JUNE 2004[1]

FUNDAMENTAL RELATIONSHIP BETWEEN AN INDEX'S VOLATILITY AND THE CORRELATION AND AVERAGE VOLATILITY OF ITS COMPONENTS

Sebastien Bossu

sebastien.bossu@jpmorgan.com

Yi Gu

yi.gu@jpmorgan.com

Equity Derivatives Institutional Marketing

JPMorgan Securities Ltd. – London

CONFIDENTIAL

JPMorgan

[1]Original publication February 2004. All data and charts as of 25 April 2004.

Overview

In this document, we derive and analyse the fundamental relationship between an index's volatility and the correlation and average volatility of its components:

$$\text{Index Volatility} \approx \sqrt{\text{Correlation}} \times (\text{Average Component Volatility}) .$$

This relationship holds in practice when correlation is above 0.15, and the number of components is above 20[2].

- Section 1 reviews the definition of realised and implied correlation.
- Section 2 derives the relationship from fundamentals.
- Section 3 evaluates the strength of the relationship using statistical methods.
- Appendix A is a short reference of statistical formulas.

[2] Additionally, weights and components' volatilities must be reasonable to avoid concentration on specific stocks.

1. Realised and Implied Correlation

In this section, the definition of realised and implied correlation is reviewed.

Realised Correlation

Realised correlation on an index is traditionally defined as the average of the realised correlation matrix between the index components, excluding the diagonal of 1's:

$$\rho_{Realised} = \frac{\sum_{i<j}^{n} w_i w_j \rho_{ij}}{\sum_{i<j}^{n} w_i w_j}$$

where n is the number of components, w's are the component weights, and ρ's are the pair-wise correlations:

$$\rho_{i,j} = \frac{COV_{i,j}}{\sigma_i \sigma_j}.$$

Hence:

$$\rho_{Realised} = \frac{\sum_{i<j}^{n} w_i w_j \frac{COV_{ij}}{\sigma_i \sigma_j}}{\sum_{i<j}^{n} w_i w_j}.$$

A slightly different way of defining realised correlation could be given as the ratio of average covariance to average 'cross-volatility':

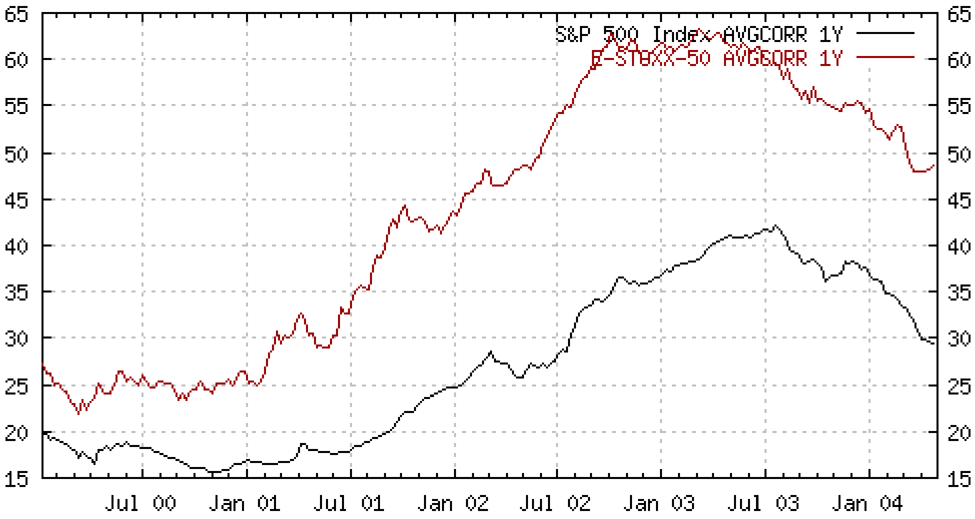
$$\rho'_{Realised} = \frac{\sum_{i<j} w_i w_j COV_{ij}}{\sum_{i<j} w_i w_j \sigma_i \sigma_j} = \frac{\sum_{i<j} w_i w_j \sigma_i \sigma_j \rho_{i,j}}{\sum_{i<j} w_i w_j \sigma_i \sigma_j}$$

This is not the market practice, but would be more consistent with the way *implied correlation* is defined.

Exhibit 1 below shows the evolution of realised correlation over a one-year rolling window for two indices: EuroStoxx 50 and S&P 500.

Exhibit 1

1Y Realised Correlation– EuroStoxx 50 and S&P 500



Source: JPMorgan– DataQuery.

Implied Correlation

Implied Correlation is the correlation parameter extracted from market option prices on an index and its components:

$$\rho_{Implied} = \frac{\sigma_{Index}^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2}{2\sum_{i<j} w_i w_j \sigma_i \sigma_j}$$

where n is the number of components, w's are the component weights, and σ's are implied volatilities.

This definition is derived from the well-known probability formula (see Appendix):

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

which translates as follows in portfolio theory:

$$\sigma_P^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + 2\sum_{i<j} w_i w_j \sigma_i \sigma_j \rho_{i,j}\ .$$

$\rho_{Implied}$ is thus closely connected to $\rho'_{Realised}$, and commonly interpreted as the market's expectation of the future realised correlation. However, the existence of a volatility surface (skew and term structure) translates into another surface for implied correlation levels. When it comes to determine the 'fair strike' of a correlation swap[3], traders will have a look at both implied and realised levels.

[3] A correlation swap is an instrument which pays off the notional multiplied by the realised correlation between inception and maturity in exchange for a fixed amount.

Exhibit 2 below shows the evolution of ATM 1Y implied correlation for two indices: EuroStoxx 50 and S&P 500.

Exhibit 2

1Y ATM Implied Correlation– EuroStoxx 50 and S&P 500

Source: JPMorgan– DataQuery.

2. Fundamental relationship between an index's volatility and the correlation and volatility of its components

In this section, a Proxy for implied and realised correlation is derived from fundamentals, leading to the relationship.

Correlation Proxy

The following mathematical equation holds for any given numbers $x_1, \dots, x_n$:

$$\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 + 2\sum_{i<j} x_i x_j \,.$$

Whence:

$$r_{Implied} = \frac{s_{Index}^{\;2} - \sum_{i=1}^{n} w_i^2 s_i^2}{\left(\sum_{i=1}^{n} w_i s_i\right)^2 - \sum_{i=1}^{n} w_i^2 s_i^2} \,.$$

For a sufficiently large number of components (in practice n > 20) and well-behaved weights and volatilities, the second term on both the numerator and denominator, $\sum_{i=1}^{n} w_i^2 s_i^2$, becomes negligible. This is most straightforwardly observed when the components are equally weighted:

$$\sum_{i=1}^{n} \frac{1}{n^2} s_i^2 \le \frac{1}{n^2} \sum_{i=1}^{n} s_{\max}^2 = \frac{s_{\max}^2}{n} \xrightarrow[n \to +\infty]{} 0$$

The limiting case gives us the proxy:

$$r \approx \frac{s_{Index}^2}{s_{Average}^2}$$

where $s_{Average} = \sum_{i=1}^{n} w_i s_i$ is the average implied volatility of the components. Note that this proxy will become inaccurate if the true implied correlation becomes low (in practice <0.15). Since most indices select their components regionally or sectorially, this should rarely be observed in practice.

The same derivations hold for the second definition of realised correlation, $r'_{Realised}$, once it has been noticed that:

$$r'_{Realised} = \frac{s_{Index}^2 - \sum_{i=1}^{n} w_i^2 s_i^2}{2\sum_{i<j} w_i w_j s_i s_j}$$

by backing out the numerator through the formula for the volatility of a portfolio and its components (see Section 1.)

Interpretation

In essence, correlation is the squared ratio of index volatility to the average volatility of the components. In other words:

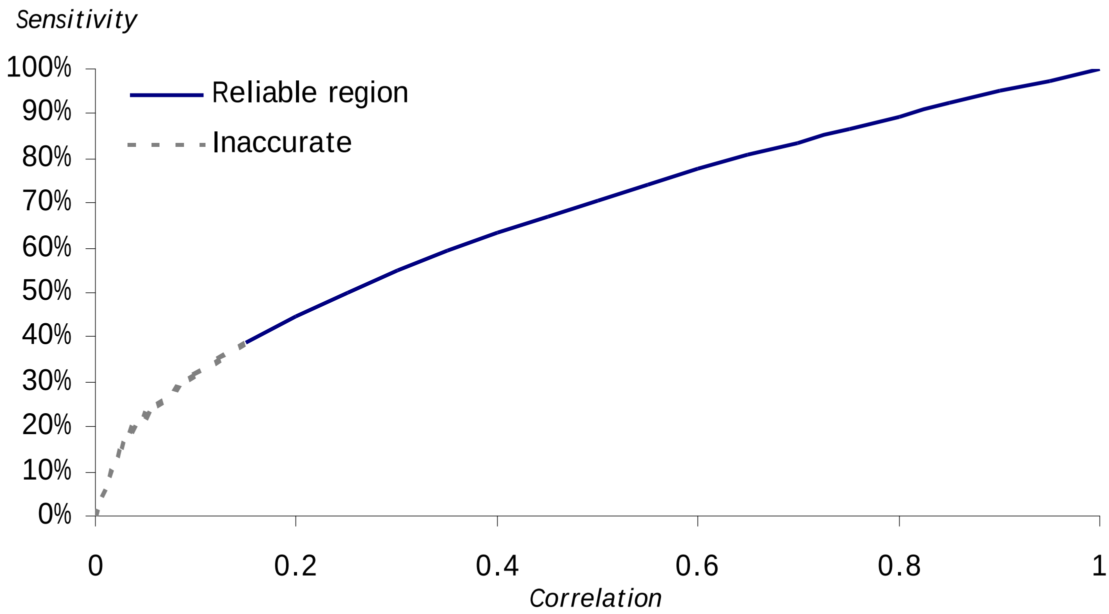
$$\text{Index Volatility} \approx \sqrt{\text{Correlation}} \times (\text{Average Component Volatility})$$

This relationship means that index volatility is less sensitive to changes in components' volatility when correlation is low. However, the relationship becomes weaker when correlation reaches the region 0.15–0.25; and becomes inaccurate below 0.15.

Exhibit 3 below shows the sensitivity of index volatility to average component volatility in function of correlation.

Exhibit 3

Sensitivity of index volatility to average component volatility.



3. Statistical Analysis

In this section, the strength of the relationship derived previously is evaluated using statistical methods on both implied and realised data.

New functions in DataQuery introduced in January 2004[4] give access to both implied and realised data for volatility, correlation and average volatility of the major indices.

Implied Data

Exhibits 4 and 5 below compare the index ATM implied volatility (IVOL) for both one-year and 3-month maturities against $\sqrt{IMPCORR} \times AVGIVOL$ on the EuroStoxx 50 between 2000 and 2004. The fundamental relationship is so accurate in both cases that the two lines are almost indistinguishable.

This is confirmed by the regression results in Exhibit 6: the R^2 is 0.99873.

Exhibit 4

1Y ATM implied volatility of EuroStoxx50 vs. fundamentals

Source: JPMorgan– DataQuery.

[4] *Equity Derivatives in DataQuery*, EDG Strategy London, 26 January 2004.

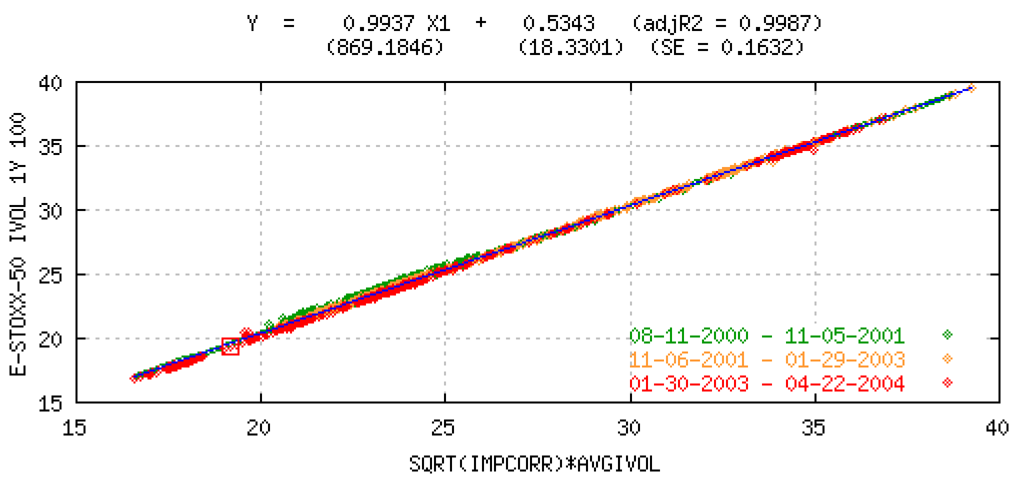
Exhibit 5

3M ATM implied volatility of EuroStoxx50 vs. fundamentals

Source: JPMorgan– DataQuery.

Exhibit 6

1Y ATM implied volatility of EuroStoxx50 vs. fundamentals– Regression



Source: JPMorgan– DataQuery.

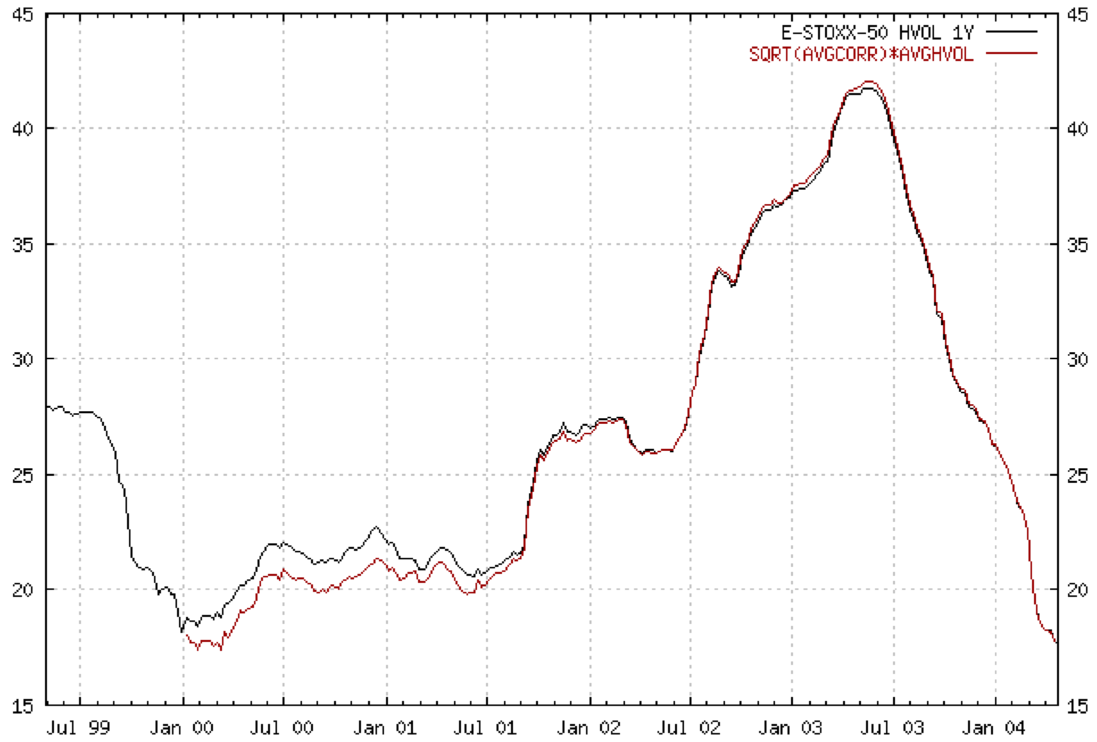
Realised Data

The same methodology is used for realised data. Exhibits 7 and 8 below compare the index realised volatility (HVOL) for both one-year and 3-month rolling windows against $\sqrt{AVGCORR} \times AVGHVOL$ on the EuroStoxx 50 between 2000 and 2004.

The relationship is also very strong (see the regression results in Exhibit 9, with a R^2 of 0.9975), but slightly less so for the period starting January 2000 and ending August 2001. Interestingly, this period coincides with a low realised correlation (see Exhibit 1 in Section 1), where the fundamental relationship is expected to be weaker.

Exhibit 7

1Y realised volatility of EuroStoxx50 vs. fundamentals



Source: JPMorgan– DataQuery.

Exhibit 8

3M realised volatility of EuroStoxx50 vs. fundamentals

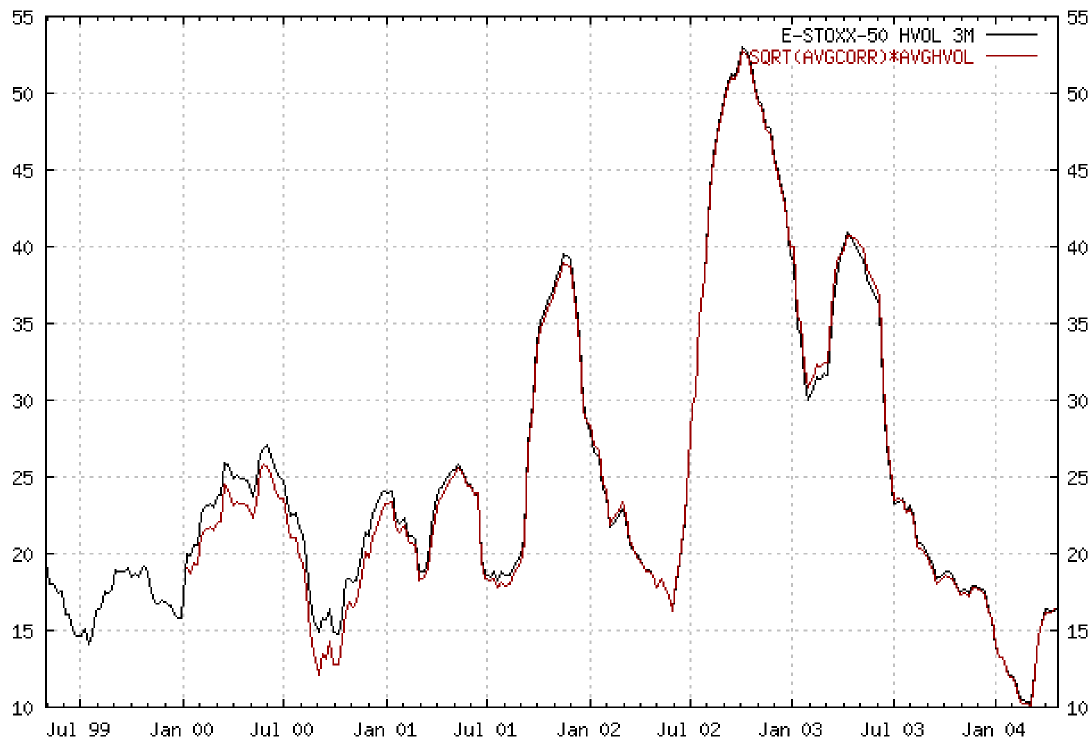
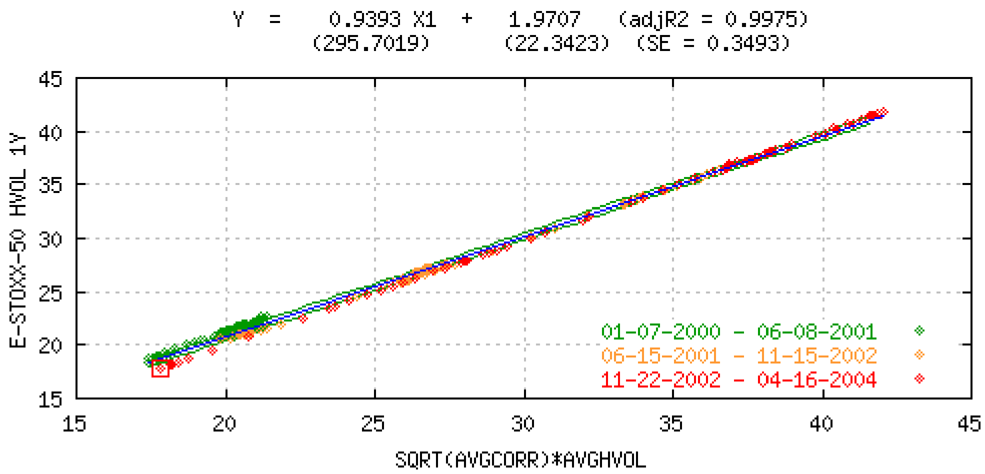


Exhibit 9

1Y realised volatility of EuroStoxx50 vs. fundamentals– Regression



Other Indices

The relationship was also tested on the S&P 500. Exhibit 10 displays the results for both implied and realised data. Again, strong accuracy was observed.

Exhibit 10

1Y implied and realised volatility of S&P 500 vs. fundamentals

Appendix A– Quick Reference of Statistical Formulas

The following formulas hold for **complete** statistical series of length N, e.g. X = (x_i) = (x_1, … , x_N). Please refer to the sub-section on Sample Estimation for formulas on incomplete series.

Mean

The **mean** of a series X is:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Variance. Standard Deviation

The **variance** of a **complete** series X is:

$$Var(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})^2$$

and the **standard deviation** is the square root of the variance (also called **volatility** in finance):

$$s_X = \sqrt{Var(X)}$$

Note: $Var(aX + b) = a^2 Var(X)$.

Covariance

The **covariance** between two **complete** series X and Y is:

$$Cov(X,Y) = N\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})$$

Covariance appears in the variance of a sum of two series:

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

and more generally, for n series:

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} Var(X_k) + 2\sum_{1\le j<k\le n} Cov(X_j, X_k)$$

Note: $Cov(aX+Y,Z) = aCov(X,Z) + Cov(Y,Z)$ (bilinearity)

$$Cov(X,X) = Var(X)$$

Correlation

The **coefficient of correlation** between two series X and Y is:

$$r_{X,Y} = \frac{Cov(X,Y)}{s_X s_Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Correlation appears in the variance of a sum of two series:

$$s_{X+Y}^2 = s_X^2 + s_Y^2 + 2 s_X s_Y r_{X,Y}$$

whence:

$$r_{X,Y} = \frac{s_{X+Y}^2 - (s_X^2 + s_Y^2)}{2 s_X s_Y}$$

and since $(s_X + s_Y)^2 = s_X^2 + s_Y^2 + 2 s_X s_Y$, we also have:

$$r_{X,Y} = \frac{s_{X+Y}^2 - (s_X^2 + s_Y^2)}{(s_X + s_Y)^2 - (s_X^2 + s_Y^2)}$$

Notes

Correlation is a statistical measure for the level of interdependence between two variables. In the case of the correlation of returns, it is a measure of the "joint directionality" of two assets: the higher the correlation, the more frequently the assets move upwards/downwards together.

A well-known property of correlation is that it is a number comprised between -1 and 1. The three perfect cases are:

Correlation	Interpretation
+1	The two variables are related with 100% certainty by a linear formula: X = a + bY, where b > 0.
0	The two variables are independent[5] from each other.
-1	The two variables are related with 100% certainty by a linear formula: X = a - bY, where b > 0.

The definition of correlation expands as follows:

$$r_{X,Y} = \frac{1}{N} \sum_{i=1}^{N} \left[\left(\frac{x_i - \bar{X}}{s_X} \right) \left(\frac{y_i - \bar{Y}}{s_Y} \right) \right]$$

where N is the length of the series (number of observation dates),

x_i, y_i are the i^{th} observations of X and Y respectively,

$\bar{x}, \bar{y}$ are the means (average returns), and

s_X, s_Y are the standard deviations (volatilities).

A closer look at this formula yields the following observations:

[5] As understood in general language. In mathematics, two independent variables must have zero correlation, but the converse is not necessarily true.

- $\left(\dfrac{x_i - \bar{x}}{s_X}\right)$ is the number of standard deviations from the mean for observation x_i: for instance, if the mean return is 5%, volatility is 10% and x_i = 20%, this observation diverges by +1.5 standard deviations from its mean
- $\left(\dfrac{x_i - \bar{x}}{s_X}\right)\left(\dfrac{y_i - \bar{y}}{s_Y}\right)$ hence measures the joint deviation of x_i and y_i: it is a positive number if x_i and y_i 'move' in the same direction, a negative one if they move in opposite directions, and close to zero otherwise.

	y_i : upwards	y_i : no deviation	y_i : downwards
x_i : upwards	Joint deviation >0	Nil	Joint deviation <0
x_i : no deviation	Nil	Nil	Nil
x_i : downwards	Joint deviation <0	Nil	Joint deviation >0

- Thus, correlation is nothing else than the average of joint deviations.

Because a single number cannot summarise the complexity of the dynamics between two statistical series, it does not replace a thorough pattern analysis, as illustrated below.

	x_i	y_i
Observation 1	1	1
Observation 2	-1	-1
Observation 3	1	-1
Observation 4	-1	1
Mean	0	0
Standard Deviation	1	1
Correlation	0	

In this example, the first two observations are perfectly correlated, while the last two are perfectly anti-correlated. This could reflect that some event between observations 2 and 3 had changed the nature of the relationship between X and Y. However, the overall correlation is 0.

Sample Estimation

The formulas above are valid when the data series are 'complete'. When dealing with samples, however, we can only *estimate* the mean and variance. A Theory of Estimation was developed accordingly, and the main finding is that an **unbiased estimate** of the variance of a sample is:

$$Var(X) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$$

and similarly for covariance.

This subtlety has little impact for large N, and has no effect on the correlation coefficient as the averaging weights cancel out:

$$r_{X,Y} = \frac{\frac{\mathbf{1}}{\mathbf{N-1}}\sum(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\frac{\mathbf{1}}{\mathbf{N-1}}\sum(x_i - \overline{X})^2}\sqrt{\frac{\mathbf{1}}{\mathbf{N-1}}\sum(y_i - \overline{Y})^2}}$$

$$= \frac{\sum(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum(x_i - \overline{X})^2}\sqrt{\sum(y_i - \overline{Y})^2}}$$

Zero-mean assumption

If the true mean is assumed to be zero, an unbiased estimate of the variance is simply:

$$Var(X) = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

Notes

- Other estimation techniques, such as Maximum Likelihood, may yield different formulas.
- The vast majority of estimation formulas are based on the assumption that observations are independent.