Correlation: Pitfalls and Alternatives

Paul Embrechts, Alexander McNeil & Daniel Straumann
Departement Mathematik, ETH Zentrum, CH-8092 Zürich
Tel: +41 1 632 61 62, Fax: +41 1 632 15 23
embrechts/mcneil/strauman@math.ethz.ch

March 1999

1 Introduction

Correlation is a minefield for the unwary. One does not have to search far in the literature of financial risk management to find misunderstanding and confusion. This is worrying since correlation is a central technical idea in finance.

Correlation lies at the heart of the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT), where its use as a measure of dependence between financial instruments is essentially founded on an assumption of multivariate normally distributed returns. Increasingly, however, correlation is being used as a dependence measure in general risk management, often in areas where the assumption of multivariate normal risks is completely untenable - such as credit risk. In using correlation as an all-purpose dependence measure and transferring CAPM thinking to general risk management, many integrated risk management systems are being built on shaky foundations.

This article will tell you when it is safe and unproblematic to use correlation in the way that you imagine you can use it, and when you should take care. In particular it will tell you about two fallacies that have claimed many victims. These traps are known to statisticians, but not, we suggest, to the general correlation-using public. We will help you avoid these pitfalls and introduce an alternative approach to understanding and modelling dependency - *copulas*.

The recognition that correlation is often an satisfactory measure of dependence in financial risk management is not in itself new. For instance, Blyth (1996) and Shaw (1997) have made the point that (linear) correlation cannot capture the non-linear dependence relationships that exist between many real world risk factors. We believe this point is worth repeating and amplifying; our aim is to provide *theoretical* clarification of some of the important and often subtle issues surrounding correlation.

To keep things simple we consider only the *static* case. That is, we consider a vector of dependent risks $(X_1, \ldots, X_n)'$ at a fixed point in time. We do not consider serial correlation within or cross correlation between stochastic processes (see Boyer, Gibson, and Loretan (1999) in this context). Nor do we consider the statistical estimation of correlation, which is fraught with difficulty. In this paper we do not go this far because enough can go wrong in the static case to fill an entire book.

We require only simple mathematical notation. A risk X_i has *marginal* distribution $F_i(x) = \mathbb{P}\{X_i \le x\}$. A vector of risks has *joint* distribution function $F(x_1, \ldots, x_n) = \mathbb{P}\{X_1 \le x_1, \ldots, X_n \le x_n\}$. The *linear* correlation of two risks is $\rho(X_i, X_j)$. For full mathematical details of the ideas we describe consult Embrechts, McNeil, and Straumann (1999).

2 When is correlation unproblematic?

As we have indicated, if the risks $X_1, \ldots, X_n$ have a *multivariate normal* distribution then everything is fine. However, it is not enough that each of the risks has a normal distribution, you must be convinced that they have *jointly* a multivariate normal distribution. We can be more general and introduce a class of multivariate distributions where the standard correlation approach to dependency is natural and unproblematic - these are the *elliptical* distributions.

The elliptical distributions (of which the multivariate normal is a special case) are distributions whose density is constant on ellipsoids. In two dimensions, the contour lines of the density surface are ellipses.

The multivariate t–distribution provides a nice example. This distribution has standard univariate t marginal distributions, which are heavy-tailed. For this reason, it is sometimes used instead of the multivariate normal as the distribution of vectors of asset returns. Interestingly, a multivariate t–distribution with uncorrelated components $X_1, \ldots, X_n$ is *not* a distribution with independent components. It provides an example where zero correlation of risks does not imply independence of risks. Only in the case of the multivariate normal can uncorrelatedness always be interpreted as independence.

If $(X_1, \ldots, X_n)'$ has an elliptical distribution, then it is natural to use the correlation matrix as a summary of the dependence structure of the constituent risks. Moreover, in the elliptical world the variance-covariance (Markowitz and CAPM) approach to optimizing portfolios of these risks makes sense. In fact, it turns out that the elliptical distributions constitute a kind of ideal environment for standard risk management where, from a methodological point of view, very little can go wrong. The use of Value-at-Risk (VaR) as a measure of risk is also unproblematic in the elliptical world. We collect these somewhat bold assertions together in a Fundamental Theorem for Risk Management. For a more precise statement of this theorem see Embrechts, McNeil, and Straumann (1999). Note that VaR denotes an abstract quantity in this result - the quantile of a distribution - and that we are not concerned with specific calculation methods such as variance-covariance, Monte Carlo or historical simulation.

Theorem

Suppose $X_1, \ldots, X_n$ have an elliptical distribution and consider the set of linear portfolios $\mathcal{P} = \{Z = \sum_{i=1}^{n} \lambda_i X_i \mid \sum_{i=1}^{n} \lambda_i = 1\}$.

1. VaR is a coherent risk measure in the sense of Artzner, Delbaen, Eber, and Heath (1997). As well as being *monotonic*, *positive homogeneous* and *translation invariant*, it fulfills most importantly the *sub-additivity* property

$$\mathrm{VaR}_\alpha(Z_1 + Z_2) \le \mathrm{VaR}_\alpha(Z_1) + \mathrm{VaR}_\alpha(Z_2),\ Z_1, Z_2 \in \mathcal{P},\ \alpha > 0.5,$$

regardless of the probability level α used in the definition of VAR.

2. Among all portfolios with the same expected return the portfolio minimizing VaR (or any other positive homogeneous, translation invariant risk measure) is the Markowitz variance minimizing portfolio.

The second statement shows that managing risk with VaR is entirely equivalent to managing risk with the variance of the portfolio. To rate the riskiness of the portfolio all we need know are the weights λ_i, the variances of $X_1, \ldots, X_n$ and their *correlation matrix*. No other information on dependence is necessary.

3 When is correlation problematic?

Outside the elliptical world correlation must be used with care. Unfortunately, most dependent real world risks do not appear to have an elliptical distribution (see again Shaw (1997)). For certain market returns multivariate normality or ellipticality might be plausible; it is a much less reasonable model for credit or operational risks. If you are not prepared to believe in elliptical risks then you must be on your guard against fallacious interpretations of correlation, as the following examples show.

You are a risk manager at a major bank, the financial fortunes of which depend critically on just two dependent risk factors X_1 and X_2. Suppose a colleague comes to you and tells you the marginal distributions of these risk factors and their correlation coefficient. Should you be satisfied with this information? Do you now understand the joint risk that X_1 and X_2 represent? Do you know enough to simulate dependent pairs of these risks in a Monte Carlo procedure? If you have answered yes to any of these questions then you have fallen for the first fallacy.

Fallacy 1. Marginal distributions and correlation determine the joint distribution.

This is true in the elliptical world. If we know that X_1 and X_2 have standard normal marginal distributions and a correlation coefficient of 70%, then there is only one possible bivariate elliptical distribution that fits with this information – the standard bivariate normal with correlation 0.7. Outside the elliptical world, there are infinitely many distributions that fit.

This is illustrated in Figure 1. We show simulated data from two distributions that are consistent with the information we have. The first distribution is the bivariate normal and the second is an entirely different joint distribution, whose construction we shall later describe. This mystery distribution has a tendency to generate extreme values of X_1 and X_2 simultaneously and is, in this sense, a more dangerous distribution for risk managers. On the basis of correlation, these distributions cannot be differentiated. Correlation tells us nothing about the degree of dependence in the tail of the underlying distribution. The bivariate normal has rather weak *tail dependence*; the mystery distribution has pronounced tail dependence.

Now suppose your colleague tells you that the logarithm of risk X_1 has a normal distribution with zero mean and unit standard deviation, whilst the logarithm of risk X_2 is normally distributed with zero mean and standard deviation $\sigma = 2$; their correlation is 70%. How should you view this information? In view of the preceding Fallacy you might be worried that the joint distribution has not been uniquely specified. However the problem this time is different. This time there are *no* bivariate distributions that fit this information. Your colleague has described a non-existent model.

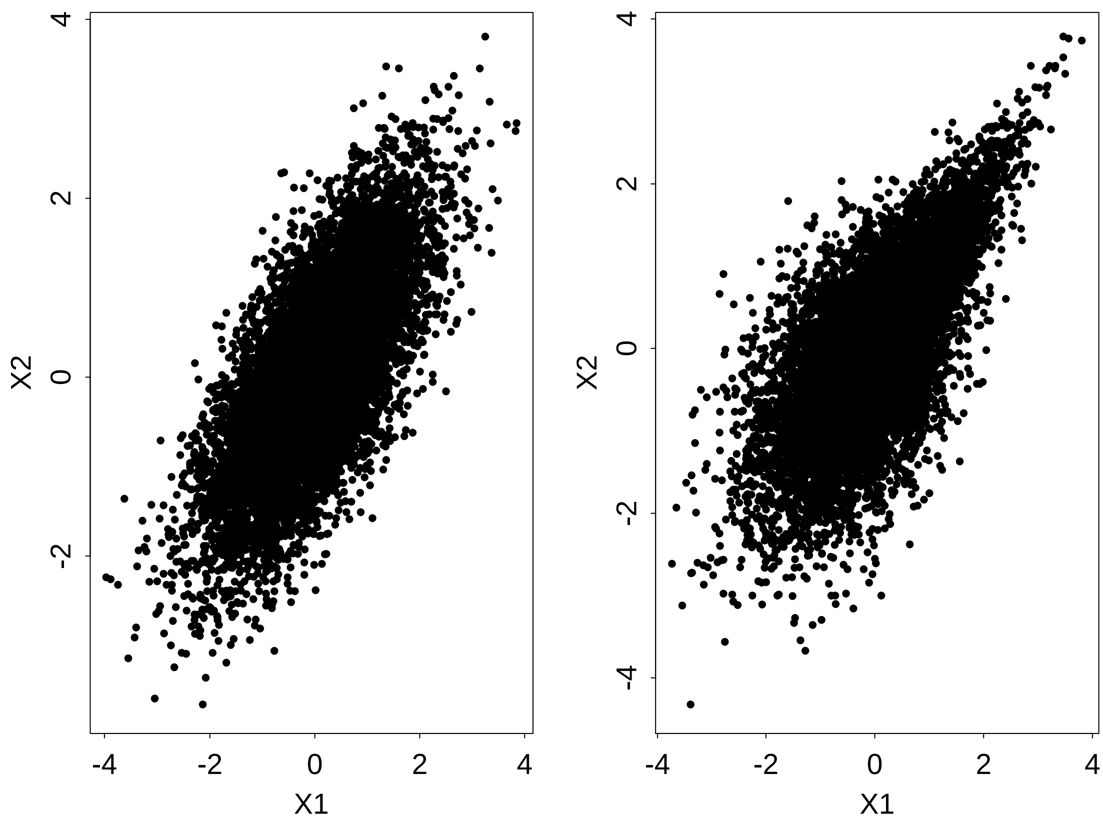

Figure 1: 10000 simulated data from two bivariate distributions with identical normal marginal distributions and identical correlation of 0.7, but different dependence structures.

Fallacy 2. Given marginal distributions F_1 and F_2 for X_1 and X_2, all linear correlations between -1 and 1 can be attained through suitable specification of the joint distribution F.

Again, this is certainly true in the elliptical world but not in general. In general, the *attainable* correlations depend on F_1 and F_2 and form a closed interval $[\rho_{\min}, \rho_{\max}]$ containing zero that is a subset of $[-1, 1]$. For every ρ in $[\rho_{\min}, \rho_{\max}]$ can find a bivariate distribution F for the vector $(X_1, X_2)'$ so that X_1 has distribution F_1, X_2 has distribution F_2 and $\rho(X_1, X_2) = \rho$. In the case of lognormal distributions with parameters as above the attainable interval is $[-0.090, 0.666]$. It is not possible to find a bivariate distribution with the prescribed marginals and a correlation of 0.7.

If we continue with the lognormal example and let σ, the standard deviation of $\log(X_2)$, take a range of values, we can summarize the attainable correlations as in Figure 2. As σ increases the attainable interval $[\rho_{\min}, \rho_{\max}]$ becomes arbitrarily small. We would, however, be very wrong in leaping to the conclusion that this means the dependence of X_1 and X_2 becomes very weak.

The upper boundary of the interval $\rho_{\max}$ always represents a situation where X_1 and X_2 are *perfectly* positively dependent, or *comonotonic*. That is, they can be represented

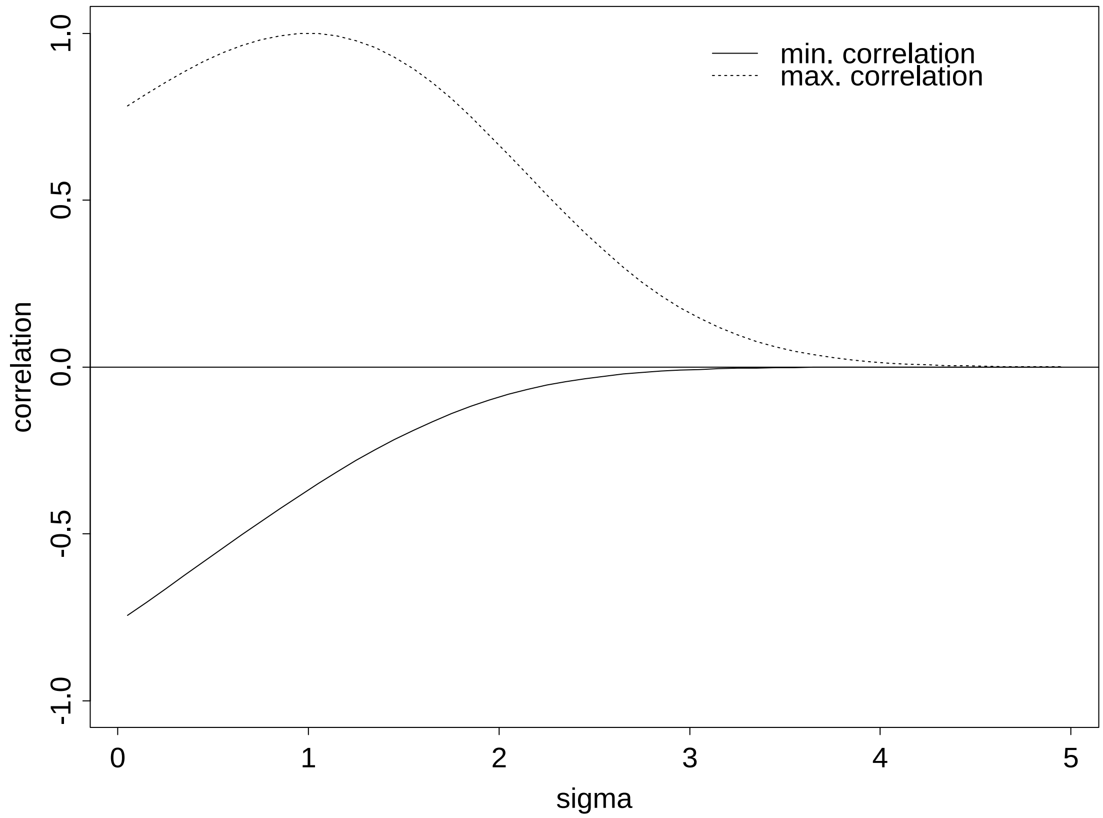

Figure 2: Maximum and minimum attainable correlations for lognormal risks X_1 and X_2, where $\log(X_1)$ is standard normal and $\log(X_2)$ has mean zero and variance σ.

as increasing deterministic functions u and v of a single underlying random risk Z: $X_1 = u(Z), X_2 = v(Z)$. (In fact, if F_1 and F_2 are continuous, they can be represented as $X_2 = w(X_1)$ for an increasing function w.) If X_1 exceeds its VaR at some probability level then X_2 is *certain* to exceed its VaR at the same probability level. The dependence of X_1 and X_2 is by no means weak. Analogously the lower boundary $\rho_{\min}$ always represents a situation where X_1 and X_2 are perfectly negatively dependent. They can be expressed as deterministic functions of an underlying single risk Z, where one function is increasing and the other decreasing.

We would like correlation to assign a value of 1 in the case of perfect positive dependence and a value of -1 in the case of perfect negative dependence. The example shows that this is not always possible. X_1 and X_2 must have identical distributions up to a change of location and scale if it is to be possible.

Here then is a list of the problems of correlation as a dependency measure that we have mentioned, and a few additional ones that you should also be aware of.

1. Correlation is simply a scalar measure of dependency; it cannot tell us everything we would like to know about the dependence structure of risks.

2. Possible values of correlation depend on the marginal distribution of the risks. All values between -1 and 1 are not necessarily attainable.

3. Perfectly positively dependent risks do not necessarily have a correlation of 1; perfectly negatively dependent risks do not necessarily have a correlation of -1.

4. A correlation of zero does not indicate independence of risks.

5. Correlation is not invariant under transformations of the risks. For example, $\log(X)$ and $\log(Y)$ generally do not have the same correlation as X and Y.

6. Correlation is only defined when the variances of the risks are finite. It is not an appropriate dependence measure for very heavy-tailed risks where variances appear infinite.

4 Alternatives to Correlation

By turning to rank correlation, certain of these theoretical deficiencies of standard linear correlation can be repaired. It does not matter whether we choose the Kendall or Spearman definitions of rank correlation. The latter defines rank correlation to be

$$\rho_S(X_i, X_j) = \rho(F_i(X_i), F_j(X_j)),$$

i.e. the linear correlation of probability-transformed variables.

Rank correlation does not have deficiencies 2, 3, 5 and 6. It is invariant under (strictly) increasing transformations of the risks; it does not require that the risks be of finite variance; for arbitrary marginals F_i and F_j a bivariate distribution F can be found with any rank correlation in the interval $[-1, 1]$. But, as in the case of linear correlation, this bivariate distribution is not the unique distribution satisfying these conditions. Deficiencies 1 and 4 remain.

Moreover, rank correlation cannot be manipulated in the same easy way as linear correlation. Suppose we consider linear portfolios $Z = \sum_{i=1}^{n} \lambda_i X_i$ of n risks and we know the means and variances of these risks and their rank correlation matrix. This does not help us to manage risk in the Markowitz way. We cannot compute the variance of Z and thus we cannot choose the Markowitz risk-minimizing portfolio. Our fundamental theorem for risk management in the idealized world of elliptical distributions is of little use to us if we only know the rank correlations. It can however be argued that if our interest is in simulating dependent risks we are better off knowing rank than linear correlations; Fallacy 2 can at least be avoided.

But, ideally, we want to move away from simple scalar measurements of dependence. An important message of this paper is that, in the absence of a model for our risks, correlations (linear or rank) are only of very limited use. On the other hand, if we have a model for our risks $X_1, \ldots, X_n$ in the form of a joint distribution F, then we know everything that is to be known about these risks. We know their marginal behaviour and we can evaluate conditional probabilities that one component takes certain values, given that other components take other values. The dependence structure of the risks is contained within F.

Copulas represent a way of trying to extract the dependence structure from the joint distribution and to extricate dependence and marginal behaviour. It has been shown by Sklar (see Schweizer and Sklar (1983)) that every joint distribution can be written as

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)),$$

for a function C that is known as the copula of F. A copula may be thought of in two equivalent ways: as a function (with some technical restrictions) that maps values in the unit hypercube to values in the unit interval or as a multivariate distribution function with standard uniform marginal distributions. If the marginal distributions of F are continuous then F has a unique copula, but if there are discontinuities in one or more marginals then there is more than one copula representation for F. In either case, it makes sense to interpret C as the dependence structure of F. As with rank correlation, the copula remains invariant under (strictly) increasing transformations of the risks; the marginal distributions clearly change but the copula remains the same.

The Sklar marginal-copula representation of a joint distribution shows how we can construct multivariate models with prescribed marginals. We simply apply *any* copula C (i.e. any multivariate distribution function with uniform marginals) to those marginals. We are now in a position to reveal the identity of the mystery distribution in Figure 1. We took the so-called Gumbel copula

$$C_\beta(u,v) = \exp\left[-\left\{(-\log u)^{1/\beta} + (-\log v)^{1/\beta}\right\}^{\beta}\right],\ 0 < \beta \leq 1$$

and applied it to standard normal marginals $F_1 = F_2 = \Phi$ to obtain the bivariate distribution $F(x_1, x_2) = C_\beta(\Phi(x_1), \Phi(x_2))$. We chose the parameter β so that the linear correlation of X_1 and X_2 was 0.7.

In our opinion copulas represent a useful approach to understanding and modelling dependent risks. In the literature (Joe 1997, Nelsen 1999) there are numerous parametric families of copulas and these may be coupled to arbitrary marginal distributions without worries about consistency. Instead of summarizing dependence with a dangerous single number like correlation we choose a model for the dependence structure that reflects more detailed knowledge of the risk management problem in hand. It is a relatively simple matter to generate random observations from the fully-specified multivariate model thus obtained, so that an important application of copulas might be in Monte Carlo simulation approaches to the measurement of risk in situations where complex dependencies are present.

Acknowledgements

The second author is Swiss Re Research Fellow at ETH Zürich and gratefully acknowledges the financial support of Swiss Re. The work of the third author was supported by RiskLab at ETH Zürich.

References

Artzner, P., F. Delbaen, J. Eber, and D. Heath (1997): "Thinking Coherently," *RISK*, 10(11), 68–71.

Blyth, S. (1996): "Out of line," *RISK*, 9(10), 82–84.

Boyer, B. H., M. S. Gibson, and M. Loretan (1999): "Pitfalls in tests for changes in correlations," International Finance Papers No. 597, Board of Governors of the Federal Reserve System.

Embrechts, P., A. McNeil, and D. Straumann (1999): “Correlation and dependency in risk management: properties and pitfalls,” preprint, ETH Zürich.

Joe, H. (1997): *Multivariate Models and Dependence Concepts.* Chapman & Hall, London.

Nelsen, R. B. (1999): *An Introduction to Copulas.* Springer, New York.

Schweizer, B., and A. Sklar (1983): *Probabilistic Metric Spaces.* North–Holland/Elsevier, New York.

Shaw, J. (1997): “Beyond VAR and Stress Testing,” in *VAR: Understanding and Applying Value at Risk*, pp. 211–224. Risk Publications, London.