Index variance arbitrage

Strategic Research for Equity Derivatives

Arbitraging component correlation

■ This study focuses on arbitrage strategies involving the variance of an index against the variances of its components. Unlike classic index arbitrage, where risks can be eliminated, index variance arbitrage always leaves a residual risk.

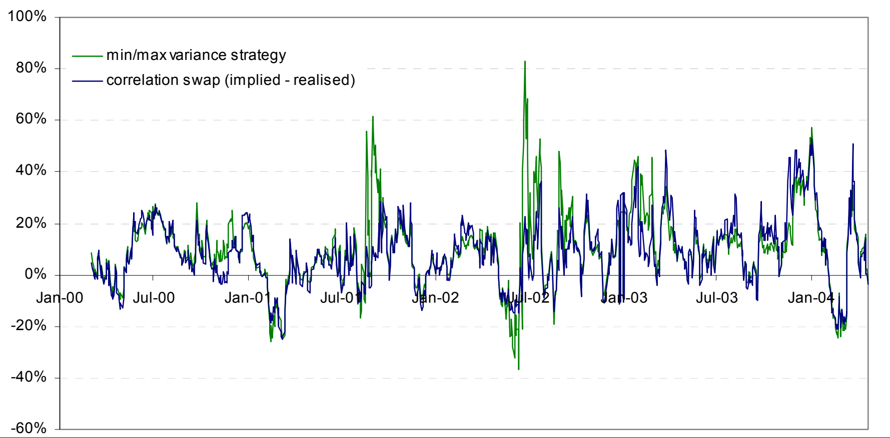
■ There are many ways to devise such arbitrage strategies. It is possible, for instance, to replicate the theoretical concepts of covariance or dispersion. Practically speaking, there is a high degree of freedom in choosing the hedge ratios.

■ In attempting to identify the strategies that are best suited for arbitrage purposes, we analyse them according to various criteria: differentiation from outright variance positions, predictability of returns, minimal residual risk. We notice that neither covariance nor dispersion meets these objectives.

■ Correlation, on the other hand, would be a much more interesting candidate in this respect. But correlation cannot be replicated through static positions in the index and components' variances. However, we find that minimising the risks of index variance arbitrage generate a residual that is in fact a good proxy for correlation exposure.

■ We explore different approaches to optimising index variance arbitrage (or, equivalently, correlation swap replication). An analytical method provides a first order hedge but diverges during volatility crises. A statistical approach, based on minimising historical residual variance, eliminates much more of the volatility risk but is complex to set up. A final analytical methodology, which we call min/max, provides the best compromise between accuracy and ease of implementation. It is, in our view, the most efficient way to set up a strategy of index variance arbitrage.

Correlation swap and the min/max proxy



Source – Strategic Research for Equity Derivatives

Nicolas Blanc
Nicolas.Blanc@bnpparibas.com
+33 (0) 1 40 14 76 75

Introduction

Trading variance is a technique that was developed thanks to the tools offered by the option markets. Initially, trading variance was done through hedging single options. This type of position, however, creates many residual risks that add to, and possibly even offset, the desired variance exposure.

More recently, variance swap contracts, pure exposure on realised variance, were introduced. Their rapid development relies on the mathematical possibility of replicating future realised variance with a static portfolio of options, thereby eliminating the drawbacks of hedging single options. Variance swaps offer numerous applications, including: directional positions, risk management and arbitrage.

This study focuses on a particular type of arbitrage strategy for which variance swaps are a building block: arbitrage between the realised variance of a basket (a stock index) and that of its components.

When we seek to extend the classic concept of index arbitrage (cash & carry) to non-linear (non delta-one) derivatives, we face the problem that risk free arbitrage in general does not exist. If we consider vanilla options, for instance, we know that an option on a basket can never be statically decomposed into a basket of options on the single components.

Variance also falls into that category: arbitrage between the realised variance of a basket and the realised variances of its components can never be risk-free. However, depending on the way the arbitrage is set up, its results create another kind of derivative: exposure on the 'co-variations' of the index components. In particular, mathematical concepts such as covariance and dispersion may be replicated through fixed variance positions.

Among the various possibilities in index variance arbitrage, both in terms of the mathematical concepts and in terms of the variety of hedge ratios (one per component), we try and identify the strategies that isolate a pure co-variation effect, as opposed to being affected by other factors, like volatility. More precisely, we would like the arbitrage strategy to be as 'volatility neutral' as possible, just as option hedging seeks market neutrality.

Average correlation is the concept that best meets this objective. In contrast, covariance and dispersion are greatly influenced by global volatility levels. When we deal with large indexes, for instance, playing covariance is not different from simply playing index variance.

Another approach to optimising the strategy is to minimise risk. We observe empirically (using the example of the EuroStoxx 50) that the strategies with minimal risk are also those that leave a residual close to correlation exposure. Correlation is thus the 'natural' concept underlying index variance arbitrage.

On the other hand, correlation is not replicable through fixed positions on variance swaps. We thus explore different ways through which correlation may be approximated by variance swaps and come up with an analytical solution that, in our view, is the best compromise between implementation costs and reliability.

The mathematical framework

In this section, we define and discuss the basic mathematical concepts arising from arbitrage between the variance of an index and the variances of its components.

Index weights are assumed constant
In the following, we consider an index I containing n stock components denoted 1,…,n. To make calculations tractable, we assume that the index returns $R_I(t)$ result from the components' returns $R_i(t)$ according to weights $w_1,\ldots,w_n$ that are **constant in time**. This assumption can be expressed formally as follows:

At any date t, $R_I(t) = \sum_{i=1}^{n} w_i R_i(t)$

Real indexes do not satisfy this assumption, because they generally reflect the value of a portfolio composed of fixed amounts of the component shares. As a result, for example, when the price of a component rises relative to the others, its weight also rises.

However, if we consider a strategy whose duration is limited in time, the relative variations of the components will be bounded and the actual weights may be considered fixed. In practice, in our simulations, the maturity of the strategies is only six weeks, so the fixed weights assumption is acceptable.

Variance, realised variance, implied variance

Throughout the study, we focus on strategies involving the variance of asset returns. Normally, returns are combined simply linearly in portfolios and variance is introduced as a risk measure for the asset returns. Here, variance is the final payoff of financial products or strategies. In the next chapter, we will discuss the techniques on which these new products are based. They belong to a "second generation" of derivatives because their replication is achieved by combining vanilla option positions and underlying asset positions.

Assuming that R_S, the return of a security S over a period Δt, is modelled as a random variable, we use the following definition for its annualised variance (boldface letters indicate the mathematical concepts, as opposed to historical or implied levels, E is the mathematical expectation):

$$\boldsymbol{\sigma}_S^2 = \frac{1}{\Delta t}\mathbf{E}\left[\mathbf{R}_S^2\right]$$

Note that we have eliminated the average return from the usual mathematical formula, for various reasons:

- the re-hedging period is assumed to be short enough for the trend term to be negligible in comparison to the volatility term
- the definition better fits the payoffs achievable through option hedging
- it is the definition usually applied in variance swap contracts

Returns may be defined as linear returns or log returns. We have chosen linear returns in our simulations.

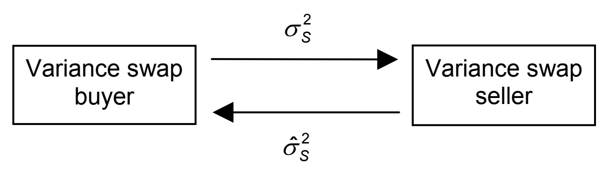
The annualised realised variance of a security S is calculated from the historical returns of the security as (the ^ sign denoting realised values):

$$\hat{\sigma}_S^2 = \frac{1}{\Delta t}\sum_{t=1}^{H}\frac{1}{H}\hat{R}_S^2(t)$$

A realised level is not different in nature from the mathematical concept (when the probability universe is the realised values and the probability levels 1/H). In some instances, the weights are chosen at 1/(H-1), which has a negligible impact in practice.

As mentioned above, the realised variance of asset returns may be the object of a swap contract through which realised variance is exchanged for a fixed level, the variance strike. If the market value of the swap is zero, the variance strike is referred to as implied variance and denoted σ_S^2 (simple letters will denote implied levels in the following).

Cash flows of a variance swap (payable at maturity)



Source – Strategic Research for Equity Derivatives

The advantages of using variance of returns as the basis for a financial product include the following:

- Realised variance forms a new “asset class”, loosely correlated to the underlying asset
- Variance is a promising variable for arbitrage, either for relative value or for dispersion trades, the object of this study
- Variance is a risk-management tool for option traders, as option books are generally exposed to realised variance
- Variance swaps are replicable through a static set of options together with a continuous hedge and therefore they enjoy good levels of pricing and liquidity
- Variance swap returns do not incur the path dependency risk present in option hedging, the other vehicle available for variance exposure

Covariance

The covariance of two asset returns is a measure of what impact combining assets in a portfolio may have on variance, as opposed to their isolated contributions.

We define the annualised covariance of two asset returns, viewed as random variables, in a similar way as for defining variance:

$$\boldsymbol{\Gamma}_{S,T} = \frac{1}{\Delta t}\frac{1}{2}\left(\mathbf{E}\left[(\mathrm{R_S} + \mathrm{R_T})^2\right] - \mathbf{E}\left[\mathrm{R_S^2}\right] - \mathbf{E}\left[\mathrm{R_T^2}\right]\right) = \frac{1}{\Delta t}\mathbf{E}\left[\mathrm{R_S R_T}\right]$$

Combining assets in a portfolio   Isolated contributions

Realised covariance is defined accordingly.

Index variance decomposition

Using our assumptions for computing index returns, there is a simple relationship between the variance of the index and the variance and covariance of the components' returns:

$$\boldsymbol{\sigma}_I^2 = \frac{1}{\Delta t}\mathbf{E}\left[\mathbf{R}_I^2\right] = \frac{1}{\Delta t}\mathbf{E}\left[\left(\sum_{i=1}^{n} w_i \mathbf{R}_i\right)^2\right] = \sum_{i=1}^{n} w_i^2 \boldsymbol{\sigma}_i^2 + \sum_{i \neq j} w_i w_j \boldsymbol{\Gamma}_{i,j}$$

The same formula applies to realised levels. We can also summarise all the covariance terms using a unique average, the index covariance Γ, such that:

$$\boldsymbol{\sigma}_I^2 = \sum_{i=1}^{n} w_i^2 \boldsymbol{\sigma}_i^2 + \sum_{i \neq j} w_i w_j \boldsymbol{\Gamma} \quad \text{or} \quad \boldsymbol{\Gamma} = \frac{\boldsymbol{\sigma}_I^2 - \sum_{i=1}^{n} w_i^2 \boldsymbol{\sigma}_i^2}{\sum_{i \neq j} w_i w_j}$$

The index covariance is the average covariance of pairs of components, weighted by their weight products in the index. It is a linear function of the index and the components' variances.

We observe that the definition of the index covariance may be extended to an implied covariance, by replacing mathematical variance by implied variance. We can then write:

$$\hat{\Gamma} - \Gamma = \frac{1}{1 - \sum_{i=1}^{n} w_i^2}\left(\hat{\sigma}_I^2 - \sigma_I^2\right) - \sum_{i=1}^{n} \frac{w_i^2}{1 - \sum_{i=1}^{n} w_i^2}\left(\hat{\sigma}_i^2 - \sigma_i^2\right)$$

Index realised covariance is thus fully replicable through positions on the variance of the index and of its components. Its fair value is the implied covariance. A long covariance exposure is a long position on the index variance hedged with short positions on the components' variances with weights equal to w_i^2.

Dispersion

The dispersion of returns within the index components may be defined as the annualised average of the cross-sectional variance of returns (i.e. the variance of returns, not in time, but among components, at a specific date):

$$\mathbf{D}_I^2 = \frac{1}{\Delta t}\mathbf{E}\left[\sum_{i=1}^{n} w_i\left(\mathbf{R}_i - \mathbf{R}_I\right)^2\right]$$

It demonstrates how, on quadratic distance, component returns differ from the index return. Dispersion is nil when all component returns are the same. We easily get:

$$\mathbf{D}_I^2 = \sum_{i=1}^{n} w_i \boldsymbol{\sigma}_i^2 - \boldsymbol{\sigma}_I^2$$

It is thus a linear function of the index's and components' variances. Using the same definition for implied and realised levels, we get:

$$\hat{D}_I^2 - D_I^2 = -\left(\hat{\sigma}_I^2 - \sigma_I^2\right) + \sum_{i=1}^{n} w_i\left(\hat{\sigma}_i^2 - \sigma_I^2\right)$$

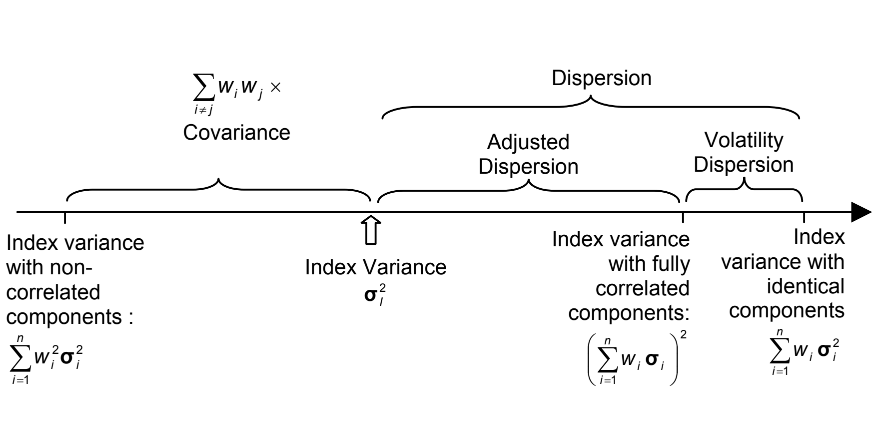
Realised dispersion is thus fully replicable through positions on the variance of the index and the components. Its fair value is the implied dispersion. A long dispersion exposure is a short position on the index variance hedged with long positions on the components' variances with weights equal to w_i.

Given their definition, we would expect covariance and dispersion exposure to yield opposite results. This is in fact not always the case, for reasons that we will develop further later on.

Covariance, dispersion and index variance

Covariance and dispersion are indicators that are naturally linked to index variance arbitrage. They indicate the position of the index variance relative to the situations where components are: 1/ non-correlated, on the one hand, or 2/ identical, on the other.

Index variance, covariance and dispersion



Source – Strategic Research for Equity Derivatives

Covariance, save for a weighting term, is the distance between the index variance and what it would be if components were non-correlated. The weighting term is a constant and is close to 1 when we deal with sufficiently diversified indexes.

Dispersion, on the other hand, may be split into two components:

- A term representing simply the dispersion of component variances. That term is nil whenever all component volatilities are the same, even if component returns are independent. That term may be written as so:

$$\sum_{i=1}^{n} w_i \boldsymbol{\sigma}_i^2 - \left(\sum_{i=1}^{n} w_i \boldsymbol{\sigma}_i\right)^2 = \sum_{i=1}^{n} w_i \left(\boldsymbol{\sigma}_i - \sum_{j=1}^{n} w_j \boldsymbol{\sigma}_j\right)^2 \geq 0$$

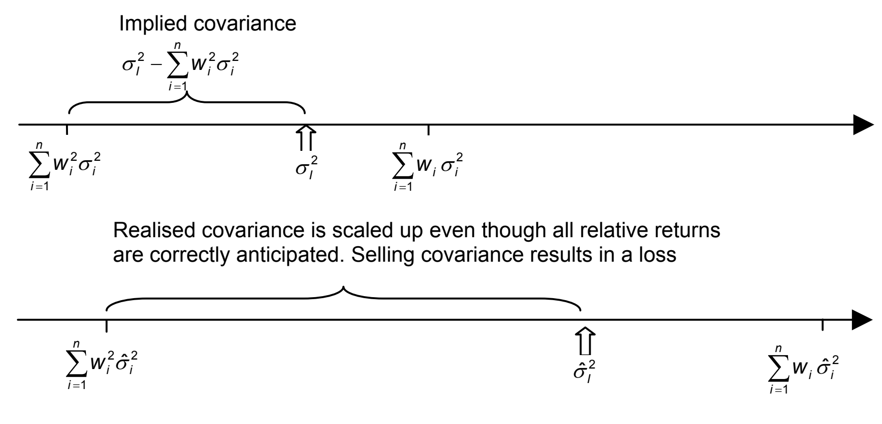
- A term corresponding to the difference between the index variance with components fully correlated and the actual index variance. We refer to it as Adjusted Dispersion. It is always positive and reaches 0 whenever the components are fully correlated.

$$\left(\sum_{i=1}^{n} w_i \boldsymbol{\sigma}_i\right)^2 - \boldsymbol{\sigma}_I^2 \geq 0$$

Adjusted dispersion may not be perfectly replicated through variance swap positions. However, we can create an 'over-hedge' for the adjusted dispersion swap (see appendix for details):

We can see that covariance and dispersion measure an absolute difference in variances. The general level of realised variance will thus impact them and not only the way returns are linked together. For example, if returns are all affected by the same scaling factor (as that anticipated in the implied levels), we may be in the following situation:

Covariance is sensitive to general volatility levels



Source – Strategic Research for Equity Derivatives

In this case, selling covariance would result in a loss, even though returns were perfectly anticipated. Selling dispersion would also result in a loss.

This justifies the use of an indicator that would not be influenced by a global volatility shift but would reflect only the position of the index variance in relative terms.

Correlation

The correlation coefficient is introduced to resolve the difficulties mentioned above. Between two random variables, it is defined as:

$$\boldsymbol{\rho}_{S,T} = \frac{\boldsymbol{\Gamma}_{S,T}}{\boldsymbol{\sigma}_S \boldsymbol{\sigma}_T}$$

The correlation coefficient of two random variables always lies within –1 and +1. It is 0 when the two variables are independent (but not only in this case) and is equal to +1 or –1 when there is a linear relation between S and T ($R_S = \alpha R_T$).

The correlation coefficient is unchanged if a scaling factor is applied to any of the variables.

Average correlation within an index

From the definition of the correlation coefficient introduced above, we note that we may break down the variance of an index further into:

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

The average correlation within the index, ρ, is defined so that it is equivalent, in terms of index variance, to the actual individual pairwise correlation coefficients:

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

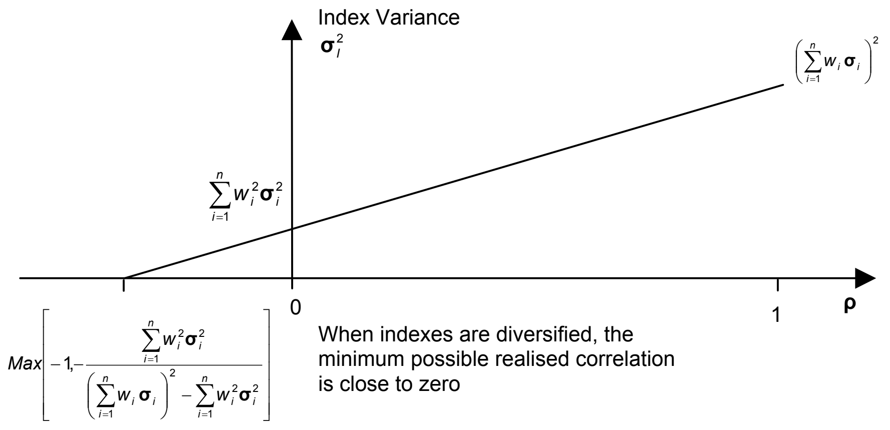
That definition also holds for realised levels. There are alternate expressions for the average correlation within an index:

$$\rho = \frac{\sigma_I^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2}{\left(\sum_{i=1}^{n} w_i \sigma_i\right)^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2} = 1 - \frac{\left(\sum_{i=1}^{n} w_i \sigma_i\right)^2 - \sigma_I^2}{\left(\sum_{i=1}^{n} w_i \sigma_i\right)^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2}$$

It may be noted in the alternate definition that when the variance of an index tends to the weighted average of the components' variance, the index correlation tends to 1.

The minimum average correlation, in contrast, is in general greater than –1. When indexes are diversified, it is close to 0. The situation is described in the following graph.

Average correlation and index variance

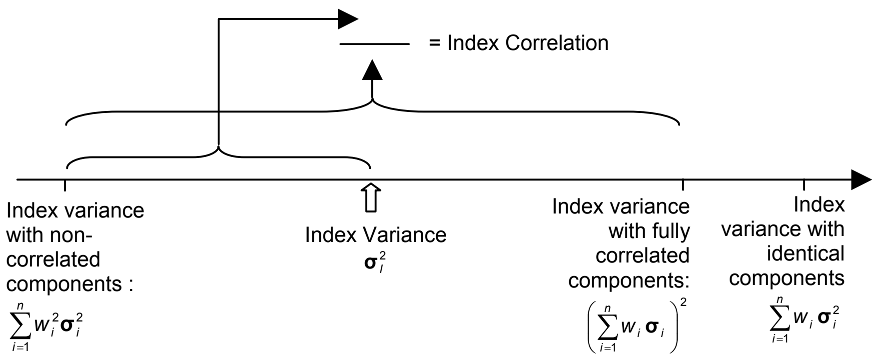


Source – Strategic Research for Equity Derivatives

Intuitively, if the number of stocks is greater than 3, we cannot have all the pairwise correlations equal to –1. For instance, if the correlation of (1,2) and (1,3) is –1, then, necessarily, the correlation (2,3) is 1.

The graphical definition of variance is shown in the following chart:

Index variance, covariance, dispersion and correlation



Source – Strategic Research for Equity Derivatives

Correlation is untouched by a scaling factor applied to returns. We may thus expect it to be less dependent on general volatility levels.

Implied correlation

Since correlation may be defined only with variance levels, we may simply extend its definition to introduce an implied level:

$$\rho = \frac{\sigma_I^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2}{\left(\sum_{i=1}^{n} w_i \sigma_i\right)^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2}$$

It should be noted that, unlike for the mathematical definition of correlation, we do not necessarily have:

$$\rho \leq 1$$

However, $\rho>1$ indicates an arbitrage opportunity. Indeed, the no-arbitrage boundary for $\sigma_I, \sigma_1, .., \sigma_n$ is defined by (see appendix for details):

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

This condition is equivalent to $\rho \leq 1$. If the no-arbitrage condition is broken, then the (or a) corresponding risk-free arbitrage is defined as:

- A short position on index realised variance
- Long positions on components variances, with the following amounts:

$$\alpha_i = \frac{\sigma_I^2}{\sum_{i=1}^{n} w_i \sigma_i} \frac{w_i}{\sigma_i}$$

A fundamental difference between correlation and the above-mentioned covariance and dispersion is that correlation is not a linear function of variances: volatility appears on the denominator of the formula. It is thus not replicable through static positions on the variances of the index and of its components. As a result, there is no reason for the correlation swap, whose payoff is:

$$\hat{\rho} - \rho ,$$

to have a zero fair market value. The strike of a correlation swap is thus not the implied correlation as we defined it.

In the course of this study, we will concentrate on strategies that involve fixed variance positions. Although correlation swaps may be replicated by dynamically hedging a portfolio of options, the risks and costs induced by the modelling and the continuous hedging when the underlying asset is a portfolio of options are such that fixed variance positions should, in our view, be strongly favoured.

How to obtain exposure to variance

Realised variance is not an achievable payoff if we consider only simple positions in the underlying asset. It is the creation of option markets that opened the way for trading variance.

First, hedging vanilla options with positions on the underlying, as described by the Black & Scholes model, generates exposure to realised variance. However, as we show in the next section, that exposure is mixed with a significant amount of undesirable "noise".

The creation products offering pure variance exposure were linked to a remarkable property: that realised variance can be replicated with a static (but continuous in strike) portfolio of vanilla options, together with an instantaneous hedge. Most of the drawbacks of the single option hedge are then eliminated. Because of this replication scheme, variance swaps enjoy better pricing and liquidity than any other volatility trading instrument.

Hedging a single vanilla option

Before variance swaps were introduced, hedging vanilla options was the only way to achieve exposure to variance. To decompose the P&L of a hedged option position, we assume that one:

- buys a call (or put) option at the level σ,
- calculates the deltas from the Black & Scholes model and uses for this purpose a hedge volatility σ_H,
- re-hedges the position every Δt, n times, up until T
- observes a realised volatility $\hat{\sigma}^2$ over the option life.

Then the P&L of the strategy may be written as follows (proofs are shown in the appendix):

$$\left(\hat{\sigma}^2 - \sigma^2\right)Tg_0 + \left(\hat{\sigma}^2 - \sigma_H^2\right)T\left(\overline{g} - g_0\right) + \sum g_i\left(R_i^2 - \hat{\sigma}^2\Delta t\right) + \varepsilon$$

$$\text{where : } \mathrm{g_i} = \frac{e^{rT}}{2}\frac{\partial^2 C}{\partial P^2}\left(P_i, \sigma_H, T - t_i\right)P_i^2 \quad , \quad \overline{\mathrm{g}} = \sum_{i=0,\dots,n-1} g_i$$

It appears thus that purchasing and hedging an option leads to a P&L that can be decomposed into:

- A "variance risk" term (VR), or variance swap exposure, for a notional amount equal to Tg_0 and a variance strike equal to the square of the option implied volatility. When the option is chosen at-the-money (forward), the amount of variance exposure per euro of option underlying may be expressed as:

$$\frac{Tg_O}{P_0} = \frac{e^{(r-d)T}\sqrt{T}}{2\sqrt{2\pi}}\frac{e^{-\frac{1}{8}T\sigma_H^2}}{\sigma_H}$$

It is a decreasing function of hedge volatility. It may seem odd that the variance exposure amount depends on hedge volatility which is arbitrarily chosen. In fact, the "true" variance nominal should be computed from future realised volatility, because the residual terms are then minimised. Future realised being unknown, one can only target a variance exposure.

■ A "gamma risk" term (GR), reflecting the fact that the gamma exposure of the position is random because it depends on the stock price path. If hedge volatility is equal to future realised, that term is nil. If not, with the function g being centred on the option strike, the term will reflect the fact that the underlying price hovers around the strike (rangy path) or, on the contrary, keeps away from it (trendy path). Controlling that term amounts to finding the best estimate for realised volatility.

■ A "convergence risk" term (CR), that depends on: the discretisation of the time interval and the divergence between the log-normal assumption and the actual process. Controlling that term amounts to choosing smaller re-hedge intervals (although jumps introduce non-hedgeable disturbances) and finding the best estimate for realised volatility.

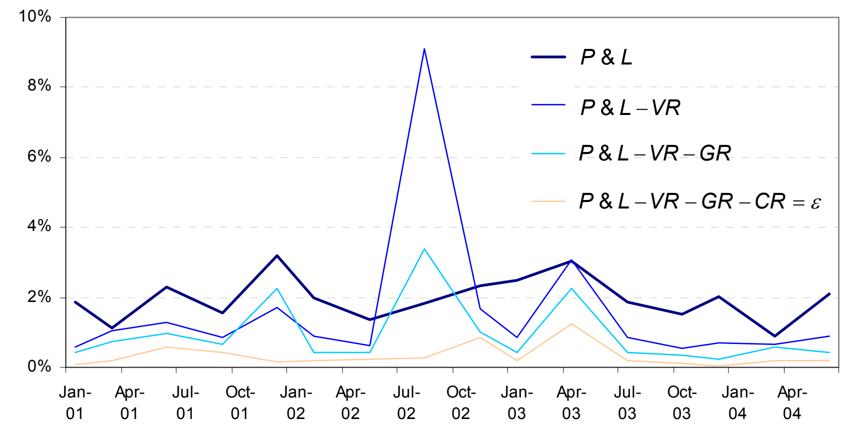
■ A residual ε, reflecting the second order approximations upon which the formula relies.

Consequently, although hedging an option does provide exposure to realised variance, there are major residual risks that cannot be eliminated efficiently (because realised variance is unknown). We could even add to the list the carry risks created by the possible occurrence of unexpected dividends.

We have simulated empirically the breakdown of hedge results on each component of the €-Stoxx 50 index and over a 3-year period, for at-the-money, three-month options. We have assumed that hedge volatility is equal to the implied level.

We present in the following graph the cross-section (i.e. among index components, grouped per date) standard errors of the decomposition.

Cross-sectional second moments of the P&L decomposition



Source – Strategic Research for Equity Derivatives

The graph shows that:

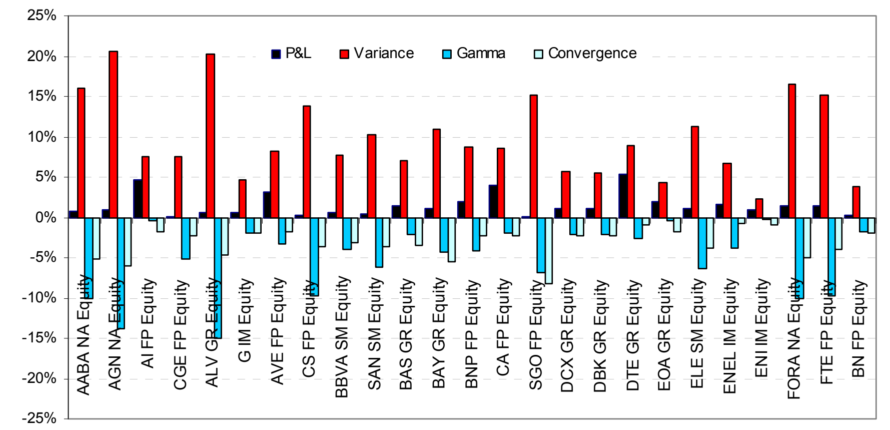
■ The decomposition presented above can be considered valid, in the sense that the norm of the residual ε is, on average, only 16% of the norm of the P&L results.

■ If we exclude three periods of volatility shocks: 9/11, summer 2002 and the Iraq war, the variance risk represents, on average, 52% of the P&L. In the best case, thus, only half of the hedge results reflect realised variance exposure, the remaining part resulting from the additional components GR, CR and ε.

■ During periods of volatility shocks, the P&L of options and the variance risks completely diverge. The hedge results do not reflect the difference between implied and realised variance, because the other risks play in opposite directions to variance risk.

To illustrate this last point, on the following graph we decompose the cross-section corresponding to the extreme divergence in the decomposition: the period May-August 2002. Only half of the sample is presented for better clarity.

P&L details for the period May-August 2002 ($\sigma_H=\sigma_I$)

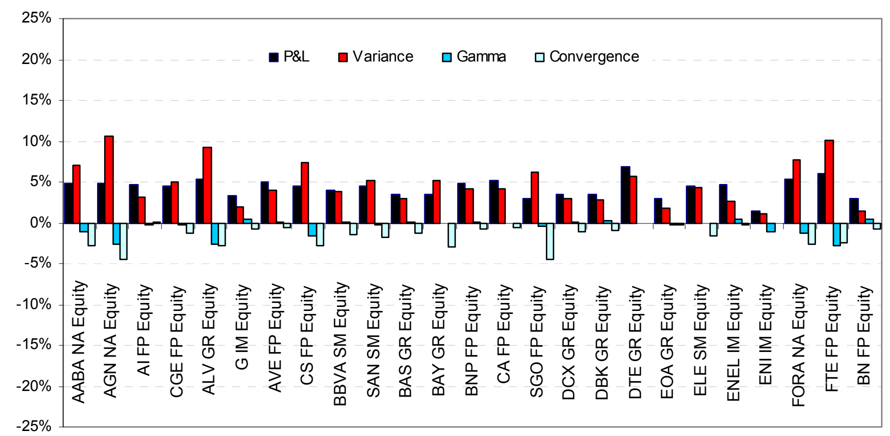


Source – Strategic Research for Equity Derivatives

During this extremely volatile period, one would expect major hedge returns for the option buyer (implied volatility was 30.7% on average, while realised reached 59.2%). However, the gamma risks offset almost all the variance effect, because the hedges were calculated at a much lower volatility than the realised volatility and the stock prices quickly and sharply diverged from the initial strike, generating a low average gamma.

To illustrate the effect of the hedge volatility choice, we have computed the same simulations with hedge volatility equal to implied volatility plus 30%.

P&L details for the period May-August 2002 ($\sigma_H=\sigma_I$+30%)



Source – Strategic Research for Equity Derivatives

One can observe that:

■ The total P&Ls are much higher on average because, during the market fall that occurred in the option period, hedge ratios decreased much less rapidly, making it possible to capture more value in the short stock hedge.

- At the same time, the expected variance gain was sharply reduced because the term g_0 was calculated from a much higher hedge volatility
- The gamma risk term was dramatically cut, because hedge volatility was much closer to realised volatility, while the convergence term was proportionally decreased by the lower 'g's.

As a consequence, expected variance exposure and realised P&L were far better matched than in the previous simulation.

In conclusion, we may say that hedging an option provides exposure to the realised variance of the underlying but that a number of other effects come into play.

Variance swaps: a continuum of vanilla options

As we show in the previous section, the exposure to variance provided by the hedging of a single option is mixed with a significant amount of "noise". To solve this issue, a very powerful analytical result has been established: by setting up a portfolio of vanilla options (continuous in strike) and maintaining a continuous hedge, it is possible to replicate, under very general assumptions, the realised variance of the asset.

Intuitively, the idea behind the variance swap decomposition is to introduce a product whose final payoff is the logarithm of the stock price in order to have constant gamma exposure in the position. Indeed, very intuitively, the disturbance terms disappear because, if we assume g is constant, then

$$(\bar{g} - g_0) = 0 \text{ and } \sum g_i \left(R_i^2 - \hat{\sigma}^2 \Delta t\right) = g \sum \left(R_i^2 - \hat{\sigma}^2 \Delta t\right) = 0$$

A rigorous presentation of the variance swap replication mathematics is presented in the appendix. We show that the realised variance of the asset may be replicated with the following positions (given a discretisation of the strike space into intervals of length ∆K):

- $\frac{2}{T}\frac{\Delta K}{K^2}$ options of strike K (call if K > forward price, put otherwise)
- $\frac{2}{T} e^{rT} \left\{ \frac{e^{rt}}{P_t} - \frac{1}{P_0} \right\}$ shares of the underlying at any date t
- Borrowing/lending of the net cash position on the money market

It appears that the fair price of the future realised variance depends on the entire skew, because all options of a given maturity are used in the replication.

It should be noted again that the decomposition does not rely on the assumptions of the Black & Scholes model, as witnessed by the expression of the hedge ratios, which are parameter-free.

This decomposition is key to the success of variance swaps. The equivalence with a static option portfolio creates a possible liquidity exchange between the two products, variance swaps being able to provide or benefit from option market liquidity.

With variance swaps, arbitrageurs can take "pure" variance exposure and set up strategies aimed solely at realised variance. In the next chapter, we analyse such strategies involving the realised variance of an index and that of its components.

Historical patterns of covariance, dispersion and correlation

In this section, we review the historical patterns of the covariance, dispersion and correlation introduced above to try and identify which should be set as the objective of an arbitrage strategy.

The computations have been performed on the EuroStoxx 50 index, over a 4-year period. The index weights were reset on each trading day. The maturity of the individual strategy is 30 business days.

For practical reasons, we use ATM implied volatility (call / put average) instead of the implied variance strike. It was not possible to constitute a database of variance swap strikes for the desired period and on the chosen universe. Absolute results are thus to be considered with caution. However, the following should be noted:

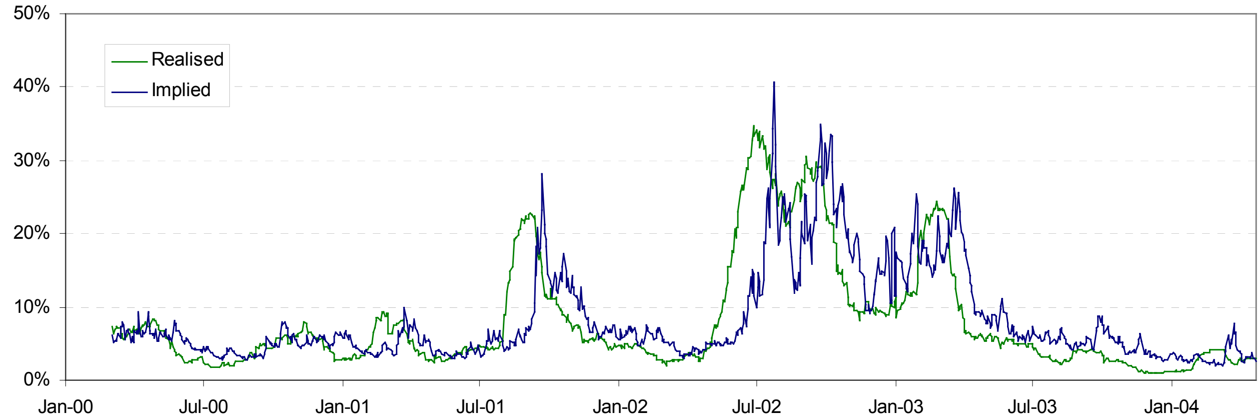
- When we consider index variance arbitrage, we distinguish between index and components variances, thus reducing the bias (the index skew being generally higher than those of its components, the short index variance / long index components arbitrage results could be underestimated in our simulations).
- Our analysis aims at comparing different index variance arbitrage strategies. We may expect the comparisons holding for ATM volatility to be true also for actual variance strikes.

In the simulations we use a correlation swap based on the above-defined implied and realised correlation (which is not fairly priced). We consider it a benchmark, only to provide an intuitive understanding of the situation. Our recommendations, however, concern only strategies achievable with fixed variance positions.

Implied and realised levels

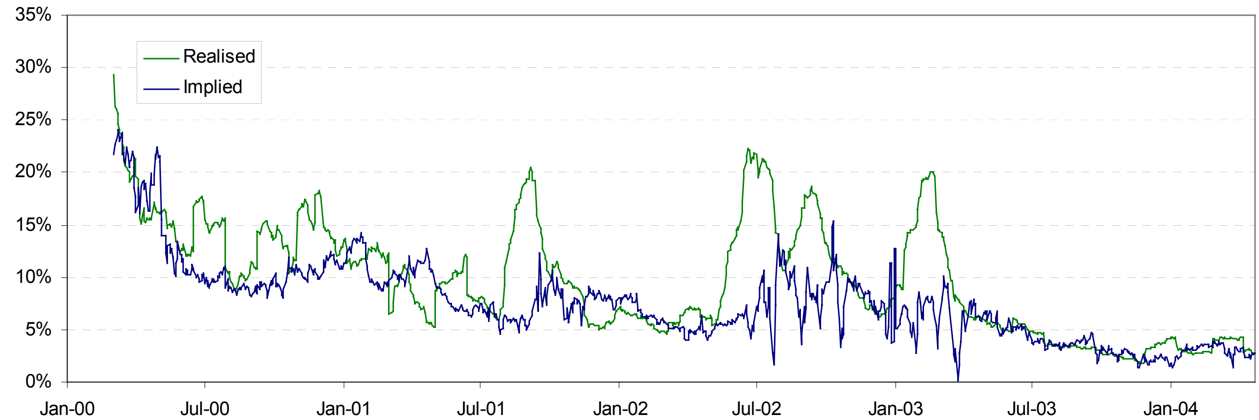
The following graphs compare implied and realised levels of covariance, dispersion and correlation within the EuroStoxx 50 index.

Implied and realised covariance



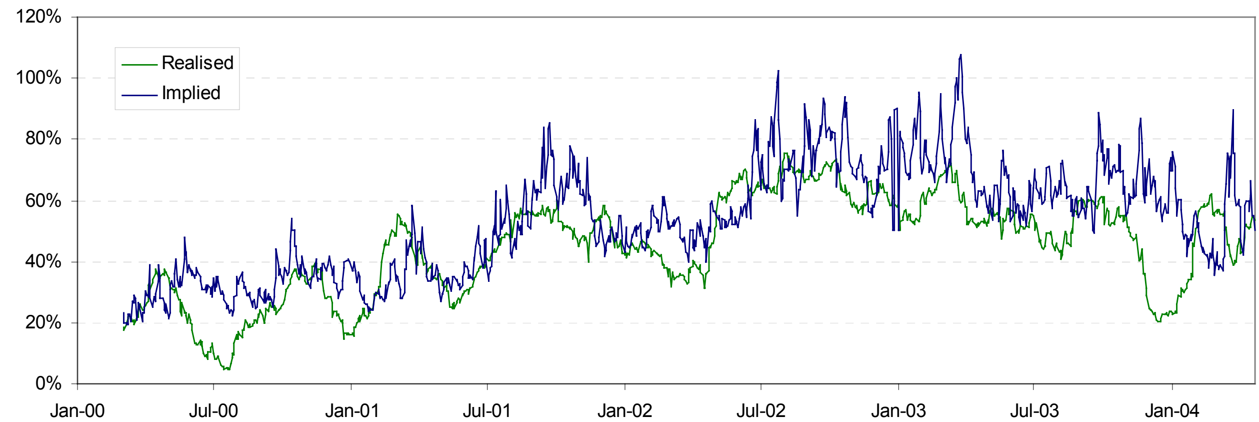
Source – Strategic Research for Equity Derivatives

Implied and realised dispersion



Source – Strategic Research for Equity Derivatives

Implied and realised correlation

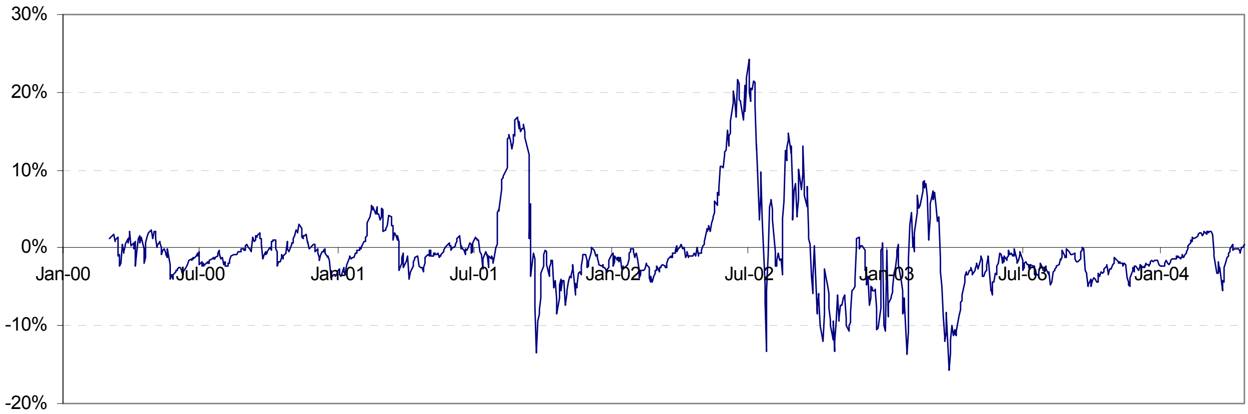


Source – Strategic Research for Equity Derivatives

Swap payoffs

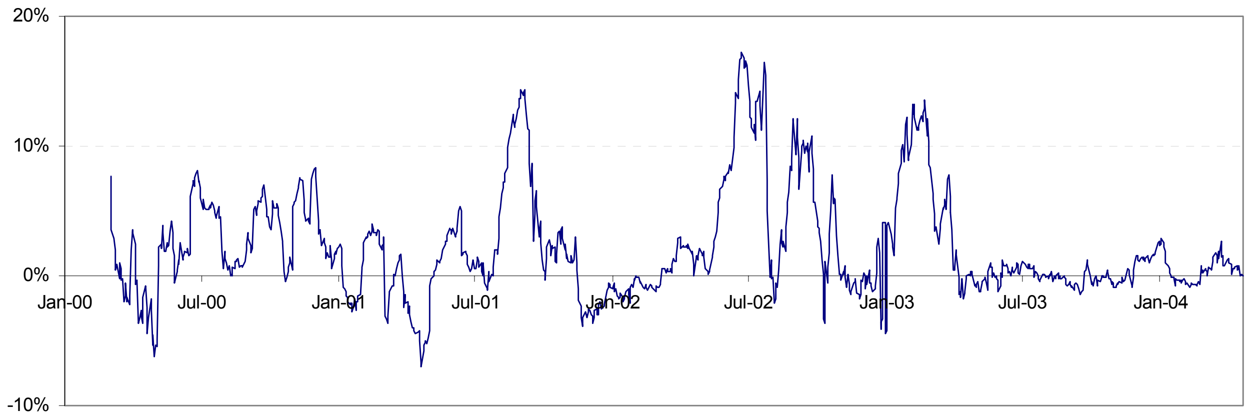
The following graphs show the payoffs of the covariance, dispersion and correlation swaps in the case of the EuroStoxx 50 index.

Long covariance (Realised – Implied)



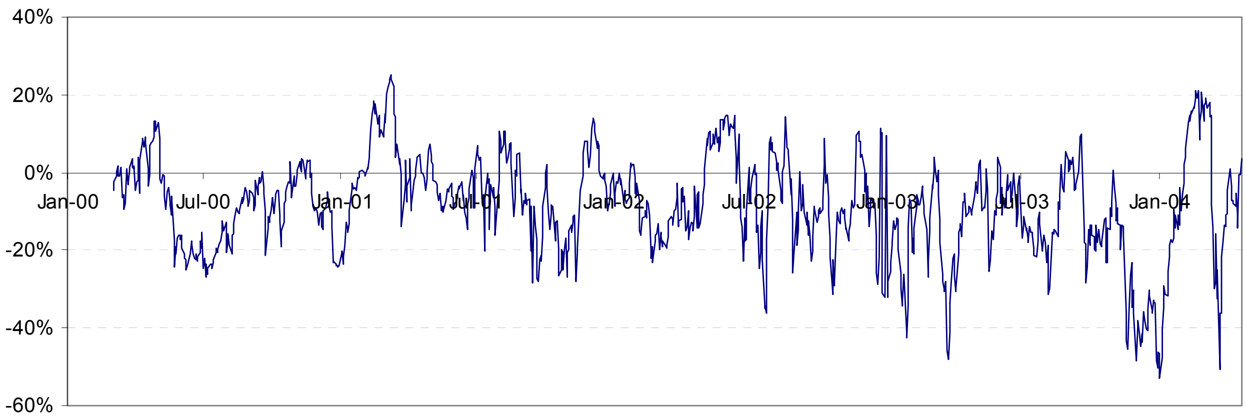
Source – Strategic Research for Equity Derivatives

Long dispersion (Realised – Implied)



Source – Strategic Research for Equity Derivatives

Long correlation (Realised – Implied)



Source – Strategic Research for Equity Derivatives

We observe the following:

- The correlation graph reveals a tendency for option markets to overestimate the realised average correlation of returns. Implied correlation generally exceeds realised correlation. In other words, index implied volatility is rich compared to individual stock implied volatility.
- Covariance and dispersion also reflect this phenomenon (covariance has negative average returns, dispersion positive average returns) but are primarily affected by volatility crises (September 11, summer 2002 and Iraq war). The impact of these volatility crises is such that covariance and dispersion then move in the same direction.
- Correlation, on the other hand, seems much less affected by general volatility levels and behaves much more homogeneously over time.

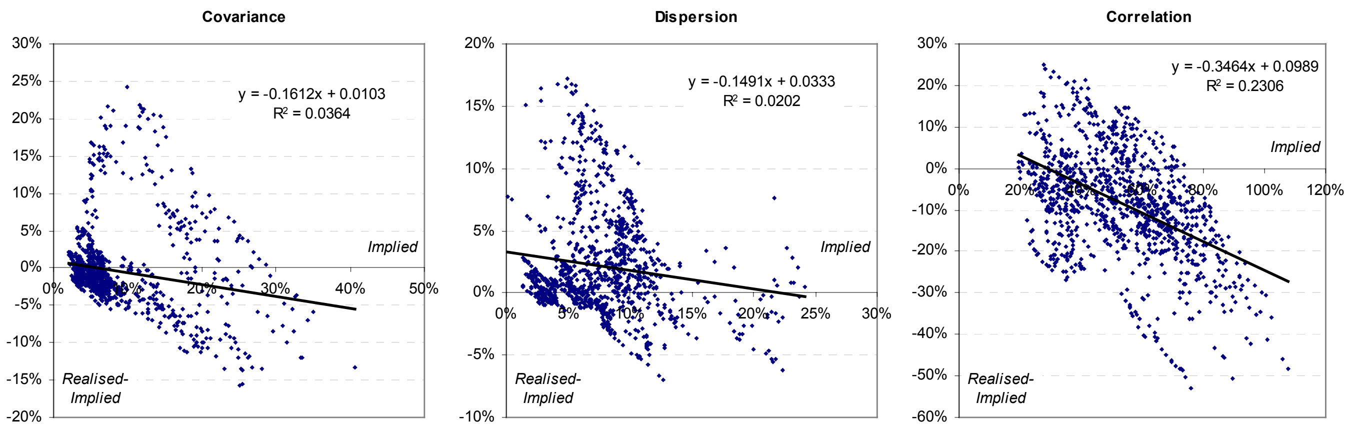
Predictability of swap payoffs

Implementing a strategy based on one of the swaps described above makes it necessary to estimate the future realised level and compare it to the known implied level.

We try here to identify a simple link between the implied level of the swap and the subsequent swap return. Specifically, we question the relevance of rules such as the following: sell a variable when the implied level is high, buy it when it is low.

To that purpose, we have in the following graphs plotted swap payoffs and the corresponding implied value (the strike of the swap):

Relationship between swap payoffs and implied levels



Source – Strategic Research for Equity Derivatives

Of the three strategies, correlation is the only one with a clear pattern: there is more chance of making money by selling implied correlation when it is higher than when it is lower. The x-coefficient and the fit of the linear regression are far more significant in the case of the prediction of correlation swap returns than for the other two instruments.

For this reason, we believe that correlation is the most suitable variable for arbitrage. It offers a degree of predictability that classic volatility does not.

Comparison with simple variance swaps

In this section, we compare the profile of the three swaps with the basic individual variance swaps, by computing the correlation coefficient between the payoffs of the three swaps and the following products:

- an index variance swap, and
- a basket of components' variance swaps (with weights equal to the index weights)

	Covariance	Correlation	Dispersion
Benchmark:			
index variance swap	1.00	0.47	0.57
components variance swaps	0.90	0.20	0.84

Source – Strategic Research for Equity Derivatives

It appears that the covariance swap is extremely close, in terms of return, to the simple index variance swap (the correlation coefficient between the two payoffs is 99.9%). In practice, on a graph, the strategy's returns can barely be distinguished from the short index variance payoffs.

Dispersion, on the other hand, behaves much like buying component volatility, although the link is not as clear as the previous one (the correlation coefficient is 84.1%).

We can also remark that the correlation swap returns are positively correlated to the variance levels, which is in accordance with intuition. When realised variance rises, the realised correlation also tends to rise.

However, among the three, only the correlation strategy can be viewed as forming an original play or a distinct asset class (correlation with the benchmark returns are limited to 46.6% and 20.2%).

At this stage, we note that covariance and dispersion are not optimal ways to set up index variance arbitrage. They are both too strongly linked to the variance swaps they are composed of. In order to find improved arbitrage strategies, we decompose index variance into correlation exposure and components' variances exposure.

How much correlation is there in index variance?

Index realised variance is linked to the level of realised average correlation, as the previous empirical analysis showed. We try here to define a linear approximation for the correlation impact on index variance.

From the definition of the mathematical tools exposed earlier, we may decompose the index variance swap payoff in the following way:

$$\left(\hat{\sigma}_I^2 - \sigma_I^2\right) = \left(\sum_{i \neq j} w_i w_j \sigma_i \sigma_j\right)\left(\hat{\rho} - \rho\right) + \sum_{i \neq j} w_i w_j \hat{\rho}\left(\hat{\sigma}_i \hat{\sigma}_j - \sigma_i \sigma_j\right) + \sum_{i=1}^{n} w_i^2 \left(\hat{\sigma}_i^2 - \sigma_i^2\right)$$

$$= \quad (1) \quad + \quad (2) \quad + \quad (3)$$

We note that

- The first term corresponds to exposure to realised correlation, for a notional amount, deterministic at the time of the trade, equal to:

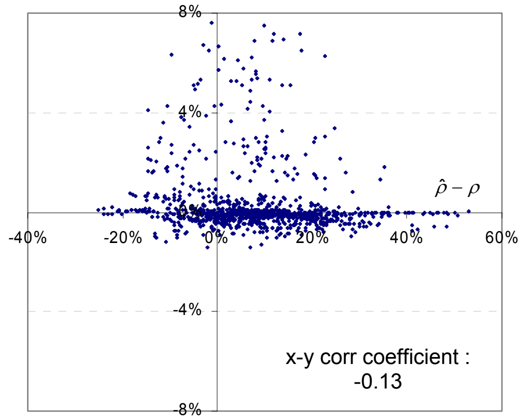
$$c = \sum_{i \neq j} w_i w_j \sigma_i \sigma_j = \left(\sum_{i=1}^{n} w_i \sigma_i \right)^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2$$

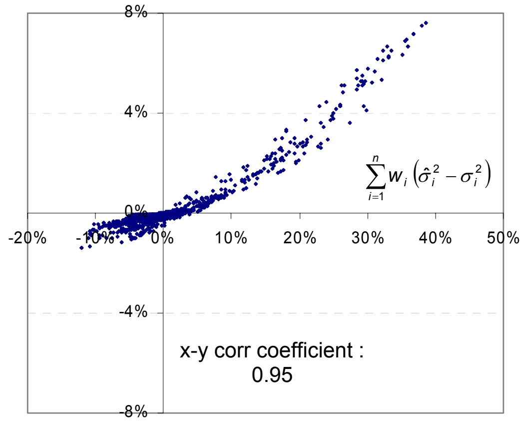
- The last term is exposure to the individual components' variance swaps.
- Whether the first term alone represents the correlation exposure depends on the behaviour of the second term.

Empirically, we find that term (2) should definitely be considered as "variance exposure" and not "correlation exposure".

On the following graphs, we compare the realised values of term (2) with, on the one hand, the correlation swap payoff and, on the other hand, the payoff of a basket of components' variance swaps.

Term (2) compared to a correlation swap and to a basket of components' variance swaps





Source – Strategic Research for Equity Derivatives

Term (2) obviously falls into components' variance exposure. We can thus rely on the analytical decomposition set out above and consider that the correlation exposure of a variance swap is equal to the c coefficient.

However, the shape of the right chart reveals also a "convexity effect" in the relationship between index variance and components' variance. When financial shocks occur (right/top points), the impact of realised individual variances seems to be of an order higher than 1, because it is multiplied by realised correlation which also peaks at these times.

Because of this convexity effect, the decomposition of index variance into a sum of a correlation swap and components' variance swap cannot be perfect. One can only try and optimise it.

Our aim below is to try and find the optimal decomposition of the residual terms (2) + (3) into positions on the individual variance swaps. Equivalently, we are trying to find the best static replication of a correlation swap.

Interestingly, we have remarked that decomposing index variance into correlation and components variance is also equivalent to minimising arbitrage residuals.

How much components variance is there in index variance?

In this section we review different practical approaches for determining the optimal hedge of an index variance swap with positions on the components. As noted earlier, this is also very close to the optimal replication of a correlation swap with static variance positions.

Specifically, we describe and test three methods:

- Analytically matching the correlation swap and its proxy to the first order (considering realised volatility and correlation as free variables). The approximation works so long as volatility remains tame. During a financial crisis, however, the proxy swap diverges sharply from the objective.
- Minimising the variance of the correlation replication error by using the empirical statistical links between implied and realised volatility and correlation. Results are far better than those obtained using the first order method. The statistical approach makes it possible to significantly reduce the divergences observed in volatility crisis situations. It is, however, complex to implement in practice because it requires the estimation of numerous statistical parameters.
- Analytically minimising the residual risk by choosing coefficients that minimise the "worst cover ratio". Although the replication results are not as precise as those of the statistical method, we find that it is an efficient and simple method.

First order approximation

When realised variance is limited, it is natural to try and determine hedge ratios through linearisation. We let:

$$\hat{R}_{\alpha,\alpha_1,\ldots,\alpha_n} = \left(\hat{\sigma}_I^2 - \sigma_I^2\right) - c\left(\hat{\rho} - \rho\right) - \sum_{i=1,..,n} \alpha_i \left(\hat{\sigma}_i^2 - \sigma_i^2\right)$$
$$= \sum_{i \neq j} w_i w_j \hat{\rho}\left(\hat{\sigma}_i \hat{\sigma}_j - \sigma_i \sigma_j\right) - \sum_{i=1,..,n} \left(\alpha_i - w_i^2\right)\left(\hat{\sigma}_i^2 - \sigma_i^2\right)$$

R is approximated at the first order around the implied level:

$$\hat{R}_{\alpha,\alpha_1,\ldots,\alpha_n}\left(\hat{\sigma}_{1,..,n}, \hat{\rho}\right) \approx \hat{R}_{\alpha,\alpha_1,\ldots,\alpha_n}\left(\sigma_{1,..,n}, \rho\right)$$
$$+ \sum \frac{\partial \hat{R}_{\alpha,\alpha_1,\ldots,\alpha_n}\left(\hat{\sigma} = \sigma\right)}{\partial \hat{\sigma}_i}\left(\hat{\sigma}_i - \sigma_i\right) + \frac{\partial \hat{R}_{\alpha,\alpha_1,\ldots,\alpha_n}\left(\hat{\sigma} = \sigma\right)}{\partial \hat{\rho}}\left(\hat{\rho} - \rho\right)$$

We have:

$$\hat{R}_{\alpha,\alpha_1,\ldots,\alpha_n}\left(\sigma_{1,..,n}\right) = 0$$

Minimising R is thus achieved on the first order by choosing the hedge so that:

$$\frac{\partial \hat{R}_{\alpha,\alpha_1,\ldots,\alpha_n}\left(\hat{\sigma} = \sigma\right)}{\partial \hat{\sigma}_i} = 0, \ (i = 1,\ldots,n) \quad \text{and} \quad \frac{\partial \hat{R}_{\alpha,\alpha_1,\ldots,\alpha_n}}{\partial \hat{\rho}} = 0$$

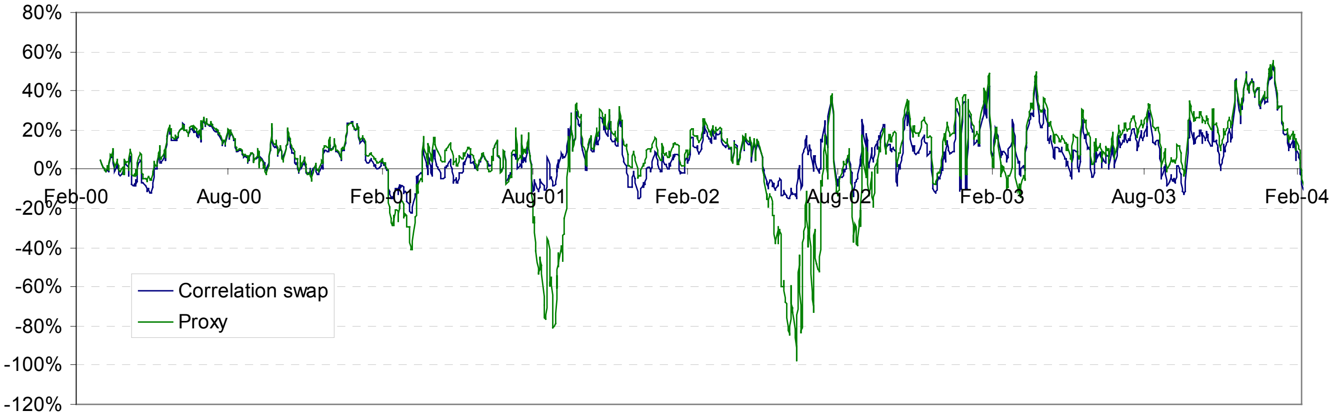
which leads to:

$$\boxed{\alpha_i = w_i^2 (1 - \frac{\rho}{2}) + w_i \frac{\rho}{2} \frac{\bar{\sigma}}{\sigma_i}}$$

Simulations

We have simulated the previous formula on the EuroStoxx 50 index (at the 30-business days maturity). In the following chart, we show the correlation swap payoff (short correlation) as well as the payoffs from the first order proxy defined above (short index variance, long components' variances):

True correlation and correlation proxy

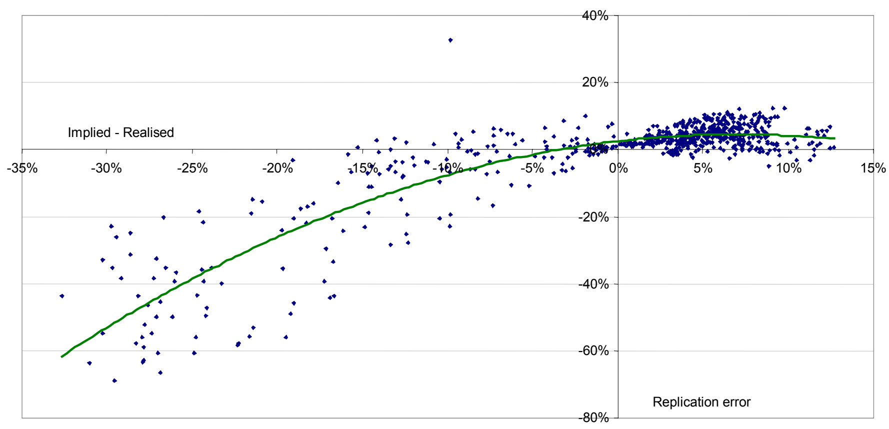


Source – Strategic Research for Equity Derivatives

The proxy for correlation replicates the objective fairly well when implied volatility is close to future realised volatility. When this does not hold, sharp divergences appear (September 11 and summer 2002 are clearly apparent on the graph).

One can also observe the second order magnitude of the errors as a function of the explicative factor on the following chart:

Replication error as a function of excess realised volatility



Source – Strategic Research for Equity Derivatives

Statistical approach

We try here to use the historical patterns of the unknown realised variables by minimising the historical variance of the residual term. Introducing the following time series:

$$u_t = \left(\hat{\sigma}_I^2(t) - \sigma_I^2(t)\right) - c(t)\left(\hat{\rho}(t) - \rho(t)\right)$$
$$v_t = \left(\hat{\sigma}_i^2(t) - \sigma_i^2(t)\right)_{i=1,\ldots,n}$$

we get: $\hat{R}_{\alpha_1,\ldots,\alpha_n} = u(t) - \alpha' v(t)$

We can then easily express the strategy minimising the variance of the residual term by introducing (the bracket signs denote historical covariance):

$$\Omega = \left\langle v_i, v_j \right\rangle_{i,j=1,\ldots,n}$$
$$\Gamma = \left\langle v_i, u \right\rangle_{i=1,\cdots,n}$$

The optimal strategy is then:

$$\boxed{(\alpha_i)^* = -\Omega^{-1}\Gamma}$$

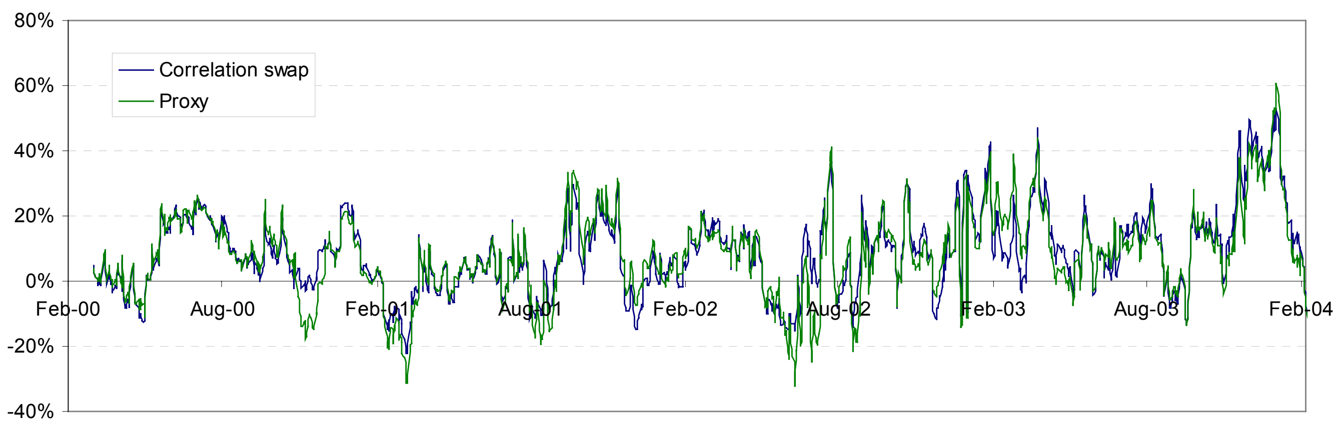
To be a possible strategy, the optimal parameters have to be estimated over a period preceding the implementation. The statistical approach's success will depend on the "persistence" of the optimal parameters, i.e. on the previously estimated parameters remaining relevant in the subsequent period. The timing of the estimation/application has to be set up in order to capture this persistence effect.

Simulations

We have simulated this statistical approach over a four-year period. We have estimated the statistical parameters (Ω and Γ) every quarter and applied them to the following quarter.

We have tested several historical periods to estimate our parameters: 6 months, 12 months, 18 months and 24 months. 6 and 12 months show the best results. We have chosen a linear combination of the four, with more weight given to the first two. The following graph depicts the actual correlation swap payoff (implied correlation minus realised correlation) as well as the statistical proxy

True correlation and correlation proxy



From the graph we can see that the replication error has been sharply reduced. The spikes due to volatility crises are much less apparent.

Volatility crises still weigh on the replication error, but maximum magnitude is reduced from 70% in the previous method to less than 20%.

The quality of the results indicates stability in the patterns of realised volatility. However, the statistical methodology has several drawbacks:

- It requires that the statistical parameters be estimated regularly
- It depends on the arbitrary choice of the historical estimation period
- It introduces negative values in the hedge ratios

Reducing the number of stocks

In order to make it easier to replicate a correlation swap through static variance swap positions, we have tested a way to limit the number of variance swaps in the replicating portfolio.

In the case of the EuroStoxx 50 index, we have limited the number of stocks to 20. The stocks are chosen individually as the 20 most efficient (should they be used alone in the replicating strategy).

Weightings are then calculated using the minimum of variance procedure, with a smaller 20 x 20 covariance matrix.

We find that the method becomes significantly less accurate when the stock number is reduced: the overall replication error is increased from 4.7% to 8.7% (standard errors are presented in a table below). As 2H01 and 1H03 show, a global variance risk re-emerges when the number of stocks in the hedge is reduced.

The min/max strategy

We consider here another approach to replication: given a short position on the index variance, we look for a portfolio of long positions on the components such that:

- the global variance strike paid on the components variance swap exactly offsets that received on the index variance swaps, and
- the worst 'payout ratio', in the context of a short index variance / long components variance arbitrage:

$$\frac{\textit{Index variance paid}}{\textit{Components variance received}}$$

is minimised.

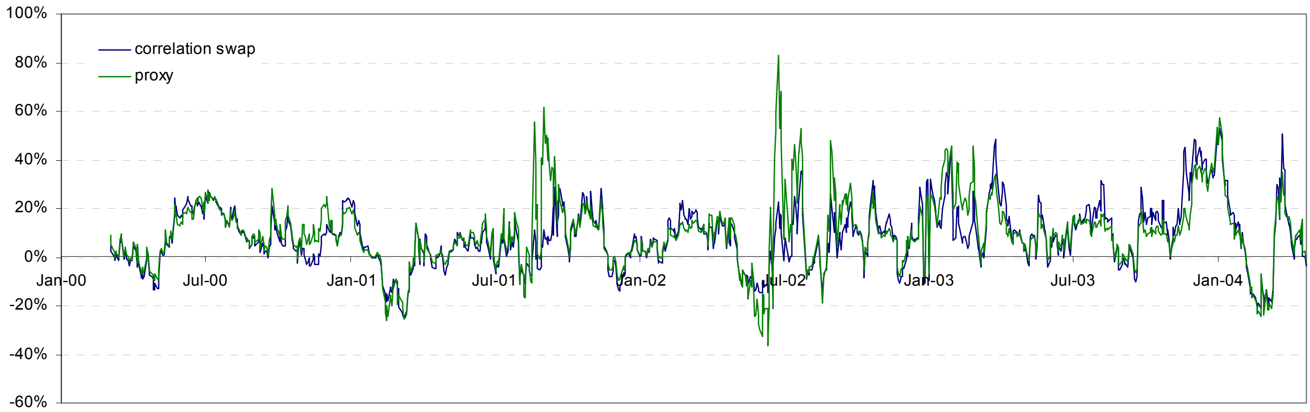
The ideas supporting the min/max hedge are: -Choosing a 0 net variance strike takes advantage of the fact that the implied strike is a good estimator for future realised variance. -Minimising the worst ratio variance paid / variance received seems a good strategy when the maximum loss is unbounded.

We have an analytical formula for this hedge strategy (see appendix for details):

$$\boxed{\alpha_i = w_i \frac{\sigma_I^2}{\sigma_i \overline{\sigma}}}$$

The following graph compares the proxy returns and the correlation swap payoff

True correlation and correlation proxy



Source – Strategic Research for Equity Derivatives

This method provides good replication results and has the advantage of relying on an analytical expression rather than a statistical estimation.

In addition, when one is selling correlation, which is the natural sign of the arbitrage, given historical behaviour, the errors generated by the proxy are:

- positive on average (1.25%),
- positively skewed (min: -26%, max: +60%)

Comparison between the different strategies

In the following, we compare the analysed index variance strategies according to various criteria and show that the initial covariance and dispersion strategies may be significantly improved.

All strategies presented in the following are of the form (natural sign for the index variance arbitrage):

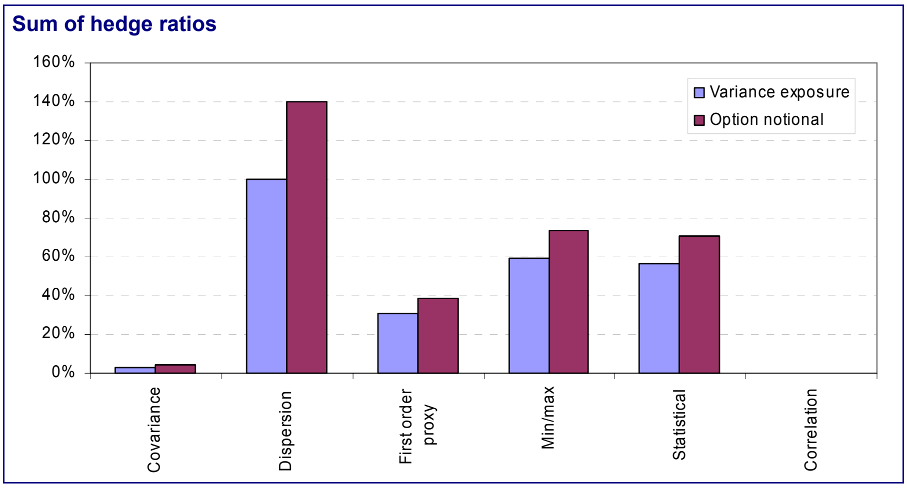
$$-\left(\hat{\sigma}_I^2 - \sigma_I^2\right) + \sum_{i=1,\ldots,n} \alpha_i \left(\hat{\sigma}_i^2 - \sigma_i^2\right)$$

Sum of hedge weights

The following table depicts the sum of the hedge ratios, averaged over 4 years, for each of the strategies. The hedge ratios are expressed either in ratios of variance exposure (the 'α's) or in ratios of option notional amounts, should the strategies be implemented by hedging vanilla options.

In the latter case, the hedge ratios, should one use at-the-money options, may be written approximately as (βi denoting the ratio of the notional amount of option i over the notional amount of index option):

$$\beta_i = \alpha_i \frac{\sigma_i}{\sigma_I}$$



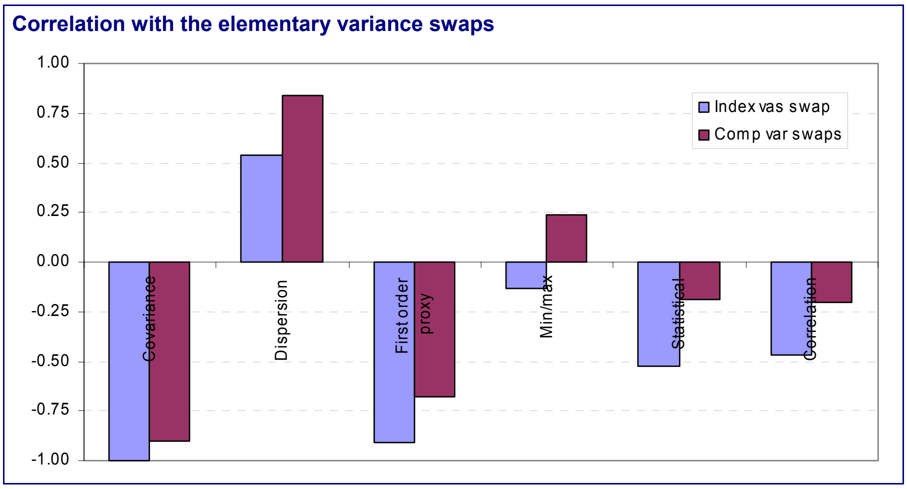
Source – Strategic Research for Equity Derivatives

By looking at the sum of weights, one could forecast the drawbacks of the covariance and dispersion strategies: the former is an obvious under-hedge while the latter, as well as the adjusted versions, are clearly over-hedged.

The min/max and the statistical strategies show an average hedge ratio of a little over 50%. It should be noted that the statistical hedge includes some negative hedge ratios.

Comparison with single variance swaps

We show in the following table the correlation coefficients between the index variance arbitrage strategies and the returns of: a simple index variance swap, a basket of components' variance swaps.



Source – Strategic Research for Equity Derivatives

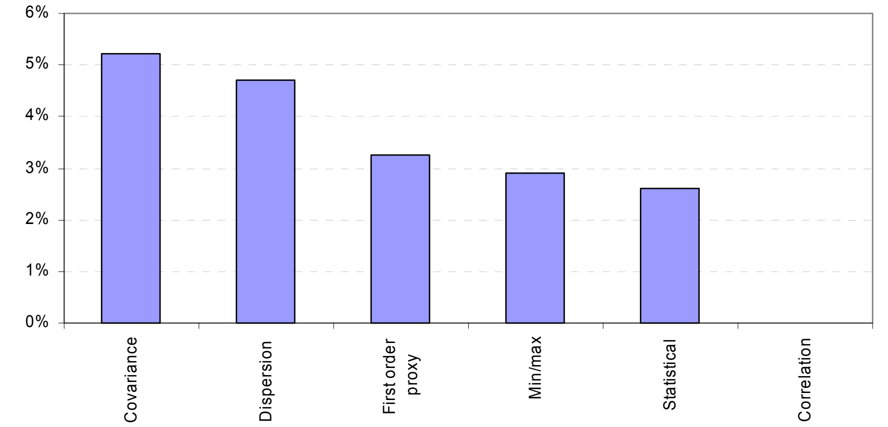
The min/max strategy appears clearly as the most original play.

Not surprisingly, the statistical hedge behaves very similarly to a correlation swap (it is designed for that purpose).

Volatility of the strategies

We show on the following graphs the volatility (not annualised) of the strategies' results over 4 years.

Volatility of the strategies' payoffs



Source – Strategic Research for Equity Derivatives

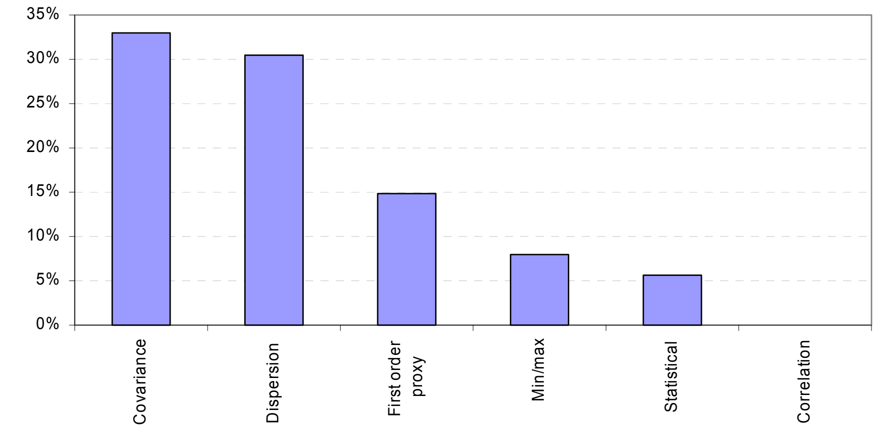
The statistical and min/max strategies clearly minimise the residual risk of index variance arbitrage. Those residuals may be broken down into a correlation exposure and a "correlation replication error". The latter is analysed below.

Correlation swap replication error

On the following graphs we show the standard error of the replication of a correlation swap with each of the strategies. Specifically, we compute (over a 4-year history):

$$StErr\left[\frac{1}{c}\left\{\left(\hat{\sigma}_I^2-\sigma_I^2\right)-\sum_{i=1,\ldots,n}\alpha_i\left(\hat{\sigma}_i^2-\sigma_i^2\right)\right\}-\left(\hat{\rho}-\rho\right)\right]$$

Standard error of the correlation swap replication



Source – Strategic Research for Equity Derivatives

Covariance, dispersion and the first order proxy are extremely inefficient strategies for taking exposure to correlation, because additional noise due to the uncovered exposure to variances far exceeds correlation exposure.

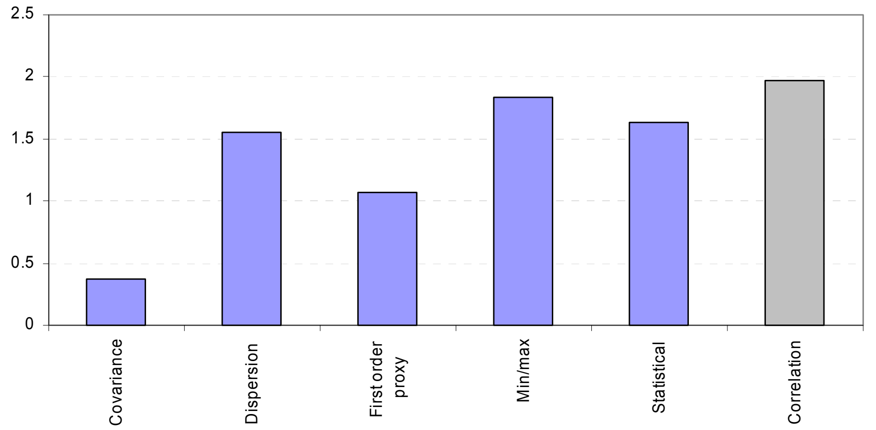
The min/max and the statistical proxies are much more efficient strategies in that respect.

With these strategies, it is possible to replicate correlation with an average error of a little over 5 points, to be compared to payoffs (realised – implied correlation) that vary within the –25% / +53% range.

Sharpe ratios

We compute the Sharpe ratios of the strategies and present them in the following graph. No adjustment is made for the overlapping swap periods since the objective is solely to compare strategies.

Sharpe ratios of the strategies



Source – Strategic Research for Equity Derivatives

Correlation (in grey because the correlation swap that we simulate is not fairly priced) is the most appealing in terms of risk/return profile. The min/max strategy is the one whose risk/return is the closest to correlation and the most appealing.

Conclusion

Index variance arbitrage, when properly set up, makes it possible to bet on the way index component returns tend to be synchronised. A variance swap payoff may be decomposed into a position on correlation and a residual term that strongly resembles exposure to the components' variances. Eliminating as much components variance risk as possible, by determining the proper hedge ratios, results in isolating correlation exposure.

Correlation is an interesting variable for arbitrage purposes. It shows a good degree of predictability, selling correlation is significantly profitable on average and correlation swap returns are loosely correlated with either market returns or market return volatility.

After reviewing several possibilities for determining hedge ratios in index variance arbitrage, we come up with an analytic solution that represents the best complexity/efficiency compromise: the min/max hedge. Through these hedge ratios, one achieves a reasonably good proxy for correlation.

The following table summarises the main fin:

Strategy	Definition	Hedge size (1) (2)	Correlation (3) (4)	Norm (5) (6)	Sharpe (7)	Comments
Covariance	$\alpha_i = w_i^2$	3.2% 4.6%	-1.00 -0.90	5.2% 33.0%	0.37	Covariance-based arbitrage returns are not different from returns from simply selling index variance. Hedge ratios are too low.
Dispersion	$\alpha_i = w_i$	100.0% 139.7%	0.54 0.84	4.7% 30.5%	1.55	Dispersion hedge ratios are too high. There is large residual long exposure to components variance
Min/Max	$\alpha_i = \frac{\sigma_i^2}{\bar{\sigma}} \frac{w_i}{\sigma_i}$	59.2% 73.4%	-0.13 0.23	2.9% 7.9%	1.83	The min/max strategy achieves low residual volatility and the best Sharpe ratio. It creates the most distinct 'asset class'.
First Order Correlation Proxy	$\alpha_i = w_i^2 (1 - \frac{\rho}{2}) + w_i \frac{\rho}{2} \frac{\bar{\sigma}}{\sigma_i}$	31.0% 38.8%	-0.91 -0.68	3.2% 14.8%	1.07	The analytical correlation proxy is close to covariance but eliminates some of the index variance exposure. Hedge ratios are still too low.
Statistical Correlation Proxy	$(\alpha_i)^* = -\Omega^{-1}\Gamma$	56.2% 70.5%	-0.52 -0.19	2.6% 5.7%	1.63	The statistical correlation proxy defines parameters by minimising historical replication error. It is the best proxy but is awkward to implement.
Correlation	Not replicable		-0.46 -0.20		1.97	Correlation is the best suited variable for arbitrage. However, it is not replicable with static variance swap positions.

Source – Strategic Research for Equity Derivatives –

(1) Expressed in variance exposure, (2) Expressed in option notional amounts, (3) Correlation of returns with an index variance swap, (4) Correlation of returns with a basket of components' variance swaps, (5) Volatility – not annualised – of swap returns, (6) Standard error of correlation replication, (7) Since ATM volatility was used instead of variance swap strikes, Sharpe ratios should not be considered individually but relative to each other.

Appendix

The P&L of a hedged vanilla option position

In order to have an analytical expression for the P&L, we assume that an option is purchased and subsequently hedged, using Black & Scholes hedge ratios, at a given volatility σ.

The hedger's portfolio is composed of three assets: the option, stock and cash. Throughout the option life, the option position is valued according to the Black & Scholes function. This is only an accounting assumption, since trading in option occurs only at the beginning and at the end of the position. At those dates, by construction, the Black & Scholes function coincides with the market prices.

In the theoretical scheme, the hedger's position is, at any time t, a combination of one option, $-\delta$ stocks and $-C+\delta P$ in cash, with a net worth constantly equal to 0. The real position will re-adjust to the theoretical hedge at every re-hedging date. However, in between, a P&L variation will appear, equal to:

$$\Delta C(P_t,\sigma,T-t)-\delta(P_t,\sigma,T-t)\Delta P_t - C(P_t,\sigma,T-t)r\Delta t+\delta(P_t,\sigma,T-t)P_t r\Delta t$$

The formula for the instantaneous deviation may be simplified thanks to:

- a Taylor expansion of the theoretical call price, giving (we keep terms in Δt, ΔS and ΔS^2):

$$\Delta C(P_t,\sigma,T-t)=\delta(P_t,\sigma,T-t)\Delta P_t+\frac{1}{2}\gamma(P_t,\sigma,T-t)\Delta P_t^{\,2}-\theta(S_t,\sigma,T-t)\Delta t+\varepsilon_t$$

- the fact that C obeys the Black & Scholes differential equation:

$$-C(P_t,\sigma,T-t)r+\delta(P_t,\sigma,T-t)rP_t+\frac{1}{2}\gamma(P_t,\sigma,T-t)\sigma^2P_t^2-\theta(P_t,\sigma,T-t)=0$$

Which yields a simple expression for the P&L variation between t and t+Δt:

$$\frac{1}{2}\gamma(P_t,\sigma,T-t)P_t^{\,2}\left\{\left(\frac{\Delta P_t}{P_t}\right)^2-\sigma^2\Delta t\right\}+\varepsilon_t$$

That formulation depends on higher order terms of the Taylor expansion being negligible and the fact that the pricing and hedging functions are solutions of the Black & Scholes equation. The accumulated variations to the Black and Scholes hedge at option maturity are thus:

$$\sum_{i=0}^{n-1}\frac{1}{2}e^{r(T-t_i)}\gamma(P_i,\sigma,T-t_i)P_i^2\left\{R_i^2-\sigma^2\Delta t\right\}+\sum_{i=0}^{n-1}\varepsilon_i=\sum_{i=0}^{n-1}g_i\left(R_i^2-\sigma^2\Delta t\right)+\varepsilon$$

where : $g_i=\dfrac{e^{r(T-t_i)}}{2}\dfrac{\partial^2 C}{\partial P^2}(P_i,\sigma\ ,T-t_i)P_i^2$ and $\varepsilon=\sum\varepsilon_i$

Now, if we distinguish between hedge volatility, σ_H and implied volatility, we get the following expression for the P&L:

$$C(P_0,\sigma_H,T)-C(P_0,\sigma_I,T)+\sum_{i=0}^{n-1}g_i\left(R_i^2-\sigma_H^2\Delta t\right)+\varepsilon$$

If hedge volatility is not too far from implied volatility, the call price may be linearised in σ^2 around σ_H:

$$\left(\sigma_H^2-\sigma_I^2\right)Tg_0+\sum_{i=0}^{n-1}g_i\left(R_i^2-\sigma_H^2\Delta t\right)+\varepsilon'$$

Now we can introduce realised volatility in the second term, to get:

$$\left(\sigma_H^2 - \sigma_I^2\right)Tg_0 + \sum_{i=0}^{n-1} g_i\left(R_i^2 - \hat{\sigma}^2\Delta t\right) + \left(\hat{\sigma}^2 - \sigma_H^2\right)\Delta t\sum_{i=0}^{n-1} g_i + \varepsilon'$$

The formula may be further decomposed into:

$$\left(\hat{\sigma}^2 - \sigma_I^2\right)Tg_0 + \left(\hat{\sigma}^2 - \sigma_H^2\right)T\left(\bar{g} - g_0\right) + \sum_{i=0}^{n-1} g_i\left(R_i^2 - \hat{\sigma}^2\Delta t\right) + \varepsilon'$$

Replicating a variance swap payoff with static positions in vanilla options

Volatility swaps may be replicated with the range of European vanilla options of the same maturity.

To derive the pricing function, we first note that if f is a smooth function, then Ito's lemma (if the price of the underlying follows a diffusion process with an instantaneous volatility process equal to σ_t) states that:

$$\int_0^T \frac{P_t^2}{2} f''(P_t)\sigma_t^2 dt = f(P_T) - f(P_0) - \int_0^T f'(P_t)dP_t$$

To value a variance swap, the idea is to devise a payoff function f so that:

$$\frac{P_t^2}{2} f''(P_t) = 1$$

We can thus choose: $f(x) = -2Log(x/P_0)$, which gives

$$\int_0^T \sigma_t^2 dt = f(P_T) - \int_0^T f'(P_t)dP_t$$, because $f(P_0) = 0$

We can thus write:

$$\hat{\sigma}_T^2 = \frac{1}{T}\int_0^T \sigma_t^2 dt = \frac{1}{T} f(P_T) - \frac{1}{T}\int_0^T f'(P_t)dP_t = \frac{-2}{T} Log\left(\frac{P_T}{P_0}\right) + \frac{2}{T}\int_0^T \frac{1}{P_t} dP_t$$

1-Pricing and replication of the term: $\frac{-2}{T} Log\left(\frac{P_T}{P_0}\right)$

Since this term is a smooth payoff, a function of the future value of the underlying asset, we may use the static replication techniques that decompose European payoffs into a forward, a bond and a continuum of vanilla options.

$$\frac{-2}{T} Log\left(\frac{P_T}{P_0}\right) = \frac{-2}{T} Log\left(\frac{P_f}{P_0}\right) + \frac{-2}{T}\frac{1}{P_f}\left(P_T - P_f\right) + \frac{2}{T}\left[\int_0^{P_f} \frac{\left(K - P_T\right)^+}{K^2} dK + \int_{P_f}^{\infty} \frac{\left(P_T - K\right)^+}{K^2} dK\right]$$

Consequently, the fair forward price of term 1 is:

$$-2r + \frac{2}{T} e^{rT}\left[\int_0^{P_f} \frac{Put(K)}{K^2} dK + \int_{P_f}^{\infty} \frac{Call(K)}{K^2} dK\right]$$

It is replicated by:

- Taking a short position in the underlying, for an amount $(2/T)(1/P_f)$

- Buying out-of-the-money (ATM=P_f) calls and puts, for the amounts $(2/T)(1/K^2)\Delta K$
- Lending/borrowing the net cash position on the money market

2-Pricing and replication of the term: $\frac{2}{T}\int_0^T \frac{1}{P_t} dP_t$

Taking a continuously re-hedged position on the underlying for the amount $(2/T)e^{-r(T-t)}(1/P_t)$, financed on the money market, leads to the following wealth at date T:

$$\int_0^T e^{r(T-t)}\left\{\frac{2}{T}e^{-r(T-t)}\frac{1}{P_t}\right\}\left(dP_t - rP_t\,dt\right) = -2r + \frac{2}{T}\int_0^T \frac{1}{P_t}dP_t$$

The fair forward value for term 2 is thus 2r, and its replication is as set out above.

Summing both terms' pricing and replicating portfolios we have:

1. A fair forward value of the realised variance equal to:

$$V_f\left[\hat{\sigma}_T^2\right] = \frac{2}{T}e^{rT}\left[\int_0^{P_f}\frac{\text{Put}(K)}{K^2}dK + \int_{P_f}^{\infty}\frac{\text{Call}(K)}{K^2}dK\right] = \underline{\text{Variance Swap Strike}}$$

2. And a realised variance replicating portfolio given by:

- $\frac{2}{T}\frac{\Delta K}{K^2}$ options of strike K (call if $K \geq P_f$, put otherwise)
- $\frac{2}{T}e^{rT}\left\{\frac{e^{rt}}{P_t} - \frac{1}{P_0}\right\}$ shares of the underlying
- Borrowing/lending of the net cash position on the money market

For the actual implementation of the variance swap replication, it may be more convenient to have the static replication of term 1 centred on an existing strike rather than on the forward. If we choose K_0 arbitrarily, we have

$$\frac{-2}{T}Log\left(\frac{P_T}{P_0}\right) = \frac{-2}{T}Log\left(\frac{K_0}{P_0}\right) + \frac{-2}{T}\frac{1}{K_0}\left(P_T - K_0\right) + \frac{2}{T}\left[\int_0^{K_0}\frac{(K-P_T)^+}{K^2}dK + \int_{K_0}^{\infty}\frac{(P_T-K)^+}{K^2}dK\right]$$

The fair forward price of the first two terms is:

$$\frac{-2}{T}Log\left(\frac{K_0}{P_f}\right) + \frac{-2}{T}Log\left(\frac{P_f}{P_0}\right) + \frac{-2}{T}\frac{1}{K_0}\left(P_f - K_0\right) = -2r - \frac{2}{T}\left\{Log\left(\frac{K_0}{P_f}\right) + \left(\frac{P_f}{K_0} - 1\right)\right\}$$

A 2nd order limited development gives:

$$-2r - \frac{1}{T}\left(\frac{P_f}{K_0} - 1\right)^2$$

So (using the same steps as above) the fair variance swap strike is given by:

$$V_f\left[\hat{\sigma}_T^2\right] = \frac{2}{T} e^{rT} \left[\int_0^{K_0} \frac{\mathrm{Put}(K)}{K^2} dK + \int_{K_0}^{\infty} \frac{\mathrm{Call}(K)}{K^2} dK \right] - \frac{1}{T} \left(\frac{P_f}{K_0} - 1 \right)^2$$

Hedge ratios are also changed accordingly.

The theory behind variance swap replication has been initially set out in the following academic research papers:

- Dupire (1994): *Pricing With a Smile* (Risk)
- Neuberger (1994): *The Log Contract: A new instrument to hedge volatility* (Journal of portfolio management)
- Carr & Madan (1998): *Towards a Theory of Volatility Trading* (Risk)

No-arbitrage condition for variance swap strikes

For a given set of components' implied variances, we compute the no-arbitrage upper bound for the index implied variance. We assume that α is an arbitrage opportunity. We then get:

$$\forall \hat{\sigma}_I, \hat{\sigma}_1, \ldots \hat{\sigma}_n : \left(\sigma_I^2 - \hat{\sigma}_I^2\right) + \sum_{i=1}^{n} \alpha_i \left(\hat{\sigma}_i^2 - \sigma_i^2\right) \geq 0$$

for all possible realised variances.

If all component returns are perfectly correlated to a variable R, which has a unit variance, we have:

$$R_i = \lambda_i R$$

We get the following arbitrage condition:

$$\forall \lambda, \sum_{i=1}^{n} \alpha_i \lambda_i^2 - \left(\sum_{i=1}^{n} w_i \lambda_i\right)^2 + \sigma_I^2 - \sum_{i=1}^{n} \alpha_i \sigma_i^2 \geq 0$$

which is equivalent to:

$$(1)\ \forall \lambda, \sum_{i=1}^{n} \alpha_i \lambda_i^2 - \left(\sum_{i=1}^{n} w_i \lambda_i\right)^2 \geq 0 \text{ and}$$

$$(2)\ \sigma_I^2 - \sum_{i=1}^{n} \alpha_i \sigma_i^2 \geq 0$$

(1) is satisfied if and only if:

$$\text{Max}\left[\frac{\left(\sum_{i=1}^{n} w_i \lambda_i\right)^2}{\sum_{i=1}^{n} \alpha_i \lambda_i^2} \middle| \forall \lambda\right] = \sum_{i=1}^{n} \frac{w_i^2}{\alpha_i} \leq 1$$

We deduce that α is an arbitrage opportunity if and only if:

$$\sigma_I^2 - \sum_{i=1}^{n} \alpha_i \sigma_i^2 \geq 0$$

$$\sum_{i=1}^{n} \frac{w_i^2}{\alpha_i} \leq 1$$

To specify the volatility conditions under which an arbitrage opportunity exists, we solve:

$$\text{Min}\left[\sum_{i=1}^{n} \frac{w_i^2}{\alpha_i} \middle| \forall \alpha, \sigma_I^2 - \sum_{i=1}^{n} \alpha_i \sigma_i^2 = 0\right]$$

We find the following minimum:

$$\text{Min}\left[\sum_{i=1}^{n}\frac{w_i^2}{\alpha_i}\middle|\forall\alpha,\sigma_I^2-\sum_{i=1}^{n}\alpha_i\sigma_i^2=0\right]=\frac{\left(\sum_{i=1}^{n}w_i\sigma_i\right)^2}{\sigma_I^2}$$

$$\text{reached at}: \alpha_i=\frac{\sigma_I^2}{\sum_{i=1}^{n}w_i\sigma_i}\frac{w_i}{\sigma_i}$$

So there is an arbitrage opportunity if and only if:

$$\text{Min}\leq 1 \Leftrightarrow \sigma_I^2 \geq \left(\sum_{i=1}^{n}w_i\sigma_i\right)^2$$

The (or an) arbitrage portfolio is then defined by:

$$\alpha_i=\frac{\sigma_I^2}{\sum_{i=1}^{n}w_i\sigma_i}\frac{w_i}{\sigma_i}$$

An over-hedge for the adjusted dispersion swap

If we choose:

$$\alpha_i = \left(\sum_{j=1}^{n} w_j \sigma_j\right) \frac{w_i}{\sigma_i}$$

then we have:

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

and

$$\sum_{i=1}^{n} \frac{w_i^2}{\alpha_i} = \left(\sum_{i=1}^{n} w_i \sigma_i\right)^{-1} \sum_{i=1}^{n} \ w_i^2 \frac{\sigma_i}{w_i} = 1$$

Consequently (see previous section):

$$\forall \hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2 : \sum_{i=1}^{n} \alpha_i \hat{\sigma}_i^2 - \left(\sum_{i=1}^{n} w_i \hat{\sigma}_i\right)^2 \geq 0$$

Thus:

$$\begin{aligned}
P \,\&\, L &= \left(\sigma_I^2 - \hat{\sigma}_I^2\right) + \sum_{i=1}^{n} \alpha_i \left(\hat{\sigma}_i^2 - \sigma_i^2\right) \\
&= \left[\sum_{i=1}^{n} \alpha_i \hat{\sigma}_i^2 - \hat{\sigma}_I^2\right] - \left[\sum_{i=1}^{n} \alpha_i \sigma_i^2 - \sigma_I^2\right] \\
&= \left[\sum_{i=1}^{n} \alpha_i \hat{\sigma}_i^2 - \left(\sum_{i=1}^{n} w_i \hat{\sigma}_i\right)^2\right] + \left[\left(\sum_{i=1}^{n} w_i \hat{\sigma}_i\right)^2 - \hat{\sigma}_I^2\right] - \left[\sum_{i=1}^{n} \alpha_i \sigma_i^2 - \sigma_I^2\right] \\
&\geq \left[\left(\sum_{i=1}^{n} w_i \hat{\sigma}_i\right)^2 - \hat{\sigma}_I^2\right] - \left[\left(\sum_{i=1}^{n} w_i \sigma_i\right)^2 - \sigma_I^2\right]
\end{aligned}$$

This over-hedge is also the cheapest one possible.

The min/max strategy

We define the min/max hedge as:

$$\left(\alpha_i^*\right): \text{ArgMin}\left[\text{Max}\left\{\frac{\hat{\sigma}_I^2}{\sum_{i=1}^{n}\alpha_i\,\hat{\sigma}_i^2}\middle|\hat{\sigma}_1^2,\ldots,\hat{\sigma}_n^2,\hat{\sigma}_I^2\right\}\middle|\forall\alpha,\sigma_I^2-\sum_{i=1}^{n}\alpha_i\,\sigma_i^2=0\right]$$

The ideas supporting the min/max hedge are:

- Choosing a 0 net variance strike takes advantage of the fact that the implied strike is a good estimator for future realised variance.
- Minimising the worst ratio variance paid / variance received seems a good strategy when the maximum loss is unbounded.

The mix/max hedging strategy results from the same program as that defining the arbitrage opportunity. The results are thus:

$$\alpha_i = \frac{\sigma_I^2}{\sum_{i=1}^{n} w_i\sigma_i}\frac{w_i}{\sigma_i}$$

Note that, in the general case:

$$\sum_{i=1}^{n}\frac{w_i^2}{\alpha_i} = \frac{\left(\sum_{i=1}^{n} w_i\sigma_i\right)^2}{\sigma_I^2} > 1$$

so the maximum loss is not bounded.

BNP Paribas is incorporated in France with Limited Liability.
Registered Office 16 boulevard des Italiens, 75009 Paris.
BNP Paribas is regulated by the FSA for the conduct of its
designated investment business in the UK and is a member of the London Stock Exchange.
BNP Paribas London Branch is registered in England and Wales under No. FC13447.
Registered Office: 10 Harewood Avenue, London NW1 6AA.
Tel: +44 (0)20 7595 2000 Fax: +44 (0)20 7595 2555 www.bnpparibas.com

BNP PARIBAS