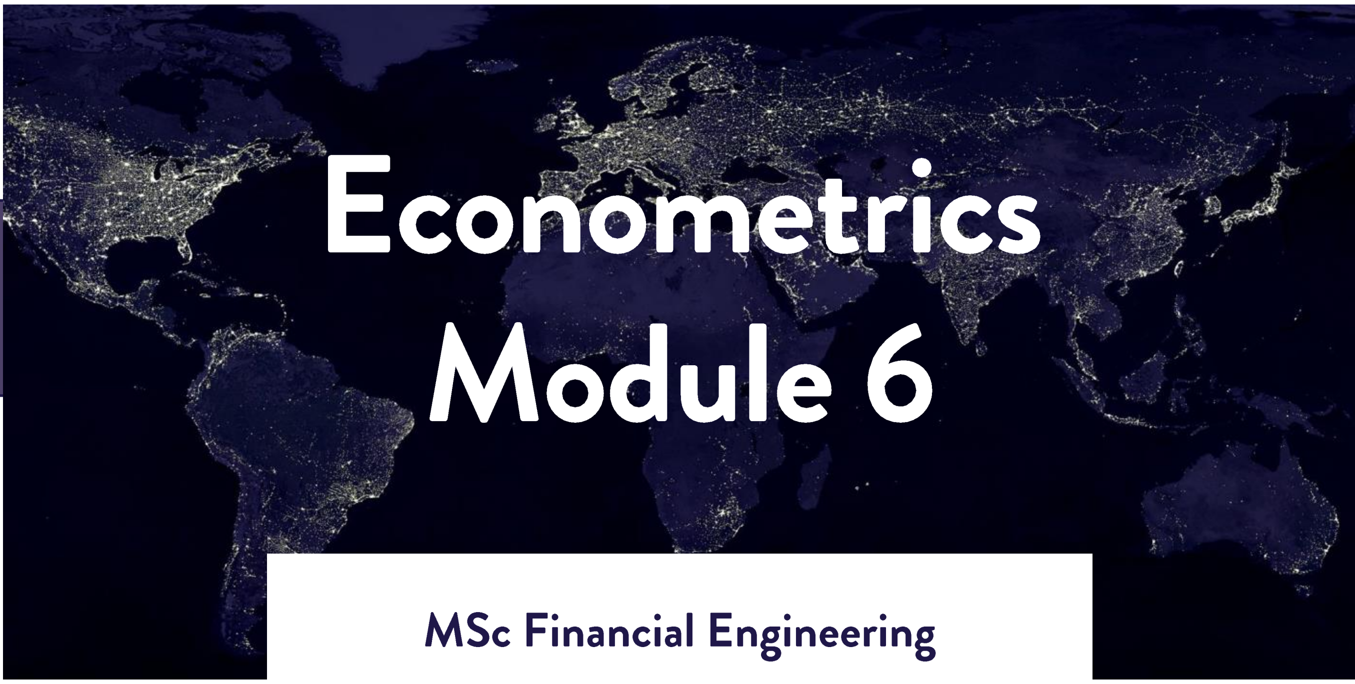


Econometrics
Module 6

MSc Financial Engineering

Table of Contents

1. Brief

This document contains the core content for Module 6 of Econometrics, entitled Introduction to Risk Management. It consists of five video lecture transcripts and three sets of supplementary notes.

2. Course Context

Econometrics is the second course presented in the WorldQuant University (WQU) Master of Science in Financial Engineering (MScFE) program. In this course, you will apply statistical techniques to the analysis of econometric data. The course starts with an introduction to the R statistical programming languages that you will use to build econometric models, including multiple linear regression models, time series models, and stochastic volatility models. You will learn to develop programs using the R language, solve statistical problems, and understand value distributions in modeling extreme portfolio and basic algorithmic trading strategies. The course concludes with a review on applied econometrics in finance and algorithmic trading.

2.1 Course-level Learning Outcomes

Upon completion of the Econometrics course, you will be able to:

1. Write programs using the R language.
2. Use R packages to solve common statistical problems.
3. Formulate a generalized linear model and fit the model to data.
4. Use graphic techniques to visualize multidimensional data.
5. Apply multivariate statistical techniques (PCA, factor analysis, etc.) to analyze multidimensional data.
6. Fit a time series model to data.
7. Fit discrete-time volatility models.
8. Understand and apply filtering techniques to volatility modeling.
9. Understand the use of extreme value distributions in modeling extreme portfolio returns.
10. Define some common risk measures like VaR and Expected Shortfall.
11. Define and use copulas in risk management.
12. Implement basic algorithmic trading strategies.

2.2 Module Breakdown

The Econometrics course consists of the following one-week modules:

1. Basic Statistics
2. Linear Models
3. Univariate Time Series Models
4. Univariate Volatility Modeling
5. Multivariate Time Series Analysis
6. Introduction to Risk Management
7. Algorithmic Trading

3. Module 6:

Introduction to Risk Management

In Module 6, we combine the results developed in the first five modules into a holistic approach to the modeling of expected returns and risk to a portfolio of assets. This is the overarching tool that a financial engineer or investor uses to manage investment returns and related risks.

3.1 Module-level Learning Outcomes

After completing this module, you will be able to:

1. Understand basic static results and time-varying application of empirical expected portfolio return.
2. Use copulas to find the joint distribution of a random vector.
3. Understand Extreme Value Theory.

3.2 Transcripts and Notes

3.2.1 Transcript: Portfolio Choice and Empirical Modeling of Expected Portfolio Return and Risk – Basic Static Results

In this module, we combine the results we developed in the first five modules of this course into a holistic approach to the modeling of the expected returns and risk to a portfolio of assets, the central tool that a financial engineer or investor uses to manage investment returns and the related risks.

Portfolio optimization

In this video we will be looking at basic static results of empirical modeling of expected portfolio risk and return.

To focus on the core ideas of portfolio choice, which you will study in detail in a later course, we only consider *long positions* in two risky assets and one safe asset. All the results we will use extend readily to more assets (which would usually be the case in a balanced, well diversified portfolio), but this highlights the most important relationships and results. If you need more detail, consult Ruppert and Matteson (2015).

Consider two risky assets with the following characteristics:

Asset 1 has expected return μ_1 with variance σ_1^2 (or standard deviation σ_1). Asset 2 has expected return μ_2 with variance σ_2^2. The correlation between returns is given by ρ_{12}.

Since the expectation operator is linear, this means a long portfolio p with weight w on asset 1 will have expected return:

$$\mu_p = w\mu_1 + (1-w)\mu_2.$$

Due to the simple formula for the variance of the sum of two random variables, the variance of the portfolio will be:

$$\sigma_p^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho_{12}\sigma_1\sigma_2.$$

With two assets, the investment problem is not very interesting: given the parameters of the returns, there is only one way to get a given expected return, which fixes the variance, and one way to get a specific variance, which fixes the expected return.

Once we add a safe or risk-free asset, however, we can ask questions about the "optimal portfolio".

To keep things simple, we can consider the optimal portfolio from two angles:

1. If the investor has a specific desired expected return from a portfolio, the optimal portfolio is the portfolio that attains that expected return with the lowest possible variance.
2. If the investor wants a specific risk or portfolio variance, then the optimal portfolio is the one that achieves that portfolio variance with the largest possible expected return.

Let's define the **safe asset** as having a fixed return μ_f and, by definition, zero variance.

The two questions we need to ask are:

1. How much of our portfolio must be invested in the safe asset? Let's denote this with $(1-v)$, so that we invest v in some portfolio of the two risky assets.

2 Given v to be invested in the two risky assets, how must we allocate this across the two assets?

The second question has a fixed answer: we always invest in the tangency portfolio that represents the optimal trade-off between risk and return across the two risky assets. We only state the result here. See Ruppert and Matteson (2015) for the derivation. The tangency portfolio is given by the relative weight w_T on asset 1:

$$w_T = \frac{(\mu_1-\mu_f)\sigma_2^2-(\mu_2-\mu_f)\rho_{12}\sigma_1\sigma_2}{(\mu_1-\mu_f)\sigma_2^2+(\mu_2-\mu_f)\sigma_1^2+(\mu_1+\mu_2-2\mu_f)\rho_{12}\sigma_1\sigma_2}.$$

Yielding risky portfolio with expected return and variance:

$$\mu_T = w_T\mu_1 + (1-w_T)\mu_2$$
$$\sigma_T^2 = w_T^2\sigma_1^2 + (1-w_T)^2\sigma_2^2 + 2w_T(1-w_T)\rho_{12}\sigma_1\sigma_2.$$

The answer to the first question depends on the goal of the investor.

The expected return μ_v and variance σ_v^2 on an arbitrary portfolio with $(1-v)$ in the safe asset is given by:

$$\mu_v = \mu_f + v\left(\mu_T - \mu_f\right)$$
$$\sigma_v^2 = v^2\sigma_T^2.$$

These equations then describe the optimal trade-offs between the final expected return and variance of all portfolios in this model: the lowest risk. The final choice of v depends only on the risk appetite of the investor.

Suppose our investor requires return equal to μ^*. Then, for any given, and usually data dependent, values of μ_f and μ_T, the optimal fraction of wealth v^* to invest in the tangency portfolio is given by:

$$v^* = \frac{\mu^* - \mu_f}{\mu_T - \mu_f}.$$

You can easily derive a value for v if the investor is targeting a specific portfolio variance instead.

This video looked at basic static results of empirical modeling of expected portfolio return and risk.

3.2.2 Notes: Minimum Variance Portfolios

Whenever two assets are not perfectly positively correlated – i.e. when $\rho < 1$ there are diversification opportunities[1] and one can find a portfolio that obtains lower variance than the least risky asset individually. The minimum variance portfolio is the portfolio that exploits all possible diversification potential. Note: depending on the values of the conditional variances and conditional covariance, this may require shorting one asset or the other.

The minimum variance portfolio is constructed in the following way:

Consider two assets with expected returns r_{1t+1} and r_{2t+1}; expected conditional variances σ_{1t+1}^2 and σ_{2t+1}^2; and expected conditional covariance $\sigma_{12,t+1}$.

As in the static case, a portfolio with weight w on asset 1 will have expected portfolio variance that depends on the weight, and is given by:

$$\sigma_{p,t+1}^2(w) = w^2\sigma_{1t+1}^2 + (1-w)^2\sigma_{2t+1}^2 + 2w(1-w)\sigma_{12,t+1}.$$

By standard calculus results, we can minimize this variance by finding the first order condition with respect to the portfolio weight, setting it to zero and solving for the optimal weight w^*:

$$\frac{d\sigma_{p,t+1}^2(w)}{dw} = 2w\sigma_{1t+1}^2 + (2w-2)\sigma_{2t+1}^2 + (2-4w)\sigma_{12,t+1}.$$

Solving $\frac{d\sigma_{p,t+1}^2(w^*)}{dw} = 0$ for w^* yields:

[1] In theory, two perfectly correlated assets can be used to obtain a riskless portfolio, but in practice this never happens with real world financial assets.

$$w^* = \frac{\sigma_{2t+1}^2 - \sigma_{12,t+1}}{\sigma_{1t+1}^2 + \sigma_{2t+1}^2 - 2\sigma_{12,t+1}}.$$

Substituting this weight into formulae above yields the expected return and conditional variance of the minimum variance portfolio.

Of course, we applied this in a naïve way to the simplest of possible investment situations, but the results extend easily to more complicated situations. The critical take-away from this section is that you should understand both (a) the approaches that convert empirical interdependences across different assets, and (b) the properties of time-series econometrics, to develop the best investment approach for a specific investor.

One dimension that you should explore on your own is how the algorithm should be adjusted for an investor with a longer investment horizon – i.e. if $k > 1$. Then you will have to make a call on what average of the predicted features of interest to use for the longer horizon investment.

3.2.3 Notes: Copulas

In Module 2 we covered the basic ideas and characteristics of joint distributions, for example, the bivariate normal and t-distributions. These readily extend to fully characterized behaviors for n normal distributed variables (or n variables following a joint t-distribution).

In practice, however, it is unlikely that a single distribution, however precisely dynamically modeled, would capture the characteristics of a diverse set of assets in a portfolio. What should we do if, say, one asset is appropriately modeled by a normal distribution and another by a t-distribution (or some other more heavy-tailed distribution)? It is very unlikely that there even exists a well-characterized joint distribution of two very disparate marginal distributions, let alone a collection of more varied distributions.

For this purpose, copula theory was developed.

A **copula** can be viewed as a "distribution function" that contains all the information of the interdependencies among a set of jointly distributed variables, but none of the information of the marginal distribution of any of the constituent variables of the joint distribution. Remarkably, this is always possible. That is, we can always decompose the joint distribution of any set of variables, regardless of the individual marginal distributions into a copula and the individual marginal distributions. The copula is a function that takes the individual marginal distributions as arguments and yields the joint distribution function as output. This result is due to a theorem by Sklar[2] (1959).

[2] The original text is in French.

Formulas

The copopula of $(X_1, X_2, \ldots., X_d)$ is defined as the joint cumulative distribution function of $(U_1, U_2, \ldots, U_d)$

$$C(u_1, u_2, \ldots, u_d) = P[U_1 \leq u_1, U_2 \leq u_2, \ldots, U_d \leq u_d].$$

$(X_1, X_2, \ldots, X_d)$ – random vector

$F_i(x) = P[X_i \leq x]$ – marginal CDFs (continuous functions)

By applying the probability integral transform to each component, the random vector $(U_1, U_2, \ldots, U_d) = \left(F_1(X_1), F_2(X_2), \ldots., F_d(X_d)\right)$ has uniformly distributed marginal.

In probabilistic terms:

$C: [0,1]^d \to [0,1]$ is a d-dimensional copula if C is a joint cumulative distribution function of a d-dimensional random vector on the unit cube $[0,1]^d$ with uniform marginals.

In analytic terms:

$C: [0,1]^d \to [0,1]$ is a d-dimensional copula if

- $C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_d) = 0$ (the copula is zero if one of the arguments is zero)
- $C(1, \ldots, 1, u, 1, \ldots, 1) = u$, (the copula is equal to u if one argument is u and all others are equal to 1)
- C is d –increasing for each hyperrectangle $B = \prod_{i=1}^{d}[x_i, y_i] \subseteq [0,1]^d$ the C-volume of B is non-negative:

$$\int_B dC(u) = \sum_{z \in \times_{i=1}^{d} x_i, y_i} (-1)^{N(z)} C(z) \geq 0,$$

where

$$N(z) = k: z_k = x_k.$$

Sklar's theorem

According to Sklar's theorem, every multivariate cumulative distribution function can be expressed using only the marginals.

$$H(x_1, \dots, x_d) = P[X_1 \leq x_1, \dots, X_d \leq x_d].$$

A random vector $(X_1, X_2, \dots, X_d)$ can be expressed by involving only the marginals as following:

$$F_i(x) = P[X_i \leq x],$$
$$H(x_1, \dots, x_d) = C\big(F_1(x_1), \dots, F_d(x_d)\big).$$

C – copula

f – density function

c – density of the copula

$$f(x_1, \dots, x_d) = c\big(F_1(x_1), \dots, F_d(x_d)\big) \cdot f_1(x_1), \dots, f_d(x_d).$$

The copula is unique on $Ran(F_1) \times \ldots \times Ran(F_d)$(the cartesian product of the ranges of the marginal Cumulative Density Functions (CDFs). If the marginals F_i are continuous, the copula is unique.

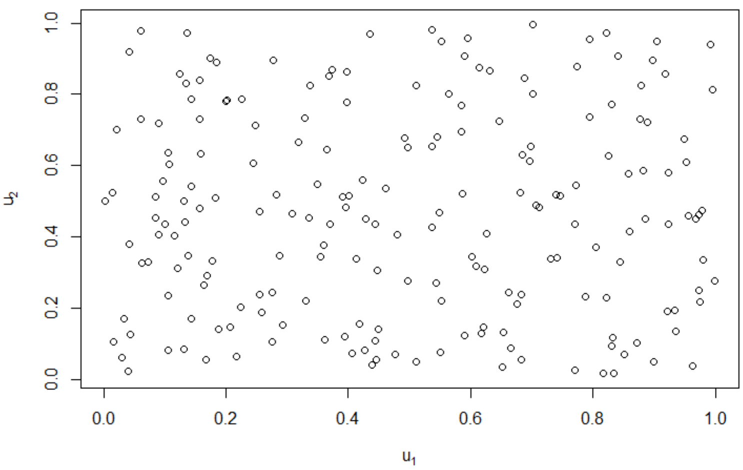
Scatterplots of random draws from some standard copulas

First, we will show the various different shapes implied for the copulas we have introduced thus far.

We will use the R package copula.

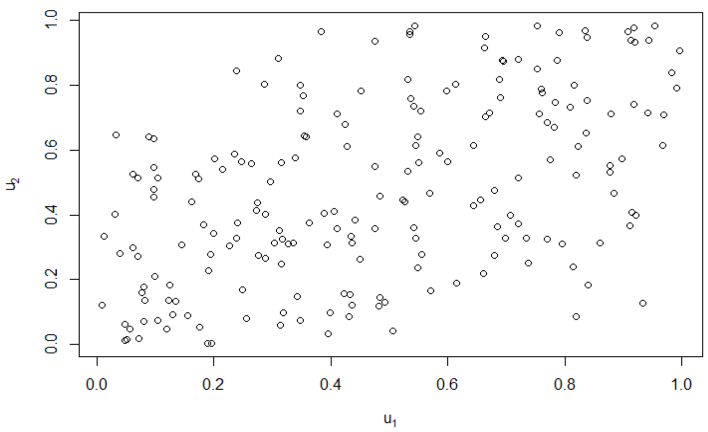
Let's study the type of relationships generated by the normal and t-copulas. For each, we generate random bivariate draws from the copula and consider their scatterplots to evaluate the dependence:

```
library(copula)
norm.cop <- normalCopula(0.0)
U0 <- rCopula(200, norm.cop)
plot(U0,xlab=expression(u[1]), ylab=expression(u[2]))
```



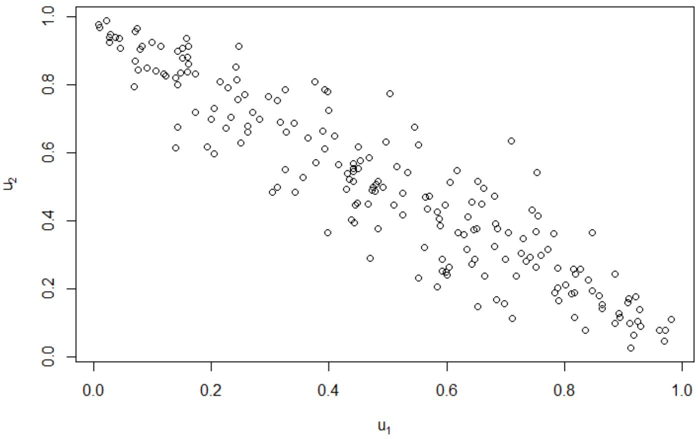
When there is no correlation in the normal copula, the random draws are clearly unrelated and uniform. They seem to "fill" the unit square with no apparent relationship.

```
norm.cop <- normalCopula(0.5)
U0 <- rCopula(200, norm.cop)
plot(U0,xlab=expression(u[1]), ylab=expression(u[2]))
```



When we use a normal copula with correlation 0.5, we begin to see a positive dependency. Note that each U is still individually uniformly distributed, jointly, the points are closer to the positive diagonal.

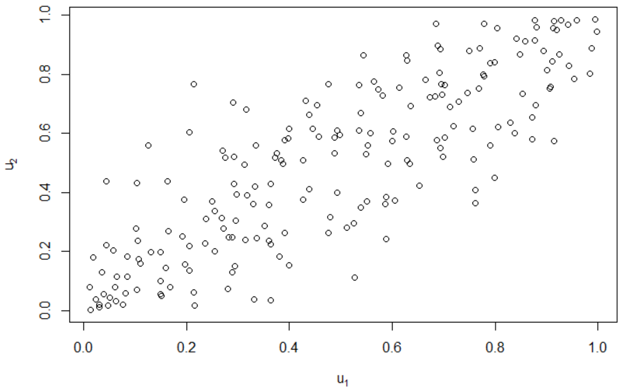
```
norm.cop <- normalCopula(-0.95)
U0 <- rCopula(200, norm.cop)
plot(U0,xlab=expression(u[1]), ylab=expression(u[2]))
```



When we use a normal copula with a strong negative correlation, -0.95, we see the clear negative tendency: the points are clustered around the negative diagonal. We can also see the elliptical nature of the distribution: both extreme ends of the distribution suggest tail dependencies: when u_1 is very low, u_2 tends to be very high, and when u_1 is very high, u_2 tends to be very low.

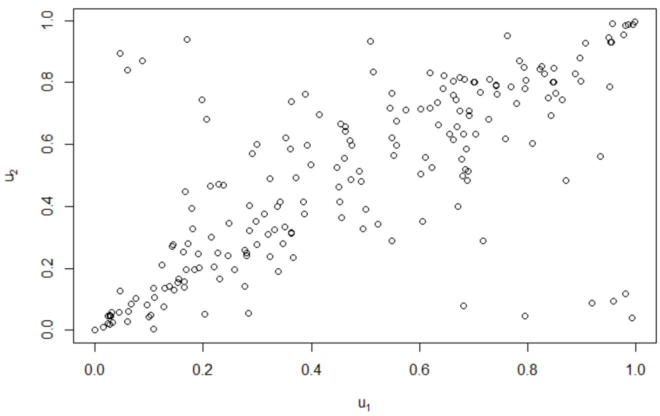
In the t-copula, we want to show the impact of the degrees of freedom/tail index, so we do two draws with same high correlation but vary the tail index.

```
t.cop <- tCopula(0.8,df=100)
v <- rCopula(200,t.cop)
plot(v,xlab=expression(u[1]), ylab=expression(u[2]))
```



When the tail index is high, we get a picture similar to the normal distribution, as expected. There is a clear positive dependence, but the tails are not particularly extremely closely clustered. We can also see the elliptical nature of the clustering of the points.

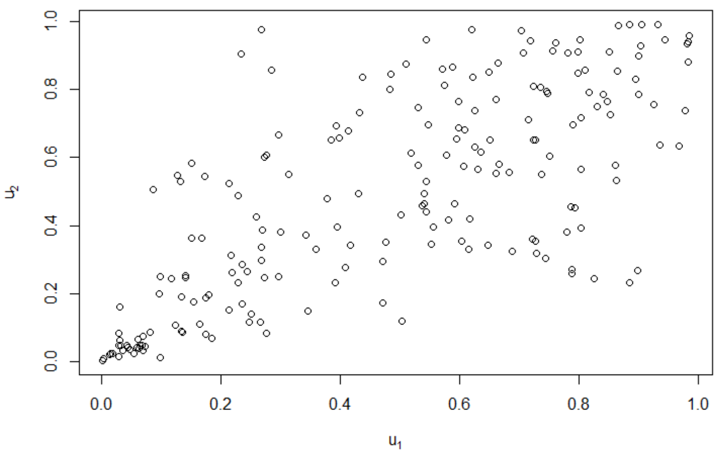
```
t.cop <- tCopula(0.8,df=1)
v <- rCopula(200,t.cop)
plot(v,xlab=expression(u[1]), ylab=expression(u[2]))
```



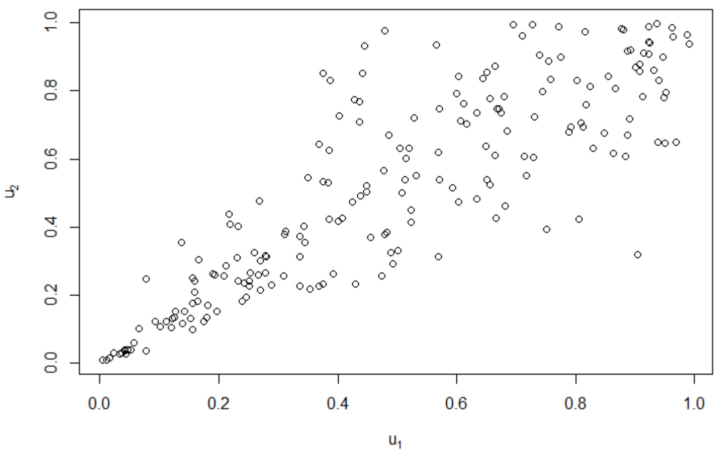
When the tail index is very low, we see strong clustering at both tails, even though there are some points far off the diagonal. This will be a sensible model for two assets that sometimes have strong tail dependencies in both tails, but not always.

Turning to the Clayton copula, we vary the θ parameter to show how this changes the shape of the distribution:

```
U <- rCopula(n=200, copula=archmCopula(family="clayton", param=2))
plot(U,xlab=expression(u[1]), ylab=expression(u[2]))
```



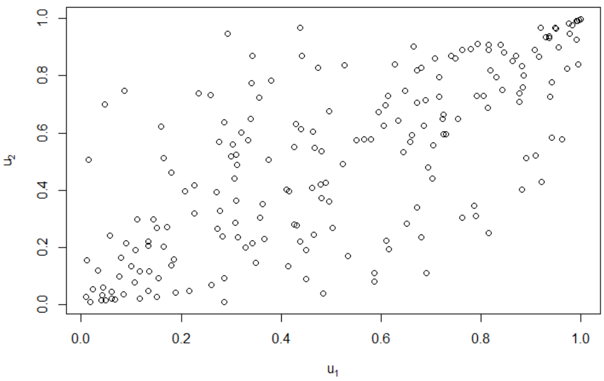

```
U <- rCopula(n=200, copula=archmCopula(family="clayton", param=4))
plot(U,xlab=expression(u[1]), ylab=expression(u[2]))
```



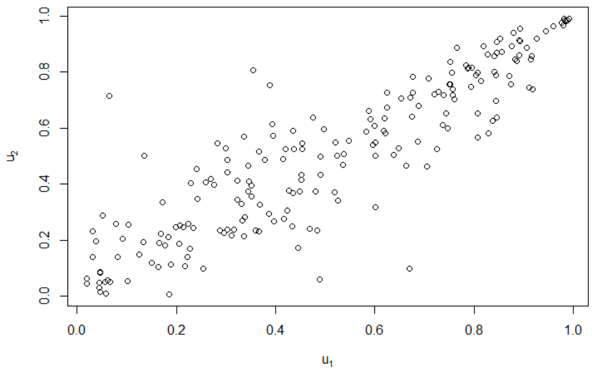
As we increase the θ from 2 to 4, we can clearly see the increase in lower tail dependency, with no obvious change in upper tail dependency (although there must be some – see below). This would thus be a copula that would describe the dependency between two assets where they tend to be highly correlated when returns are low but much less correlated when returns are high. We can also clearly observe that this is not an elliptical distribution like the normal or t-copulas. You should experiment to show that as θ grows very large (for example, 100) the distribution collapses to the positive diagonal. For instance, u_1 and u_2 are essentially the same random variable.

Lastly, we consider the Gumbel copula.

```
U <- rCopula(n=200, copula=archmCopula(family="gumbel", param=2))
plot(U,xlab=expression(u[1]), ylab=expression(u[2]))
```




```
U <- rCopula(n=200, copula=archmCopula(family="gumbel", param=4))
plot(U,xlab=expression(u[1]), ylab=expression(u[2]))
```



As we increase the θ from 2 to 4, we can clearly see the increase in *upper* tail dependency, with much slower increase lower tail dependency. Thus, this would be a copula that would describe the dependency between two assets where they tend to be highly correlated when returns are high but much less correlated when returns are low. We can also clearly observe that this is, again, not an elliptical distribution like the normal or t-copulas. You should experiment to show that as θ grows very large (for example 100) the distribution also collapses to the positive diagonal. An example of this is the co-monotonicity copula.

Correlated random variables

Copulas can be used for the simulation of loss distributions of credit portfolios. In this section, we show how to use correlated random variables in copulas. We use the following steps:

- Simulate a pair of correlated random variables using a Gaussian copula (we initially indicate the correlation)
- Simulate a pair of correlated random variables using a t-copula (we initially indicate the correlation)
- Estimate the Gaussian copula parameter using the Maximum Likelihood Estimation

We can use historical datasets instead of simulated data.

```
#load copula package in R software
library(copula)

#declare a Gaussian copula class with 0.63 correlation
normal_copula<-normalCopula(0.63)

#indicate seed for generating pseudo-random numbers
set.seed(100)

#generate 400 realizations of two uniformly distributed random
variables with the Gaussian copula dependency structure
v1<-rCopula(400, normal_copula)
#declare a t-copula with an 0.63 correlation and 4 degrees of freedom
t.copula<-tCopula(0.63, df=4)

#set seed at 100 as previously
set.seed(100)

#generate 400 realizations of pairs of random variables with t-copula
dependence
v2<-rCopula(400, t.copula)

#the two graphs will be placed next to one other
par(mfcol=c(1,2))
```

```
plot(v1)
plot(v2)
fit.ml <- fitCopula(normal_copula, v1, method = "ml")
fit.ml

#Results obtained
#The estimated correlation at about 0.59657
```

```
fitCopula() estimation based on 'maximum likelihood'
and a sample of size 400.
      Estimate Std. Error z value Pr(>|z|)
rho.1  0.59657    0.02753   21.67   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
The maximized loglikelihood is  89.35
Optimization converged
Number of loglikelihood evaluations:
function gradient
      15        4
```

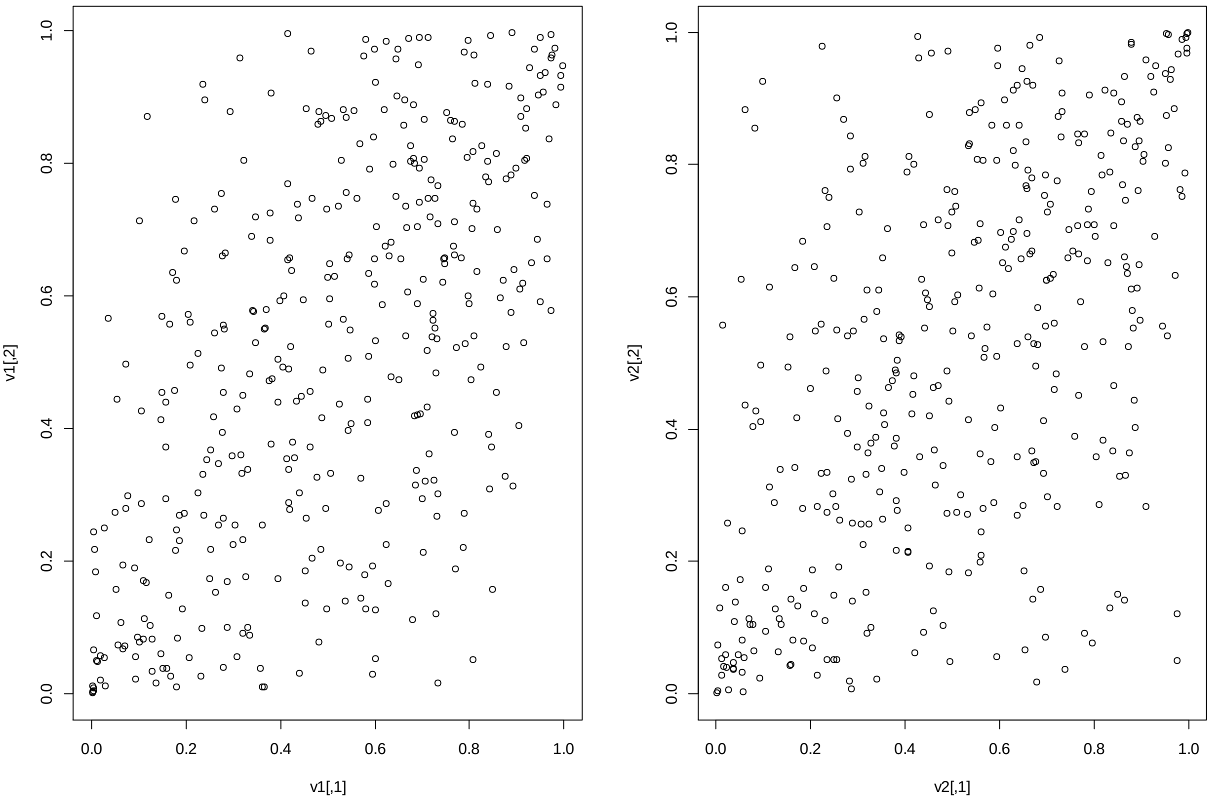


Figure 1: Scatter plot of random variable pairs generated by Gaussian copula (left). Scatter plot of random variable pairs generated by t-copula (right).

Copula in finance

Copula is used in:

- Risk management
- Credit scoring
- Default risk modeling
- Derivative pricing
- Asset allocation

Consider the following when using copula for risk management:

- The 2008-2009 global crisis is said to have been driven by the extensive use of Gaussian copulas that did not correctly calculate the collapse of the financial systems. Experts criticized the simplicity of Gaussian copulas that cannot adequately model the complex dependencies in a portfolio.
- Currently, there is a substantial amount of literature that presents copula or new extensions of this topic.
- During periods of upward movement, investors tend to buy riskier assets (equities, derivatives or real estate assets), while in periods of financial crisis, investors tend to invest more in cash or bonds (known as the "flight-to-quality effect").
- Research shows that equities tend to be more correlated during a downward movement of the market compared to an upward movement.
- Negative news has a more significant impact on stock prices as compared to positive news. Copula can be used to perform stress-tests and robustness checks during periods of financial crisis / downward movement / panic.
- The correlation coefficient cannot tell us the entire story behind asset interactions. Therefore, copula can provide better information regarding asset dependencies.
- Gaussian and Student-t copulas only model elliptical dependence structures and do not allow for correlation asymmetries where correlations differ on the upside or downside regimes.
- Vine copula (pair copula) allow us to flexibly model the dependencies in large dimension portfolios.

- Panic copula estimated by Monte Carlo simulation quantifies the effect of panic in financial markets on portfolio losses – you will learn more about this in the Computational Finance course.
- Copulas are used on a large scale to model Collateralized Debt Obligations (CDOs).

Practice exercise

1. Simulate a pair of correlated random variables using a Gaussian copula (the correlation = 0.75, number of simulations = 600, seed = 110).
2. Simulate a pair of correlated random variables using a t-copula (the correlation = 0.75, number of simulations = 600, seed = 110).
3. Estimate the Gaussian copula parameter using the Maximum Likelihood Estimation.

3.2.4 Transcript: Special Copulas

In this video we will be looking at special copulas. Since copulas are supposed to capture the dependencies between variables, there are three special copulas that capture the "extremes" of possible interdependencies and one that captures the absence of interdependencies. These are useful for calibration of empirical copulas, as they define the space we need to consider when estimating any copula. For any type of copula, and there are many, there is a version that captures independence, and any parameterization of any copula must lie between the two extreme copulas we define here.

The independence copula

If a set of k uniformly distributed variables are mutually independent, the independence copula C_0 is itself uniform on $[0,1]^k$ and given by:

$$C_0(u_1, u_2, \dots, u_k) = \prod_{i=1}^{k} u_i.$$

The co-monotonicity copula

If a set of k uniformly distributed variables are perfectly positively dependent, it means they all have the same marginal distribution $F_{X_i} = F$ they are the same random variable. Thus, the co-monotonicity copula C_+ is just constructed of k copies of the same distribution:

$$\begin{aligned} C_+(u_1, u_2, \dots, u_k) &= \Pr\left[F_{X_1} \leq u_1 \text{ and } F_{X_2} \leq u_2 \text{ and} \dots \text{and } F_{X_k} \leq u_k\right], \\ &= \Pr[F \leq u_1 \text{ and } F \leq u_2 \text{ and} \dots \text{and } F \leq u_k], \\ &= \Pr[F \leq \min(u_1, u_2, \dots, u_k)], \\ &= \min(u_1, u_2, \dots, u_k). \end{aligned}$$

The co-monotonicity copula forms an *upper bound* for all copulas (on the same domain).

The counter-monotonicity copula

If two[3] uniformly distributed variables, X_1 and X_2, are perfectly negatively dependent, then if X_1 is distributed according to F, X_2 is distributed according to $1 - F$. Their interdependence is captured by the counter-monotonicity copula C_{-}:

$$\begin{aligned} C_{-}(u_1, u_2) &= \Pr[F \leq u_1 \text{ and } 1 - F \leq u_2], \\ &= \Pr[1 - u_2 \leq F \leq u_1], \\ &= \max\{u_1 + u_2 - 1, 0\}. \end{aligned}$$

The counter-monotonicity copula forms a *lower bound* for all bi-variate copulas. This can be extended to the Fréchet-Hoeffding lower bound on k-dimensional copulas:

$$C_{-}(u_1, u_2, \dots, u_k) = \max\left\{1 - k + \sum_{i=1}^{k} u_i \, , 0\right\}.$$

This is then the lower bound corresponding to the co-monotonicity copula as the upper bound. Note, however, it is no longer the "counter-monotonicity" copula, as that is only defined for two variables.

This video looked at special copulas, namely the independence copula, the co-monotonicity copula and the counter-monotonicity copula.

[3] It is a good exercise to show that it is not possible for more than two random variables to be perfectly negatively correlated / dependent.

2.3.5 Notes: Introduction to Extreme Value Theory (EVT)

The 2007 global financial crisis showed that rare and extreme events can generate huge losses for financial institutions. The collapse of the global markets generated a slump of asset values held by companies and brought huge losses. Can we forecast such an event, and can we determine the maximum losses generated by such an event? Classic statistics cannot provide answers in this field as it relies on quantifying "normal behavior". The solution comes from Extreme Value Theory (EVT) which is concerned with the statistical analysis of extreme events.

Applications of Extreme Value Theory in finance:

- Large insurance losses
- Equity losses
- Day to day market risk
- Distribution of income
- Re-assurance
- Credit losses
- Operational losses

X – the random loss that we want to model

$F(x)$ – distribution function

u – a given threshold

Y – excess loss over the threshold

$$F(x) = P(X \leq x),$$
$$Y = X - u.$$

For a given threshold u, the excess loss over the threshold Y has the following distribution function:

$$F_u(y) = P(X - u \leq y \,\middle|\, X > u) = \frac{F(y+u) - F(u)}{1 - F(u)}.$$

$F_u(y)$ distribution of excess losses over a high threshold u converges to a Generalized Pareto Distribution (GPD).

The cumulative distribution function of GPD:

$$G_{\xi,\beta}(y) = \begin{bmatrix} 1 - \left(1 + \frac{xy}{\beta}\right)^{\frac{1}{x}} & if\ x \neq 0 \\ 1 - \exp\left(-\frac{y}{\beta}\right) & if\ x = 0 \end{bmatrix}.$$

ξ – shape parameter

β – scale parameter

$$F_u(y) = G_{\xi,\beta}(y).$$

The loss distribution function:

$$F(x) = [1 - F(u)]G_{\xi,\beta}(x - u) + F(u).$$

$F(u)$ – estimated empirically

Mean excess function – average excess of the random variable X over the threshold u

$$e(u) = E[X - u|X > u].$$

Modeling insurance claims

We use the `evir` package from R to model insurance claims.

Data: large industrial fire insurance claims from Denmark

Period: 1980 – 1990

The data shows all fire losses exceeding one million Danish kroner.

```
#install evir package
```

```
install.packages("evir")

#use evir library
library(evir)

#use Danish insurance claims data
data(danish)

#number of observations.
length(danish)

 [1] 2167
#the maximum insurance claim= 263.25 million Danish Kroner
max(danish)

 [1] 263.2504
#the minimum insurance claim=1 million Danish Kroner
min(danish)

 [1] 1
#the mean insurance claim = 3.385 Danish Kroner
mean(danish)

 [1] 3.385088


#Plot the complementary cumulative distribution function
#emplot function - plots the complementary cumulative
distribution function(ccdf).
```

This shows the empirical probability of the claims exceeding any given threshold. alog="xy" command makes logarithmic scales for both axes.

`#ccdf - is not linear` (see Graph 1).

```
#ccdf - is linear if we use logarithmic scales for both axes.
```

This shows the fat-tailed nature of the data. It also shows that the claims may follow a Pareto-type distribution (see Graph 2).

```
emplot(danish)
emplot(danish, alog="xy")

#Plot the mean excess function of Danish fire insurance claims
#meplot function - plots the sample mean excesses over
increasing thresholds
#omit argument - eliminates the indicated number of upper
points

meplot(danish, omit=5)

#Determining the threshold
#We can use the visual inspection of the plot to determine a
threshold.
#between 0 and 10 - there is a slight curve indicating that
smaller losses might be modeled using a different law
#between 10 and 20 - linear graph
#higher than 20 - sparse data
#Considering the information presented above, 20 can be a
good threshold (see Graph 3)


#Fitting a gpd distribution to the tails

gpdfittedparameters <- gpd(danish, threshold = 20)

gpdfittedparameters

#R provides the following results.
#$n represents the total number of observations (in our
case 2167)
```

#$threshold=20 we selected it initially

#$n.exceed 36- the number of insurance claims that exceed the threshold

#$method “ml”- Maximum Likelihood Estimation method used for estimating xi and beta parameters of the GPD distribution

#$par.ests - xi and beta estimated parameters

#$par.ses - standard errors

```
$n
[1] 2167

$data
 [1]  26.21464  21.96193 263.25037  34.14155  20.96986  56.22543
[16]  29.02604  23.28386  32.46753  29.03711  27.82931  38.15439
[31]  20.86367 152.41321  32.38781  20.82673 144.65759  28.63036

$threshold
[1] 20

$p.less.thresh
[1] 0.9833872

$n.exceed
[1] 36

$method
[1] "ml"

$par.ests
       xi      beta
0.6840479 9.6316941

$par.ses
       xi      beta
0.2749542 2.8958268
```

```
$varcov
             [,1]       [,2]
[1,]  0.07559979 -0.4202254
[2,] -0.42022538  8.3858127

$information
[1] "observed"

$converged
[1] 0

$nllh.final
[1] 142.1845

attr(,"class")
[1] "gpd"
```

Calculating the expected loss (see Graph 4)

```
library(evir)
data(danish)
gpdfittedparameters <- gpd(danish, threshold = 20)
tp <- tailplot(gpdfittedparameters)
#Non-parametric estimation of Value of Risk (VaR) using
quantiles
#the estimated 99.9 % quantile is 102.18 million Danish kroner
gpd.q(tp, pp = 0.999, ci.p = 0.95)
```

```
Lower CI  Estimate  Upper CI
63.27843 102.18226 310.68768
```

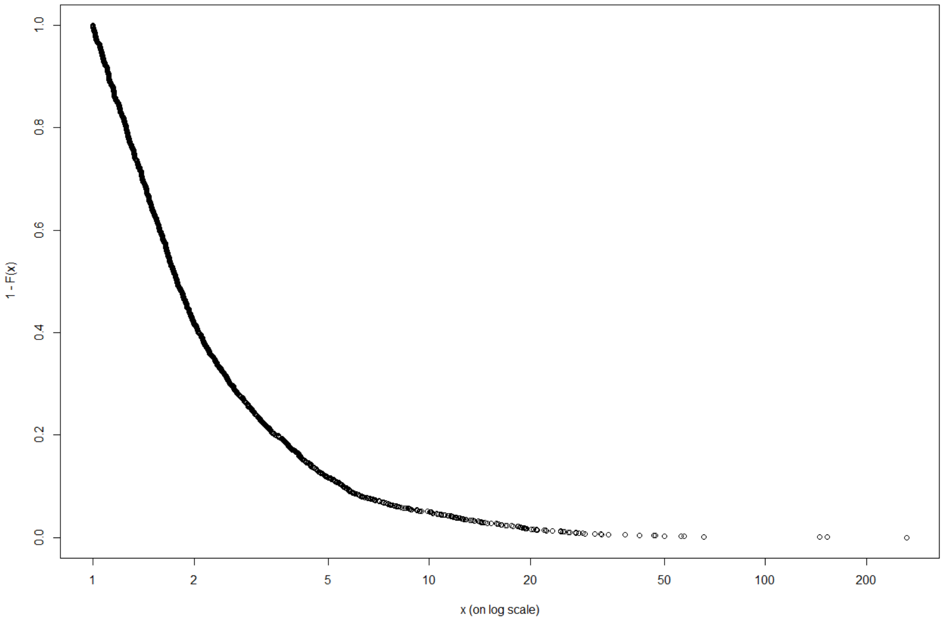
```
tp <- tailplot(gpdfittedparameters)

#The estimated 99 % quantile is 25.845 million Danish kroner
#Lower Confidence Interval: 23.44 million Danish kroner
#Upper Confidence Interval: 29.80 million Danish kroner
gpd.q(tp, pp = 0.99)
```

```
Lower CI Estimate Upper CI
23.44016 25.84510 29.79587
```

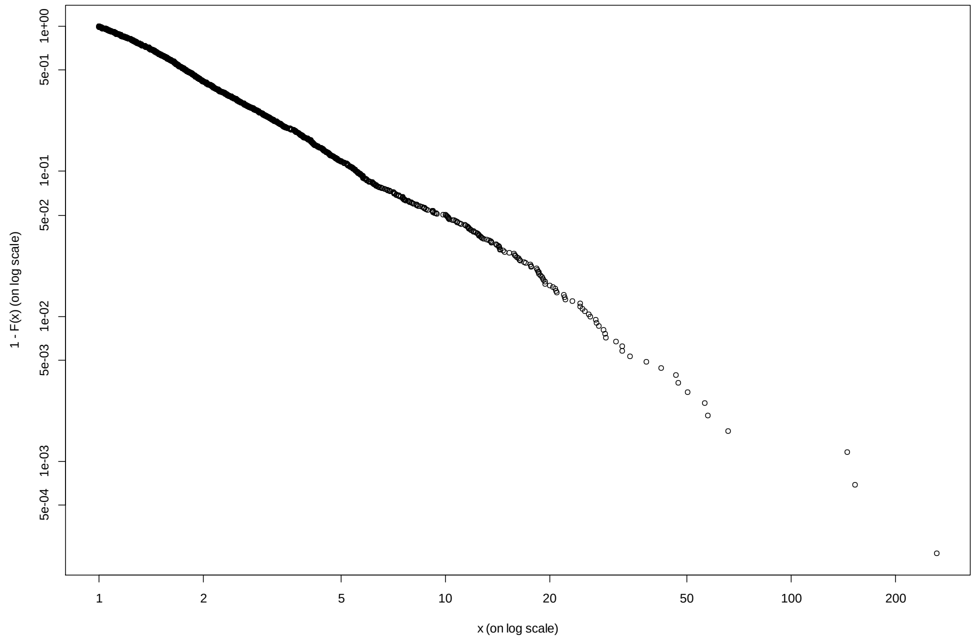
```
#Assuming a 99 % quantile level of 25.8451 million Danish
kroner is exceeded, the expected loss is 68.98463 million
Danish kroner
gpd.sfall(tp, 0.99)
```

```
Lower CI  Estimate  Upper CI
42.16106  68.98463 394.87555
```

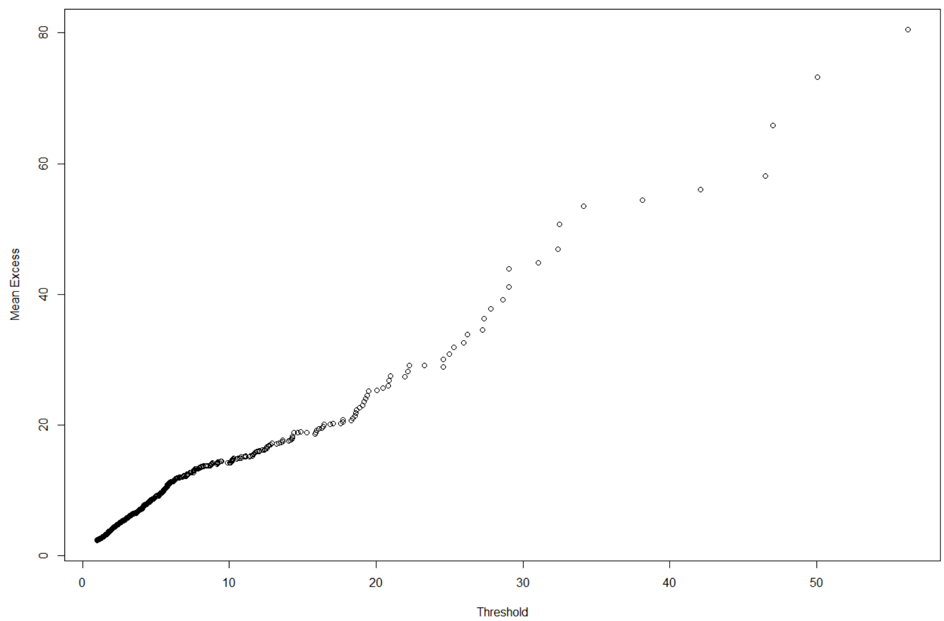
Graphs



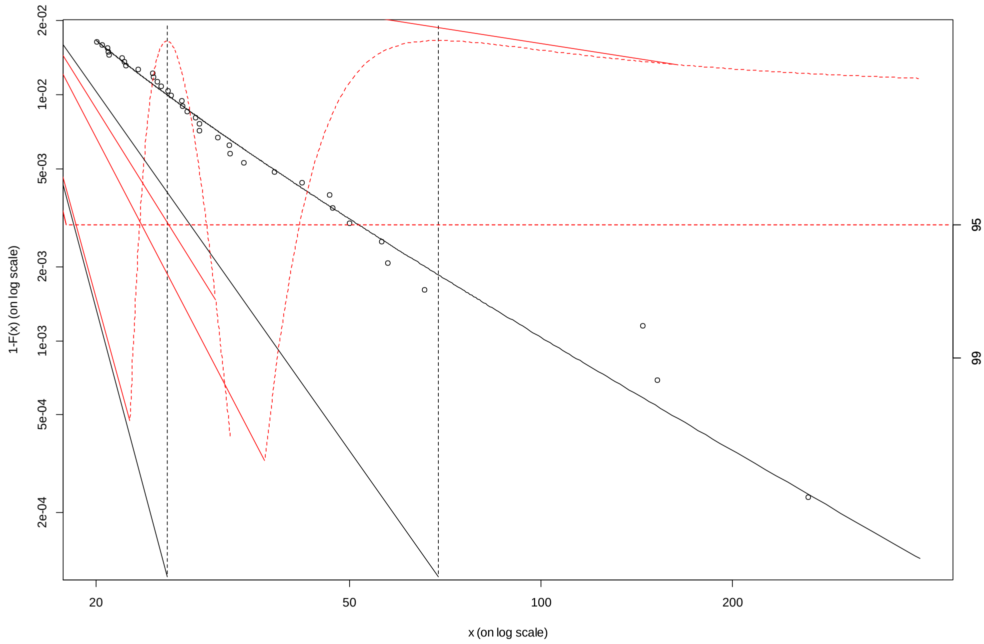
Graph 1: Empirical CCDF of Danish fire insurance claims



Graph 2: Empirical CCDF of Danish fire insurance claims – logarithmic scale for both axes



Graph 3: Mean excess function of Danish fire insurance claims



Graph 4. Tail of the Danish fire loss data with estimated 99.9% quantile

For Graph 4:

- The first vertical dashed line – 99% quantile or VaR
- The first curved dashed red line – likelihood curve
- The second vertical dashed line – 99% expected shortfall
- The second curved dashed red line – likelihood curve

3.2.6 Transcript: Extreme Value Theory - Theorectical Results

Extreme Value Theory considers the asymptotic behavior of $[F(q)]^n$ as $n \to \infty$.

As you may recall from Module 2, the OLS estimator based on a sample of size n, $\hat{\beta}_{OLS}^{(n)}$ converges to the true parameter in the population β as the sample size increases:

$$\lim_{n\to\infty} \left(\hat{\beta}_{OLS}^{(n)} - \beta\right) = 0.$$

Here the random variable $\hat{\beta}_{OLS}^{(n)}$ converges to a constant (a degenerate distribution with a point mass on one constant value, i.e. zero, variance). However, by the Central Limit Theorem (CLT), a suitably re-weighted version of this difference converges to a normally distributed *random* variable (i.e. a non-degenerate distribution with positive variance):

$$\sqrt{n}\left(\hat{\beta}_{OLS}^{(n)} - \beta\right) \underset{n\to\infty}{\sim} N\left(0, Var\left(\hat{\beta}_{OLS}\right)\right).$$

An asymptotic inference on the maximum value of a sample is based on a result that is analogous to the CLT. This is called the **Extremal Types Theorem.**

We define the maximum of a sample of n values of some random variable with the common distribution $F(q)$ as:

$$m_n = \max_i \{x_i\}_{i=1}^n.$$

Suppose that the maximum value that x can ever take is given by x_{max}, then for any $q < x_{max}$:

$$\lim_{n\to\infty} [F(q)]^n = 0.$$

However, for standard distributions, one can find sequences of constants a_n and $b_n > 0$ with well-defined limits $a_n \to a$ and $b_n \to b$, such that the normalized maximum value $m_n^* = \frac{m_n - a_n}{b_n}$ has the following asymptotic distribution:

$$\lim_{n\to\infty} \Pr[m_n^* \leq q] = G(q),$$

where $G(q)$ is a non-degenerate distribution function of one of *only* three families of distribution functions:

1 The Gumbel distribution:

$$G(q) = \exp\left(-\exp\left[-\frac{q-a}{b}\right]\right), \quad -\infty < q < \infty.$$

2 The Fréchet distribution:

$$G(q) = \begin{cases} 0, & q \leq a \\ \exp\left[-\frac{q-a}{b}\right]^{-\alpha}, & q > a \text{ and } \alpha > 0 \end{cases}.$$

3 The Weibull distribution:

$$G(q) = \begin{cases} \exp\left(-\left[-\frac{q-a}{b}\right]^{\alpha}\right), & q < a \text{ and } \alpha > 0 \\ 1, & q \geq a \end{cases}.$$

The names and functional forms of these distributions are not the focus of this course; the names and functional forms, as well as their characteristics are easily estimated by standard techniques available in R.

What you should focus on is the remarkable nature of the result: no matter the underlying distribution $F(q)$, the maximum of a sample converges to only one of these three distributions. This gives us a powerful asymptotic tool to study extreme outcomes.

The three distributions have some simple characteristics to bear in mind:

1. The Gumbel distribution has full support; it does not have lower or upper limits. The max of a sample of normally distributed variables will have an asymptotic Gumbel distribution.
2. The Fréchet distribution has a lower limit. The max of a sample of Fréchet distributed variables will have an exact Fréchet distribution.
3. The Weibull distribution has an upper limit. The max of a sample of uniformly distributed variables will have an asymptotic Weibull distribution.

For further illustration, here are some example graphs of the density of each of these distributions (where the parts of the Fréchet and Weibull densities that are 0 have been suppressed for exposition, to illustrate that the Gumbel distribution has no upper or lower bounds – i.e. no part of the Gumbel density is exactly equal to 0):

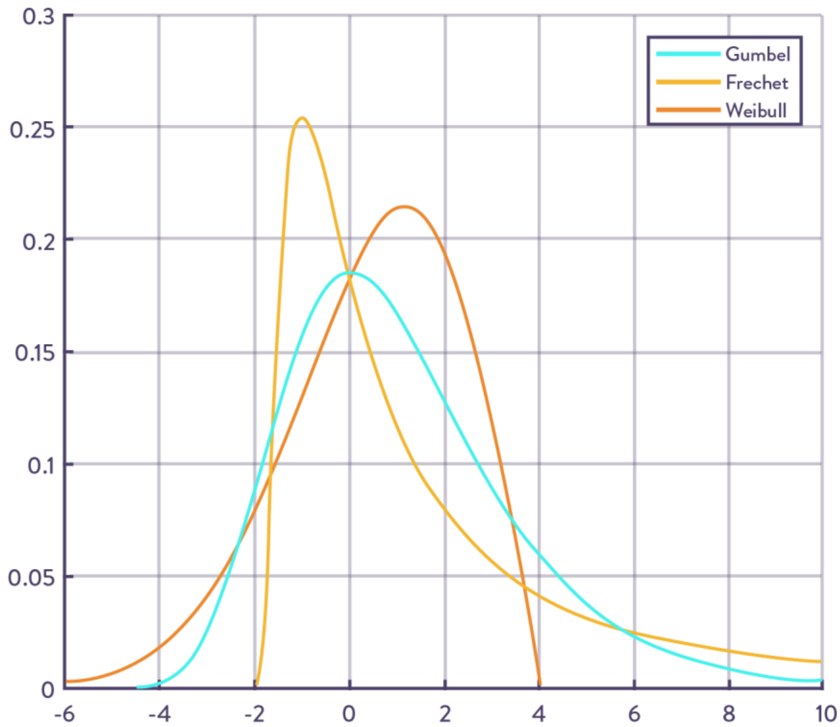


Figure 2: Different types of GEV distributions

We can capture all three of these distribution types in terms of the generalized extreme value (GEV) distribution. The GEV distribution has three parameters: a location parameter a (which shifts the density left or right); a scale parameter b (which increases or decreases the width/variance of the density); and a *shape* parameter ξ which determines which one of the three distributions occur, and the nature of their shape.

The GEV distribution is given by:

$$G(q) = \exp\left\{-\left[1 + \xi\left(\frac{q-a}{b}\right)\right]^{-\frac{1}{\xi}}\right\}.$$

Where $G(q)$ is defined for all q that satisfy $1 + \xi\frac{(q-a)}{b} > 0$, with $-\infty < a < \infty$, $b > 0$ and $-\infty < \xi < \infty$.

The "general" in the GEV name is because of the following result (which you should check): we obtain the Gumbel distribution as $\xi \to 0$, the Fréchet distribution if $\xi > 0$, and the Weibull distribution if $\xi < 0$.

A graphical interpretation of the different types of GEV distributions

The quantile q_p has the interpretation that, in the next investment period of length τ, the probability that the process x_t will have a maximum larger than q_p is p.

We can thus define quantile q_p of the GEV distribution as the inverse of the distribution function evaluated at some probability p.

$$q_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{for } \xi \neq 0 \\ \mu - \sigma\log\{-\log(1-p)\} & \text{for } \xi = 0 \end{cases}.$$

A way of visualizing the analytical content of a GEV distribution is constructed as follows. Define $y_p = -\log(1-p)$. Now consider a plot of the quantiles q_p with $p \in [0,1]$ as a function of $\log y_p$. If the GEV distribution is Gumbel ($\xi = 0$), the resulting plot will be linear. If the GEV distribution is Fréchet $(\xi > 0)$ the resulting plot will be that of a convex function. Lastly, if the GEV distribution is Weibull ($\xi < 0$) the resulting plot will be that of a concave function.

We will use this, called a return level plot, as a diagnostic of a fitted GEV distribution below.

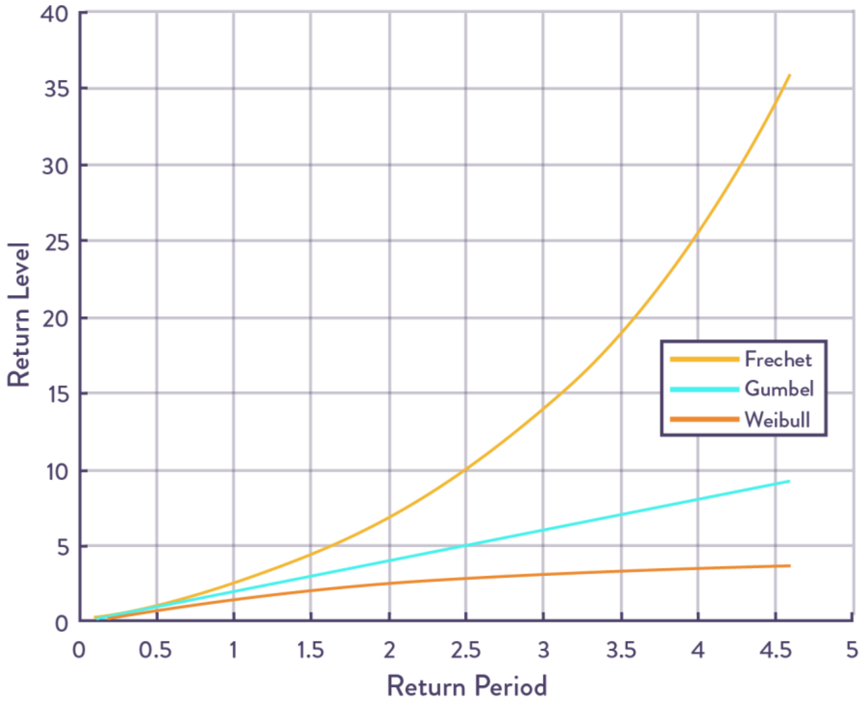


Figure 3: Return level plot

A scatter plot of the actual return levels against estimated return periods gives a view of how regularly the GEV distribution captures the pattern of observed maxima.

3.2.7 Transcript: Interpretation of an Estimated GEV Distribution: Value at Risk and Expected Shortfall

Once estimates of parameters have been obtained, the most informative ways of evaluating what the estimation means comes from studying the quantiles of the extreme value distribution.

These are simply obtained by inverting the distribution function. The p^{th} quantile, which we denote q_p, is defined as the value of the extreme that yields:

$$G(q_p) = 1 - p,$$

so that:

$$q_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{for } \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\} & \text{for } \xi = 0 \end{cases}.$$

Why is this useful?

The quantile q_p has the interpretation that in the next investment period of length τ, the probability that the process x_t will have a maximum larger than q_p is p. Hence, we have a quantification of the likelihood of an extreme observation that exceeds a known value. This will hold for minima as well, so this is exactly what we needed. If we are willing to accept a probability p risk of a severe loss, then q_p is the value-at-risk (VaR) at probability p, where q_p is the p^{th} quantile of the GEV distribution fitted to the block maxima of process $-x_t$ (equivalently, the block minima of process x_t).

Another interpretation of quantile p is that it is the return level associated with return period $\frac{1}{p}$. By this terminology we mean that the process x_t will exceed the level q_p in expectation once every $\frac{1}{p}\tau$ periods.

One weakness of VaR is that it may underestimate the actual loss that occurs if the process exceeds the quantile.

The concept of expected shortfall addresses this issue and is simply the expected value that the process will take, conditional on exceeding the quantile of interest.

If x_t is a normally distributed variable (say, a log return process) with conditional expectation μ_t conditional variance[4] σ_t^2, and we are interested in VaR q_p, then

$$E(x_t | x_t > q_p) = \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{q_p^2}{2}\right)}{p}.$$

The expected shortfall for quantile q_p is then defined as:

$$ES_{1-p} = \mu_t + E(x_t | x_t > q_p)\sigma_t^2.$$

Thus, we take the conditional expectation and add our best estimate of the variance of the process at the time that we make the forecast, inflated by the expectation of the process conditional on exceeding the quantile of interest. Note that a larger variance or a smaller probability – i.e. a more unlikely realization – increases the expected shortfall.

The following example is taken from Tsay (2010):

> If $p = 0.05$ and $x_t \sim N(0,1)$ then $VaR_{0.95} \approx 1.645$ and $ES_{0.95} = 2.0627\sigma_t^2$.
> If $p = 0.01$ then $ES_{0.99} = 2.6652\sigma_t^2$.

[4] Example obtained from an ARCH model for the process x_t.

3.2.8 Transcript: Maximum Likelihood Estimation of GEV Distribution

In this video we will be estimating the maximum likelihood of GEV distribution using R.

For this exercise we use the `evir` package in R. The output has been edited for concise presentation.

```
# Load packages and data, store the log returns in a variable:
library(tidyverse)
library(readxl)
library(evir)
SnP500 <- read_excel("C:/<your path>/<your file name>",
                      col_types = c("date", "numeric",
"numeric"))
# store log returns in a variable:
LogReturns = SnP500$LogReturns
# to fit the GEV to minima, we take the negative of the
variable:
NegativeLogReturns = -LogReturns
# the following call finds the coefficients of the GEV
distribution for blocks of 21 days (the approximate number of
trading days in a month
gev(NegativeLogReturns, block = 21)
```

The output has been edited for concise presentation. Text following # was added to explain the output:

```
$n.all
[1] 2434
# the sample has 2434 observations
$n
[1] 116
# this yields 116 blocks of length 21
$data
  [1] 0.024587038 0.020584529 0.018279807 0.031376343 0.023396064 0.030
```

```
  [7] 0.092002411 0.094695145 0.093536559 0.030469134 0.050368620 0.047741
…
[115] 0.041842561 0.013413908
# the data used in the estimation is the 116 block maxima
$block
[1] 21
```

Suppose we are instead interested in blocks of length 63, the approximate number of trading days in a quarter:

```
gev(NegativeLogReturns, block = 63)
```

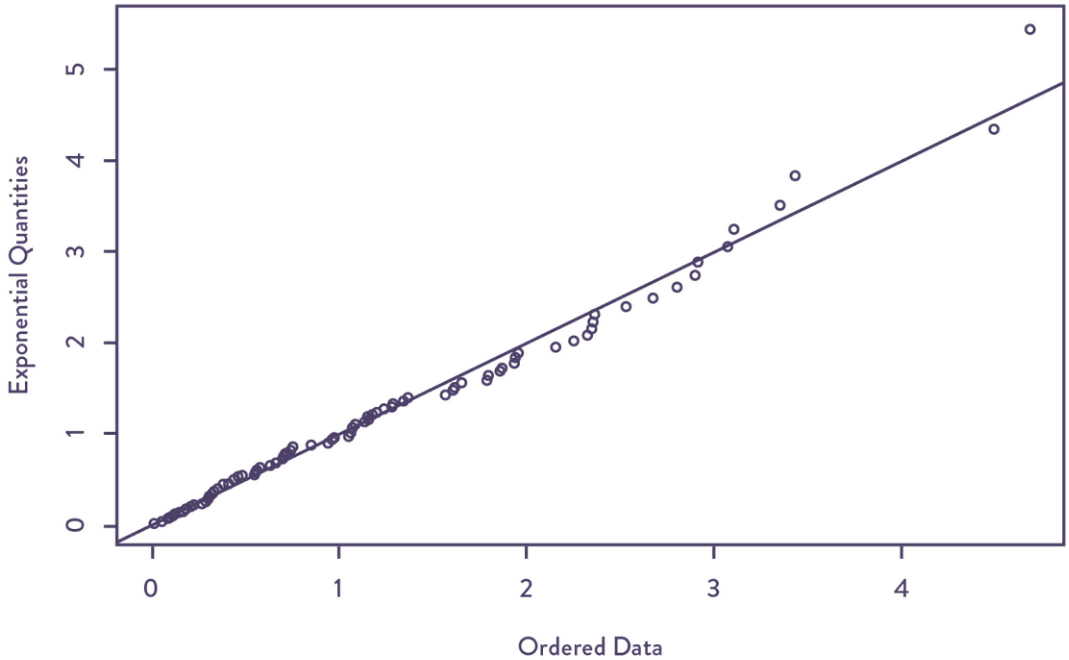
Output:

```
$n.all
[1] 2434

$n
[1] 39

# now we have only 39 block maxima to use to uncover the parameters

$block
[1] 63

$par.ests
         xi       sigma          mu
0.137495327 0.009590904 0.020706169

$par.ses
         xi       sigma          mu
0.103596861 0.001184601 0.001689171

# The low number of data points clearly lead to a severe deterioration of
the statistical significance of our parameter estimates. We would be
hesitant to attach too much value to the point estimate
```

Unfortunately, the `evir` package does not automatically provide the standard plots that we are interested in, but you should be able to generate these with careful use of the methods in this package and the standard methods in, for example, the `stats` package, which gives the standard statistical methods such as histograms. The `evir` package does at least provide the quantile plot:

```
GEV<-gev(NegativeLogReturns, block = 21)
plot(GEV)
```

Now you have to make a selection. Type "2" to obtain the quantile plot, as pictured below:



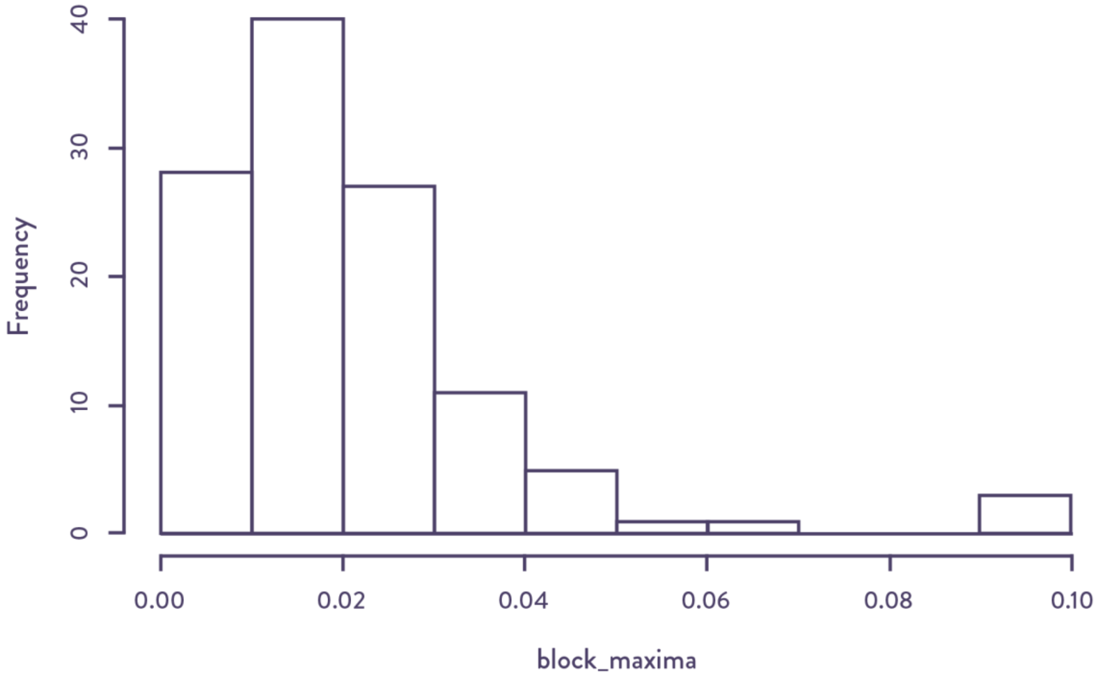
From this plot we conclude that for the smaller end of the block maxima we found, the GEV distribution does a reasonable job at capturing the behavior of smaller extrema. For the largest extrema, the distribution of the extrema deviates further from the perfect line. Thus, the most extreme negative returns of the S&P500 are less well modeled by the GEV approach.

We can also obtain a histogram of the extreme values to evaluate the closeness of the GEV distribution that we estimated to the empirical density of our data:

```
GEV<-gev(NegativeLogReturns, block = 21)
```

```
block_maxima = GEV$data
hist(block_maxima)
```

This yields the following graph:



Again, we reach the same conclusion: While the lower end of the distribution seems to conform to the possible shapes of the GEV distribution (see Figure 2), the very extreme values seem to deviate from this model.

References

Bensalah, Y. (2010). *Steps in Applying Extreme Value Theory to Finance: A Review*. [online] Bank of Canada. Available at: https://www.banqueducanada.ca/wp-content/uploads/2010/01/wp00-20.pdf.

Daróczi, G et al. (2013). 'Chapter 2: Portfolio Optimization', in *Introduction to R for quantitative finance*. Birmingham: Packt Publishing, pp.27 - 42.

Daróczi, G., et al. (2013). 'Chapter 8: Extreme Value Theory', in *Introduction to R for quantitative finance*. Birmingham: Packt Publishing, pp.113 - 124.

Embrechts, P., Resnick, S. and Samorodnitsky, G. (2000). Extreme Value Theory as a Risk Management Tool. *Derivatives Use Trading & #38 Regulation*, [online] 6(1), pp.30-41. Available at: https://www.researchgate.net/publication/2640395_Extreme_Value_Theory_Potential_And_Limitations_As_An_Integrated_Risk_Management_Tool.

Embrechts, P. (2009). Copulas: A Personal View. *Journal of Risk and Insurance*, [online] 76(3), pp.639-650. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6975.2009.01310.x.

Fernández, V. (2003). Extreme Value Theory: Value at Risk and Returns Dependence Around the World. *Documentos de Trabajo*, [online] 161. Available at: https://ideas.repec.org/p/edj/ceauch/161.html.

Gilli, M. and Këllezi, E. (2006). An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics*, [online] 27(2-3), pp.207-228. Available at: https://link.springer.com/article/10.1007/s10614-006-9025-7.

Marimoutou, V., Raggad, B. and Trabelsi, A. (2006). *Extreme Value Theory and Value at Risk: Application to Oil Market*. [online] Available at: https://halshs.archives-ouvertes.fr/halshs-00410746/document.

McNeil, A. (1997). Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. *ASTIN Bulletin*, [online] 27(01), pp.117-137. Available at: https://www.cambridge.org/core/journals/astin-bulletin-journal-of-the-iaa/article/estimating-the-tails-of-loss-severity-distributions-using-extreme-value-theory/745048C645338C953C522CB29C7D0FF9.

McNeil, A., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. New Jersey: Princeton University Press.

Ruppert, D. and Matteson, D. (2015). *Statistics and Data Analysis for Financial Engineering*. New York: Springer.

Sklar, A. (1959). Fonctions de Répartition à n Dimensions et Leurs Marges. *Institut Statistique de l'Université de Paris*, 8, pp.229-231.

Tsay, R. (2010). *Analysis of financial time series*. Hoboken, N.J.: Wiley.