

Quantitative Risk Management

A brief course

Giovanni Puccetti

Università degli Studi di Milano

Chapter 1

Correlation: the most common mistakes made in Risk Management practice

Pearson's correlation is the most widely used *and misused* measure of stochastic dependence between two random variables. This chapter aims to highlight some of the most common mistakes in the use of correlation that are made in risk management practice, as both a legacy and an update of the milestone pitfalls paper Embrechts et al. [6], and taking a numerical approach similar to what done in Section 2.6.1 of Hofert et al. [11].

First of all, we recall the definition of Pearson's linear correlation coefficient which we will simply refer to as *correlation* in the remainder. On a given non-atomic probability space, let X and Y be two square integrable random variables (i.e., having finite second moments: $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$). Pearson's correlation coefficient is defined by

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}},$$

where $\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ is the *covariance* between X and Y , $\text{var}(X) = \text{cov}(X, X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ is the *variance* and $\mathbb{E}[X]$ the *expected value* (mean) of X (analogously for Y). In the following, all equalities involving random variables are intended with probability one. The *quantile* F^{-1} of a distribution function F is defined as

$$F^{-1}(u) = \inf \{x \in A : F(x) \geq u\}, \quad u \in (0, 1],$$

and $F^{-1}(0) = \inf \{x \in A : F(x) > 0\}$. If F is continuous and strictly increasing, F^{-1} coincides with its unique inverse; see Embrechts and Hofert [5].

The original scope of Pearson's correlation as pursued for instance in the early studies of Hoeffding [9] (translated in Hoeffding [10]) and Fréchet [7], was to provide a measure of similarity (association) between random variables. It is useful then to see how a higher correlation in fact corresponds to a higher similarity. Suppose that two square integrable random variables X and Y have fixed distribution F and, respectively, G (in what follows we denote this by $X \sim F$ and $Y \sim G$). Their covariance can be written as

$$\text{cov}(X, Y) = \mathbb{E}[XY - X\mathbb{E}[Y] - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Since the first moments $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ are given by the fixed distributions (as well as $\text{var}(X)$ and $\text{var}(Y)$), the only correlation term that depends on the similarity of the two random variables is $\mathbb{E}[XY]$. Why the expectation of the product plays such a role? This can be explained via the formula

$$\mathbb{E}[(X - Y)^2] = \mathbb{E}[X^2] - 2\mathbb{E}[XY] + \mathbb{E}[Y^2], \quad (1.1)$$

from which it follows that the maximization of the expected product of two random variables is equivalent to the minimization of the average squared Euclidean distance between them (this holds analogously in multivariate spaces; see Puccetti [14]). The higher the expected product, the less the expected distance between the two random variables. This is why correlation in principle works as a measure of similarity.

We believe that most misunderstandings about correlation originate from the fact that correlation is always told to assume values between -1 and 1 , these bounds identifying the case of perfect negative and perfect positive *linear* dependence. Though this statement is formally correct, confusion stems from the fact that these bounds are not always achieved and do not represent in general the minimum and maximal attainable values of correlation.

Moreover, it is well known that if X and Y are square integrable and stochastically independent, one has $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and correlation is zero. However, the converse is not true, and in a strong sense. In the absence of further assumptions, a value of (sample) correlation close to zero does not have any relevant implication, and might derive from independence as well from perfect positive or negative dependence.

Correlation finds its ideal environment in a family of fully specified statistical models closed under linear combinations, such as the multivariate Normal (Gaussian) or Student's t distributions. In the lack of such (strong) assumptions, mistakes and wrong interpretations of correlation values may arise.

To avoid some of the potential mistakes and misinterpretations that we discuss in what follows, we strongly suggest a straightforward alternative to correlation, which is *rank correlation*. To be safe in real practice, one should not compute correlation but correlation between *ranks*, something can always be done at very low computational extra effort. Rank correlation does not fix all problems of correlation, but it overcomes many: it is a quick, viable and better alternative to plain correlation.

1.1 Common mistakes

Linear correlation has an appealingly simple definition, and can be easily computed or estimated. However, its definition is based on the assumption that the random variables involved have finite second moments, that is their variances (and therefore their means) have a finite value. Overlooking this assumption is one of the most common mistakes made in practice: as all assumptions in mathematics, square integrability might be satisfied or not.

Mistake 1: “We can always estimate correlation”

First notice that requiring the finiteness of the variances implicitly implies that $\rho(X, Y)$ depends on the marginal distributions of X and Y , something that is intrinsically against the scope of correlation, which should measure association of two random variables regardless of their distribution or of the scale upon which they have been measured. More on this will follow in Mistake 2!

Second, it is not always statistically obvious whether data are coming from a distribution having a finite variance or not. Suppose that one has to estimate the correlation between two random variables $X \sim F$ and $Y \sim G$ based on a number N of independent joint observations

$$(x_i, y_i), i = 1, \dots, N,$$

coming from the random vector $(X, Y)'$. How can one realize whether $\rho(X, Y)$ actually exists? Computing sample correlation, an immediate operation in R, does not answer the question since sample correlation is always finite, even when data are coming from a infinite variance distribution. A more in-depth investigation is then called for.

In order to check whether the moments of the underlying distributions exist or not, a very handy mathematical tool is the so-called *Mean-Excess (ME) function*. If $\mathbb{E}[X]$ is finite, the mean excess function $e(u)$ is defined by

$$e(u) = \mathbb{E}[X - u \mid X > u],$$

that is as the mean of exceedances of a random variable X above a given threshold u . The ME function is a distinctive mark of a distribution, like a fingerprint: each distribution has its own specific ME function and they all look different. For the x -data x_1, \dots, x_N , the sample version of the ME function is easily computed as

$$e_N(u) = \frac{\sum_{i=1}^N (x_i - u) I_{\{x_i > u\}}}{\sum_{i=1}^N I_{\{x_i > u\}}},$$

which is simply the sum of the exceedances above the threshold u divided by the number of data points exceeding u . The scatter plot of the ordered data

$$\{x_{(i)}, e_N(x_{(i)}), i = 1, \dots, N - 1\}$$

is called the *ME-plot*. In practice, instead of averaging the observations in a sample, the ME function/plot *averages their exceedances* with respect to a given threshold.

From the ME-plot of the data one can quickly retrieve relevant information for the random variable under study in the blink of an eye. A simple rule-of-thumb arises from Figure 1.1, which shows the shape of the ME function for some commonly used distributions. If the ME-plot is decreasing or constant, then the all moments of the distribution are likely to be finite (like in the Normal and Exponential cases); if the ME-plot is increasing or looks uncertain, it is necessary to perform a deeper statistical investigation before estimating moments-based statistics.

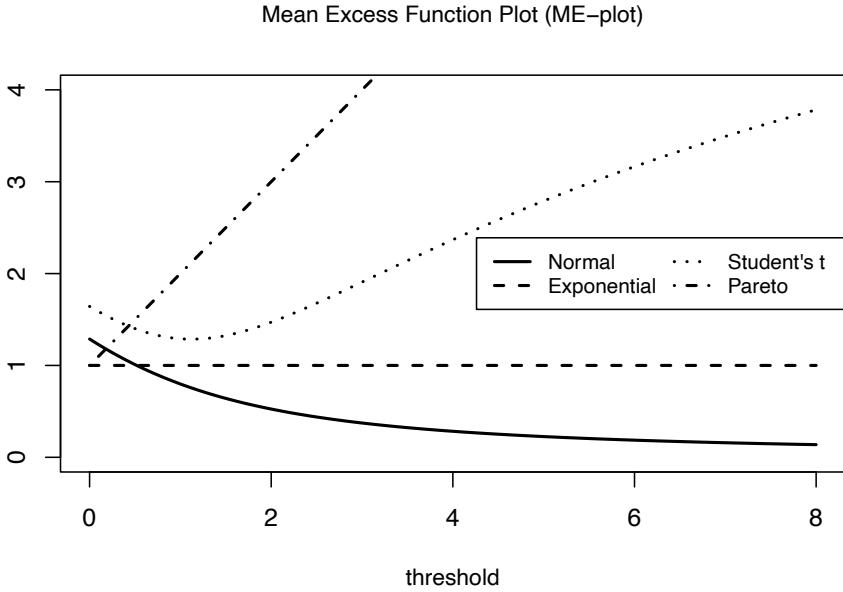


Figure 1.1: Mean Excess (ME) function for some distributions having mean 1. The t and the Pareto distributions considered here do not possess a finite variance.

As an example, we generate a set of data (x_i, y_i) from two independent distributions: a Normal distribution for the x -data and a Student's t distribution for the y -data.

```
set.seed(100) # seed for reproducibility
N=500 # number of simulations
## Generate samples from X and Y s.t.
## mean and sample variance are approximately the same
Y=rt(N,df=2) # t sample
X=rnorm(N,mean=0,sd=sqrt(var(Y))) # Normal sample
## Compute sample mean of X and Y
c(mean(X),mean(Y))
```

```
## [1] -0.02258291  0.22845815
```

```
## Compute sample variance of X and Y
c(var(X),var(Y))
```

```
## [1] 11.18967 11.91335
```

```
## Compute sample correlation between X and Y
cor(X,Y)
```

```
## [1] 0.02138129
```

The two data samples have approximately the same sample mean and sample variance, and sample correlation close to 0. Even if the two random variables generating the sample are stochastically independent, we will see in this example (and in the following ones) that a value of sample correlation close to zero does not have any practical implication on correlation.

Figure 1.2 illustrates the corresponding MEplot for the x - and y -data. The increasing ME-plot for the y -data raises a warning on the possibility that the corresponding distribution might not possess a finite variance.

The x -sample has been generated from a standard Normal $N(0, 1)$ distribution, for which all moments (not only the first and the second) exist, whereas the y -sample is coming from a Student's t with $\nu = 2$ degrees of freedom, a distribution which does not possess a finite variance. While one, based on sample evidence, would be prompted to estimate a zero correlation between X and Y , $\rho(X, Y)$ actually *does not exist*. This is a message that holds for any statistics (in particular for the mean of a distribution):

Key message: Before estimating something, make sure that something exists!

```
library(QRM) # we use QRM library for the MEplot function
## Set scale for MEplots (three largest observations are omitted)
m=0 # lower axis limit
M=max(sort(X)[N-3],sort(Y)[N-3]) # upper axis limit
## Create ME plots for x- and y- data
par(mfrow = c(1, 2),pty="s",font.main=1)
MEplot(X,xlim=c(0,M),ylim=c(0,M),xlab="",ylab="",main="Normal",
col = "black",lwd = 2,cex.lab = 1.2,cex.main = 0.9,cex.lab = 0.9)
MEplot(Y,xlim=c(0,M),ylim=c(0,M),xlab="",ylab="",main="Student's t",
col = "black",lwd = 2,cex.lab = 1.2,cex.main = 0.9, cex.lab = 0.9)
```

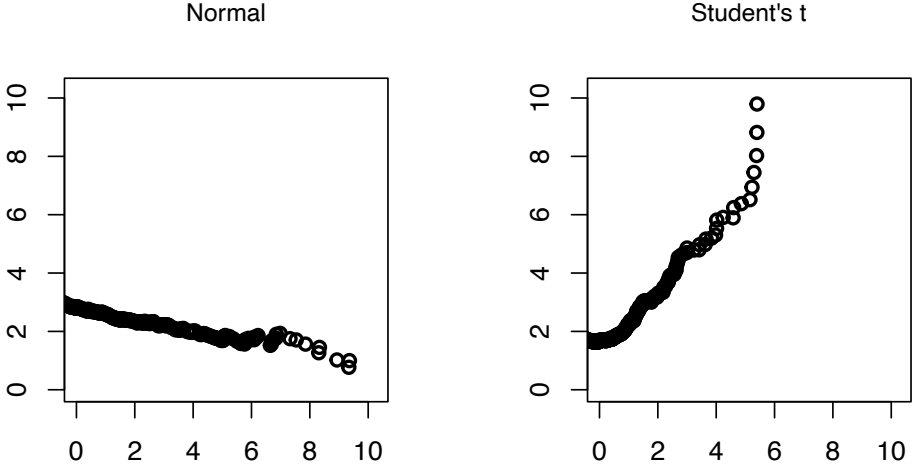


Figure 1.2: Mean Excess plot for two set of simulations having approximately the same sample mean and sample variance. The three largest observations are omitted.

Mistake 2: “We estimated a moderate value of correlation so the risks are moderately dependent”

What are the minimum and maximum values attainable by correlation? Apparently there is an immediate answer to this question: -1 and 1 ! Unfortunately, this is not (always) correct. In general, these values depend on the marginal distributions of the random variables considered and, surprisingly enough, they can be arbitrarily close to zero.

We give a very simple but enlightening example. Consider a random variable $X_1 \sim F$ and a second random variable Y_1 being a linear function of X_1 , namely $Y_1 = a X_1$ for some $a > 0$. In this case, the linear correlation between them is given by

$$\begin{aligned} \rho(X_1, Y_1) &= \frac{\text{cov}(X_1, aX_1)}{\sqrt{\text{var}(X_1)}\sqrt{\text{var}(aX_1)}} = \frac{a \text{cov}(X_1, X_1)}{\sqrt{\text{var}(X_1)}\sqrt{a^2 \text{var}(X_1)}} \\ &= \frac{a \text{var}(X_1)}{|a|\sqrt{\text{var}(X_1)}\sqrt{\text{var}(X_1)}} = 1. \end{aligned} \quad (1.2)$$

In fact, it is easy to show that $\rho(X, Y) = 1$ if and only if Y is a positive linear function of X (as in the previous case) and $\rho(X, Y) = -1$ if and only if Y is a negative linear function of X (e.g. if in (1.2) one has $a < 0$). These lower and upper bounds represent the case of perfect positive and negative *linear* dependence and justify the name of *linear correlation*. In the following example, this is confirmed when X_1 is a Normal random variable, but one could choose an arbitrary square integrable distribution.

```

set.seed(100) # seed for reproducibility
N=10^6 # number of simulations
X1=rnorm(N) # standard Normal N(0,1) sample
a=3 # standard deviation of Y
Y1=a*X1 # Y1 is a times X1 so has df N(0,a^2)
## Compute sample correlation between X1 and Y1
cor(X1,Y1)

```

```
## [1] 1
```

Correlation is able to detect perfect linear dependence. On the other hand, correlation is short-sighted when non-linear dependency are brought into play. To see this, consider the random variables $X_2 = \exp(X_1)$ and $Y_2 = \exp(Y_1)$ (the exponential can be replaced by an arbitrary strictly increasing *non linear* function $f(X_1)$ and $f(Y_1)$). Applying the exponential to X_1 and Y_1 simply means that they are now measured on a different scale, so one should pretend or at least expect that the correlation between them remains unchanged. Unfortunately, this is not true and the correlation coefficient might be arbitrarily close to 0.

```

X2=exp(X1) # change of scale for X1 (which becomes LogNormal)
Y2=exp(Y1) # change of scale for X2 (which becomes LogNormal)
## Compute sample correlation between X2 and Y2
cor(X2,Y2)

```

```
## [1] 0.2782197
```

A relatively small value of correlation is intuitively interpreted as a modest association between the random variables, which instead are in this case perfectly dependent (via a strictly increasing function!).

This misunderstanding originates from the naive assumption that, since correlation always lies between -1 and 1 , then a moderate value of correlation (like the one above computed) should reflect a moderate dependence between the risks. This is unfortunately not true, unless it is possible to establish a linear dependence between the random variables. This is true for two Normal variables, but not for two LogNormal ones (apart from when they have exactly the same distribution).

So, what are the minimum and maximal values of linear correlation, and when are they attained? According to the formula (1.1), the higher the expected product $\mathbb{E}[XY]$, the less the expected distance between the two random variables X and Y . The question then becomes, for $X \sim F$ and $Y \sim G$, when does $\mathbb{E}[XY]$ attain its minimal and maximal values? The answer goes back to the milestone result given in Theorem 368 in Hardy et al. [8], where it is proved that the scalar product of two vectors is maximal when the components of the two vectors are similarly ordered (i.e., they are monotonic

in the same sense, or comonotonic), and minimal when they are oppositely ordered (i.e., they are countermonotonic).

If F and G are continuous, via approximations of continuous distributions with discrete ones, one obtains that: $\mathbb{E}[XY]$ is maximized when Y is (almost surely) an increasing function of X (we say that X and Y are *comonotonic*); minimized when Y is (a.s.) a decreasing function of X (we say that X and Y are *countermonotonic*); see for instance Puccetti and Wang [15] for more details and a history of this result.

Theorem 1.1. *Assume that $X \sim F$ and $Y \sim G$, with F and G continuous and square integrable. Then we have that:*

- a) $\mathbb{E}[XY]$ is maximized when $Y = G^{-1}(F(X))$;
- b) $\mathbb{E}[XY]$ is minimized when $Y = G^{-1}(1 - F(X))$.

As a consequence, the range of possible values of $\rho(X, Y)$ forms a closed interval $[\rho_{\min}, \rho_{\max}]$ with $\rho_{\min} < 0 < \rho_{\max}$. Since the distributions F, G , and the corresponding quantile functions F^{-1}, G^{-1} , are increasing, we have that:

- the maximum correlation ρ_{\max} is attained if and only if X and Y are comonotonic;
- the minimum correlation ρ_{\min} is attained if and only if X and Y are countermonotonic.

Based on Theorem 1.1, linear correlation (in case exists) always attains its minimum value under countermonotonicity and maximum value under comonotonicity. The attained extreme values, however, are equal to -1 and 1 *only in the case* when the functional relationships $G^{-1}(F)$ and $G^{-1}(1 - F)$ are *linear*. This is true for instance when F and G are assumed to belong to the same family of distributions closed under affine transformations, like in the case when the vector (X, Y) is assumed to have a multivariate Normal distribution.

In the general case when F is not continuous, Theorem 1.1 holds analogously. In general, X and Y are said to be comonotonic if they are both increasing function of a common random factor U and they admit the representation $X = F^{-1}(U)$, $Y = G^{-1}(U)$ for a uniform random variable $U \sim U(0, 1)$. Analogously, X and Y are said to be countermonotonic if and only if X is an increasing, Y a decreasing function of U : $X = F^{-1}(U)$, $Y = G^{-1}(1 - U)$; see Theorem 2 in Dhaene et al. [4].

Based on Theorem 1.1, we have that the maximum value of correlation is given by

$$\rho_{\max}(X, Y) = \int_{\mathbb{R}} x G^{-1}(F(x)) f(x) dx = \int_0^1 F^{-1}(u) G^{-1}(u) du, \quad (1.3)$$

where the second integral comes from the change of variable $u = F(x)$. Analogously, the minimal value of correlation is given by

$$\rho_{\min}(X, Y) = \int_{\mathbb{R}} x G^{-1}(1 - F(x)) f(x) dx = \int_0^1 F^{-1}(u) G^{-1}(1 - u) du. \quad (1.4)$$

As it is illustrated in Figure 1.3, in the case of a Normal and a Student's *t* distribution correlation *never* attains the values -1 and 1 , and both maximal and minimal correlation goes to zero when the number of degrees of freedom of the *t* distribution approaches 2 from above.

```
## Parameters of X
meanx=0 # mean of X
varx=1 # variance of X
invx<-function(x){qnorm(x)} # Normal quantile function of X
N=100 #resolution of the correlation plot
dof=seq(2,3,len=N) # points at which to compute correlation
maxcor=seq(0,0,len=N) # vector to store max correlation
mincor=seq(0,0,len=N) # vector to store min correlation
## Loop to compute coordinates to plot
for(i in 1:N) {
  ## Parameters of Y
  meany=0 # mean of Y
  dofy=dof[i] # dof of Y
  vary=dofy/(dofy-2) # variance of Y
  invy<-function(x){qt(x,dofy)} # t quantile function of Y
  ## Compute max/min correlation via numerical integration
  f<-function(u){invx(u)*invy(u)}
  g<-function(u){invx(u)*invy(1-u)}
  maxcor[i]=(as.numeric(integrate(f, lower = 0,
  upper=1)[1])-meanx*meanx)/(sqrt(varx*vary))
  mincor[i]=(as.numeric(integrate(g, lower = 0,
  upper = 1)[1])-meanx*meanx)/(sqrt(varx*vary))
}
## Max/Min correlation plot
matplot(dof,cbind(maxcor,mincor),xlim=c(2,3),ylim=c(-1,1),
type="l",lty=c(1,2),xlab="dof",ylab=" ",
main="Attainable Correlations",col = "black",
lwd = 2,font.main=1,cex.main=0.9,cex.lab=0.9)
abline(h=1,lty=5)
abline(h=-1,lty=5)
abline(h=0,lty=5)
legend("right",legend=c("Max correlation","Min correlation"),
lty =c(1,2),col="black")
```

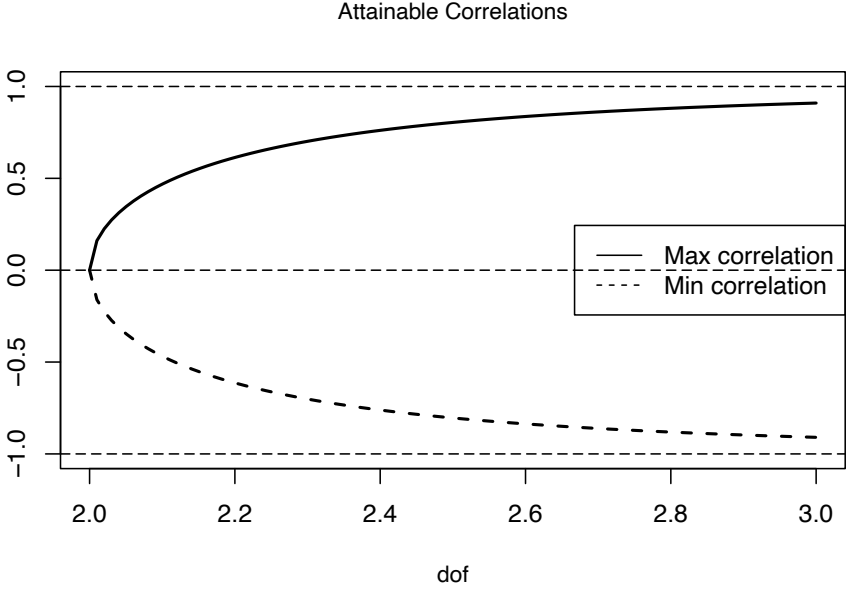


Figure 1.3: Maximal (solid curve) and minimal (dashed curve) values attainable by correlation between a standard Normal and a standard t distribution, computed versus the number of degrees of freedom of the t distribution.

Example 1.1. The example where X_2 and Y_2 are LogNormal distributions can be treated analytically as done in McNeil et al. [13] (Example 7.29).

Assume that $X_1 \sim F = N(0, 1)$ and $Y_1 = \sigma X_1$ for some $\sigma > 0$, hence $Y_1 \sim G = N(0, \sigma^2)$. The two distributions are *of the same type*, that is it is possible to establish a linear relationships between them. Either a positive one, by setting $Y_1 = \sigma X_1$ and in this case $\rho(X_1, Y_1) = 1$, either a negative one, by setting $Y_1 = -\sigma X_1$, and in this case $\rho(X_1, Y_1) = -1$; see (1.2). As a consequence of the symmetry of the standard Normal distribution, correlation in this case truly belongs to the interval $[-1, 1]$.

If a change of scale occurs, things change. Assume that the random variables X_1 and Y_1 are transformed into $X_2 = e^{X_1} \sim \text{LogN}(0, 1)$ and $Y_2 = e^{Y_1} = e^{\sigma X_1} \sim \text{LogN}(0, \sigma^2)$.

Even if the dependence structure between the two random variables is unchanged, the different scale makes it impossible to establish liner relationships as in the previous case. The only case where this is possible is $\sigma = 1$, when the two random variables X_2 and Y_2 are identically distributed and it is possible to set $Y_2 = X_2$ (but *not* $Y_2 = -X_2$). As a result, for $\sigma \neq 1$, the correlation range will be strictly smaller than $[-1, 1]$.

To calculate the attainable value of correlation, observe that Y_1 and Y_2 are still comonotonic. Clearly, $\rho_{\max} = \rho(e^{X_1}, e^{\sigma X_1})$ and, by a similar argument, $\rho_{\min} = \rho(e^{X_1}, e^{-\sigma X_1})$.

Analytical calculations easily follow (see Exercise 1.1) and yield

$$\rho_{\max} = \frac{e^{\sigma} - 1}{\sqrt{(e - 1)(e^{\sigma^2} - 1)}}, \quad \rho_{\min} = \frac{e^{-\sigma} - 1}{\sqrt{(e - 1)(e^{\sigma^2} - 1)}}. \quad (1.5)$$

Figure 1.4 shows the range of attainable correlation values for different values of σ . As previously noted, for $\sigma = 1$ one has $\rho_{\max} = 1$ and $\rho_{\min} > -1$.

Also, the range rapidly narrows to a neighborhood of zero as σ gets larger. Table 1.1 shows, for $\sigma = 5$, that the maximal value attained by correlation is $\rho_{\max} = 0.0004191$, while the minimal is $\rho_{\min} = -0.0000028$.

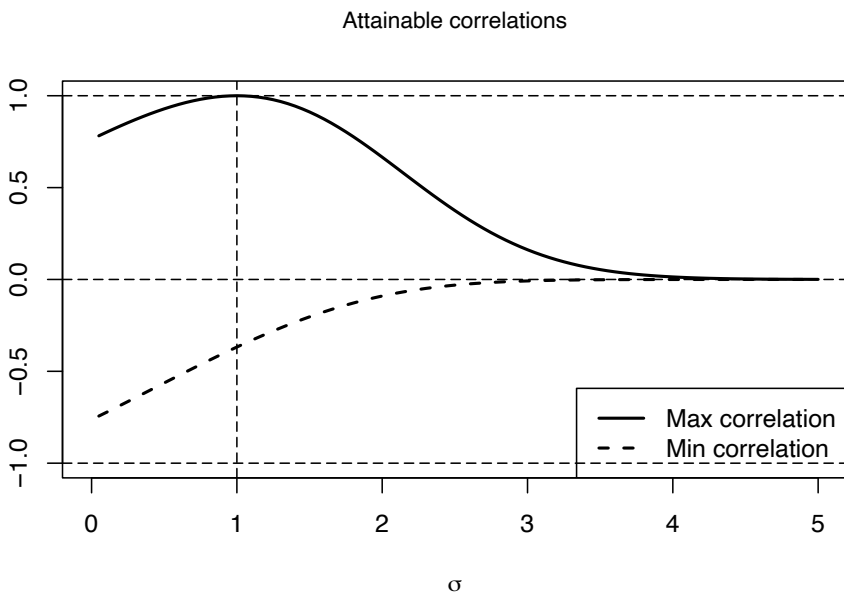


Figure 1.4: Maximal (solid curve) and minimal (dashed curve) values attainable by correlation between two LogNormal random variables $Y_1 \sim \text{LogN}(0, 1)$ and $Y_2 \sim \text{LogN}(0, \sigma^2)$, computed versus the value of σ .

Figure 1.5 summarizes the above results: the maximum value of correlation is attained in the case of perfect positive dependence (comonotonicity), and the minimum value of correlation is attained in the case of perfect negative dependence (countermonotonicity). However, the corresponding attained values are -1 and 1 if and only if these dependency are linear, otherwise the minimal value will be strictly larger than -1 and the maximal value strictly smaller than 1 .

Table 1.1: Minimal and maximal attainable correlation between two Log-normal random variables $Y_1 \sim \text{LogN}(0, 1)$ and $Y_2 \sim \text{LogN}(0, \sigma^2)$ for some values of σ .

σ	Minimal Correlation	Maximal Correlation
1	-0.3678794	1.0000000
2	-0.0901002	0.6657553
3	-0.0080533	0.1617554
4	-0.0002512	0.0137166
5	-0.0000028	0.0004191

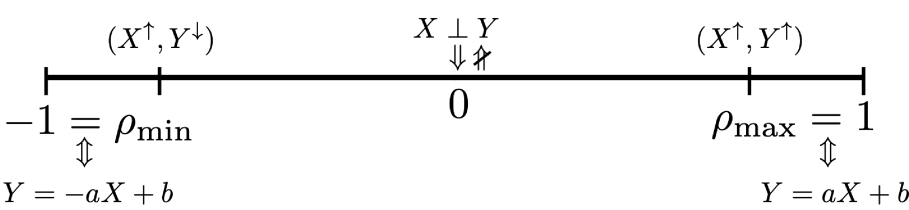


Figure 1.5: Range of attainable correlation ($a > 0$).

Fortunately, there is a quick remedy to correct the scale of attainable values of correlation. If one wants to preserve the value of correlation against increasing transformations of the random variables involved, it is sufficient to compute sample correlation *between ranks*. This corresponds to the estimation of Spearman rank correlation coefficient; see Hofert et al. [11] for a definition.

Rank correlation will always exist and be between -1 and 1 , and the two bounds will correspond to countermonotonicity and comonotonicity. In the previous example, when encountering a moderate value of rank correlation one can at least say that the random variables involved are not perfectly associated (something which would remain uncertain if plain correlation is used).

In principle, the original sin of correlation lies in its denominator, which standardizes correlation only with respect to the value assumed in case of perfect linear dependence. When linear dependence cannot be attained, the correct scaling is lost. Another way of using the full range of values in $[-1, 1]$ for correlation values would be that of standardizing $\text{cov}(X, Y)$ by the values (1.3) and (1.4) attained by correlation under comonotonicity and countermonotonicity (though in general these require integration and they are not symmetric): this is the approach taken in Ai [1]. Also for random vectors in a multivariate space, computing correlation between ranks seems a safe alternative; see Puccetti [14].

```
## Compute correlation between ranks
Y1<-rank(Y1) # ranks of Y1
Y2<-rank(Y2) # ranks of Y2
```

```
## Compute sample correlation between ranks
cor(Y1,Y2)
```

```
## [1] 1
```

Key message: Do not use correlation, use correlation *between ranks*!

Mistake 3: “We estimated an almost zero correlation so the risks are independent”

It is well known that if X and Y have finite variances and are independent, then $\rho(X, Y) = 0$. This implication is typically a source of confusion, because one naively associates a sample correlation close to zero with independence. However, the converse implication is false: uncorrelatedness in general does not imply independence and, actually, does not deliver a particular information unless a full joint model for (X, Y) is specified.

The classical example is to assume that $X \sim N(0, 1)$ and $Y = X^2$: X and Y are functionally dependent since the value of X determines the value of Y , nevertheless $\rho(X, Y) = 0$. We illustrate this in the following pedagogical example.

```
set.seed(100) # seed for reproducibility
N=10^6 # number of simulations
X=rnorm(10^6) # standard Normal sample
Y=X^2 # Y is X squared (so not independent from X)
cor(X,Y) # sample correlation between X and Y
```

```
## [1] -0.002683984
```

In general, we can give the following proposition as a more general counterexample.

Proposition 1.1. *Let the random variable X be symmetric (around zero) and assume that $\mathbb{E}[X^{2k+1}]$ is finite for some integer k . Then*

$$\rho(X, X^{2k}) = 0.$$

Proof. Being X symmetric, notice that $\mathbb{E}[X^{2i+1}] = 0$, $0 \leq i \leq k$. Hence

$$\begin{aligned} \text{cov}(X, X^{2k}) &= \mathbb{E}[(X - \mathbb{E}X)(X^{2k} - \mathbb{E}[X^{2k}])] = \mathbb{E}[X(X^{2k} - \mathbb{E}[X^{2k}])] \\ &= \mathbb{E}[X^{2k+1}] - \mathbb{E}[X]\mathbb{E}[X^{2k}] = \mathbb{E}[X^{2k+1}] = 0, \end{aligned}$$

therefore we have that $\rho(X, X^{2k}) = 0$. □

Unfortunately, not even computing rank correlation would help in the above case.

```
cor(rank(X),rank(Y)) # sample rank correlation between X and Y
```

```
## [1] 0.002258707
```

Indeed, while independence between two random variables is sufficient for correlation and for rank correlation to vanish, it is not necessary. It would be desirable to have a dependence measure being -1 if and only if the random variables are countermonotonic, 1 if and only if the random variables are comonotonic, and 0 if and only if the random variables are independent (notice that only the last “only if” is not satisfied by rank correlation). Unfortunately, as shown in Scarsini [17] and Section 4.2 in Embrechts et al. [6], a sensible dependence measure with such properties does not exist.

As a consequence, one should not produce implications from having a low value of sample correlation. A sample correlation value close to zero could have been produced by independence, or by perfect positive dependence, or even by perfect negative dependence. At least, with rank correlation, a value smaller than 1 (bigger than -1) implies that the random variables considered are not perfectly positively (negatively) dependent.

Key message: Correlation is zero, so . . . what?

Mistake 4: “The VaR for the sum of Normal risks is maximal under maximal correlation”

Value-at-Risk (VaR) is one of the most popular and widely used risk measures in banking, finance and insurance. The *Value-at-Risk* of a loss random variable (risk) $X \sim F$, computed at the probability level $\alpha \in (0, 1)$, is the α -quantile of its distribution F , that is

$$\text{VaR}_\alpha(X) = F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}.$$

If F is continuous and strictly increasing, then $q = \text{VaR}_\alpha(X)$ is the unique real number such that $F(q) = \alpha$, so $\text{VaR}_\alpha(X)$ is usually interpreted as the minimum amount of money that will be sufficient (under the prescribed model F) to cover the losses coming from X with probability α . Coherently with this interpretation, in risk management regulation the confidence level α is usually chosen close to 1 , e.g. $\alpha = 0.95, 0.975, 0.99, \dots$, depending on the typology of the underlying risk X .

The VaR for a sum is straightforward to compute in the case the random variables are comonotonic. It is a simple probability exercise to show the following theorem (which holds analogously for the sum of an arbitrary number of random variables).

Theorem 1.2. Assume that $X \sim F$ and $Y \sim G$ are comonotonic. Then, for $\alpha \in (0, 1)$, we have that

$$\text{VaR}_\alpha(X + Y) = \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) = F^{-1}(\alpha) + G^{-1}(\alpha). \quad (1.6)$$

Proof. Assume that $X \sim F$ and $Y \sim G$ are comonotonic. Then, they admit the representation $X = F^{-1}(U)$, $Y = G^{-1}(U)$, with $U \sim U(0, 1)$. Since F^{-1} and G^{-1} are increasing, then also the function $\phi = F^{-1} + G^{-1}$ is increasing. Recalling that $\text{VaR}_\alpha(\phi(U)) = \phi(\text{VaR}_\alpha(U))$ for any increasing function ϕ , one obtains that

$$\begin{aligned} \text{VaR}_\alpha(X + Y) &= \text{VaR}_\alpha(F^{-1}(U) + G^{-1}(U)) = \text{VaR}_\alpha(\phi(U)) = \phi(\text{VaR}_\alpha(U)) \\ &= \phi(\alpha) = F^{-1}(\alpha) + G^{-1}(\alpha) = \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y). \end{aligned}$$

An analogous proof holds for an arbitrary number of comonotonic random variables. \square

Since the comonotonic case corresponds to maximal correlation, it is a common (and naive) conclusion to believe that the above formula provides the maximal attainable VaR. Unfortunately, this is in general not true.

Another concern regards the set-up of a risk management framework. Very often, the assumption of having two Normal random variables $X \sim N(0, 1)$ and $Y \sim N(0, 1)$ (“We have Normal risks”) is confused with the much stronger assumption that the random vector $(X, Y)'$ is Normal (“We have a Normal *vector*”). This latter assumption, compared to the first, includes the typology of (Normal) dependence between the risks. If one considers the weaker assumption of having two random variables individually distributed like a standard Normal distribution, formally $X \sim N(0, 1)$ and $Y \sim N(0, 1)$, the maximum attainable VaR and the joint distribution attaining it can be computed analytically (see Makarov [12], Rüschendorf [16], and Wang and Wang [18] for more than two random variables) or numerically via the *Rearrangement Algorithm*. Denoting by Φ the standard Normal distribution one has that the worst VaR for the sum of two standard Normal risk is equal to

$$\text{VaR}_\alpha(X^* + Y^*) = 2\Phi^{-1}\left(\frac{\alpha + 1}{2}\right),$$

while the comonotonic VaR, according to (1.6), is smaller and given by

$$\text{VaR}_\alpha(X^C + Y^C) = 2\Phi^{-1}(\alpha);$$

see a numerical comparison in Table 1.2.

Table 1.2: Comonotonic VaR and worst-case VaR for the sum of two standard Normal random variables, computed at some typically used (high) values of α .

α	Comonotonic VaR	Worst-case VaR
0.900	2.563103	3.289707
0.950	3.289707	3.919928
0.990	4.652696	5.151659
0.995	5.151659	5.614067

In Figure 1.6, left, we represent the bivariate distribution of (X, Y) which maximizes $\text{VaR}_\alpha(X + Y)$ for $\alpha = 0.90$.

In order to maximize the α -quantile of a sum, only the largest $(1 - \alpha)$ part of the marginal probability mass is relevant, so it does not matter where the smallest α probability mass is spread on the left part of the figure, provided it is coherent with the assumed margins.

The relevant $(1 - \alpha)$ part of the probability mass, spread on the right part of the figure, is used to make the sum of the random variables be (as close as possible to) a constant. In fact, the probability of exceeding the α -quantile of a distribution is $(1 - \alpha)$. If more probability is concentrated around (or exactly in) a single point, this will help to pull the quantile further in the right part of the distribution. In the case of Normal marginals, it is not possible to obtain exactly a constant in this part so the one in the figure is the joint distribution closest to it.

As a result, the solution in Figure 1.6 is a mix of positive (on the left) and negative (on the right) dependence: it is then not surprising that the implied value of correlation between X and Y is not maximal, i.e., is smaller than one (in this case of identical marginals, the maximal attainable correlation is truly 1). Moreover, since the smallest α probability mass on the left of the figure can be spread arbitrarily, one could also achieve smaller values of correlations.

The solution in Figure 1.6 is typically regarded by practitioners as “too weird to be true”. However, from a mathematical perspective (but are others possible?), the maximization of VaR is an optimization problem. Unless further assumptions are imposed, Figure 1.6 represents its solution.

Figure 1.6 also remarks the fact that assuming the individual marginals of X and Y , and fixing the value of correlation $\rho(X, Y)$, is not sufficient to univocally identify a joint distribution. This is exactly why it is so important to clearly specify the assumptions used within a risk management framework.

Assuming that the vector $(X, Y)'$ (notice the brackets) has a *bivariate* Normal distribution (Figure 1.6, right) would immediately wipe off the previous worst-case solution. However, even with this stronger assumption, we still need an extra hypothesis to guarantee that the VaR is maximized under comonotonicity: the confidence level α has to be larger than 0.5, as the following example shows.

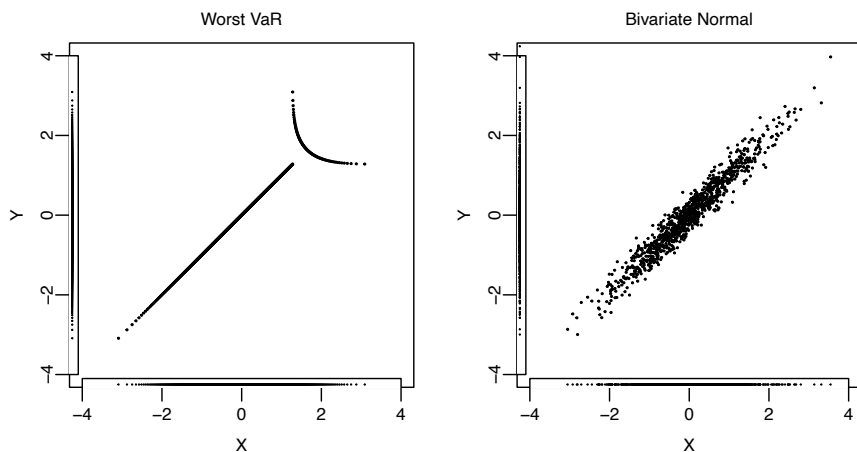


Figure 1.6: 2000 simulated points from two different bivariate distributions with the same standard Normal marginals and the same value of correlation. The left picture shows data from the distributions attaining the maximum VaR for the sum of its components at the confidence level $\alpha = 0.90$; the right picture shows a bivariate Gaussian distribution with the same value of correlation between its components.

Example 1.2. Suppose that the vector $(X, Y)'$ has a bivariate Normal distribution with standard Normal marginals Φ . If X and Y are assumed to be comonotonic, then correlation is maximal and equal to 1. In this case $X = Y$ and one has $X + Y = 2X \sim N(0, 4)$. Under comonotonicity then

$$\text{VaR}_\alpha(X^C + Y^C) = 2\Phi^{-1}(\alpha);$$

see also (1.6). If X and Y are assumed to be independent, then correlation is null and $X + Y \sim N(0, 2)$ (under independence the variance of a sum is the sum of variances). Under independence then

$$\text{VaR}_\alpha(X^\perp + Y^\perp) = \sqrt{2}\Phi^{-1}(\alpha).$$

Since $\Phi^{-1}(\alpha)$ is negative when $\alpha < 0.5$, one has that

$$\text{VaR}_\alpha(X^\perp + Y^\perp) > \text{VaR}_\alpha(X^C + Y^C) \quad \text{for all } \alpha < 0.5,$$

i.e. the VaR of a sum under independence is bigger than the VaR under comonotonicity.

There are other examples of “regular” assumptions for which the sum of the marginal VaRs (the value attained under comonotonicity) is not the worst possible VaR. Figure 1.7 shows that for the sum of two Pareto distributed risks with infinite mean, independence is more dangerous than comonotonicity, for *any* value of $\alpha \in (0, 1)$ (see Chen et al. [3] for a formal proof), yet not being the worst-possible value.

```

set.seed(200) # seed for reproducibility
theta=0.9 # Pareto tail parameter (theta<1 means infinite mean)
N=10^5 # number of simulations
alpha=seq(0.01,0.99,0.001) # values at which to draw the plot
qp=function(x){(1-x)^(-1/theta)-1} # Pareto quantile function
U1=runif(N) # U(0,1) simulation
U2=runif(N) # U(0,1) simulation
I=qp(U1)+qp(U2) # sum of two independent Pareto
VaR_ind=quantile(I,probs=alpha,type=1) # VaR calculation
C=qp(U1)+qp(U1) # sum of two comonotonic Pareto
VaR_com=quantile(C,probs=alpha,type=1) # VaR calculation
## Analytical computation of worst VaR
VaR_max=2*qp((alpha+1)/2) # VaR calculation
## Plot VaR against confidence level
par(pty="s",mai=c(0.8,0.4,0.4,0.4),font.main=1)
plot(alpha,VaR_ind,type="l",col="black",xlab=expression(alpha),
ylab="VaR",ylim=c(0,20),
main="VaR of the sum of two Pareto risks with infinite mean",
lwd=2,lty=2,cex = 0.25,cex.main=0.9,cex.lab=0.9,cex.axis=0.9)
lines(alpha,VaR_com,type="l",col="black",lty=1,lwd = 2)
lines(alpha,VaR_max,type="l",col="black",lty=3,lwd = 2)
legend("topleft",legend=c("Worst-case", "Independence",
"Comonotonicity"),col = "black",lty=c(3,2,1), lwd=c(2,2,2))

```

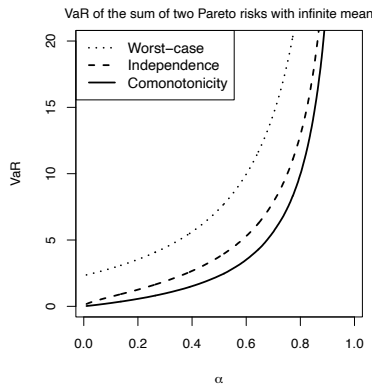


Figure 1.7: VaR of the sum of two risks identically distributed like a $\text{Pareto}(\theta)$, i.e. $F(x) = 1 - (1+x)^{-\theta}$, $x \geq 0$, with tail parameter $\theta = 0.9$ (infinite-mean model), under the assumptions of independence, comonotonicity, and worst-case scenario.

Key message: In general, an higher correlation does *not* imply a higher VaR.

Mistake 5: “We can freely stress-test correlation matrices”

For a multivariate random vector $(X_1, \dots, X_d)'$, the *covariance matrix* $\Sigma = \sigma_{ij}$ is a square matrix showing the covariance $\sigma_{ij} = \text{cov}(X_i, X_j)$, $1 \leq i, j \leq d$, between each pair of marginal components:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \dots & \sigma_{dd} \end{pmatrix}.$$

The diagonal elements of Σ , $\sigma_{ii} = \text{cov}(X_i, X_i) = \text{var}(X_i)$, $1 \leq i \leq d$, are the d variances of the marginal components. In the case these variances are standardized to 1, we speak of a *correlation matrix* (in this latter case we have that $\sigma_{ij} = \rho(X_i, X_j)$).

It immediately follows from the definition that a covariance (correlation) matrix is symmetric, but it is important to notice that not all the symmetric matrices can be covariance (correlation) matrices. To be such, a matrix has to satisfy a special algebraic property, that is has to be positive semidefinite. For any further detail about this requirement, we refer to the classical reference Bhatia [2].

Theorem 1.3. *A symmetric matrix Σ is a covariance matrix if and only if it is positive semidefinite, that is if and only if*

$$a' \Sigma a \geq 0, \text{ for all } a' = (a_1, \dots, a_d) \in \mathbb{R}^d, \quad (1.7)$$

The meaning of this requirement is illustrated in the following example.

Example 1.3. Consider a d -variate random vector $X = (X_1, \dots, X_d)'$ having finite correlation matrix Σ . Let the random vector $Y = a'X$ be a linear combination of X . For example, if X are the daily joint log-returns of a given portfolio of assets at a specific trading desk, the random variable

$$Y = a_1 X_1 + \dots + a_d X_d$$

may represent the linearized profit/loss operator for such portfolio. The variance of Y can be easily computed by the formula

$$\text{var}(Y) = \text{cov}(Y, Y) = a' \Sigma a.$$

Since such variance must be non-negative for any possible choice of weights, one has

$$a' \Sigma a \geq 0, \text{ for all } a' = (a_1, \dots, a_d) \in \mathbb{R}^d,$$

which is the requisite in (1.7).

Care has therefore to be taken when one manipulates covariance matrices: if the covariance matrix of X fails to be positive semidefinite, it becomes possible to find a specific set of weights a_1^*, \dots, a_d^* , such that the variance of the linear combination $a_1^*X_1 + \dots + a_d^*X_d$ is negative. As a consequence, any probabilistic model based on a matrix Σ not satisfying (1.7) has to be deemed as not consistent.

Fortunately, there exist a straightforward computational method to check whether (1.7) is satisfied.

Theorem 1.4. *A symmetric matrix Σ with real entries is positive semidefinite if and only if all its eigenvalues are non-negative.*

Figure 1.8 shows a correlation matrix widely used in Risk Management, as given in the Annex IV, point (1), of the *Solvency II* directive for the calculation of the Solvency Capital Requirement (SCR) Standard Formula. We have a random vector consisting of $d = 5$ risks $(X_1, \dots, X_5)'$ namely: *market*, *counterparty default*, *life* underwriting, *health* underwriting, *non-life* underwriting risk modules. The Solvency prescribed correlation matrix Σ is given in Figure 1.8.

	Market	Default	Life	Health	Non Life
Market	1.00	0.25	0.25	0.25	0.25
Default	0.25	1.00	0.25	0.25	0.50
Life	0.25	0.25	1.00	0.25	0.00
Health	0.25	0.25	0.25	1.00	0.00
Non Life	0.25	0.50	0.00	0.00	1.00

Figure 1.8: Solvency correlation matrix for the calculation of the Solvency Capital Requirement (SCR) Standard Formula, which displays in each cell the correlation between two variables of the Solvency II regulation risk module vector composed by: non-life underwriting risk, life underwriting risk, health underwriting risk, market risk, counterparty default risk.

One might be tempted to provide some naive conclusions from the null terms of the matrix, like the fact that Non-life and Life risks are independent. Ignoring whether this was also the aim of the regulator, we remark again that, in the absence of further assumptions about the joint distribution of $(X_1, \dots, X_5)'$, one can only state that Non-life and Life risks are uncorrelated (if this can be of some use). This highlights once

more the limited use of a correlation matrix outside the context of a well-specified, possibly linear, probabilistic model.

The Solvency Capital Requirements are the amount of funds that insurance and reinsurance companies are required to hold under the European Union’s Solvency II directive in order to have a 99.5% yearly confidence level, so that they could expect to survive the most extreme losses over the course of 500 years; in other words, the VaR_α of the aggregate loss $X_1 + \dots + X_5$ computed at the probability level $\alpha = 0.995$.

A common practice within risk management departments, in order to determine more conservative capital requirements, consists in arbitrarily increasing the correlation between some risks X_i and X_j of the Solvency correlation matrix. This practice is likely based on the wrong belief that by increasing the values of the correlation matrix, the VaR of the final position (sum) will be automatically augmented. As noticed in Mistake 4, this is not true.

Moreover, one cannot arbitrarily change the values in a correlation matrix: even preserving the symmetry of the matrix and the positiveness of its entries, slightly increasing a few correlation values might infringe the condition of positive semidefiniteness, as the following example shows.

Example 1.4. Suppose that an insurance company, for the purpose of stress-testing, changes the correlation between the health underwriting risk module and the market, default, and life underwriting risk module from 0.25 to 0.7, with the naive aim of obtaining stricter Solvency Capital Requirements (this potential outcome very much depends on the full model used). Before estimating the new regulatory capital coming from the model change, the company should definitely check whether the new matrix is positive semidefinite, by computing its eigenvalues (for instance via the R function `eigen`). If the eigenvalues of the new correlation matrix are all non-negative, then the matrix is positive semidefinite; see Theorem 1.4. If just one eigenvalue is negative, the model (and the implied regulatory capital) should be disregarded as not consistent.

Table 1.3: Eigenvalues - stress test 1.

Eigenvalue 1	Eigenvalue 2	Eigenvalue 3	Eigenvalue 4	Eigenvalue 5
2.564775	1.231365	0.75	0.5270761	-0.0732159

One of the eigenvalues of the correlation matrix displayed in Figure 1.9 is negative. As a consequence, the new correlation matrix is no more positive semidefinite. In Figure 1.10 we show other stress tests in which the newly obtained matrices are not positive definite: the probabilistic models underlying such matrices are simply wrong.

The bad habit not to check whether a matrix is positive definite (that is if it has non-negative eigenvalues) likely originates from the fact that in a 2×2 correlation matrix, the off-diagonal element $\rho_{12} = \rho_{21}$ can be any number between -1 and 1 (the matrix will remain positive semidefinite). As the next pedagogical example shows, for a 3×3 case symmetric matrix having positive entries between -1 and 1 , this is no more true.

	Market	Default	Life	Health	Non Life
Market	1.00	0.25	0.25	0.70	0.25
Default	0.25	1.00	0.25	0.70	0.50
Life	0.25	0.25	1.00	0.70	0.00
Health	0.70	0.70	0.70	1.00	0.00
Non Life	0.25	0.50	0.00	0.00	1.00

Figure 1.9: Solvency Capital Requirement correlation matrix - stress test 1.

	Market	Default	Life	Health	Non Life		Market	Default	Life	Health	Non Life
Market	1.00	0.25	0.25	0.70	0.75	Market	1.00	0.80	0.25	0.80	0.75
Default	0.25	1.00	0.25	0.70	0.75	Default	0.80	1.00	0.25	0.70	0.75
Life	0.25	0.25	1.00	0.70	0.00	Life	0.25	0.25	1.00	0.70	0.00
Health	0.70	0.70	0.70	1.00	0.00	Health	0.80	0.70	0.70	1.00	0.00
Non Life	0.75	0.75	0.00	0.00	1.00	Non Life	0.75	0.75	0.00	0.00	1.00

Figure 1.10: Two stress-tested false correlation matrices.

Example 1.5. We build a 3×3 symmetric matrix having positive random entries between -1 and 1 and not being positive definite, and we show it in Figure 1.11.

```
library(corrplot)
set.seed(100) # seed for reproducibility
## Generate a symmetric matrix with random entries in (0,1)
## until finding one not being positive semidefinite
success<-FALSE
while(!success){
  a=round(runif(3),3) # draw three random entries in (0,1)
  ## Create a symmetric 3x3 matrix with such entries
  A=matrix(c(1, a[1], a[2],
            a[1], 1, a[3],
            a[2], a[3], 1), nrow = 3, byrow = TRUE)
  rownames(A)=c(" ", " ", " ")
  colnames(A)=c(" ", " ", " ")
  eigen_values=eigen(A)$values # compute eigenvalues
  ## Repeat until one eigenvalue is negative
  success<-length(eigen_values[eigen_values<0])>0
}
## Display the false correlation matrix
corrplot(A,method="number",type="full",tl.col="black",
col=color_matrix,addCoef.col="black",cl.pos="n",
order="original",number.cex=3)
```

1.00	0.17	0.62
0.17	1.00	0.88
0.62	0.88	1.00

Figure 1.11: A symmetric matrix with entries between 0 and 1 which is not positive semidefinite.

Another important issue here is that if the marginals are specified (as often in SCR), then positive semidefiniteness is no longer sufficient for a matrix to be a covariance matrix with the specific marginals (unless it is in a family of distributions like the multivariate Normal). This is a notoriously difficult problem. Even the determination of covariance matrices of standard uniform marginals (i.e., rank correlation matrices) seems to be out of computational reach and we know that it differs from positive semidefiniteness in dimension 12 or higher; see Wang et al. [19].

Key message: Check the eigenvalues of your correlation matrix before use!

Acknowledgements

The mistakes illustrated in this chapter have been collected from students and practitioners attending the Risk Management courses held by the author. A special thank goes to Giacomo Cagliani who wrote some of the R codes accompanying this chapter and presented some of the mistakes in his Master thesis dissertation. An excerpt of this chapter is available as a standalone paper at *SSRN*.

Bibliography

- [1] Ai, X. (2024). Adjust pearson's ρ to measure arbitrary monotone dependence. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.
- [2] Bhatia, R. (2007). *Positive Definite Matrices*. Princeton University Press.
- [3] Chen, Y., Embrechts, P., and Wang, R. (2025). Technical note—an unexpected stochastic dominance: Pareto distributions, dependence, and diversification.
- [4] Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., and Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: theory. *Insurance Math. Econom.*, 31(1):3–33.
- [5] Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Math. Methods Oper. Res.*, 77(3):423–432.
- [6] Embrechts, P., McNeil, A. J., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk management: Value at Risk and Beyond*, pages 176–223. Cambridge Univ. Press, Cambridge.
- [7] Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon. Sect. A. (3)*, 14:53–77.
- [8] Hardy, G. H., Littlewood, J. E., and Pólya, G. (1934). *Inequalities*. Cambridge, at the University Press.
- [9] Hoeffding, W. (1940). Massstabinvariante korrelationstheorie. *Schriften Math. Inst. Univ. Berlin*, 5(5):181–233.
- [10] Hoeffding, W. (1994). Scale-invariant correlation theory. In Fisher, N. and Sen, P., editors, *The Collected Works of Wassily Hoeffding*, pages 57–107. Springer New York. This is the english translation of Hoeffding [9].
- [11] Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. (2018). *Elements of Copula Modeling with R*. Springer, Cham.
- [12] Makarov, G. D. (1981). Estimates for the distribution function of the sum of two random variables with given marginal distributions. *Theory Probab. Appl.*, 26:803–806.

- [13] McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management. Concepts, Techniques and Tools. Revised edition*. Princeton University Press.
- [14] Puccetti, G. (2022). Measuring linear correlation between random vectors. *Information Sciences*, 607:1328–1347.
- [15] Puccetti, G. and Wang, R. (2015). Extremal dependence concepts. *Stat. Science*, 30(4):485–517.
- [16] Rüschendorf, L. (1982). Random variables with maximum sums. *Adv. in Appl. Probab.*, 14(3):623–632.
- [17] Scarsini, M. (1984). On measures of concordance. *Stochastica*, 8(3):201–218.
- [18] Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivariate Anal.*, 102(10):1344–1360.
- [19] Wang, B., Wang, R., and Wang, Y. (2019). Compatible matrices of spearman’s rank correlation. *Statistics & Probability Letters*, 151:67–72.