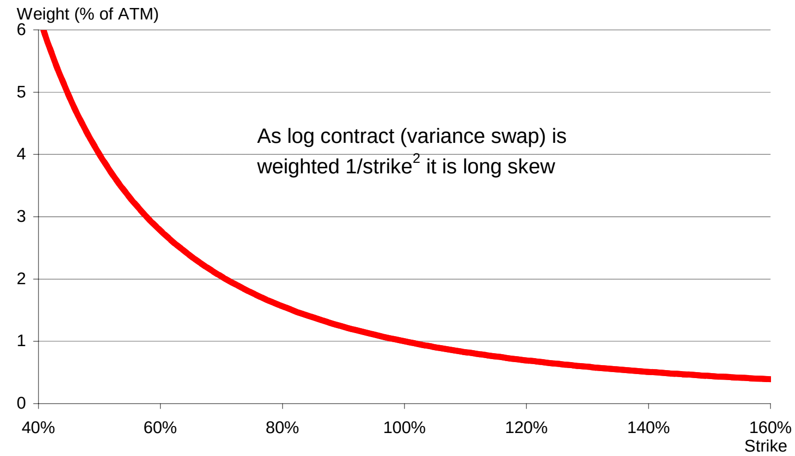
Source: Santander Investment Bolsa.

__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[7].

Figure 33. Weight of Options in Log Contract (Variance Swap)



Source: Santander Investment Bolsa.

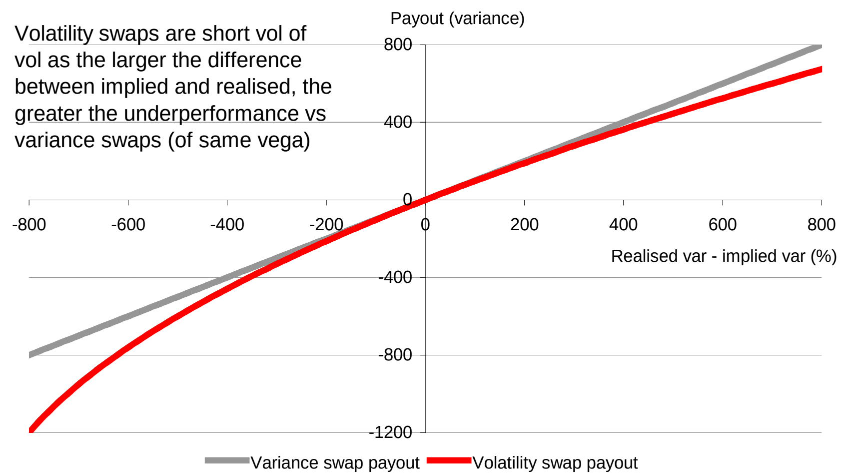
VOLATILITY SWAPS CAN BE HEDGED WITH VARIANCE SWAPS

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 34 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)

[7] 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).

Figure 34. Payout of Variance Swap and Volatility Swap



Source: Santander Investment Bolsa.

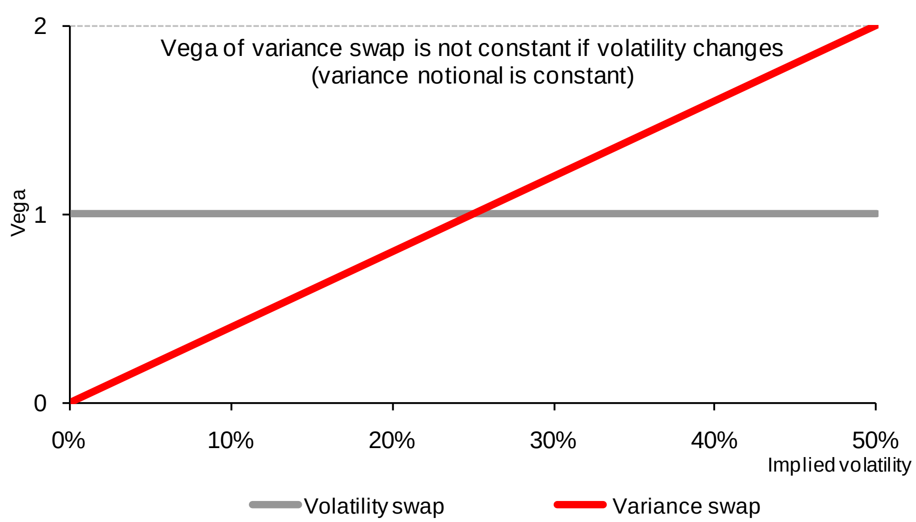
VOLATILITY SWAPS ARE SHORT VOL OF VOL

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 $1/(2\sigma)$. If a trader is long a volatility swap and has hedged with a short variance swap position weighted $1/(2\sigma)$, then as volatility decreases more variance swaps have to be sold (as σ decreases, $1/(2\sigma)$ 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.

Figure 35. Vega of Variance Swaps and Volatility Swaps (for identical vega at 25% vol)



Source: Santander Investment Bolsa.

VARIANCE SWAP VEGA IS NOT CONSTANT IF VOLATILITY CHANGES

We note that although the vega profile of a variance swap against spot is a flat line, this value is not constant and it moves with volatility (variance swap vega = variance notional × 2σ). The vega profile of a volatility swap against volatility is, of course, a constant flat line (as vega is equal to the volatility notional). Therefore, variance swaps have constant vega for changes in spot (but not changes in volatility), while volatility swaps have constant vega for changes in volatility and spot.

DIFFERENCE BETWEEN VAR AND VOL 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 36.

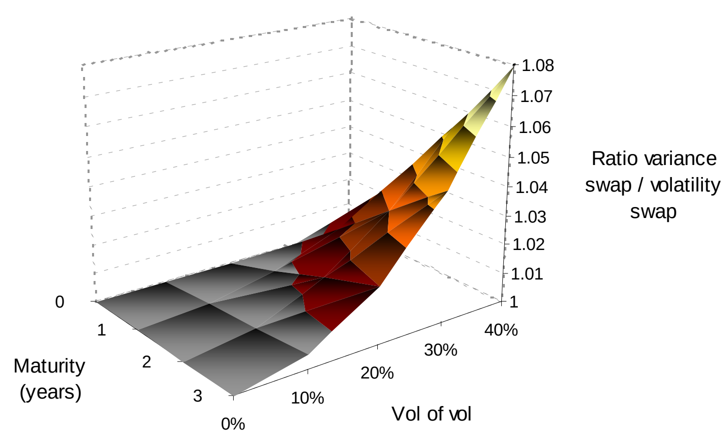
$$c \approx \frac{1}{6}\omega^2 T \sqrt{\text{variance swap price} \times e^{rT}}$$

where:

v = variance swap price

ω = volatility of volatility

Figure 36. 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 37 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 37. 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).

γ = gamma = number of shares bought (or sold) per €1 spot move

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

$\gamma \times S / 100$ = 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

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

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:

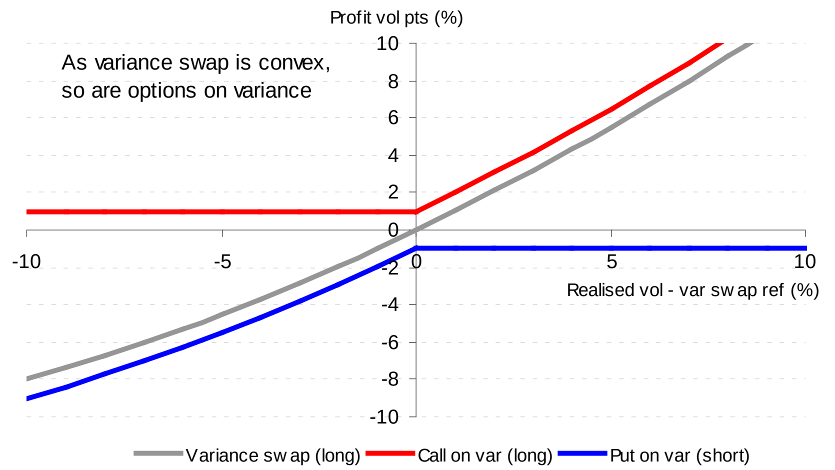
$\text{Premium}_{\text{variance points}} = 2\sigma_S \times \text{Premium}_{\text{volatility points}}$ where σ_S = variance swap reference

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 38. Variance Swap, ATM Call on Variance and ATM Put on Variance



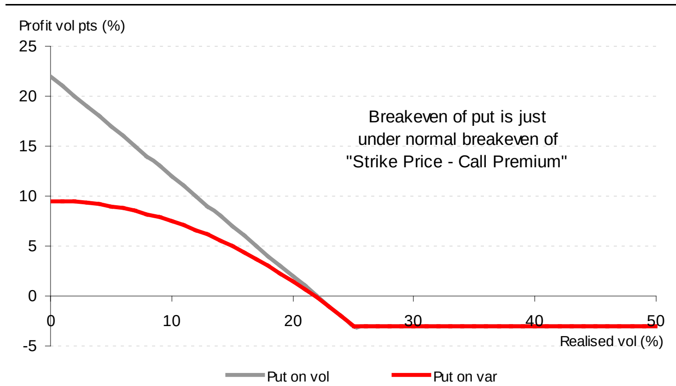
Source: Santander Investment Bolsa.

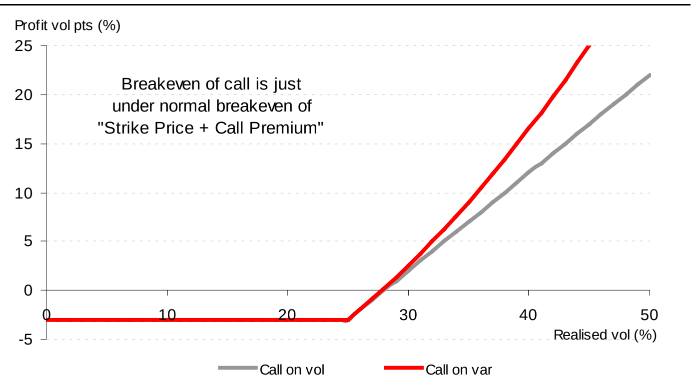
CONVEX PAYOUT MEANS BREAKEVENS ARE NON-TRIVIAL

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 39). 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 39. 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: $\mathrm{P} \times \mathrm{Vega} = (\sigma_{\text{Call Breakeven}}{}^2 - \sigma_K{}^2) \times \dfrac{\mathrm{Vega}}{2\sigma_S}$

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

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

OPTIONS ON VARIANCE HAVE POSITIVE SKEW

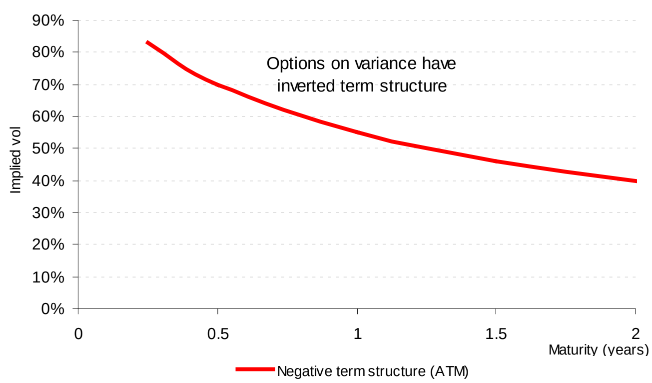
Volatility (and hence variance) is relatively stable when it is low, as calm markets tend to have low and stable volatility. Conversely, volatility is more unstable when it is high (as turbulent markets could get worse with higher volatility, or recover with lower volatility). For this reason, options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

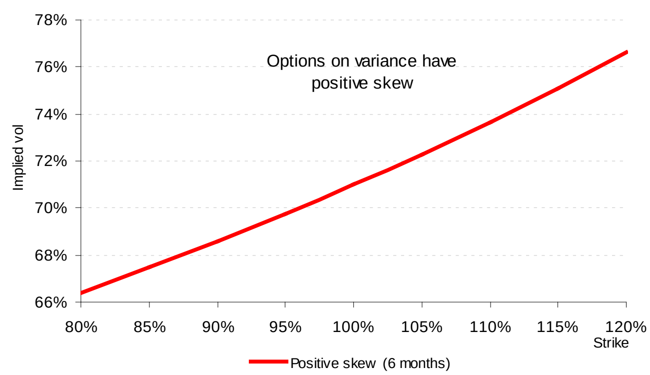
Skew and term structure of options on variance are opposite to options on equity

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 40. 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

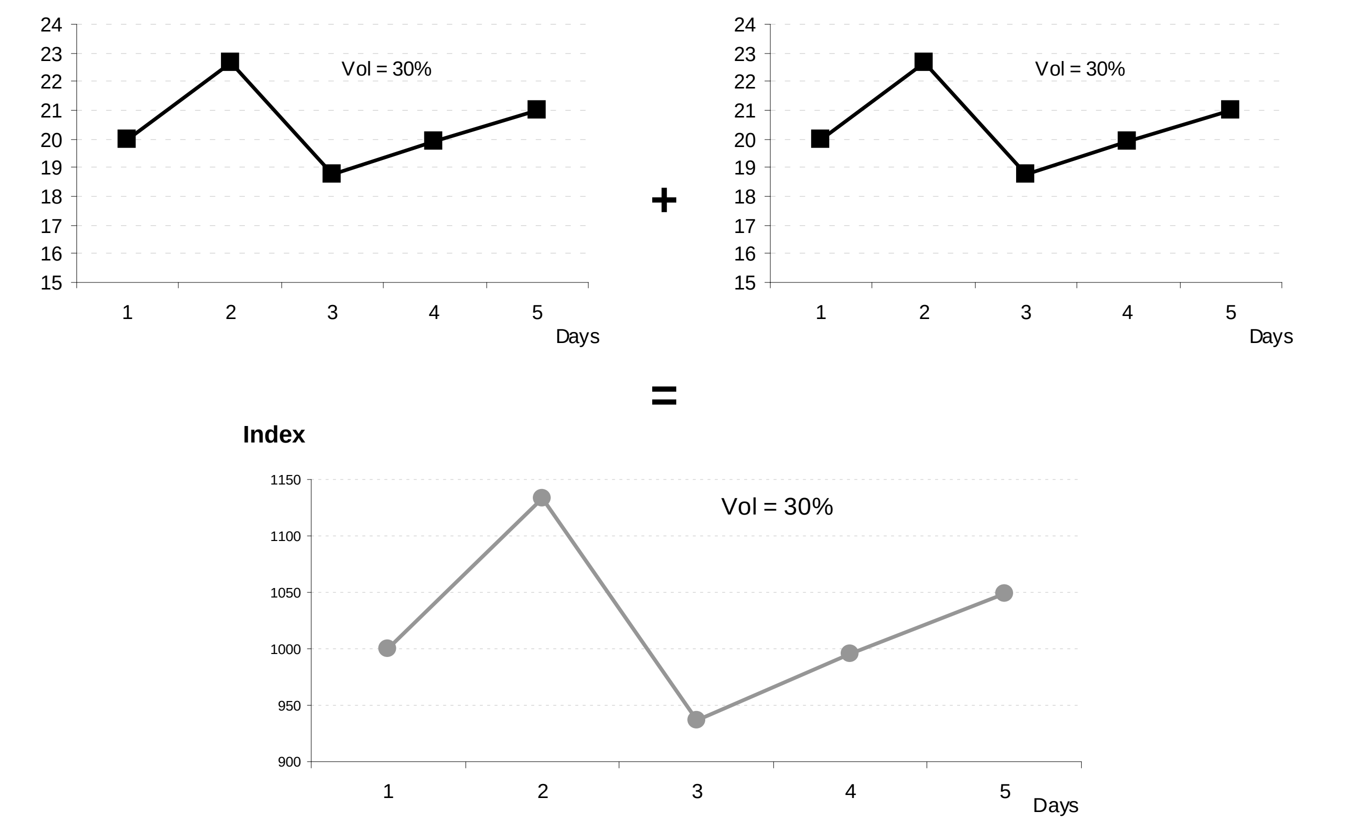
The volatility of an index is capped at the weighted average volatility of its constituents. Due to diversification (or less than 100% correlation), the volatility of indices tends to trade significantly less than its constituents. The flow from both institutions and structured products tends to put 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 41. Stock 1 **Stock 2 (100% correlation to stock 1)**

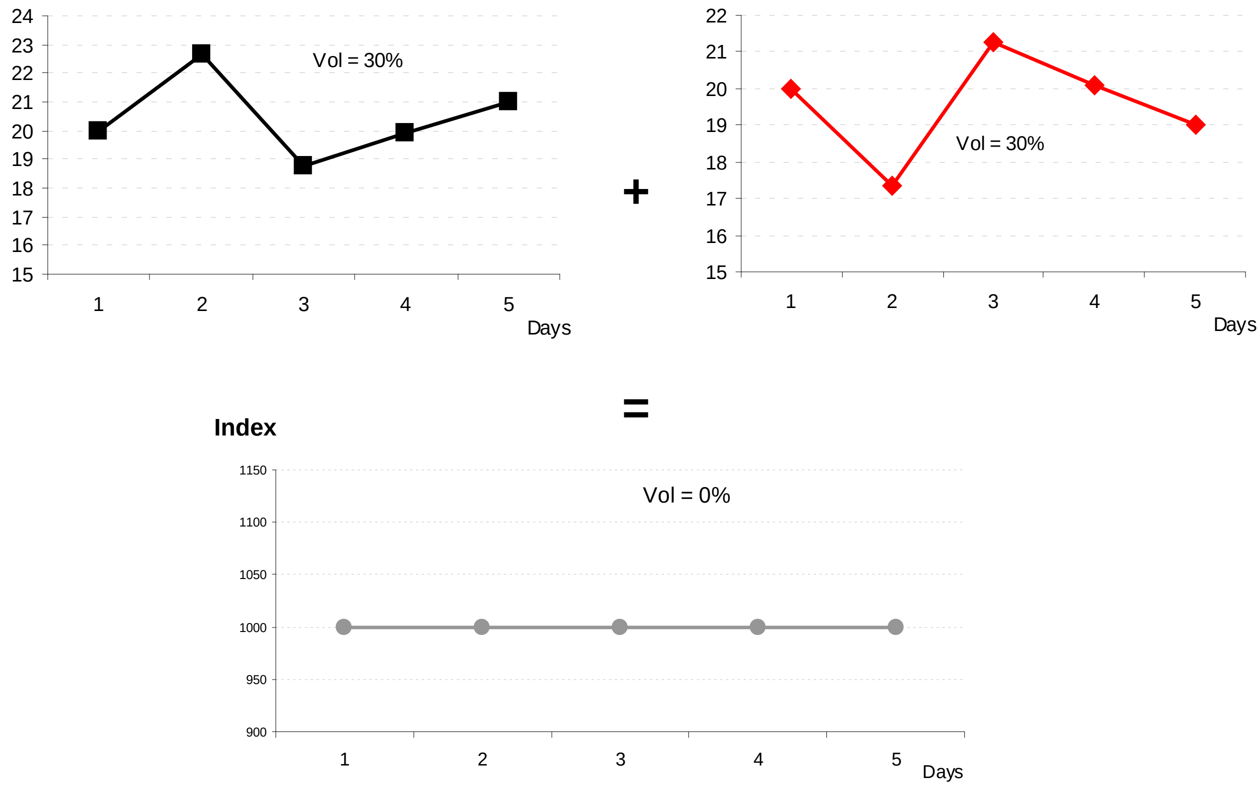


Source: Company data and Santander Investment Bolsa estimates.

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 42. 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.

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

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%, ρ≈64%.

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

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

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 43. 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)	$\dfrac{\sigma_I^{\,2} - \sum_{i=1}^{n} w_i^{\,2}\sigma_i^{\,2}}{\sum_{i=1,\, j\neq i}^{n} w_i w_j \sigma_i \sigma_j} \approx \dfrac{\sigma_I^{\,2}}{\sum_{i=1}^{n} w_i \sigma_i^{\,2}}$
c10pts above implied correlation (as not short volga)		
Correlation swap (market value weighted)	Pairwise realised correlation (market value weighted)	$\dfrac{\sum_{i=1,\, j>i}^{n} w_i w_j \rho_{ij}}{\sum_{i=1,\, j>i}^{n} w_i w_j}$
c5pts above market value weighted correlation		
Correlation swap (equal weighted)	Pairwise realised correlation (equal weighted)	$\dfrac{2}{n(n-1)}\rho_{ij}$

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 44. CBOE Implied Correlation Tickers and Expiries

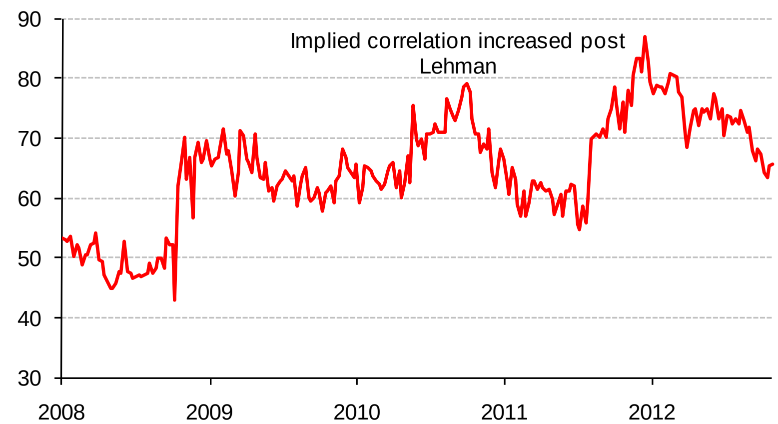
Expiry S&P500	Expiry Top 50 Stocks	Ticker	Start Date	End Date
Dec-09	Jan-10	ICJ	Nov-07	Nov-09
Dec-10	Jan-11	JCJ	Nov-08	Nov-10
Dec-11	Jan-12	KCJ	Nov-09	Nov-11
Dec-12	**Jan-13**	**ICJ**	**Nov-10**	**Nov-12**
Dec-13	**Jan-14**	**JCJ**	**Nov-11**	**Nov-13**
Dec-14	**Jan-15**	**KCJ**	**Nov-12**	**Nov-14**

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 45. 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

$(\rho_K - \rho) \times \text{Notional}$

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[8]. 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).

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

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.

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

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

Need to decide on weighting scheme for dispersion trades

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[9] 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 46. Greeks of Dispersion Trades with Different Weightings

Greeks	Theta-Weighted	Vega-Weighted	Dollar Gamma-Weighted
Theta	0	Short/pay	Very short/pay a lot
Vega	Short	0	Long
Gamma	Very short	Short	0
Ratio single stock vega to index vega	$\sigma_{Index} / \sigma_{Single\ Stock}$	1	$\sigma_{Single\ Stock} / \sigma_{Index}$
Total single-stock vega	**Less than index**	**Equal to index**	**More than index**

Source: Santander Investment Bolsa.

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

Figure 47. 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 47 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[10], which supports our view of vega-weighted dispersion being the best.

[10] 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[11].

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

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.

[11] 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[12], 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

[12] 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[13]. Because of this, attempts were made to create a covariance swap market, but liquidity never took off.

Correlation swap payoff

[Covariance(A,B) - $K_{covariance}$] × Notional

where

Notional= notional paid (or received) per covariance point

ρ = correlation between A and B

σ_i = volatility of i

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

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

[13] 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[14]. 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

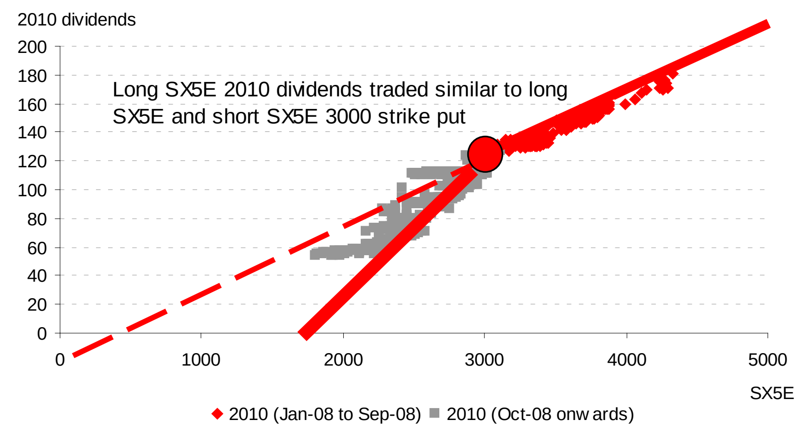
In 140 years of US data, there has never been a larger decline in index realised dividends than the index itself. This is because in a bear market certain sectors are affected more than others, and it is the companies in the worst affected sectors that cut dividends. For example, in the 2000-03 bear 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.

[14] 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 48 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 48. 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

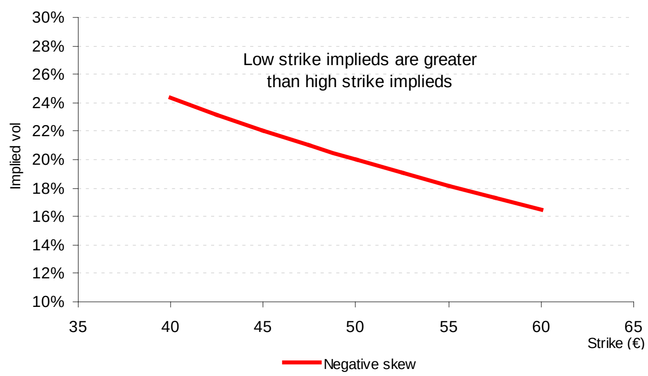
As it is not possible to hedge an option on dividend with realised dividend (they are not traded), the volatility of the underlying implied dividend is the key driver of an option on dividend's value. While realised dividends are less volatile than equities, implied dividends can be more volatile than 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).

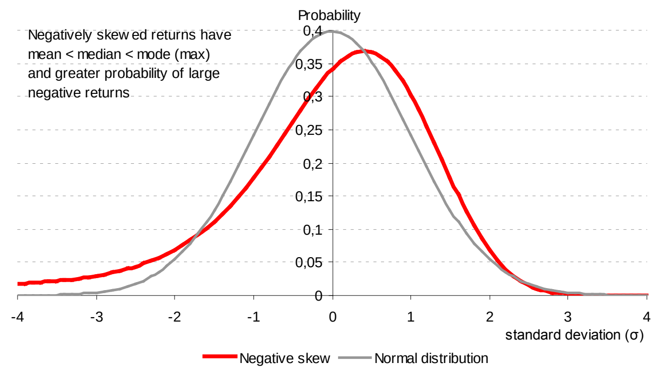
DIVIDENDS SHOULD HAVE HIGHER SKEW THAN EQUITIES

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.

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

Figure 49. Implied Volatility with Negative Skew **Probability Distribution of Negative Skew**





Source: Santander Investment Bolsa estimates.

This page left blank intentionally.

OPPORTUNITIES, IMBALANCES AND MYTHS

OVERPRICING OF VOL IS PARTLY AN ILLUSION

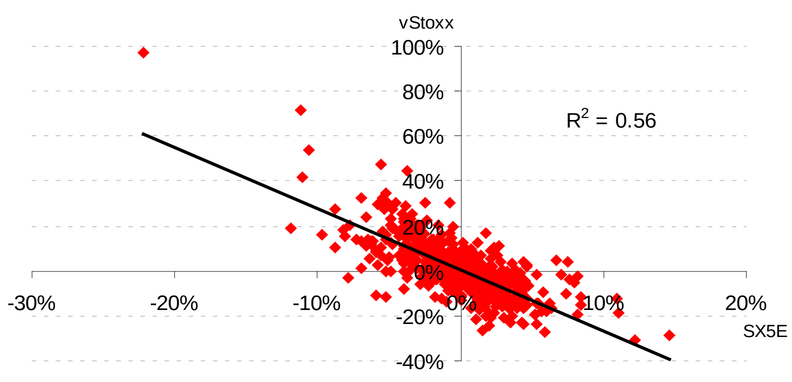
Selling implied volatility is one of the most popular trading strategies in equity derivatives. Empirical analysis shows that implied volatility or variance is, on average, overpriced. However, as volatility is negatively correlated to equity returns, a short volatility (or variance) position is implicitly 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 50. 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[15]. 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.

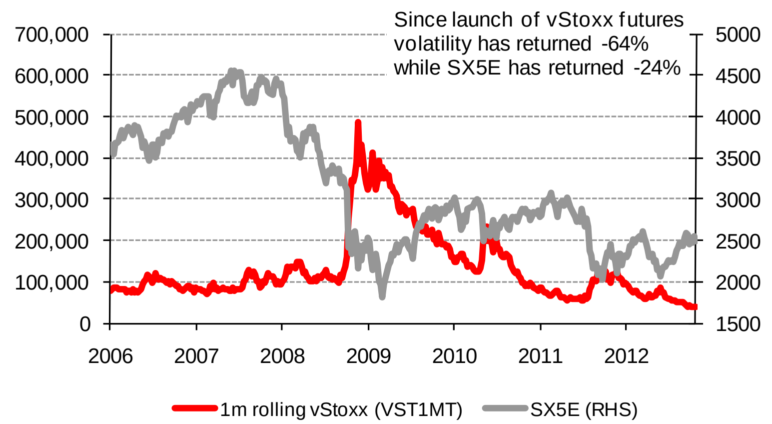
SHORT-DATED VOLATILITY INDEX FUTURES ARE A POOR HEDGE

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 51 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.*

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

Figure 51. 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:

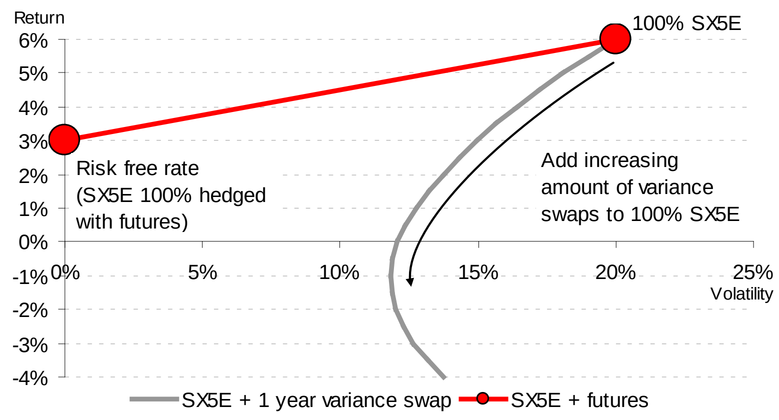
- **Correlation with equity market.** As equity markets are expected to return an equity risk premium over the risk-free rate, strategies that are implicitly long equity risk should similarly outperform (and strategies that are implicitly short equity risk should underperform). As a long volatility strategy 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 52 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 52. 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

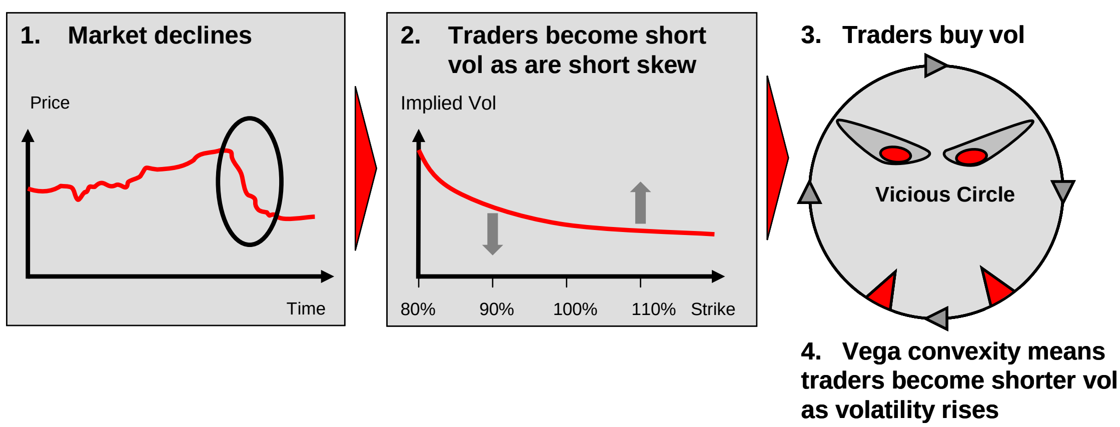
The sale of structured products leaves investment banks with a short skew position (eg, short an OTM put in order to provide capital-protected products). Whenever there is a large decline in equities, this short skew position causes the investment bank to be short volatility (eg, as the short OTM put 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[16]. 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 53 below.

Figure 53. Four Stages Towards Implied Volatility Overshoot



Source: Santander Investment Bolsa.

(1) EQUITY MARKET DECLINES

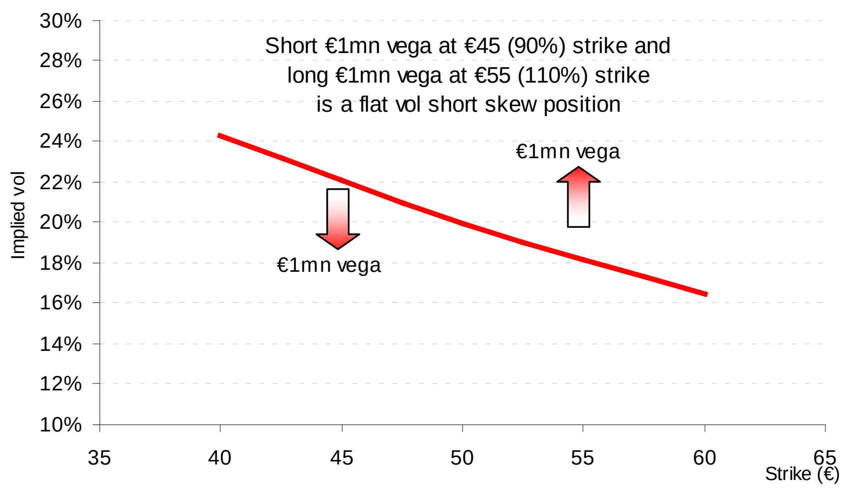
While implied volatility moves – in both directions – are exaggerated, for this example we shall assume that there is a decline in the markets and a rise in implied volatility. If this decline occurs within a short period of time, trading desks have less time to hedge positions, and imbalances in the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

[16] 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)[17].

Figure 54. Short Skew Position Due to 90%-110% Risk Reversal (initially vega flat)



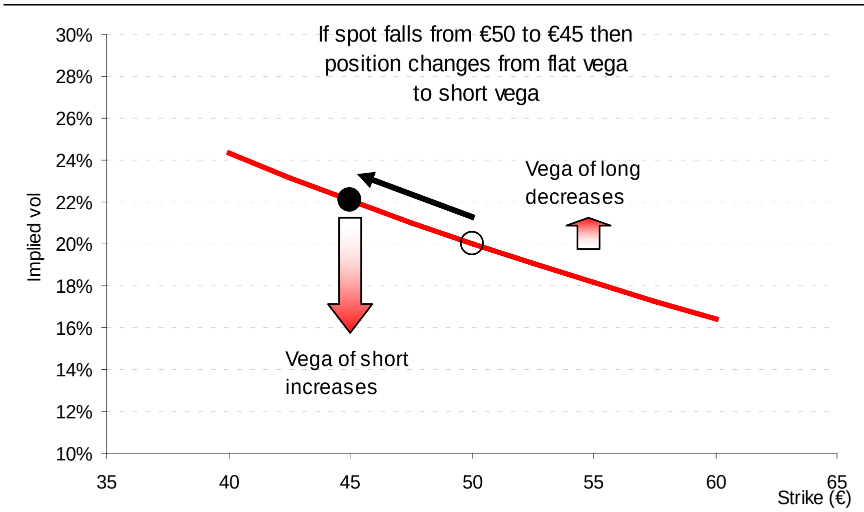
Source: Santander Investment Bolsa.

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



Source: Santander Investment Bolsa.

[17] 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

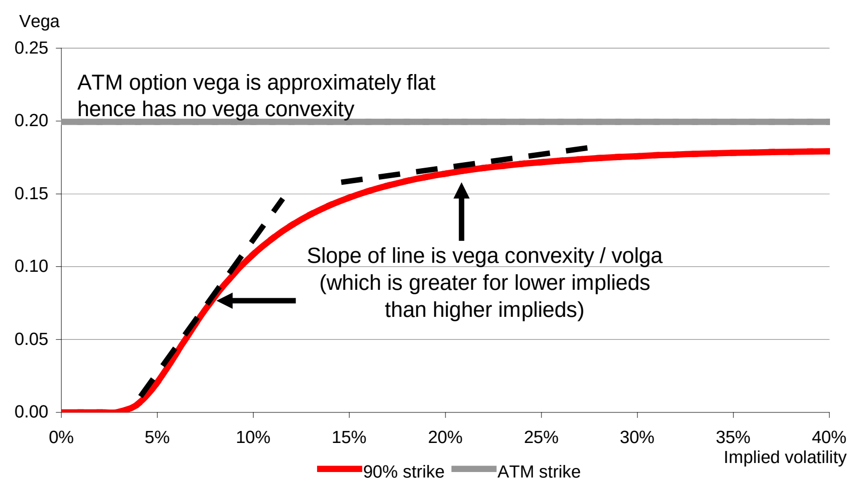
Vanna can be thought of as the size of the skew position (in a similar way that vega is the size of a volatility position), while skew (eg, 90%-100% skew) measures the value of skew (in a similar way that implied volatility measures the value of a volatility position). For more details on different Greeks, 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 56. 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 Figure 56). 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 56 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 VOLATILITY INDICES

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 futures 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)[18]. 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[19] during the forward starting period.

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 (forward starting options have a theoretical delta of zero). The rationale is similar to the argument that variance swaps have negative shadow delta due to skew.

[18] We shall assume for this section that the investor wishes to be long a forward starting product.
[19] 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.

FIXED DIVIDENDS ALSO CAUSES SHADOW DELTA

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).

Proportional dividends reduce volatility of underlying

Options, variance swaps and futures on volatility indices gain in value if dividends are fixed, as proportional dividends simply reduce the volatility of an underlying.

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:

(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) **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).

(3) **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.

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 straddles (or very OTM strangles if they are liquidity enough, as they also have zero delta and gamma). A long straddle has to be purchased on the expiry date of the option, while a short straddle has to be sold on the strike fixing date. As spot moves the strikes will need to be rolled, which increases costs (which are likely to be passed on to clients) and risks (unknown future volatility and skew) to the trader.

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

Similarly, the hedging of futures on volatility indices is not trivial, as (like volatility swaps) they require a volatility of volatility model. While the market for futures 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[20]. 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[21]). 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 57. 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.

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

[21] 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 58. 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 59. 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.

(2) 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 ($N = \text{vega} \div 2\sigma$). 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

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

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

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

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

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

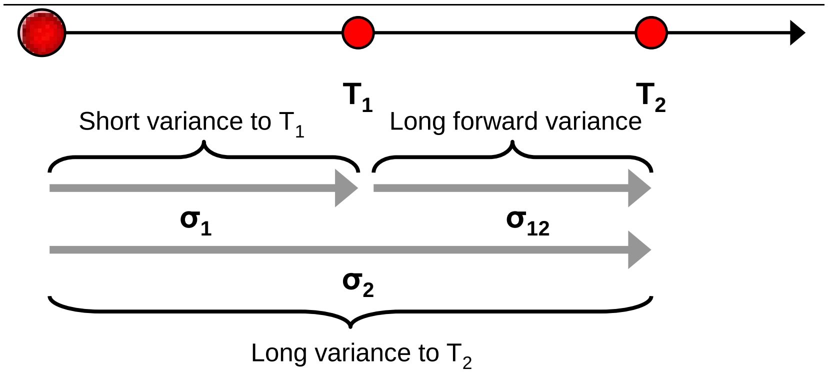
$$\Rightarrow \quad 0 < N_1 < N_2 \ \text{(as } T_1 > T_2\text{)}$$

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 60. 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).

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

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

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).

(3) FUTURE ON VOLATILITY INDEX

Futures on volatility indices have become one of the most popular forward starting products. For more details on both volatility indices, and futures on those indices, please see the following two sections.

VOLATILITY INDICES

While volatility indices were historically based on ATM implied, most providers have swapped to a variance swap-based calculation. The price of a volatility index will, however, typically be 0.2-0.7pts below the price of a variance swap of the same maturity as the calculation of the volatility index typically chops the tails to remove illiquid prices. Each volatility index provider has to use a different method of chopping the tails in order to avoid infringing the copyright of other providers.

THERE ARE TWO WAYS OF CALCULATING A VOLATILITY INDEX

Historically, volatility indices (old VIX and VDAX) were based on ATM implied volatility. This level is virtually identical to the fair price of a volatility swap (as volatility swaps ≈ ATMf implied). This methodology has the advantage that it uses the most liquid strikes, and it is still used by some providers in less liquid markets for this reason. Due to the realisation that variance, not volatility, was the correct measure of deviation, on September 22, 2003, the VIX index moved away from using ATM implied towards a variance-based calculation (and also moved from using the S&P100 to the S&P500). While the calculation is variance-based, the index is 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 swaps and volatility indices should always trade above ATM options).

Volatility indices based on ATM implied usually average eight different options

The old VIX, renamed VXO, took the implied volatility for the S&P100 strikes above and below spot for both calls and puts. As the first two-month expiries were used, the old index was calculated using eight implied volatility measures as 8 = 2 (strikes) × 2 (put/call) × 2 (expiry). Similarly, the VDAX index, which was based on DAX 45-day ATM-implied volatility, has been superseded by the V1X index, which, like the new VIX, uses a variance-based calculation.

MOST VOLATILITY INDICES NOW USE VARIANCE-BASED CALCULATIONS

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

Variance-based calculations have also been used by other volatility index providers. 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 (due to illiquidity of OTM options on the Mexican index). While the formula for a variance is a mathematical formula and hence not subject to copyright, if this formula is modified to exclude tails (eg, requiring a non-zero bid and/or offer price, excluding strikes too far away from spot, etc), then this calculation becomes proprietary and is subject to copyright. This is the reason why different volatility index providers have chosen different calculation methods.

DIFFERENCES BETWEEN VOLATILITY INDICES AND VARIANCE SWAPS

Differences between variance swap and volatility index calculation methods cause volatility indices to be lower than variance swaps

While the calculation of a volatility index might be based on a variance swap calculation, the price of a volatility index will typically be lower than that of a variance swap. The magnitude of the difference depends on the calculation itself, the number of strikes with available prices and the difference in width between strikes.

- **Excluding very high and very low strikes.** To increase the stability of the calculation, volatility indices exclude the implied volatility of options with very high or very low strikes. Given the importance of low strike implied volatility to variance swap pricing, chopping the wings of low strike implieds has a greater impact than removing high strike implieds, hence the level of a volatility index is below the variance swap price (typically between 0.2 and 0.7 volatility points).

- **Discrete sampling by using only listed strikes.** When pricing a variance swap, the value of a parameterised volatility surface is used. This surface is completely continuous and, hence, is not subject to errors due to using discrete data. As a volatility index has to rely on data from listed strikes, this introduces a small error which causes the level of the implied volatility index to be slightly below the variance swap price.

- **Noise due to rolling expiries.** If a volatility index does not interpolate between expiries then the implied volatility will ‘jump’ when the maturity rolls from one expiry to another. This difference can be c2 volatility points. Some indices only interpolate over a few days and take an exact maturity the rest of the time, which smoothes this effect (but does not fully remove it). Indices calculated by the CBOE move from interpolation to extrapolation which will cause similar noise, but has a much smaller effect than rolling. The average value from a volatility index that uses rolling is below the value of a variance swap as term structure is normally positive.

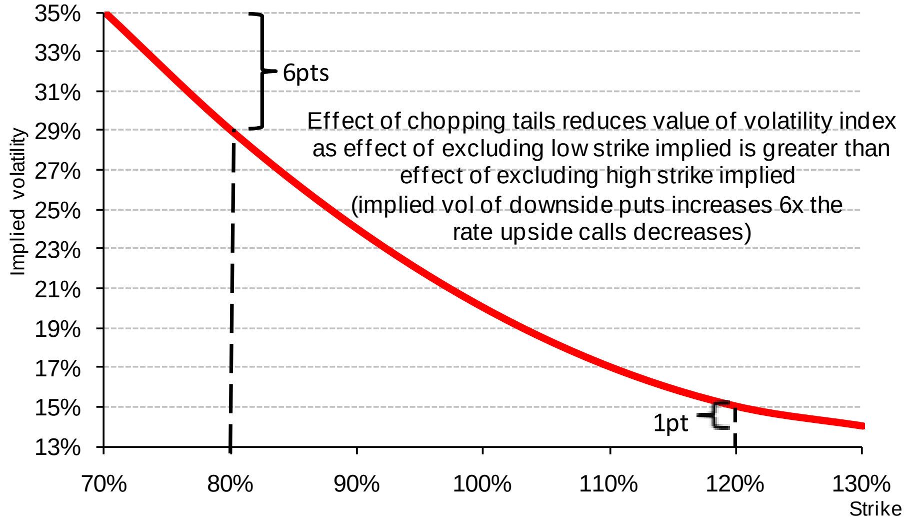
- **Linear interpolation between expiries (should be square root of time).** Linearly interpolating between expiries assumes a flat volatility term structure. In reality, a volatility surface follows a ‘square root of time’ rule, which means that the slope of term structure is steeper for near-dated maturities than for far dated ones. As a volatility surface is normally upward sloping this means a volatility index is on average below the level of a variance swap (up to c0.8 volatility points).

Excluding very high and very low strikes lowers the value of a volatility index

A volatility index can trade almost a volatility point below a variance swap

ATM options are the most liquid, as they have the most time value. For very OTM options, not only is liquidity typically poor but a small change in price can have a large effect on the implied volatility. To improve reliability of calculation, the very high and very low strikes are excluded. This is either done via a fixed rule (ie, only use strikes between 80% and 120%) or by insisting on a bid price above zero. Requiring the existence of both bid and offer prices implicitly chops the wings as well. Excluding the tails excludes high implied volatility low strike options; which causes the level of the volatility index to be below the fair price of variance swaps. The difference depends on the size of the tail that is chopped. If only strikes between 80% and 120% are used this can cause a discount of c0.7 volatility points between the 1-month volatility index and 1-month variance swaps. If all strikes with a liquid price are used then typically prices can be available for strikes between c60% and c120%, as downside puts are more liquid than upside calls (due to increased demand from hedging and due to the higher premium value given higher implied volatility). Using strikes between c60% and c120% has a small discount to variance swaps of c0.2 volatility points. The VIX requires a non-zero bid and an offer, and stops when two consecutive options have no price.

Figure 61. Chopping Tails of Volatility Surface Is More Important for Low Strikes than High Ones



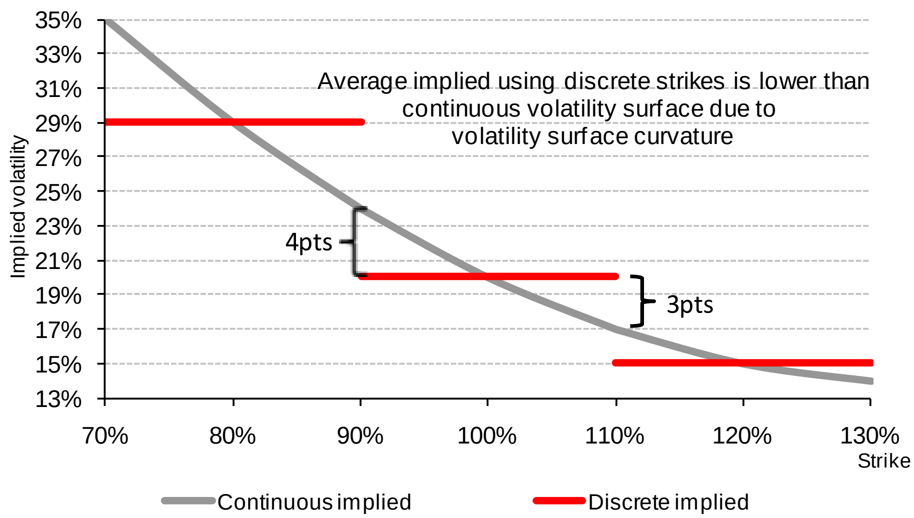
Source: Santander Investment Bolsa.

Discrete sampling by using only listed strikes also lowers the value of a volatility index

Volatility surface curvature causes a volatility index to be less than variance (due to chopping of tails and discrete sampling)

Even if prices were available for all strikes, a volatility index would give a slightly lower quote than variance swaps due to discretely sampling the implied volatility. The effect of discretely sampling a volatility surface can be modelled as a continuous volatility surface whose implied volatility is flat near the listed strikes (and jumps in between the listed strikes). Due to volatility surface curvature, this effect causes the value of a volatility index to be lower than a variance swap (as can be seen in Figure 62 below). The effect of discretely sampling depends on the number of strikes available (ie, the price difference between strikes), but is very small compared to the effect of chopping the tails (but both are caused by volatility surface curvature).

Figure 62. Discretely Sampling a Volatility Surface



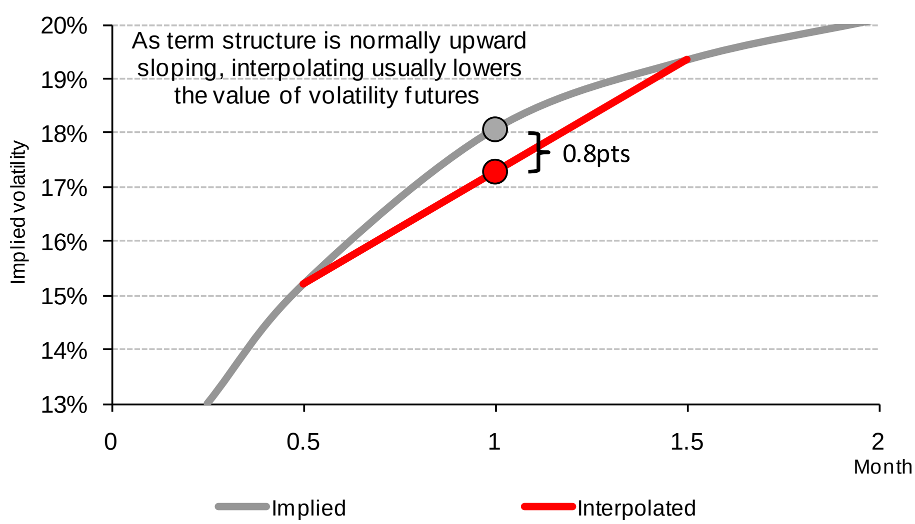
Source: Santander Investment Bolsa.

Noise due to rolling depends on the calculation method

Volatility futures settlement does not need to interpolate between expiries

There are volatility indices that instead of linearly interpolating between expiries roll from one maturity to the other. For similar reasons as to why linearly interpolating a volatility index usually gives a lower figure than variance swap (as there is a greater difference between 0.5-month and 1-month implied than between 1-month and 1.5-month implied), the average value for a volatility index that rolls is similarly too low. There will, however, be greater volatility for the index, due to the jump when the maturity rolls from one expiry to another. The difference between the implied volatility of the front two expiries can be c2 volatility points. Some indices smooth this effect by interpolating for a few days, while having a fixed un-interpolated value the rest of the time. Even a smoothed calculation will have a higher volatility over rolling than a fully interpolated based calculation. The CBOE ignores the front-month expiry in the final week before expiry and extrapolates from the second and third expiry. While this is a fully interpolated based calculation, jumping from using interpolation between the first and second expiry to extrapolation between the second and third can add some noise (but less noise than a roll-based calculation). We note that, as 30-day volatility futures expire exactly 30 days before the vanilla expiry, no interpolation by maturity is necessary for settlement.

Figure 63. Interpolating between Maturities



Source: Santander Investment Bolsa.

Interpolation between expiries usually lowers the value of a volatility index

A volatility index can be higher than a variance swap if term structure is negative

The slope of near-dated implieds is typically steeper than far-dated implieds, as volatility surfaces often move in a 'square root of time' manner (near-dated implieds fall more than far-dated implieds when volatility declines). Given the steeper slope of near-dated implieds, linearly interpolating underestimates the implied volatility for positive term structure (and similarly overestimates it for negative term structure). Typically, the demand for long-dated hedges and risk aversion causes far-dated implieds to be greater than near-dated implieds, hence term structure is normally positive. The effect of linearly interpolating between maturities for a fixed maturity volatility index therefore causes a volatility index to normally underestimate a variance swap level. This effect will be greatest when the (typically 1-month) maturity of the volatility index is exactly in between listed expiries, and an extreme example is given in the graph above where the difference is c0.8 volatility points. We note that should the maturity of a volatility index lie close to a listed maturity the error due to interpolating between expiries will be close to zero.

FUTURES ON VOLATILITY INDICES

While futures on volatility indices were first launched on the VIX in March 2004, it has only been since the more recent launch of structured products and options on volatility futures that liquidity has improved enough to be a viable method of trading volatility. As a volatility future payout is based on the square root of variance, the payout is linear in volatility not variance. The fair price of a future on a volatility index is between the forward volatility swap, and the square root of the forward variance swap. Volatility futures are, therefore, short vol of vol, just like volatility swaps. It is therefore possible to get the implied vol of vol from the listed price of volatility futures.

PRICE IS BETWEEN FORWARD VARIANCE AND FORWARD VOLATILITY

A future on a volatility index functions in exactly the same way as a future on an equity index. However, as volatility future 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.

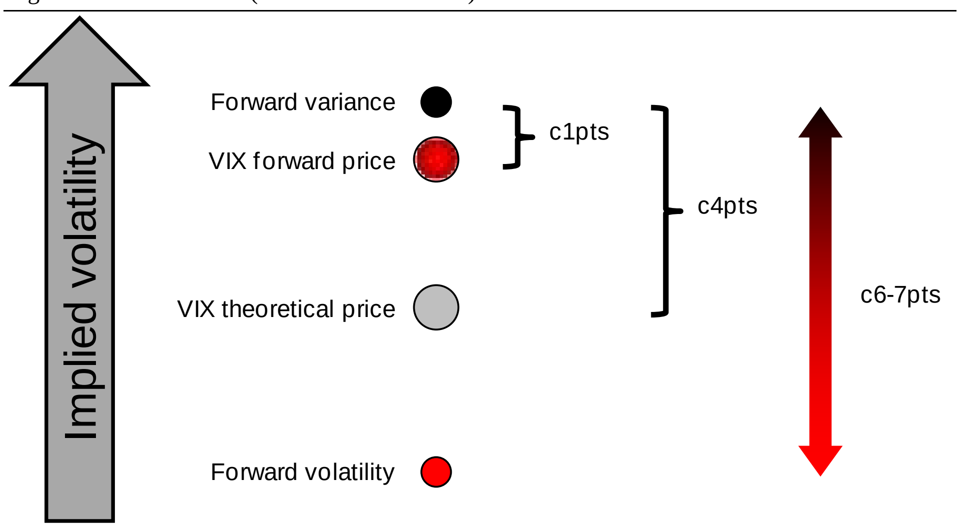
⇨ $\sigma_{\text{Forward volatility swap}} \leq$ Future on volatility index $\leq \sigma_{\text{Forward variance swap}}$

FUTURES ON VOLATILITY INDICES ARE SHORT VOL OF VOL

Prices of futures on volatility indices are linear in volatility (not variance)

A variance swap can be hedged by delta hedging a portfolio of options (the 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 model. As a variance swap is more convex than a volatility swap (variance swap payout is on volatility squared), a volatility swap is short convexity compared to a variance swap. A volatility swap is, therefore, short volatility of volatility (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).

Figure 64. Theoretical (stochastic local vol) and Actual Prices of 6-Month VIX Futures



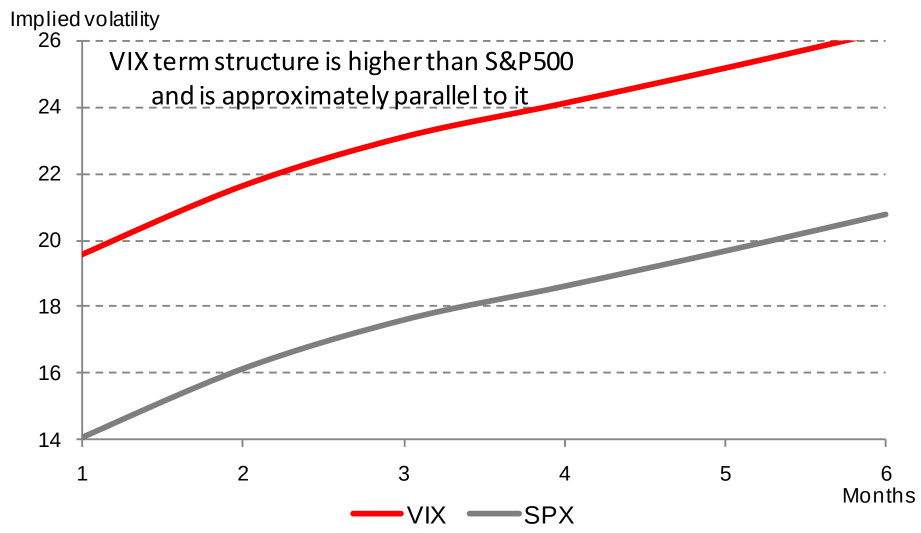
Source: Santander Investment Bolsa.

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

Being short volatility futures and long forward variance is a popular trade

While the price of volatility futures should be well 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. Using a stochastic local volatility model, we found that VIX futures[22] should trade roughly half way between a variance future and a volatility future (in fact, slightly closer to forward volatility, as is to be expected for a product linear in volatility). While this means VIX futures should be c4pts below forward variance, they appear to only trade c1pt below. Similarly, VIX futures should be only c2-3pts above ATMf implied volatility, but during 2012 they were c5-6pts above ATMf implied (see Figure 65 below). 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.

Figure 65. VIX and S&P500 Average Term Structures During 2012



Source: Santander Investment Bolsa.

VIX and S&P500 term structures are roughly parallel to each other

We note that while the VIX futures term structure lies above S&P500 term structure, they are approximately parallel to each other for the same reason variance term structure is parallel to ATMf term structure. While variance swaps are long skew and skew is lower for far-dated implieds, as OTM options gain more time value as maturity increases, these effects cancel each other out (hence, in the absence of supply and demand imbalances, variance and implied volatility term structure should be roughly parallel to each other). As equities typically have an upward sloping term structure, volatility futures term structure is typically upward sloping as well. Volatility futures, like variance swaps, are also long volatility surface curvature. Volatility futures will have the same seasonality of vol as the underlying security (eg, dips over Christmas and year-end).

VOL OF VOL CAN BE BACKED OUT FROM VOLATILITY FUTURE PRICES

A forward on a volatility future is short vol of vol. This means it is possible to back out the implied vol of vol from the price of this volatility future. This implied vol of vol can be used to price options on variance or even options on volatility futures themselves[23].

[22] We assume a volatility index calculation matches that of a variance swap, ie, no chopping of tails.
[23] Assuming the volatility of volatility is log normally distributed.

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. More recently, futures based on the Russell 2000 traded from 2007 until their delisting in 2010.

Figure 66. Volatility Indices with Listed Futures

Ticker	Underlying
V2X	SX5E
VIX	SPX
VXD	DJIA
VXN	NDX
VXTH	VIX Tail Hedge (long SPX and long 30 delta VIX calls)

Source: Santander Investment Bolsa.

VOLATILITY FUTURES EXPIRE ON THIRD OR FOURTH WEDNESDAY

Normal options expiry is on the third Friday of the month (but not always; for example, Nikkei uses the second Friday), to be close to the end of the month and on a day that does not usually have a bank holiday (sometimes, however, Good Friday will fall on the third Friday). For US markets, the expiration is on the Saturday after the third Friday (to give more time for the administration of expiration); however, as the last trade date is the Thursday before (expiration is based on opening prices on the Friday), this is irrelevant for cash-settled derivatives such as volatility futures. We note the extra time value due to expiry being on Saturday rather than Friday could be relevant for physical delivery, as an investor could use the performance of different markets to estimate the weekend movements and hence the likely opening price on Monday. As the only volatility futures currently listed are for volatility indices whose maturity is 30 days, the expiry of volatility futures is either the third or fourth Wednesday of the previous month (so the 30-day VIX calculation for settlement price is based on only one maturity, rather than an interpolation between two maturities). This ensures that the settlement price is not subject to interpolation errors (see previous section) that affect volatility indices on other days.

VIX SETTLEMENT PRICE CALCULATION COULD BE MANIPULATED

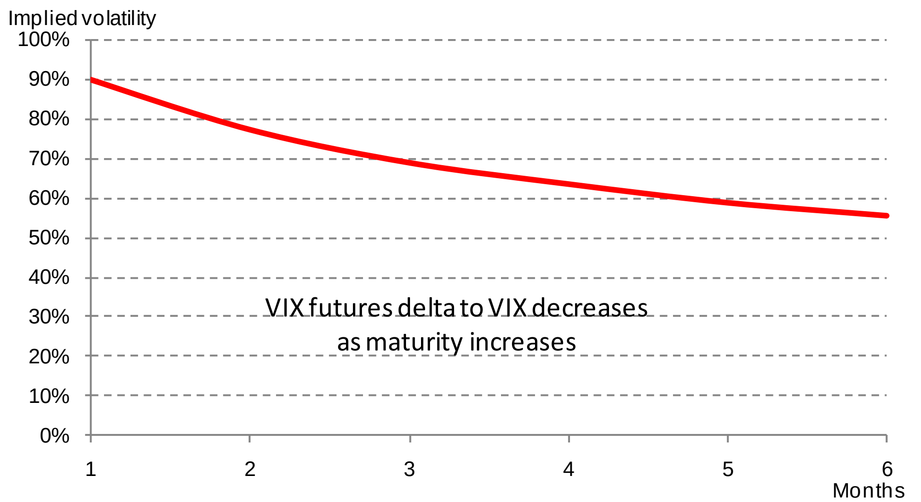
VIX settlement could be manipulated, but vStoxx calculation is robust

As the settlement price for the VIX is based on opening trades on S&P500 options, it is far easier to manipulate than the settlement price for the vStoxx, which is based on a 30-minute average ending midday (CET). As shown earlier, the payout of a variance swap is based on a portfolio of options of all strikes weighted $1/K^2$, where K is the strike of the option. This means the calculation of variance swaps is very sensitive to the price of downside puts. Typically, there are offers for downside puts of all strikes at the tick value (ie, the smallest possible non-zero price), as these puts have near zero theoretical value. As the VIX calculation requires a non-zero bid, these offers are usually excluded for strikes below c50%-60%. By entering the minimum bid of US$0.05 (= tick value), these prices will be included in the calculation and could lift the settlement price by c1pt (as the implied volatility for these low strike puts will be very large). There have been times when the VIX settlement (on the open) has been significantly different to both the close of the day, and the close of the previous day.

MEAN REVERSION MEANS VOLATILITY FUTURES HAVE DELTA <1

Unlike normal futures, volatility futures are not linear in the underlying index (as the mean reversion of volatility has an effect) and, hence, have deltas significantly lower than 100%. An equity future has near a 100% delta. While the front month VIX future has a high 90% delta (delta vs the VIX), the 6-month VIX future has a lower 55% delta. The lower delta is due to the mean reversion of volatility, as 6-month VIX futures will not trade at 80% even if the VIX trades at 80% (as the VIX only briefly went above 80% post Lehman bankruptcy and swiftly declined, it is highly unlikely to still be at 80% in 6 months' time). The empirical deltas of VIX futures by maturity are shown in Figure 67 below. These values decline in a similar way but less rapidly than they would if volatility solely obeyed a square root of time rule, which is to be expected as volatility surfaces sometimes move in parallel.

Figure 67. VIX Futures Delta to VIX by Maturity, 2004-12



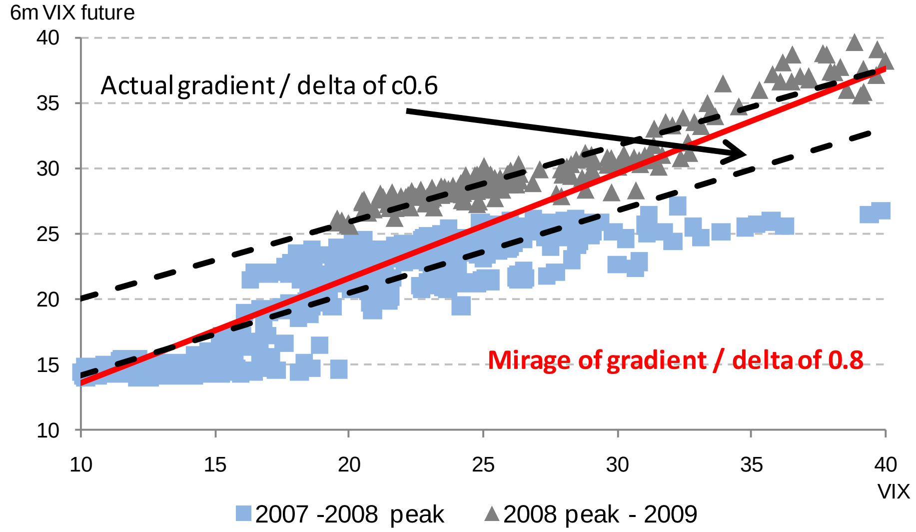
Source: Santander Investment Bolsa.

VIX futures delta is both time and volatility dependent

Delta of volatility futures declines for high levels of volatility

While Figure 67 above gives the average delta of VIX futures over a 9-year period, over a shorter time period the delta can be significantly different. For values of the VIX above 40% the delta of a 6-month VIX future can be close to zero, as the market does not expect the high 40+% volatility environment to last six months as it is likely to be a temporary spike. For values of the VIX below 30%-40%, there is a far higher delta, as moves in volatility within this range are more likely to be part of a volatility regime change that could be longer lasting. When estimating the delta of a volatility future care has to be taken when choosing the period of time used to find the estimate. Figure 68 below shows how the delta for a 6-month VIX future over the 3-year period 2007-09 could be estimated to be 80%, however, in fact, the delta is actually c60% as there was a regime shift assuming higher future volatility post the 2008 peak in volatility post the Lehman bankruptcy. Note also how in the lead-up to the 2008 peak in volatility the 6-month VIX future appeared capped in the high 20s (and had near zero delta for values of VIX above 30%), whereas afterwards no such cap existed (and had high delta for value of VIX above 30%).

Figure 68. VIX 6-Month Future Sensitivity (Delta) to VIX, 2007-09



Source: Santander Investment Bolsa.

R^2 of VIX futures declines as maturity increases

The length of time to mean revert, and the level of volatility to which the VIX will mean revert, changes over time. Hence the R^2 between the VIX and a VIX future decreases as maturity increases. While the front-month VIX future has a high R^2 of 0.97, the 6-month VIX future has a lower R^2 of 0.70.