EDHEC-Risk Institute

393-400 promenade des Anglais
06202 Nice Cedex 3
Tel.: +33 (0)4 93 18 32 53
E-mail: research@edhec-risk.com
Web: www.edhec-risk.com

Correlation vs. Trends: A Common Misinterpretation

April 2011

Francois-Serge Lhabitant
Chief Investment Officer, Kedge Capital Fund Management Ltd.
and Professor of Finance, EDHEC Business School

EDHEC BUSINESS SCHOOL

Abstract

Two common beliefs in finance are that (i) a high positive correlation signals assets moving in the same direction while a high negative correlation signals assets moving in opposite directions; and (ii) the mantra for diversification is to hold assets that are not highly correlated. We explain why both beliefs are not only factually incorrect, but can actually result in large losses in what are perceived to be well diversified portfolios.

The concept of correlation was first introduced by Sir Francis Galton (1886) in the context of a very specific biometric problem – analysing the relationship between the average height of mothers and fathers with those of their offspring in order to develop an evolutionary theory of heredity. Galton essentially drafted the first semi-graphic scatter plot from which correlation was a somewhat ingenious deduction. It was only a decade later that Karl Pearson (1986) published the first rigorous treatment of correlation and regression. In particular, he introduced an index called the Pearson's product moment correlation coefficient to measure the extent of a linear relationship between two random variables. Today, this index is often referred to as "Pearson's r" or simply put, "the correlation coefficient".

Correlation coefficients were introduced in finance by Harry Markowitz (1952) when developing Modern Portfolio Theory. Markowitz illustrated that the variance of a portfolio's return was a weighted average of the correlation coefficients of the returns of its component assets. Since all the weights in this average were positive, the obvious solution to reduce the portfolio variance was to search for uncorrelated assets or even negatively correlated assets if possible. This has since become one of the fundamental pillars of portfolio construction.

The difficulty with this approach is that correlations are not directly observable and tend to vary over time. As a result, an extensive body of the financial literature has focused on how best to estimate correlations, model their variations over time or during market shocks, and/or forecast their future evolution. While the usefulness of this research cannot be contested, its technical nature makes it difficult to apprehend. In fact, one has to acknowledge that there is a widening gap between academic models and practitioners beliefs regarding correlation. This gap starts at a surprisingly low level – the correct interpretation of what correlation coefficients effectively measure. Most investors and even some high-level academics appear to misinterpret the true meaning of a correlation coefficient, with potentially grave structural consequences for portfolio construction and risk management.

In this brief note, we aim to provide an intuitive tutorial-level introduction to the true nature of correlation coefficients. In particular we will explain why some common beliefs about the interpretation of correlation coefficients and their signs are factually incorrect.

Revisiting the definition of correlation

Let us start by reviewing the technical definition of correlation. Consider two random variables X and Y, for instance the returns of two different stocks with finite and positive variance. The *linear correlation* between these two variables is defined as the covariance between the two variables divided by the product of the standard deviations of each variable.

$$\text{Correlation}(X,Y) = \frac{\text{Covariance}(X,Y)}{\sqrt{\text{Variance}(X) \times \text{Variance}(Y)}} \qquad (1)$$

Equation (1) defines the population correlation coefficient. A consistent estimator of the population correlation coefficient between two time series $\{x_t\}_{t=1}^{N}$ and $\{y_t\}_{t=1}^{N}$ is given by:

$$\hat{\rho}_{X,Y} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sum_{t=1}^{T}(y_t - \bar{y})^2}} \qquad (2)$$

where x and y are the sample mean of X and Y, respectively. This, or some simple algebraic variant, is the usual formula found in most introductory statistic textbooks. Using the Schwartz inequality, one can easily show that the absolute value of the numerator is less than or equal to the denominator. Therefore, correlation coefficients are by construction bounded between -1 and +1, making them easy to communicate and discuss with investors.

This conceptual simplicity is unfortunately the source of a major misinterpretation. Investors with no strong statistical background often too quickly conclude that a positive correlation signals a tendency for the two random variables to move in the same direction while a negative correlation signals the opposite. As an illustration, a recent Morgan Stanley (2009) research note stated: "If the correlation between two assets is +1, they are said to be perfectly correlated. Their returns always move in the same direction at the same time and by the same amounts. If correlation is -1, the assets are said to be negatively correlated. Their returns always move in opposite directions, by exactly opposite amounts." Worse, this belief has also been explicitly validated by several high quality textbooks. For instance, Alexander (2001) stated: "strong positive correlation indicates that upward movements on one return time series tend to be accompanied by upward movements in the other, and similarly downward movements of the two series tend to go together".

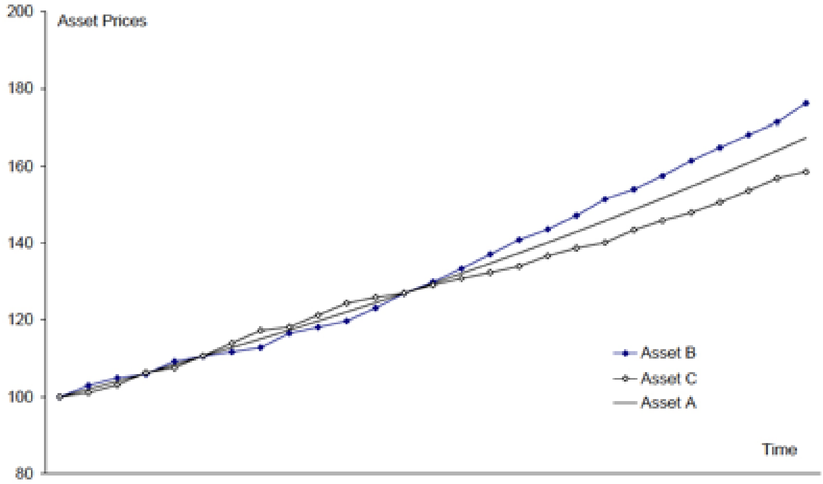
Unfortunately, this interpretation is false. A careful examination of Equation (2) reveals that the correlation coefficient $\hat{\rho}_{X,Y}$ is calculated from *deviations from the means* and not from the original raw data. As a result, any inference derived from the sign or value of the correlation coefficient can therefore only be on the *deviations from the mean* of the respective time series. For instance, one could say: "strong positive correlation indicates that upward deviations from the mean on one return time series tend to be accompanied by upward deviations from the mean in the other and similarly downward". This is obviously very different from the original statement discussed above, which was related to the series themselves.

Losing information about the original data when calculating correlations is usually not a concern for Modern Portfolio Theory where risk is exclusively defined as volatility, i.e. deviations from the mean. However, we believe that it can quickly become a major concern for investors who also care about the mean and its sign. For instance, a portfolio made of assets that all lose money at the same time, but with some deviations around their downward trend, should not be seen as very well diversified.

To illustrate the pitfalls in these common but incorrect interpretations of correlation, let us introduce two numerical examples. For the sake of simplicity, we will assume that asset prices follow geometric Brownian motions. That is, asset prices are driven by the combination of a predictable component (a long-term expected trend, which we will assume to be constant) and an unpredictable component (some short-term unexpected variations or "uncertainties" around this trend).

Example 1: Identical trends, opposite deviations

Figure 1: Price path of assets with perfectly negatively correlated returns.



We first consider the example of assets whose prices have exactly the same long term trend, but different short term deviations. Say asset A is a purely deterministic asset and its price grows at a constant rate. Asset B and C are stochastic – their prices share the same growth trend as asset A's price, but with a small random deviation at each period. This random variation has zero mean and is identical in value but of opposite signs for B and C. That is, if B outperforms the trend over a given period, C will underperform it by the same amount, and vice versa.

Figure 1 illustrates one simulated path for the three asset prices. As expected, the prices of assets B and C prices seem to oscillate around asset A. The trend component dominates in the long run, as the stochastic variations around the trend tend to revert over time due to their zero mean. Most investors looking at this figure would conclude that assets B and C are almost perfectly correlated. Wrong! In fact, their returns are perfectly negatively correlated, because their *deviations* from the mean are identical, but of opposite signs.

Example 2: Opposite trends, identical deviations

Let us now consider the example of two assets with a new set of very specific characteristics. Asset D and E are stochastic – their prices are driven by a trend component plus a small random deviation at each period. This random variation is identical in value and sign for D and E. However, the trend is of an opposite sign. That is, if the price of D grows at a certain rate, then the price of E shrinks at the same rate.[1]

Figure 1: Price path of perfectly positively correlated assets.

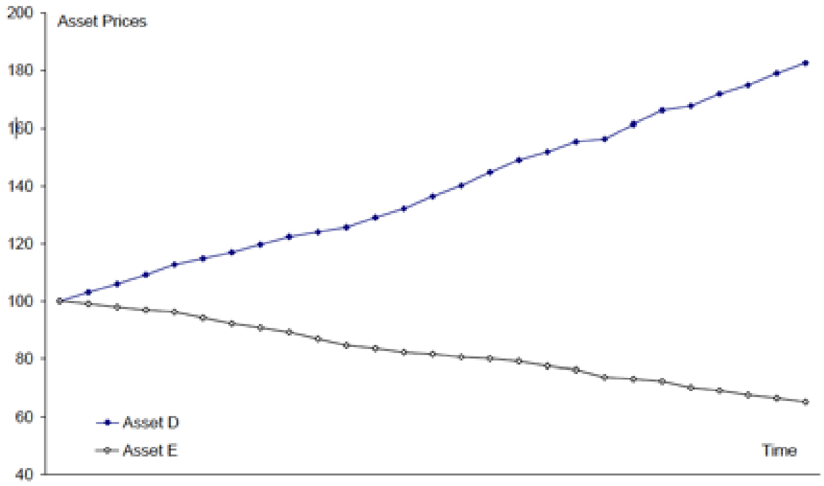


Figure 2 illustrates one simulated path for the three asset prices. As expected, the prices of D and E seem to oscillate around two very different trends. These trend components dominate in the long run, as the variations around the trend tend to mean revert over time. Most investors looking at this figure would conclude that assets D and E are almost perfectly negatively correlated. Wrong again! In fact, their returns are perfectly positively correlated, because their deviations from their respective mean are identical, and the sign of the trend component does not matter in the calculation of the correlation coefficient.

Modern Portfolio Theory suggests that the relationship between correlation and proper portfolio diversification is typically inversed. That is, when the correlation between portfolio constituents increases, diversification benefits decrease and when the correlation decreases, diversification benefits increase. Consequently, most investors follow Markowitz (1991) and believe that "to reduce risk, it is necessary to avoid a portfolio whose securities are all highly correlated with each others". This is true when risk is exclusively measured in terms of variance, but can be problematic as soon as trends come into play. Assets B and C are negatively correlated securities, but their

1 - For the sake of simplicity, we will assume hereafter that the growth rate, respectively shrink rate, is constant over time.

combination would not bring much to a portfolio in terms of effective diversification. Assets D and E are good examples of perfectly correlated securities that could be combined to create a structurally well-diversified and more stable portfolio.

Conclusions

One of the common misuses of statistical jargon is the use of the word "correlation" to describe any variable that increases as another variable increases, particularly in risk management and asset management. While intuitive and convenient, this practice can turn out to be dangerous, because Pearson's correlation coefficients say nothing about the trend of asset prices. Investors relying exclusively on correlation coefficients to build a diversified portfolio might therefore see all their underlying assets sharing the same trend, despite low or even negative correlations. Our opinion is therefore that additional indicators such as trend gaps, or the difference between the returns of different assets or between two portfolios, should also be taken into consideration when assessing diversification.

References

• Alexander C. (2001), "Market models: a guide to financial data analysis", John Wiley and Sons, England.

• Markowitz H. (1952), "Portfolio selection", *The Journal of Finance*, vol. 7, no. 1, pp. 77-91

• Markowitz H. (1991), "Foundations of Portfolio Theory", Les Prix Nobels 1990, 292, Nobel Foundation, Stockholm.

• Morgan Stanley Smith Barney (2009), "Investment diversification using asset allocation: What investors should consider", Briefing Note, June

• Galton F. (1886), "Regression towards mediocrity in hereditary stature", Journal of the Anthropological Institute of Great Britain and Ireland, vol. 15, pp. 246-263

• Pearson K. (1896), "Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia", Philosophical Transactions of the Royal Society of London, 187, 253318