# Multivariate Risk:
# From Univariate to High-Dimensional Graphical Models

*A Thesis Submitted in Fulfilment of the Requirements for the Degree of*
Master in Statistics, 15 ECTS

*by*

**Erik Oldehed**

Supervised by: Marie Kratz, Sreekar Vadlamani


LUND
UNIVERSITY

Department of Statistics
LUND UNIVERSITY

2020

STAN40 (2020) 1–69 Submitted 18/6

# Multivariate Risk: From Univariate to High-Dimensional Graphical Models

**Erik Oldehed** ERIK.OLDEHED@ICLOUD.COM
*Department of Statistics*
*Lund University*
*220 07, Lund, Sweden*

## Abstract

We present a comparison of different univariate and multivariate extreme value risk models. Our focus is on exploring how these can be used to model financial risk. We use simulated as well as real data and compare deterministic and cross-validation threshold selection methods for the GP model to a GEV model. For comparison, we carry out a bivariate analysis using copulas. Finally, an undirected graphical lasso model using $n = 45$ block maxima of the log-returns from 95 of the stocks in the FTSE 100 index is combined with copulas and PCA to model the extreme loss risk within the FTSE 100 index. The contribution of this study lies in exploring some ideas on risk models in multivariate high-dimensional settings.

**Keywords:** Block Maxima, Mean Excess Plot, Tail Risk, Cross-Validation Threshold Selection, Graphical Lasso, Nonparanormal Distribution.

## Contents

## 1. Introduction

In an increasingly uncertain world, the ability to assess risk may very well be one of the core capabilities of a society. One of the earliest theories of risk is credited to Cramér (1930). Since then, the modern portfolio theory of Markowitz (1952) was considered an industry standard for many years with variance as a central measure of risk. The limits of variance in quantifying extreme unexpected losses caused a wide range of quantile-based measures to emerge as a response.

One such measure is value-at-risk (VaR). It was intended as a measure of extreme risk where the variance had previously failed. Initially available only to a small group of quantitative analysts, the RiskMetrics publication made the popularity of VaR to surge (see Longerstaey and Spencer, 1996). The Basel II Accord further positioned VaR as the main measure of risk (Basel Committee on Banking Supervision, 2004). Expected shortfall (ES) was introduced by Rappoport (1993) at JP Morgan but was initially neglected in favour of VaR. Although Artzner et al. (1999) formalised the properties of risk, pointing to the limits of VaR, it was not until the Basel III Accord was published in the wake of the financial crisis 2008 that ES became a widely accepted risk measure (Basel Committee on Banking Supervision, 2010).

As a consequence of globalisation and technical advancements, the presence of volatility clustering has spurred a wide interest in risk models rather than risk measures. The autoregressive conditional heteroskedasticity (ARCH) model, credited to Engle (1982), was created to address this problem. The ARCH model and its extensions mainly model volatility. On the other hand, extreme value theory (EVT) is entirely concerned with the tail risk. Methods to model extreme events go back to the pioneering work of Fréchet, 1927. Rather than focusing on volatility, extreme value models describe the behaviour in the tails of a distribution. It has been shown that extreme value models are better suited than the stochastic volatility models in modelling VaR (Abad et al., 2014).

Even if our knowledge on risk models in the univariate case is fairly well understood, the behaviour of risk in higher dimensions is still not entirely clear. Considering bivariate risk, copulas provide a reliable way to model tail dependencies. As extreme events rarely occur, when the dimension scales up, multivariate extreme value models suffer from the curse of dimensionality. In particular, when the number of variables is larger than the sample size, expressing the joint distribution is often very challenging. A promising suggestion is to utilise undirected graphical models to estimate sparse covariance matrices. One such method is the *nonparanormal* approach (Liu et al., 2009). While this method previously has been applied to stock returns it has not yet been tested in modelling the multivariate dependencies in extreme losses of stock index components.

The aim of this study is to compare some common methods to model financial risk in the univariate and bivariate case while proceeding to model multivariate risk combining a *nonparanormal glasso* model with copulas and principal component analysis.

This study is organised in two parts. Section 2 introduces univariate risk models while Section 3 concerns the multivariate equivalent. We begin with a description of extreme value theory and then review some methods of univariate risk modelling. A simulation study is then performed after which a case study of the risk measures of data corresponding to pre- and post-Brexit is presented. Progressing into the multivariate case, a recap of some classic and some newer theories of multivariate risk are presented. We then describe the methods used in this study to model multivariate risk. Finally, a simulation study is performed followed by a case study of the dependence

between the FTSE 100 and S&P 500 stock indexes. We then model the extreme dependencies of 95 stocks in the FTSE 100 index. Finally, some concluding remarks are presented.

### 1.1 Stock Returns

Before defining what we mean with risk we must get an understanding of the variable that we wish to model the risk from. Since this study is concerned with financial risk on the stock market our focus is on stock returns. The choice in calculating stock returns comes down to using simple or log-returns.

The advantages of using log transformed prices can be tied to the seminal work of Samuelson (1965). Security prices are then assumed to follow a Markovian process. Under this assumption the current price incorporates all useful information in forecasting future values, implying a weak form of market efficiency. It is reasonable to believe that the increments of stock prices are independent. Hence, a natural choice would be to model stock prices as a Brownian motion. However, a Brownian motion is defined for negative values while prices are always positive. Furthermore, it does not depend on the process value itself, causing small prices to move in increments of similar size as larger prices. These are unreasonable assumptions. A more reasonable premise is that log-returns follow a Brownian motion. If this holds true, stock prices move according to a geometric Brownian motion, and it can be shown that the log-returns are normally distributed.

Let the closing value for the asset at time $t$ be denoted $S(t)$. Then the log-returns for the series are defined as

$$R_{\Delta t}(t) := \log\left(\frac{S(t)}{S(t-\Delta t)}\right)$$

where $\Delta t$ is the interval between the last closing value and the present. For our purpose, an interval is chosen such as $\Delta t = 1$ day. The losses $L$ are defined as $L := -\{R : R < 0\}$.

### 1.2 Risk

Although techniques for measuring risk are central to risk management there is no unique definition of risk. Measuring risk in terms of probability distributions might sound like an attractive idea. However, mapping from spaces of probability distributions into real numbers provides risk assessment in terms of a single capital amount representing unexpected future losses. Artzner et al. (1999) argue that instead of defining risk as the change in value between two dates, risk is related only to the variability of the future value of a position. A better measurement of risk is said to assess whether the future value belongs to a subset of "acceptable risks", as decided by a supervisor. A suggested criterion for choosing a risk measure is coherence. Four axioms are given, which if fulfilled imply that a risk measure is coherent. Following Artzner et al. (1999) we let the number $\rho(L)$ assigned by the measure $\rho$ to the risk $L$, be the minimum additional cash an agent has to add to the risky position $L$ to continue his investments. A negative value means that the cash amount $-\rho(L)$ can be withdrawn from the position.

**Definition 1.1.** A risk measure $\rho$, satisfying the following axioms, is called coherent:

**Axiom 1.** Positive homogeneity. For all loss variables $L$ and constants $h \geq 0$ it holds that $\rho(hL) = h\rho(L)$

**Axiom 2.** Subadditivity. For all loss variables $L_1$ and $L_2$ it holds that $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$

**Axiom 3.** Monotonicity. For all loss variables $L_1$ and $L_2$ it holds that $L_1 \leq L_2 \Rightarrow \rho(L_1) \leq \rho(L_2)$

**Axiom 4.** Translation invariance. For all loss variables $L$ and $\alpha \in \mathbb{R}$ it holds that $\rho(L - \alpha) = \rho(L) - \alpha$

One of the most commonly used risk measures is Value-at-Risk (VaR).

**Definition 1.2.** (Longerstaey and Spencer, 1996) The Value-at-Risk (VaR) of loss variable $L$ at level $\alpha$ is given by

$$\mathrm{VaR}_\alpha(L) = \inf\{x \in \mathbb{R} : \mathbb{P}(L \leq x) \geq \alpha\}, \qquad \alpha \in (0,1).$$

However, VaR has received criticism for not adequately capturing the expected loss since, for a given quantile $\alpha$, it only conveys information on the minimum loss (McNeil et al., 2005). In most circumstances, VaR is not subadditive meaning that the risk of a sum of portfolio positions can be larger than the sum of the portfolio positions individual risk.

**Definition 1.3.** (Rappoport, 1993) The Expected Shortfall of $L$ at level $\alpha$ is given by

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)du, \qquad \alpha \in (0,1).$$

While the mentioned shortcomings of VaR are addressed in the coherent risk measure ES, more validation data is required to backtest this measure compared to VaR (Emmer et al., 2015).

After giving a summary of definitions in Section 1.3 we will present the theory necessary to extend risk from being a scalar measure into a model. As our focus is on tail risk, Section 2 will give an introduction to extreme value theory.

### 1.3 Notation

We begin with introducing some notation that is used throughout the report.

Table 1: List of Abbreviations and Symbols

| Abbreviation or symbol | Description |
|---|---|
| $a$ | Parameter of the maximal data information prior. |
| $\alpha$ | In VaR/ES: the particular quantile of choice. In extreme value distributions: the shape parameter and the reciprocal of the tail index $\xi$, i.e. $\alpha = 1/\xi$ (not to be confused with the order of the quantile). |
| $a_n$ | Normalising constant in the GEVD. |
| $a.s.$ | Almost surely. |
| $A^2$ | The Anderson-Darling test statistic. |
| $A^{*2}$ | The modified Anderson-Darling test statistic. |

| | |
|---|---|
| $\simeq$ | Asymptotic equivalence. |
| $B$ | The number of bootstrap samples. |
| $b$ | The block size. |
| $b_n$ | Normalising constant in the GEVD. |
| $\beta$ | The scale parameter. |
| $\beta(u)$ | The scale parameter in the generalised Pareto distribution as a function of the threshold $u$. |
| Brexit referendum | The United Kingdom European Union membership referendum 2016-06-23. |
| $c$ | The density of a copula $C$. Clique in the context of graphs. |
| $\mathcal{C}$ | Set of maximal cliques. |
| $\mathbb{C}$ | Covariance. |
| cdf | Cumulative distribution function. |
| CI | Confidence interval. |
| CV | Coefficient of variation. |
| $d$ | Dimension. |
| d.f. | Distribution function. |
| := | Defined as. |
| $\Delta$ | Change. |
| $\delta_n$ | Truncation parameter in the Winzorized estimator of the cdf, used to estimate high dimensional graphical models. |
| Differentiable function | Function for which the derivative exists for every point in its domain. |
| $\mathbb{E}(\cdot)$ | Expectation. |
| $\mathcal{E}$ | Set of edges in a graph. |
| $e$ | Inverse of the natural logarithm. In the context of graphs $e_i$ denotes the $i$th node. |
| $\varepsilon$ | An arbitrary small number. |
| $\eta_k$ | The $k$th moment. |
| $e(u)$ | Mean excess function. |
| EDF | Empirical distribution function. |
| ES | Expected Shortfall. |
| $\exists$ | Logic notation, "exists". |
| exp | Inverse of the natural logarithm. |
| $F$ | Cumulative distribution function. |
| $\forall$ | Logic notation, "for all". |
| $\mathcal{G}$ | Graph. |
| $\Gamma$ | Gamma function. |
| $G_{\xi,\beta(u)}$ | The generalised Pareto distribution cdf. |
| GEV | Generalised Extreme Value. |
| GEVD | Generalised Extreme Value Distribution. |
| glasso | Graphical least absolute shrinkage and selection operator. |
| GP | Generalised Pareto. |
| GPD | Generalised Pareto Distribution. |
| $H$ | Distribution function of the generalised extreme value distribution. |

| | |
|---|---|
| $H_k$ | The Hill estimator of the tail index parameter for the $k$th upper order statistics. |
| Identifiable density | Density where the true values of the parameters are possible to derive from an infinitely large number of observations. |
| $\iff$ | If and only if. |
| i.i.d. | Independent and identically distributed. |
| $\in$ | Is an element of. |
| i.o. | Infinitely often. |
| $k$ | The $k$th upper order statistic. |
| $\kappa$ | The shape parameter of the Pareto distribution. |
| $k(n)$ | The $k$th upper order statistic associated with a particular threshold. |
| $L$ | The losses of a random variable. |
| $\Lambda$ | Covariance of the uniform r.v.'s corresponding to the copula of the non-paranormal distribution. |
| $\lambda$ | The coefficient of tail dependence in the context of copulas. The penalty parameter in the context of glasso models. The rate parameter for the exponential distribution. |
| $\lambda_l$ | The lower tail dependence coefficient. |
| $\lambda_u$ | The upper tail dependence coefficient. |
| lasso | The least absolute shrinkage and selection operator. |
| lim | Limit. |
| lim sup | Limit superior. |
| log | The natural logarithm. |
| $MDA$ | Maximum domain of attraction. |
| ML | Maximum likelihood. |
| $M_n$ | The maximum of $n$ random variables. |
| $\mapsto$ | Maps to. |
| MSE | Mean square error. |
| $\mu$ | The mean of a normally distributed random variable. |
| $n$ | The number of observations. |
| $\nabla$ | Gradient. |
| $\nu$ | The degrees of freedom of a Student t random variable. |
| $\mathcal{O}(\cdot)$ | Big-O expresses the limiting value of a function when the argument $\cdot$ tend to a value. |
| $o(\cdot)$ | Little-o expresses the limiting behaviour of a function and is a stronger statement than $\mathcal{O}(\cdot)$. |
| $\Omega$ | Precision matrix. |
| $p$ | Probability density function. |
| $p_u$ | The probability of exceeding $u$ in the binomial-GP model. |
| $\mathbb{P}$ | Probability distribution function. |
| $\partial$ | Partial derivative. |
| PC | Principal component. |
| PCA | Principal component analysis. |
| $\perp\!\!\!\perp$ | Perpendicular, denoting the independence of two random variables. |
| $\phi$ | The density of a standard normal random variable. |

| | |
|---|---|
| $\Phi$ | The cumulative distribution function of a standard normal random variable. |
| $\bar{\Phi}$ | The complement of the cumulative distribution function of a standard normal random variable. |
| $\pi_{Ma}$ | Generalized maximal data information prior distribution. |
| Post-Brexit | Period after the Brexit referendum (2016-06-24–2019-05-31). |
| Pre-Brexit | Period before the Brexit referendum (1984-01-03–2016-06-23). |
| $\propto$ | Proportional to. |
| $\Psi$ | Potential function. |
| $Q$ | Sparsity measure. |
| $\mathbb{R}$ | Real line. |
| $R$ | Log-return. |
| $R_{n_i}^{(j)}$ | The rank of random variable $X_i^{(j)}$. |
| $R_{\Delta t}(t)$ | Log-returns at time $t$. |
| $\rho$ | The parameter of a bivariate copula. |
| r.v. | Random variable. |
| $S$ | Sample covariance matrix or the closing value of the FTSE 100 index. |
| $\Sigma$ | Covariance matrix. |
| $\sigma$ | The standard deviation of a normally distributed random variable. |
| $\sigma(u)$ | In the context of GPD: a positive measurable function such that Theorem 6 holds, $\sigma(u) = \beta(u) + \xi u$. |
| sgn | Sign. |
| : | Set builder notation, such that. |
| sup | Supremum. |
| $t$ | Index for a given time point. |
| $\top$ | Transpose. |
| $\tau$ | The population parameter of Kendall's tau. |
| $\theta$ | Vector of parameters in the context of delta method intervals and Anderson-Darling test statistic. |
| $\to$ | Tends to. |
| $u$ | The threshold in the GPD model. The inverse CDF in the Anderson-Darling test statistic. A uniform random variable in the context of copulas. |
| $u_0$ | The chosen threshold for a GPD using deterministic threshold selection. |
| $u_0^*$ | The chosen threshold for a GPD using cross-validation threshold selection. |
| $\uparrow$ | Approaches from below. |
| $\mathbb{V}$ | Variance. |
| $\mathcal{V}$ | Set of nodes in a graph. |
| VaR | Value-at-Risk. |
| $X_{\setminus\{i,j\}}$ | All $X$s except for $X_i$ and $X_j$. |
| $\xi$ | The tail index. |
| $Z$ | A standard normal random variable. |
| $Z_\alpha$ | The quantile of a standard normal random variable. |

## 2. Univariate Risk Models

Univariate risk models are mainly concerned with analysing extreme events. In the following section, we review the theory of extreme values.

### 2.1 Univariate Extreme Value Theory

Extreme value theory can be categorised into two main branches. Classical extreme value theory concerns the asymptotic distributional theory for maxima of i.i.d. random variables (Leadbetter et al., 1983). Later developments focus on exceedances rather than maxima (see e.g. Kratz, 2019). There are two fundamental limit theorems, one for the mean and the second one for the maxima. Next, we will give an overview of the latter.

To facilitate the reading experience, we first introduce some definitions used throughout the study. We begin by defining a sequence of $n$ random variables $X = \{X_i\}_{i=1}^n$. To avoid excessive use of notation we apply this definition to i.i.d. random variables from which extreme values are obtained if otherwise is not stated. Similarly, we let $F$ be the distribution function of this sequence of random variables.

We consider the limiting behaviour of the maximum of $n$ random variables $X$,

$$M_n := \max_{1 \le i \le n} X_i.$$

The probability distribution of a maximum can be expressed as a function of the probability distribution of $X_i$, where $X_i$'s are assumed to be i.i.d. with common distribution function $F$. The resulting degenerative behaviour follows from the expression below

$$\mathbb{P}\Big[\max_{1\le i\le n} X_i \le x\Big] = \prod_{i=1}^{n} \mathbb{P}\Big[X_i \le x\Big] = F^n(x) \underset{n\to\infty}{\to} \begin{cases} 0 \text{ if } F(x) < 1 \\ 1 \text{ if } F(x) = 1 \end{cases}.$$

However, returning to the initial argument, one might be able to find a linear transformation such that this degeneracy is avoided. Indeed, it can be proved that there exists a linear transformation, i.e. one can find sequences $\{a_n\}$ and $\{b_n\}$ such that

$$\mathbb{P}\left[\frac{\max X_i - b_n}{a_n} \le x\right] = \mathbb{P}\left[\max_{1\le i\le n} X_i \le a_n x + b_n\right] = F^n(a_n x + b_n) \underset{n\to\infty}{\simeq} H(x),$$

where the normalising constants $a_n > 0$ and $b_n \in \mathbb{R}$ are the normalising sequence, and $H(x)$ is a non degenerated distribution of the rescaled sample extreme.

**Definition 2.1.** (Pickands, 1967) Given a nondegenerate distribution function $H$, its maximum domain of attraction $MDA(H)$ is defined as the set of all distributions $F$, such that for i.i.d $X_i \sim F$ and $M_n = \max_i X_i$, we have

$$\lim_{n\to\infty} \mathbb{P}[a_n^{-1}(M_n - b_n) \le x] = H(x),$$

for all $X$ on the continuity set of $H$.

### 2.1.1 LIMIT DISTRIBUTIONS OF THE MAXIMA

In contrast to the central limit theorem, there is not one but three different asymptotic distributions of the initial distribution $F$. These distributions are described in the theorem below.

**Theorem 1** (The 'three-types theorem'; Fréchet, 1927; Fisher and Tippett, 1928; Gnedenko, 1943)**.** *The rescaled sample extreme (max renormalised) has a limiting distribution H that can only be of three types:*

$$\begin{array}{lll} \textbf{Type 1 (Fréchet)}: & H_{1,\alpha}(x) := \exp\{-x^{-\alpha}\}\mathbb{I}_{(x>0)}, & \alpha > 0 \\ \textbf{Type 2 (Weibull)}: & H_{2,\alpha}(x) := \mathbb{I}_{(x\geq 0)} + \exp\{-(-x)^{\alpha}\}\mathbb{I}_{(x<0)}, & \alpha > 0 \\ \textbf{Type 3 (Gumbel)}: & H_{3,0}(x) := \exp\{-e^{-x}\}, & \forall x \in \mathbb{R} \end{array}$$

*where $\alpha$ is the inverse of the tail index.*

The distributions used in this study and their respective MDA limiting distributions are summarised in Table 2. We can see that for these initial distributions the limiting distributions are of two kinds, mainly the Gumbel and Fréchet distributions with light and heavy tails, respectively.

Table 2: Maximum Domain of Attraction Examples

| Sample cdf | MDA |
|---|---|
| Exponential | Gumbel |
| Gaussian | Gumbel |
| Student t | Fréchet |
| Pareto | Fréchet |

*Note:* We refer to the exponential distribution with parameter $\lambda$ for $x > 0$, and distribution function $F(x) = 1 - e^{-\frac{x}{\lambda}}$. Likewise, the Pareto distribution is specified with parameter $\beta > 0$ for $x \geq 1$, and distribution function $F(x) = 1 - x^{-\beta}$.

Let us go back historically to one of the first results of Pickands (1967). As an example, we show how the limiting distribution for the maximum of standard i.i.d. Gaussian sequences can be defined. Let $X$ be a stationary Gaussian stochastic process with $\mathbb{E}(X_n) = 0$, $\mathbb{E}(X_n^2) = r_0 = 1$ and $\mathbb{E}(X_i X_{i+n}) = r_n$. The following theorems are important to show that the limiting distribution of a standard Gaussian maximum is indeed $\exp\{-e^{-x}\}$, the Gumbel distribution function (for a proof, see Appendix A).

**Theorem 2** (Gnedenko, 1943)**.** *Let $\{X_n\}$ be an i.i.d. sequence. Let $0 \leq \zeta \leq \infty$ and suppose that $\{m_n\}$ is a sequence of real numbers such that*

$$n(1 - F(m_n)) \to \zeta \text{ as } n \to \infty. \tag{1}$$

*Then*

$$\mathbb{P}(M_n \leq m_n) \to e^{-\zeta} \text{ as } n \to \infty. \tag{2}$$

*Conversely, if* (2) *holds for some $\zeta$, $0 \leq \zeta \leq \infty$, then so does* (1)*.*

**Theorem 3** (Pickands, 1967)**.** *Let $X_i$ be i.i.d. standard normal, then the asymptotic distribution of $M_n$ satisfies*

$$\frac{M_n}{\sqrt{2\log n}} \underset{n\to\infty}{\to} 1 \text{ a.s.}$$

*A proof of Theorem 3 can be found in Appendix A.1.*

**Theorem 4** (Gnedenko, 1943; Cramér, 1946)**.** *If $\{X_i\}$ is an i.i.d. standard normal sequence of r.v.'s then the asymptotic distribution of $M_n = \max(X_1, \ldots, X_n)$ is of type $H_{3,0}$. Specifically*

$$\mathbb{P}(a_n^{-1}(M_n - b_n) \leq x) \to \exp(e^{-x}), \text{ as } n \to \infty, \tag{3}$$

*where*

$$a_n = (2\log n)^{-1/2},$$

*and*

$$b_n = (2\log n)^{1/2} - \frac{1}{2}(2\log n)^{-1/2}(\log\log n + \log 4\pi).$$

*A proof of Theorem 4 can be found in Appendix A.2.*

### 2.1.2 The Generalised Extreme Value Distribution

The Fréchet, Weibull and Gumbel distributions mentioned in Theorem 1 are generalised into a parametric distribution referred to as the generalised extreme value distribution (GEVD).

**Theorem 5** (Jenkinson, 1955; von Mises, 1936)**.** *If $F \in MDA(G)$ then necessarily, $G$ is of the same type as the GEV cdf, $H_\xi$(i.e. $G(x) = H_\xi(ax + b)$, $a > 0$), defined by*

$$H_\xi(x) = \begin{cases} \exp\left[-(1+\xi x)_+^{-\frac{1}{\xi}}\right] & \text{if } \xi \neq 0 \\ \exp\left(-e^{-x}\right) & \text{if } \xi = 0 \end{cases} \tag{4}$$

*where $y_+ = \max(0,\ y)$. The parameter $\xi$ is referred to as the tail or extreme-value index and determines the nature of the tail distribution: if $\xi > 0$ then $H_\xi$ is Frechét, if $\xi = 0$ then $H_\xi$ is Gumbel and if $\xi < 0$ then $H_\xi$ is Weibull.*

We define the three parameter cdf as

$$H_{\xi,\mu,\beta} := H_\xi((x-\mu)/\beta) \tag{5}$$

where $\mu \in \mathbb{R}$ is the location parameter and $\beta > 0$ the scale parameter (see McNeil et al., 2005).

### 2.1.3 The Generalised Pareto Distribution

The Generalised Pareto Distribution is commonly used to model tail behaviour in a broader sense than just its maximum. This concept is referred to as 'threshold exceedances' where all data exceeding a certain threshold is utilised (see e.g. Embrechts et al., 1997).

**Theorem 6** (Pickands, 1975; Balkema and de Haan, 1974)**.** *If $F$ belongs to one of the maximum domains of attraction (i.e. the limiting distribution of $\max X_i$ is a GEV), then for a sufficiently high threshold $u$, $\exists \beta(u) > 0$ and $\xi$ real number such that the Generalised Pareto Distribution (GPD) $G_{\xi,\beta(u)}$, defined as*

$$\bar{G}_{\xi,\beta(u)}(y) := 1 - G_{\xi,\beta(u)}(y) = \left(1 + \xi\frac{y}{\beta(u)}\right)^{-1/\xi} \mathbb{I}_{(\xi\neq 0)} + e^{-y/\beta(u)}\mathbb{I}_{(\xi=0)}, \tag{6}$$

*is a very good approximation of the excess cdf $F_u(\cdot) := \mathbb{P}[X - u \leq \cdot \mid X > u]$*

$$\lim_{u\uparrow x_F^+} \sup_{0\leq y\leq x_F^+ - u} \mid F_u(y) - G_{\xi,\sigma(u)}(y) \mid = 0,$$

*$x_F^+$ denoting the upper endpoint of $F$.*

## 2.2 Methodology

Next, the methods of the block maxima and mean excess plot are described in Section 2.2.1 and 2.2.2. The Hill estimator is defined in Section 2.2.3. In Section 2.2.4 we present the univariate risk models while Section 2.2.5 and 2.2.6 give the methods for point estimates and intervals. Finally, the evaluation methods are summarised in Section 2.2.7.

### 2.2.1 The Method of Block Maxima

The GEVD defined in section 2.1.2 forms the basis of a model for the distribution of the block maxima (Coles et al., 2001). The block maxima (BM) procedure consists in partitioning the data into blocks of equal length. The maximum value in each block is then extracted. In the second step, GEV parameters are fitted to a series of block maxima.

However, choosing a block size can be problematic as it reflects a trade-off between bias and variance (Coles et al., 2001). Furthermore, this method only makes use of information on maxima whereas observations below the largest value are disregarded. Nevertheless, there are situations when the BM method is particularly valuable. This technique has proved to be useful when modelling extremes in periodic time series where dependence appears within but not between blocks (Ferreira and de Haan, 2015). Moreover, in many situations block periods appear naturally such that the BM method may be easier to implement. Annual dependencies in financial time series data have been known for a long time, such as the January effect (see Rozeff and Kinney Jr, 1976). Hence, we partition the data into blocks $b$ of size $n/250$, where $n$ is the size of the return losses and 250 corresponds to the number of trading days per year. This amounts to a block size of 16 for the pre-Brexit period and 17 for the full sample period.

### 2.2.2 The Mean Excess Plot

The mean excess plot (MEP) is attributed to Davison and Smith (1990). An MEP is used to pick a threshold $u$ and to validate a generalised Pareto distribution for the excess distribution. For a r.v. $X$ the ME function is defined as

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

The mean excess function can be evaluated for any r.v. $X$ with finite expectation. If $X$ is GPD with parameters $\xi < 1$ and $\beta > 0$ where $\mathbb{E}(X) < \infty$, the ME function is given by

$$e(u) = \left[ \frac{\beta}{1-\xi} + \frac{\xi}{1-\xi} u \right] \mathbb{I}_{(\beta + u\xi > 0)}.$$

If the data from which a threshold is to be determined is from a GPD, by Theorem 6 the MEP should be roughly linear. The empirical mean excess function is defined such as

$$\hat{e}(u) = \frac{\sum_{i=1}^{n} (X_i - u) \mathbb{I}_{[X_i > u]}}{\sum_{i=1}^{n} \mathbb{I}_{[X_i > u]}}, \qquad u \geq 0.$$

The procedure starts with ordering the data into order statistics $X_{(1)} \geq X_{(2)} \geq, \ldots \geq X_{(n)}$. Then, one proceed plotting $\{(X_{(k)}, \hat{e}(X_{(k)})) : 1 < k \leq n\}$ determining for which value of $u$ the MEP is approximately linear. Finally, a maximum likelihood estimation is performed to evaluate the tail index $\xi$ and scale parameter $\beta$.

The threshold selection is a crucial choice in extreme value modelling. Much like the choice of block size in the block maxima method, the choice of threshold $u_0$ boils down to a bias-variance trade-off. A low threshold leads to a biased estimate while a too high cause the variance of the estimates to be too large. Picking the threshold $u_0$ in a deterministic manner, as the observation $\sqrt{n}$ in the ordered set of sample values $X_{(1)} \geq X_{(2)} \geq, \ldots, \geq X_{(n)}$ has proved to be a suitable choice for financial data (Ferreira et al., 2003; Blum and Dacorogna, 2002).

Northrop et al. (2017) suggested an alternative method to address the difficulties of threshold selection, utilizing Bayesian cross-validation methods on wave height data. Although there are plenty of suggested threshold diagnostic methods, a common shortcoming of these methods is that these often assume one single chosen threshold. As such, they ignore the uncertainty in the threshold selection. Instead, the threshold can be viewed as a parameter included in the model. The number of exceedances above a threshold $u$ is assumed to follow a binomial distribution $Bin(n, p_u)$ where $p_u$ is the probability of exceeding $u$. This translates into a binomial-GP model $BGP(p_u, \beta(u), \xi)$ (see Coles et al., 2001). A leave-one-out cross validation approach is then performed. To increase the computational efficiency, importance sampling is used to estimate the cross validation predicted densities. The package *threshr* is used for applying this methodology in the statistical computing environment r (R Core Team, 2019; Northrop and Attalides, 2019).

The prior for $p_u$ is chosen as,

$$p_u \sim \text{beta}(1/2, 1/2),$$

and a generalised maximal data information prior is used for the GP parameters,

$$\pi_{\text{M}a}(\beta(u), \xi; a) \propto \beta(u)^{-1} a \exp\{-a(\xi + 1)\},\ \beta(u) > 0,\ \xi \geq -1,\ a > 0,$$

where $a = 0.6$.

This prior showed good properties using simulated and real data as the chances of obtaining unreasonable estimates for high thresholds was lower than for other priors (see Northrop et al., 2017). We compare both threshold selection methods to see which performs better in financial risk analysis.

### 2.2.3 The Hill Estimator

A popular estimator of the tail index $\xi$ whenever $\xi > 0$ (i.e. Fréchet domain of attraction), is the Hill estimator (see e.g. Hill, 1975; Embrechts et al., 1997). Under the condition that $k(n) \to \infty$ and $n/k(n) \to \infty$ it is defined as

$$\widehat{H}_{k(n)} = \frac{1}{k(n)} \sum_{i=1}^{k(n)-1} \log\left(\frac{X_{(k(n)-i)}}{X_{(k(n))}}\right). \tag{7}$$

The form of (7) is the same as the ML estimate of $\xi$, allowing the threshold $u_0$ to be a random variable. For a large enough number of exceedances and a sufficiently high threshold, the Hill estimate $\widehat{H}_k$

converges to $\xi$. When $\xi > 0$ the GPD is equal to a Pareto distribution with scale $\kappa = \beta(u)/\xi$ and letting $\kappa = u$ we can find the estimate of the scale as $\hat{\beta}(u) = u\hat{H}_k$.

### 2.2.4 RISK MODELS

Three types of univariate extreme value models are compared, the GEVD model and the GPD model using a deterministic and cross-validation threshold. Table 3 list the extreme value models used in the preceding section and the abbreviations we use for these models in figures and tables.

Table 3: List of extreme value models.

| Model | Estimation | Additional parameters | Notation |
|---|---|---|---|
| GEV | Maximum likelihood | Block size (deterministic) | $b$ |
| GPD | Hill estimator | Threshold (deterministic) | $u_0$ |
| GPD$^*$ | Maximum likelihood | Threshold (cross-validation) | $u_0^*$ |

The risk is then modelled as VaR and ES using Definition 1.2 and 1.3. Quantiles are chosen as $\alpha = 0.99$ for VaR estimates and $\alpha = 0.975$ for ES estimates such as advocated by the Basel III regulation (Basel Committee on Banking Supervision, 2013).

The GEV models are defined according to eq. (5) for a location $\mu \in \mathbb{R}$ and scale $\beta > 0$ and using Definition 1.2, VaR for the GEV model is calculated as

$$\widehat{\mathrm{VaR}}_{\mathrm{GEV},\alpha}(L) = \hat{G}^{-1}(\alpha) = \begin{cases} \hat{\mu} - \frac{\hat{\beta}}{\hat{\xi}}\left[1 - (-n\log(\alpha))^{-\hat{\xi}}\right], & \text{for } \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\beta}\log(-n\log(\alpha)), & \text{for } \hat{\xi} = 0, \end{cases} \tag{8}$$

where $n$ is the number of block maxima used when fitting the model. Through Definition 1.3 we deduce the following expression for ES,

$$\widehat{\mathrm{ES}}_{\mathrm{GEV},\alpha}(L) = \frac{1}{1-\alpha}\int_\alpha^1 \widehat{\mathrm{VaR}}_{\mathrm{GEVD},v}(L)dv = \begin{cases} \hat{\mu} - \frac{\hat{\beta}}{\hat{\xi}} + \frac{\hat{\beta}}{\hat{\xi}}\frac{1}{1-\alpha}\int_\alpha^1 (-n\log(v))^{-\hat{\xi}}dv, & \text{for } \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\beta}\frac{1}{1-\alpha}\int_\alpha^1 \log(-n\log(v))dv, & \text{for } \hat{\xi} = 0. \end{cases} \tag{9}$$

The GPD models are defined according to (6) for a threshold $u$ and scale $\beta(u) > 0$. Using Definition 1.2, VaR for the GPD model is calculated as

$$\widehat{\mathrm{VaR}}_{\mathrm{GPD},\alpha}(L) = \bar{G}^{-1}_{\hat{\xi},\hat{\beta}}(\alpha) = \begin{cases} u + \frac{\hat{\beta}(u)}{\hat{\xi}}\left[\left(\frac{1-\alpha}{\hat{\bar{F}}(u)}\right)^{-\hat{\xi}} - 1\right], & \text{for } \hat{\xi} \neq 0,\ \hat{\beta}(u) > 0 \\ u - \hat{\beta}(u)\log\left(\frac{1-\alpha}{\hat{\bar{F}}(u)}\right), & \text{for } \hat{\xi} = 0,\ \hat{\beta}(u) > 0, \end{cases} \tag{10}$$

and we estimate $\bar{F}(u) = \mathbb{P}(X > u)$ as $\hat{\bar{F}}(u) = n/N$ where $N$ is the total number of observations (negative returns) and $n$ is the number of observations larger than the threshold $u$. An expression of ES is given by

$$\widehat{\mathrm{ES}}_{\mathrm{GPD},\alpha}(L) = \frac{1}{1-\alpha}\int_\alpha^1 \widehat{\mathrm{VaR}}_{\mathrm{GPD},v}(L)dv = \begin{cases} u + \frac{\hat{\beta}(u)}{\hat{\xi}}\left[\frac{1}{1-\hat{\xi}}\left(\frac{1-\alpha}{\hat{\bar{F}}(u)}\right)^{-\hat{\xi}} - 1\right] & \text{for } \hat{\xi} \neq 0,\ \hat{\beta}(u) > 0 \\ u - \hat{\beta}(u)\left[\log\left(\frac{1-\alpha}{\hat{\bar{F}}(u)}\right) - 1\right] & \text{for } \hat{\xi} = 0,\ \hat{\beta}(u) > 0. \end{cases} \tag{11}$$

### 2.2.5 Parameter Estimation

As a first step, we perform data analysis using descriptive statistics. While most of these measures need no introduction, we give a brief note on how skewness and kurtosis are calculated. The sample coefficient of skewness $\hat{\eta}_3$ is estimated as follows. We let

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k,$$

and

$$s^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(x_i - \bar{x})^2.$$

Using the unbiased estimator of the third moment and unbiased estimator for variance, the sample skewness is defined as

$$\hat{\eta}_3 = \frac{n^2}{(n-1)(n-2)}\frac{m_3}{s^3}.$$

The sample coefficient of excess kurtosis $\hat{\eta}_4$ is calculated using the unbiased estimator of the fourth moment and unbiased estimator for variance,

$$\hat{\eta}_4 = \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)}m_4 - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)}s^4 - 3.$$

For simplicity, we refer to excess kurtosis as kurtosis in the further sections. We use the *EnvStats* library for these calculations (see Millard, 2018).

The GEV and GP models are fitted to simulated and real data using maximum likelihood estimation as well as the Hill estimator for the GP model. For the ML estimates we use the *r* package *fExtremes* (Diethelm Wuertz, 2017).

### 2.2.6 Interval Estimates

Asymptotic normal intervals for $\xi$ can be constructed since the ML estimator is approximately normal (see Coles et al., 2001). The interval is then

$$\hat{\xi} \pm Z_{0.975}\sqrt{\mathbb{V}(\hat{\xi})},$$

where

$$\mathbb{V}(\hat{\xi}) = \text{MSE}(\hat{\xi}).$$

For the Hill estimator the interval is,

$$\widehat{H}_k \pm Z_{0.975}\sqrt{\mathbb{V}(\widehat{H}_k)},$$

where

$$\mathbb{V}(\widehat{H}_k) = \frac{\widehat{H}_k^2}{k}.$$

Asymptotic normal confidence intervals for VaR and ES in the GEV case can be found using the delta method (see e.g. Coles et al., 2001). If $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, \Sigma)$, where $n$ is the number of observations and $\Sigma$ is a symmetric positive covariance matrix and $\theta = \{\mu, \beta, \xi\}$ then by Taylor approximation, $\text{VaR}(\hat{\theta}) \approx \text{VaR}(\theta) + \nabla\text{VaR}(\theta)^\top(\hat{\theta} - \theta)$, where

$$\nabla\text{VaR}(\theta) = \left[\frac{\partial\text{VaR}(\theta)}{\partial\mu}, \frac{\partial\text{VaR}(\theta)}{\partial\beta}, \frac{\partial\text{VaR}(\theta)}{\partial\xi}\right]^\top.$$

The variance of $\text{VaR}(\hat{\theta})$ can then be approximated as,

$$\mathbb{V}(\text{VaR}(\hat{\theta})) \approx \nabla\text{VaR}(\theta)^\top\Sigma_\theta\nabla\text{VaR}(\theta).$$

Similarly, the variance of $\text{ES}(\hat{\theta})$ can be estimated as,

$$\mathbb{V}(\text{ES}(\hat{\theta})) \approx \nabla\text{ES}(\theta)^\top\Sigma_\theta\nabla\text{ES}(\theta).$$

In the GPD case we must also estimate the uncertainty in $\hat{\hat{F}}(u)$ (see Coles et al., 2001). We simply exchange the parameter vector to $\theta = \{\zeta_u, \beta, \xi\}$, where $\hat{\zeta}_u = \hat{\hat{F}}(u)$. Then $\mathbb{V}(\hat{\zeta}_u) = \hat{\zeta}_u(1 - \hat{\zeta}_u)/n$. We assume that there is no covariance between $\zeta_u$ and the remaining parameters and the threshold $u$ is considered a constant.

Confidence intervals are then calculated as,

$$\text{VaR}(\hat{\theta}) \pm Z_{0.975}\sqrt{\nabla\text{VaR}(\hat{\theta})^\top\hat{\Sigma}_\theta\nabla\text{VaR}(\hat{\theta})}$$

and

$$\text{ES}(\hat{\theta}) \pm Z_{0.975}\sqrt{\nabla\text{ES}(\hat{\theta})^\top\hat{\Sigma}_\theta\nabla\text{ES}(\hat{\theta})}.$$

In the standard ML case, we use *fExtremes* to estimate the covariance matrix $\Sigma_\theta$ (Diethelm Wuertz, 2017). For the Hill estimator we construct $\hat{\Sigma}_\theta$ by deriving,

$$\mathbb{V}(\hat{\beta}) = \mathbb{V}\left(u\widehat{H}_k\right) = \frac{(u\widehat{H}_k)^2}{k}$$

and

$$\mathbb{C}(\hat{\beta}, \widehat{H}_k) = \mathbb{C}(u\widehat{H}_k, \widehat{H}_k) = \frac{u\widehat{H}_k^2}{k}.$$

Then,

$$\hat{\Sigma}_\theta = \begin{bmatrix} \frac{(u\widehat{H}_k)^2}{k} & \frac{u\widehat{H}_k^2}{k} \\ \frac{u\widehat{H}_k^2}{k} & \frac{\widehat{H}_k^2}{k} \end{bmatrix}.$$

These asymptotic normal intervals are not suitable when the sampling distribution of the estimator is not normal. An alternative is bootstrap methods.

Bootstrap is a method for estimating the error of a statistic using resampling (see Efron, 1982). In naive methods, a sample of size $n$ is drawn from an original i.i.d. sample $\mathbf{X}_n = \{X_1, \ldots, X_n\}$ with replacement $B$ times. The sampling distribution of $\mathbf{X}_n$ can then be approximated by drawing $\mathbf{X}_n^* = \{X_1^*, \ldots, X_n^*\}$ samples with replacement from the original sample. There are however examples

where naive bootstrap approaches fail. For instance, for the maximum order statistics of a random variable the probability of obtaining the maxima is equal to $1 - e^{-1}$, and bootstrap sampling fail to work (see Bickel and Freedman, 1981). Another example arises when bootstrap is applied to the mean of a variable with infinite variance or no fourth moment (Athreya, 1987). An improvement in cases where the original bootstrap failed was discovered when the resample size $m$ is smaller than the original sample size $n$ (Swanepoel, 1986). Choosing $m$ as $2n/3$ was reducing the coverage error in percentile-$t$ intervals. The procedure of resampling less than $n$ observations is formalised as $m$-out-of-$n$ bootstrap in Bickel et al. (1997) and is found to work well in most examples of bootstrap failure. We make use of $m$-out-of-$n$ percentile intervals letting $m = 2n/3$ for the estimates of the descriptive statistics. The empirical sample-based estimates are given with $m$-out-of-$n$ percentile bootstrap intervals using the package *distillery* as this bootstrap method is suitable for heavy-tailed data (see Gilleland, 2019).

For a description of our software implementation *extreme.risk* with a link to the code on GitHub see Appendix B.

### 2.2.7 Model Evaluation

Goodness-of-fit (GoF) tests are performed to evaluate the fit of our estimated models.

Classical empirical distribution function (EDF) test statistics are constructed as the distance between an empirical distribution function $F_n$ and a modelled distribution function $\hat{F}$. One such test statistic is the Anderson-Darling test (Anderson and Darling, 1952).

We are interested in testing the composite null hypothesis

$$H_0 : F_n \in F_\theta(x),$$

where $F$ is known while $\theta$ is unknown. If the distance is too large we reject $H_0$.

While there are many EDF tests, the Anderson-Darling test statistic employs a distance of the form,

$$A^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F_{\hat{\theta}}(x))^2}{F_{\hat{\theta}}(x)(1 - F_{\hat{\theta}}(x))} dF_{\hat{\theta}}(x). \tag{12}$$

With some analytical work, one can express (12) as

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} \left( (2i - 1) \log(u_i) + (2n + 1 - 2i) \log(1 - u_i) \right), \tag{13}$$

where $u_i = \Phi((y_i - \hat{\mu})/\hat{\sigma})$, $y_i = \Phi^{-1}(v_i)$ and $v_i = F_{\hat{\theta}}(x_i, \hat{\theta})$.

Finally, a modification of the test statistic in eq. (13) is performed according to Stephens (1986).

$$A^{*2} = A^2(1 + 0.75/n + 2.25/n^2) \tag{14}$$

We rely on the modified Anderson-Darling test in eq. (14), since a goodness-of-fit test based on this test statistic put a larger weight on the observations in the tails and has shown good power properties when applied to extreme value distributions (Laio, 2004). An *r* implementation of this test can be found in the package *gnFit* (Saeb, 2018).

A graphical evaluation is carried out using the mean excess plot. We use the mePlot function in the *fExtremes* package for *r* to extract the mean excesses for the corresponding thresholds and plot the results (Diethelm Wuertz, 2017). To assess the fit for high quantiles cdfs are plotted for $\alpha \geq 0.975$. The empirical cdfs are fitted using the *ecdf* function in the *stats* package. These are plotted along with the fitted cdfs using the parameter estimates of our *extreme.risk* script with the *pgpd* and *pgev* functions in the *fExtremes* package.

## 2.3 Simulation Study I

### 2.3.1 Description of Simulated Data

We perform 50 000 simulations each from four distributions. Table 4 shows the distributions simulated and the functions used to generate the random numbers.

Table 4: Description of the simulated distributions.

| Distribution | Parameters | Function | *r* Package | Author |
|---|---|---|---|---|
| Half-normal | $\{\mu = 0.399, \theta = 0.301\}$ | rtruncnorm | *rtruncnorm* | Mersmann et al., 2018 |
| Half-*t* | $\nu = 2$ | rtrunct | *stats** | R Core Team, 2019 |
| Exponential | $\lambda = 2$ | rexp | *stats* | R Core Team, 2019 |
| Pareto | $\{\kappa = 2, \beta = 2\}$ | rpareto | *EnvStats* | Millard, 2018 |

** A function truncating the simulation at zero was created.*

*Note:* If $X \sim N(0, \sigma^2)$ and $Y = |X|$, then $Y \sim HN(\mu, \theta)$ where $\mu = \sigma\sqrt{2\pi^{-1}}$ and $\theta = \sqrt{\sigma^2(1 - 2\pi^{-1})}$. Similarly, if $X \sim t(\nu)$ where $\nu$ denotes the degrees of freedom, then $Y \sim HT(\nu)$. The density of the exponential distribution has the form $f(x) = \lambda e^{-\lambda x}$ where $\lambda$ is the ratio. We use a Pareto density on the form $f(x) = \beta\kappa^{\beta} x^{-(\beta+1)},\ \kappa > 0,\ \beta > 0,\ x \geq \kappa$ where $\beta$ is the scale and $\kappa$ is the shape.

The first of these is the half-normal distribution, generated from a normal distribution with $\mu = 0$ and $\sigma = 0.5$, where $\mu$ and $\sigma$ are the mean and standard deviation of the normal distribution. Then we simulate from a half-*t* distribution with $\nu = 2$ degrees of freedom. Lastly, we simulate from the exponential distribution with rate, $\lambda = 2$ and Pareto distribution with scale and shape equal to two. This allows us to see how our models compare for different types of data. Table 5 show a summary of descriptive measures for the simulations.

We begin with a comparison of the Student t and Pareto distributions. The theoretical distribution of the Student t and Pareto extremes is the Fréchet distribution ($\xi > 0$). The half-*t* and Pareto distributions with degrees of freedom and shape equal to 2 have a variance that is infinitely large in theory. Both of these distributions have a finite theoretical first moment. We expect that the order of the GEV tail index $\xi$ is larger than 0 and less than 1 as the k*th* moment exist when $\xi < 1/k$ for positive r.v.'s (see McNeil et al., 2005). The sample estimates of the standard deviation are 2.6 (half-*t*) and 5.7 (Pareto). Compared to the Student t sample, the estimated kurtosis (1289) and skewness (25.96) is larger for the Pareto sample while the coefficient of variation (1.41) is slightly lower. The Pareto sample maximum value (400.1) is slightly larger than the half-*t* (137.7) but more information can be gained through studying the quantiles of the distributions. The VaR (20.1) and ES (25.4) measures show that the tails of the Pareto distribution are indeed heavier than the tails of the half-*t* distribution with a VaR of 9.9 and ES 11.8.

We continue to compare the sample estimates of the half-normal and exponential distributions. The half-normal and exponential distributions both have finite moments of all positive orders and are in the maximum domain of attraction of the Gumbel distribution ($\xi = 0$). The half-normal has standard deviation $\theta = \sqrt{\sigma^2(1 - 2\pi^{-1})}$ and mean $\mu = \sigma\sqrt{2\pi^{-1}}$. We note that the half-normal maxima (1.953) is close to the theoretical maxima of a normal random variable with a variance and mean equal to $\theta$ and $\mu$, $a_n = \sqrt{2\log n} \cdot \theta + \mu = \sqrt{2\log(50000)} \cdot \sqrt{0.5^2(1 - 2\pi^{-1})} + 0.399 = 1.801$ (see Appendix A.1). Although the Gumbel class of MDA is associated with light tails, a large variety in the shape of the tails can be observed. The normal distribution is lighter in the tails than the exponential while both are considered in the MDA of the Gumbel distribution. We can see that the kurtosis and skewness indicate that the exponential sample has heavier tails than the half-normal (Table 5). The VaR and ES are larger for the exponential distribution at approximately 2.3. We expect the tail index of the exponential sample to be larger than for the half-normal sample but close to zero.

Table 5: Descriptive statistics of simulated data.

| | Half-normal | | Half-$t$ | | Exponential | | Pareto | |
|---|---|---|---|---|---|---|---|---|
| Mean | 0.399 | [0.396,0.402] | 1.393 | [1.367, 1.422] | 0.498 | [0.492,0.503] | 4.001 | [3.941, 4.066] |
| Median | 0.336 | [0.332,0.340] | 0.819 | [0.807, 0.829] | 0.343 | [0.338,0.349] | 2.838 | [2.824, 2.852] |
| Max | 1.953 | | 137.664 | | 5.496 | | 400.094 | |
| Std. Dev. | 0.302 | [0.300,0.305] | 2.574 | [2.288, 2.868] | 0.498 | [0.490,0.506] | 5.650 | [4.576, 6.758] |
| CV | 0.757 | [0.751,0.763] | 1.848 | [1.667, 2.042] | 1.001 | [0.992,1.012] | 1.412 | [1.162, 1.669] |
| Skewness | 1.019 | [0.991,1.045] | 15.585 | [10.253, 19.720] | 2.002 | [1.918,2.090] | 25.962 | [13.745, 33.206] |
| Kurtosis | 0.921 | [0.804,1.043] | 490.402 | [215.112,749.742] | 5.916 | [5.202,6.680] | 1289.003 | [348.888,1847.689] |
| $\text{VaR}_{.99}$ | 1.300 | [1.280,1.319] | 9.912 | [9.378, 10.398] | 2.314 | [2.269,2.371] | 20.121 | [19.030, 21.217] |
| $\text{ES}_{.975}$ | 1.304 | [1.288,1.319] | 11.842 | [11.111, 12.672] | 2.346 | [2.299,2.394] | 25.364 | [23.389, 27.413] |

*Note:* Estimates are calculated for 50 000 observations generated from a normal distribution with $\mu = 0$ and $\sigma = 0.5$, Student t distribution with $\nu = 2$, exponential distribution with $\lambda = 2$ and Pareto distribution with shape and scale equal to 2. The abbreviation CV denotes the coefficient of variation. The normal and Student t simulations are truncated at zero. Bootstrap percentile intervals using $B = 1000$ resamples with a resample size of $m = 33\,333$ and confidence level 0.95 are given in squared brackets.

### 2.3.2 Estimated Extreme Value Models

We fit three extreme value models to the simulated data using the *extreme.risk* script (see Appendix B). Table 6 show the fit of the GEV models. While the exponential sample has a tail index close to zero, the half-normal sample is negative. This is clearly not correct as we know that the true tail index of the underlying distribution is zero. However, the $A^{*2}$ statistic shows that the fitted distribution function is significantly different from the empirical one. The tail index of the half-$t$ and Pareto simulations are positive while not as large as we expected in our initial analysis. The tail index of the Pareto sample is slightly larger than that of the Student t sample.

The fit of the GPD models with a deterministic threshold selection using the Hill estimator is given in Table 7. The tail index of the exponential sample is larger and there is no significant difference between the fitted and empirical distribution function. The half-normal sample has a tail index that is positive and closer to zero, but the fitted GP distribution function is significantly different from

the empirical one. The tail index of the half-$t$ is larger than that of the Pareto sample. However, the difference is not large enough to capture any substantial differences in the tails.

A summary of the fit of the GPD models with a cross-validation threshold selection and ML estimation is represented in Table 8. The tail index of the half-normal sample is negative and there is no significant difference between the fitted density and the empirical one. This is unexpected since it corresponds to a distribution with a finite right tail. The exponential sample tail index is close to zero and the test statistic shows that the fitted density is close to the empirical one. The tail indexes of the half-$t$ and Pareto samples differ in size to a greater extent than those of the GEV and GPD model with deterministic threshold selection.

To assess the fit of the GPD models we study the MEP in Figure 1. The GPD model using deterministic threshold selection and Hill estimator appear to be roughly linear except for the exponential sample. The GPD$^*$ model, using cross-validation threshold selection and ML estimation appears to be linear for all samples. The threshold is so low that the mean excesses used for the model fit are containing the majority of observations in the sample except for the half-normal case.

To get a clearer picture of the fit in the tails we compare the fitted cumulative distribution functions (cdf) to the empirical one (Figure 2). The empirical cdf (grey) is fitted to the block maxima for the GEV case, and exceedances in the GPD and GPD$^*$ models. The empirical cdf is almost identical for the GEV and GPD models. We can see that the distance of the fitted distributions to their empirical counterparts is smallest for the GPD$^*$ model for all samples. The GEV distribution and the GPD models appear to have a smaller distance to their empirical cdfs in the Pareto case. For the half-normal and exponential samples, the GEV model is closer to the empirical cdf. The Student t sample appears to be better modelled with the GPD than the GEV distribution. One should note that the fitted distribution's closeness to their empirical distributions is not a perfect measure of how well the tail is modelled. A perfect fit of the main body of the distribution does not necessarily translate into a good fit of the extremes. And the answer is greatly determined by what we define as extremes to begin with. A suggestion is to analyse the tail distance between the fitted and empirical distribution function over a particular threshold to gain additional insights on the fit of the tail of the distribution.

Table 6: GEV models.

| | $\mu$ | $\beta$ | $\xi$ | $A^{*2}$ | Sig. |
|---|---|---|---|---|---|
| Half-normal | 1.445 | 0.160 | −0.212 | 424.36 | .000 |
| Half-$t$ | 14.701 | 7.189 | 0.416 | 491.46 | .000 |
| Exponential | 2.696 | 0.456 | 0.030 | 541.28 | .000 |
| Pareto | 31.624 | 16.357 | 0.429 | 2707.62 | .000 |

*Note:* Models are fitted using ML estimation. A block size of $b = 224$ and a sample size of $n = 224$ block maxima was used for all models. Significance levels of the $A^{*2}$ statistics are given for each model.

Table 7: GPD models with deterministic threshold selection.

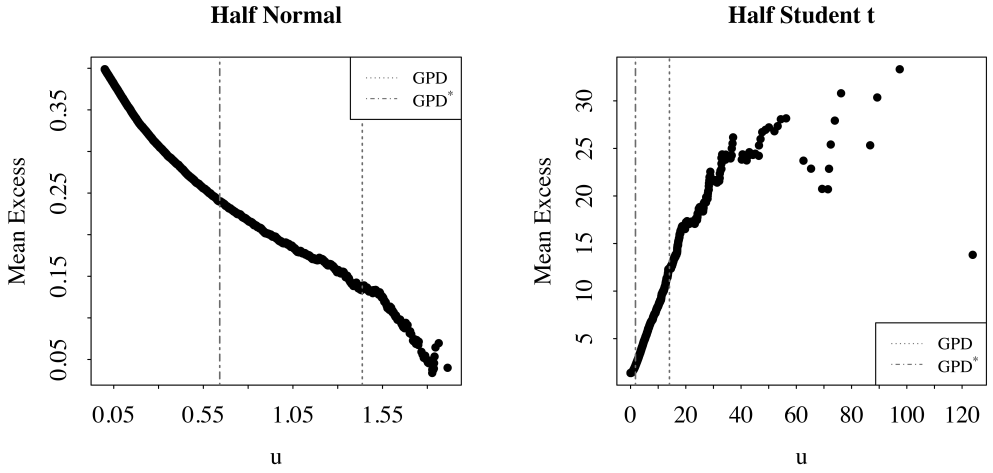
| | $u_0$ | $\beta$ | $\xi$ | $A^{*2}$ | Sig. |
|---|---|---|---|---|---|
| Half-normal | 1.437 | 0.126 | 0.088 | 1.186 | .004 |
| Half-$t$ | 14.076 | 6.970 | 0.495 | 0.461 | .260 |
| Exponential | 2.719 | 0.406 | 0.149 | 0.289 | .616 |
| Pareto | 30.920 | 14.788 | 0.478 | 0.193 | .895 |

*Note:* Models are fitted using the Hill estimator. A sample size of $n = 224$ exceedances is used to fit the models. The threshold is denoted $u_0$. Significance levels of the $A^{*2}$ statistics are given for each model.

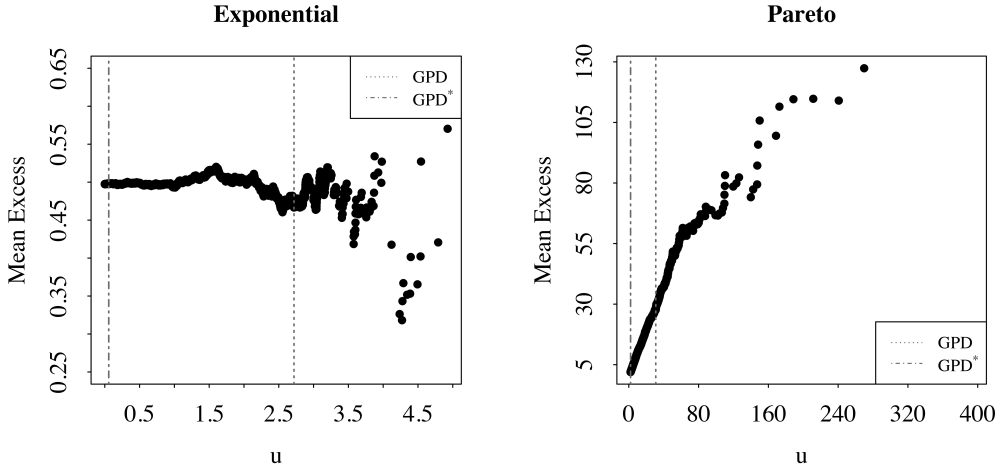
Table 8: GPD models with cross-validation threshold selection

| | $u_0$ | $\beta$ | $\xi$ | $A^{*2}$ | Sig. | $n$ |
|---|---|---|---|---|---|---|
| Half-normal | 0.642 | 0.278 | −0.158 | 0.278 | .651 | 9904 |
| Half-$t$ | 1.778 | 1.215 | 0.439 | 0.448 | .279 | 10894 |
| Exponential | 0.057 | 0.498 | 0.001 | 0.270 | .677 | 44550 |
| Pareto | 2.020 | 1.014 | 0.499 | 0.556 | .151 | 49005 |

*Note:* Models are fitted using ML estimation. The threshold is denoted $u_0$. The sample size $n$, is the number of exceedances used when fitting the model. Significance levels of the $A^{*2}$ statistics are given for each model.

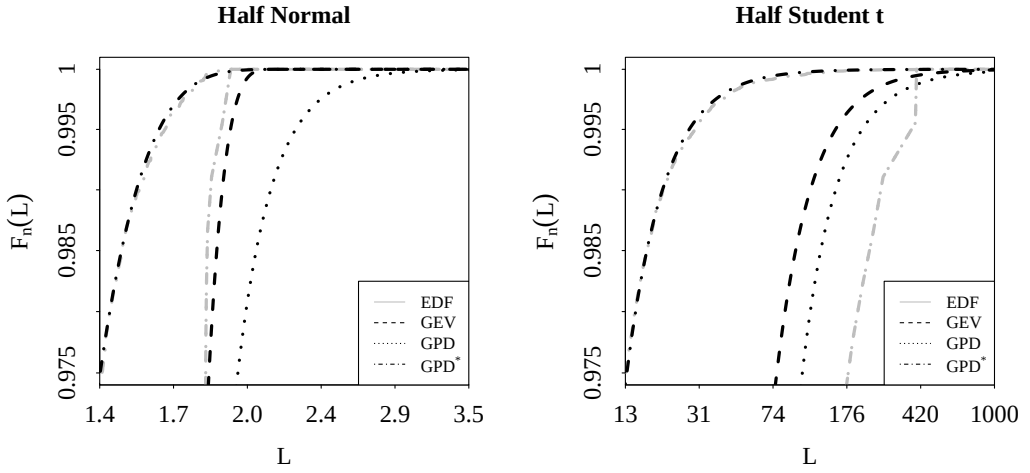



(a) Mean excesses of half-normal distributed r.v's. (b) Mean excesses of half-$t$ distributed r.v's.



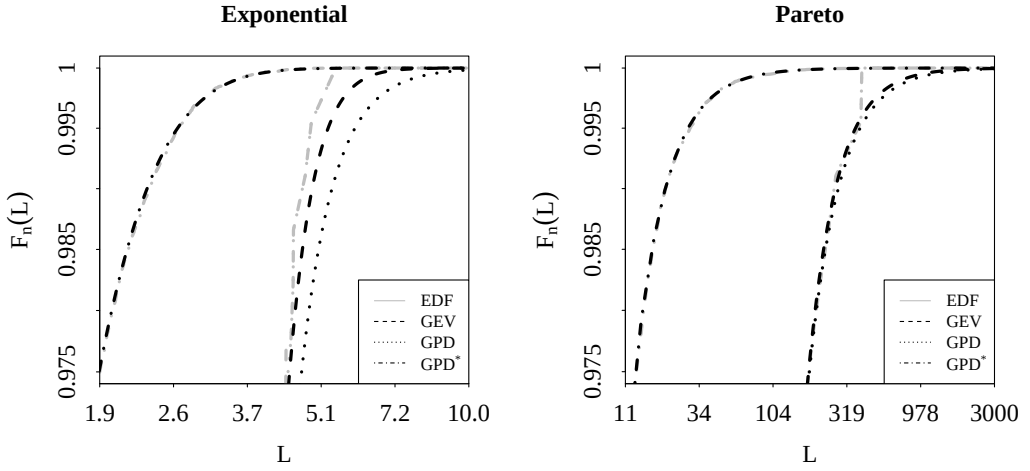

(c) Mean excesses of exponential distributed r.v's. (d) Mean excesses of Pareto distributed r.v's.

Figure 1: Mean excess plot of fitted GPD with deterministic threshold $u_0$ selection (dotted) and cross-validation threshold $u_0^*$ selection (dash-dotted). The GPD model using threshold $u_0^*$ is denoted GPD$^*$.




(a) Fitted models to simulated half-normal distributed r.v's.

(b) Fitted models to simulated half-$t$ distributed r.v's.




(c) Fitted models to simulated exponential distributed r.v's.

(d) Fitted models to simulated pareto distributed r.v's.

Figure 2: Empirical distribution functions (grey), with fitted cdfs (black). Fitted GEV (dashed), GPD with deterministic threshold $u_0$ selection (dotted) and cross-validation threshold $u_0^*$ selection (dash-dotted) in log-log scale. The GPD model using threshold $u_0^*$ is denoted GPD$^*$. The extremes of 50 000 simulated data points from each distribution are fitted to the three extreme value models.

### 2.3.3 Risk Models

One way to assess these models' abilities to capture the tail behaviour of the sampled distributions is to study quantile measures such as VaR and ES. This could give us a clearer picture of what to expect in a worst-case scenario. While not disclosed in this study, an even deeper understanding of the fit of the tails can be obtained through graphical methods such as the QQ plot.

Table 9 shows the shape parameter, tail index, VaR and ES for the normal sample. The tail index estimate is significantly different from zero in all of the models according to the interval estimates. As observed in the previous section, only in the GPD$^*$ case the fitted cdf is matching the empirical one. These results are surprising since Gaussian extremes should be distributed according to a Gumbel distribution. The incorrect tail index estimates for the GEV and GPD models could be explained by these models sensitivity to the choice of block size and threshold. Proceeding to manual analysis to fine-tune these parameters should result in tail index estimates in line with the true parameter value ($\xi = 0$). Despite unexpected tail index estimates the GEV and GPD$^*$ VaR and ES interval estimates both cover the empirical estimates.

The risk models of the Student t sample appears to capture the empirical risk well (Table 10). The tail index estimates are significantly different from zero and have the right sign. All of the VaR and ES interval estimates cover the empirical estimate. Even though the GEV model is significantly different from the empirical distribution (see Table 6), the GEV model VaR and ES estimate come closer to the empirical estimates.

In the exponential case, the tail index is as suspected not significantly different from zero in the GEV and GPD$^*$ models (Table 11). Even if the GEV model is not significant (see Table 6), it manages to model tail loss fairly well. The most surprising finding is that the GPD model tail index estimate is significantly different from zero when we know that in theory, it should be indistinguishable from zero. Further fine-tuning is needed to assure that the tail index estimate is coherent with the theoretical one ($\xi = 0$). While the VaR intervals cover the empirical estimate, there is a significant difference between the GPD model ES estimate and the empirical one.

The Pareto sample tail index estimates are closer in size than for the other distributions (Table 12). All of the tail index estimates are significantly different from zero and none of the VaR and ES model estimates is significantly different from the empirical estimate. The GPD model estimates of VaR and ES deviate from the empirical estimate more than the other model estimates do.

As a concluding remark, we observe in our study that the estimates are not good for the GPD model in the case of the Gumbel domain of attraction ($\xi = 0$). This is also true for the GPD$^*$ model of the exponential extremes. Further study could be performed, in particular, implementing manually the MEP method to compare the results with those obtained using the deterministic and R-package threshold selection rules. A more precise inference on the tail index using extreme value-based tests rather than asymptotic normal intervals might lead to results more in line with the theory. Finally, we note that the risk estimates, despite the incorrect tail index estimates, are overall fairly close to the empirical ones. Further analysis is needed to assure that these results are valid.

Table 9: Summary of estimates, half-normal distribution.

| | $\xi$ | $\alpha$ | $\text{VaR}_{.99}$ | $\text{ES}_{.975}$ |
|---|---|---|---|---|
| GEV | −0.212 | −4.708 | 1.303 | 1.301 |
| | [ −0.316, −0.109 ] | | [ 1.277, 1.329 ] | [ 1.225, 1.377 ] |
| GPD | 0.088 | 11.374 | 1.339 | 1.355 |
| | [ 0.076, 0.099 ] | | [ 1.319, 1.359 ] | [ 1.336, 1.373 ] |
| GPD* | −0.158 | −6.312 | 1.303 | 1.305 |
| | [ −0.175, −0.142 ] | | [ 1.292, 1.314 ] | [ 1.293, 1.317 ] |
| Empirical | | | 1.300 | 1.304 |
| | | | [ 1.280, 1.319 ] | [ 1.288, 1.319 ] |

*Note:* Intervals at a 0.95 level for $\xi$, $\text{VaR}_{0.99}$ and $\text{ES}_{0.975}$ are given in square brackets below each estimate. Tail index $\xi$ is fitted with asymptotic normal intervals. Intervals for model VaR and ES are performed using the delta method. Intervals for the empirical sample are fitted using m-out-of-n percentile bootstrap with $B = 1000$ resamples and a resample size of $m = 33\,333$.

Table 10: Summary of estimates, half-$t$ distribution.

| | $\xi$ | $\alpha$ | $\text{VaR}_{.99}$ | $\text{ES}_{.975}$ |
|---|---|---|---|---|
| GEV | 0.416 | 2.402 | 9.751 | 11.846 |
| | [ 0.303, 0.530 ] | | [ 9.047, 10.454 ] | [ 10.916, 12.776 ] |
| GPD | 0.495 | 2.020 | 9.458 | 11.901 |
| | [ 0.430, 0.560 ] | | [ 8.673, 10.244 ] | [ 11.105, 12.697 ] |
| GPD* | 0.439 | 2.280 | 9.708 | 11.760 |
| | [ 0.412, 0.465 ] | | [ 9.375, 10.041 ] | [ 11.143, 12.376 ] |
| Empirical | | | 9.912 | 11.842 |
| | | | [ 9.378, 10.398 ] | [ 11.111, 12.672 ] |

*Note:* Intervals at a 0.95 level for $\xi$, $\text{VaR}_{0.99}$ and $\text{ES}_{0.975}$ are given in square brackets below each estimate. Tail index $\xi$ is fitted with asymptotic normal intervals. Intervals for model VaR and ES are performed using the delta method. Intervals for the empirical sample are fitted using m-out-of-n percentile bootstrap with $B = 1000$ resamples and a resample size of $m = 33\,333$.

Table 11: Summary of estimates, exponential distribution.

| | $\xi$ | $\alpha$ | $\text{VaR}_{.99}$ | $\text{ES}_{.975}$ |
|---|---|---|---|---|
| GEV | 0.030 | 33.532 | 2.331 | 2.375 |
| | [ −0.076, 0.136 ] | | [ 2.270, 2.392 ] | [ 1.244, 3.505 ] |
| GPD | 0.149 | 6.706 | 2.349 | 2.473 |
| | [ 0.130, 0.169 ] | | [ 2.240, 2.458 ] | [ 2.418, 2.528 ] |
| GPD* | 0.001 | 1880.287 | 2.294 | 2.336 |
| | [ −0.009, 0.010 ] | | [ 2.261, 2.328 ] | [ 2.300, 2.373 ] |
| Empirical | | | 2.314 | 2.346 |
| | | | [ 2.269, 2.371 ] | [ 2.299, 2.394 ] |

*Note:* Intervals at a 0.95 level for $\xi$, $\text{VaR}_{0.99}$ and $\text{ES}_{0.975}$ are given in square brackets below each estimate. Tail index $\xi$ is fitted with asymptotic normal intervals. Intervals for model VaR and ES are performed using the delta method. Intervals for the empirical sample are fitted using m-out-of-n percentile bootstrap with $B$ = 1000 resamples and a resample size of $m$ = 33 333.

Table 12: Summary of estimates, Pareto distribution.

| | $\xi$ | $\alpha$ | $\text{VaR}_{.99}$ | $\text{ES}_{.975}$ |
|---|---|---|---|---|
| GEV | 0.429 | 2.331 | 20.415 | 25.323 |
| | [ 0.321, 0.538 ] | | [ 18.829, 22.000 ] | [ 23.274, 27.373 ] |
| GPD | 0.478 | 2.091 | 21.061 | 26.043 |
| | [ 0.416, 0.541 ] | | [ 19.371, 22.750 ] | [ 24.384, 27.702 ] |
| GPD* | 0.499 | 2.004 | 20.001 | 25.273 |
| | [ 0.486, 0.512 ] | | [ 19.379, 20.623 ] | [ 24.159, 26.387 ] |
| Empirical | | | 20.121 | 25.364 |
| | | | [ 19.030, 21.217 ] | [ 23.289, 27.413 ] |

*Note:* Intervals at a 0.95 level for $\xi$, $\text{VaR}_{0.99}$ and $\text{ES}_{0.975}$ are given in square brackets below each estimate. Tail index $\xi$ is fitted with asymptotic normal intervals. Intervals for model VaR and ES are performed using the delta method. Intervals for the empirical sample are fitted using m-out-of-n percentile bootstrap with $B$ = 1000 resamples and a resample size of $m$ = 33 333.

## 2.4 Case Study I: The Univariate Risk of the FTSE 100 Index

### 2.4.1 THE FTSE 100 DATA

A description of the data used in this study is presented in Table 13. To gain insights on the market risk in the United Kingdom we use the FTSE 100 index. The FTSE 100 is consisting of the 100 companies with the highest market capitalisation on the London Stock Exchange.

A large data set is required to have a sufficient number of extremes to model. The full series spans from 1984-01-03 to 2019–05–31 while the subset comprising the period before Brexit ranges from

1984-01-03 to 2016-06-23. A difficulty in studying the period after Brexit is the limited sample size. Hence, we compare the pre-Brexit and the full series to help us draw conclusions on the impact of the Brexit referendum on the tail risk.

Table 13: Description of the FTSE 100 stock index data.

| Data | Type | $n$ | |
|---|---|---|---|
| | | pre-Brexit | full sample |
| $S$ | Closing Value | 8283 | 9027 |
| $R_{\Delta t}$ | Log-Return | 8282 | 9026 |
| $L$ | Return Loss | 3886 | 4240 |

*Note:* The table shows the number of data points for the series $S$ of closing values for FTSE 100, log-returns $R$ and the number of losses in the log-returns $L$. The number of observations is given for the period before the Brexit referendum (1984-01-03 – 2016-06-23) as well as the full series (1984-01-03 – 2019–05–31).

### 2.4.2 Volatility of The British Stock Market

A sometimes overlooked property of securities is that the price volatility does not reflect the return risk. A change in price between two successive trading days reflects an absolute change in the cumulative geometric returns. Returns are defined as the relative change in price between two succeeding days and are assumed to be independent of preceding returns. It is problematic to assess risk based on prices as risk is measured as the deviation from the expectation. As such, returns yield more reliable information on return risk. The FTSE 100 index is used to illustrate this circumstance (Figure 3).

We inspect the upper graph of Figure 3 to compare the movements in closing values for different events. The losses in closing values at the Great Recession and around the Brexit referendum seems larger than during the Black Monday. Moreover, the loss at the Dot-com bubble appears to be larger than during the Great Recession.

The lower graph of Figure 3 is showing the log-returns for the FTSE 100 index from 1984 to 2019. At the time of the Black Monday, we see the largest negative spike of the entire period. The largest drawdown in relation to the early 90s Recession occurs after the event. At the year of the Brexit referendum, the largest drawdown does not take place at the time of the referendum. A loss of around $-0.05$ is present at the end of March. After the referendum, the return series reverts into a less volatile state. This implies that the behaviour of the market at various events is not constant. Bräutigam et al. (2018) calculated a rolling window annualised volatility of the FTSE all-share index. Their one-year rolling window volatility shows patterns close to ours.

### 2.4.3 The British Stock Market Losses

We carry out a descriptive analysis of the losses to get a preliminary view of the shape of the left tail. As the losses are positive, the distribution of the tails can only be of the type Fréchet or Gumbel. Thus, whether the tails are considered as heavy ($\xi > 0$) or light ($\xi = 0$) reduces to testing if the tail index $\xi$ is larger than zero. Before conducting a formal test, we analyse the data to get an idea of


Figure 3: Closing values and log-returns of FTSE 100 starting at 1984–01–03 and ending at 2019–05–30.

the shape of the tail. As the extremes of a normal distribution converge to a Gumbel distribution, a simulation is performed for reference. Table 14 shows a summary of the descriptive statistics for the return losses.

The central tendency estimates are larger for the pre-Brexit period compared to the full sample. However, differences are negligible as indicated by the confidence intervals. Neither the estimates of the mean nor the median is within the interval of the truncated normal sample. This indicates a different central tendency compared to a normal distribution.

The maximum value is the same for both periods and it occurs before the Brexit referendum. The size of the maxima for both periods is substantially larger than the truncated normal maxima. This suggests that the distribution of the return losses have a longer tail than the losses of a normally distributed random variable.

While the standard deviation is larger in the pre-Brexit sample compared to the full sample, the coefficient of variation is lower. The estimates of the coefficient of variation for the pre-Brexit and full sample are roughly equal while differing from the simulated normal estimates. This might indicate a slightly heavier tail since the density is more dispersed around its mean.

The skewness and kurtosis estimates are larger for the full sample compared to the pre-Brexit sample. The estimates are not differing by a considerable amount between the sample periods. As the kurtosis is larger than zero, the distributions of the return loss in both series are leptokurtic. Both skewness and kurtosis differ from the simulated normal sample by a substantial amount. This indicates that the tails of the return loss samples are heavier than that of the simulated truncated normal distribution.

The VaR estimate for the pre-Brexit sample is slightly larger than the full sample estimate. The intervals show that the differences are negligible. As these estimates exceed the upper bound of the simulated truncated normal distribution VaR, a comparably heavier tail is expected. The ES estimates result in a similar analysis.

To conclude, most of these estimates suggest a tail index slightly larger than zero for both return loss samples. A Fréchet distribution is likely ($\xi > 0$), while not certain without further analysis. Differences in the tail shape for the pre-Brexit and full series are not consistent. Judging by the confidence intervals these differences are negligible. The estimates of the mean, standard deviation, skewness and kurtosis imply the existence of four moments. For the GEV we know that the $k$th moment exits if $\xi < 1/k$. Consequently, we expect a tail index of $\xi < 0.25$. Altogether this analysis suggests the presence of moderately heavy tails in both return loss samples. To determine if the return losses follow a Fréchet distribution we conduct a formal test in the following section.

Table 14: Descriptive Statistics of Losses And Simulated Normal r.v.'s.

| | Pre-Brexit | | Full | | Normal (truncated) | |
|---|---|---|---|---|---|---|
| | Est. | 95% CI | Est. | 95% CI | Est. | 95% CI |
| Mean | 0.801 | [ 0.770, 0.835 ] | 0.780 | [ 0.752, 0.814 ] | 0.865 | [ 0.841, 0.887 ] |
| Median | 0.576 | [ 0.547, 0.613 ] | 0.558 | [ 0.530, 0.586 ] | 0.739 | [ 0.707, 0.764 ] |
| Max. | 13.029 | | 13.029 | | 3.923 | |
| Std. Dev. | 0.842 | [ 0.762, 0.933 ] | 0.822 | [ 0.744, 0.914 ] | 0.647 | [ 0.627, 0.667 ] |
| CV | 1.051 | [ 0.973, 1.142 ] | 1.054 | [ 0.977, 1.142 ] | 0.748 | [ 0.728, 0.767 ] |
| Skewness | 3.605 | [ 2.290, 4.858 ] | 3.632 | [ 2.343, 4.974 ] | 0.992 | [ 0.887, 1.094 ] |
| Kurtosis | 27.906 | [ 8.176, 46.805 ] | 28.611 | [ 8.224, 48.196 ] | 0.909 | [ 0.456, 1.368 ] |
| $\text{VaR}_{.99}$ | 3.875 | [ 3.453, 4.741 ] | 3.689 | [ 3.271, 4.339 ] | 2.747 | [ 2.623, 2.868 ] |
| $\text{ES}_{.975}$ | 4.155 | [ 3.655, 4.681 ] | 4.064 | [ 3.609, 4.567 ] | 2.786 | [ 2.663, 2.902 ] |
| Sample size | 3886 | | 4240 | | 4240 | |

*Note:* Descriptive statistics are calculated from the losses $L$, at the period before the Brexit referendum (1984-01-03 – 2016-06-23) as well as the full series (1984-01-03 – 2019–05–31). Descriptives of simulated normal r.v's are supplied for comparison reasons. The simulation is performed using a mean and standard deviation similar to the full return series and the left tail is truncated at zero. The abbreviation CV denotes the coefficient of variation. The calculations of kurtosis and skewness are defined in Section 2.4.1. Estimates (Est.) of VaR and ES are calculated as described in Section 1.2. Percentile bootstrap confidence intervals (CI) are given in brackets. Intervals are calculated using 1000 bootstrap resamples at a resample size of $m$ = 2591 for the pre-Brexit sample and $m$ = 2827 for the full and truncated normal samples.

### 2.4.4 Modelling the British Extreme Return Losses

Extreme value models are fitted to determine if there is a difference in extreme losses during the pre-Brexit period when compared to the post-Brexit period. Three types of models are fitted to the FTSE 100 return losses pre-Brexit referendum and the full sample period. First, a GEV model is fitted using the block maxima method. We then fit a GPD model using deterministic threshold selection to the two sample periods. Finally, a GPD model using cross-validation threshold selection is fitted, denoted GPD$^*$ (Section 2.2.4). Before presenting the estimates, we proceed with a graphical analysis of the model fits.

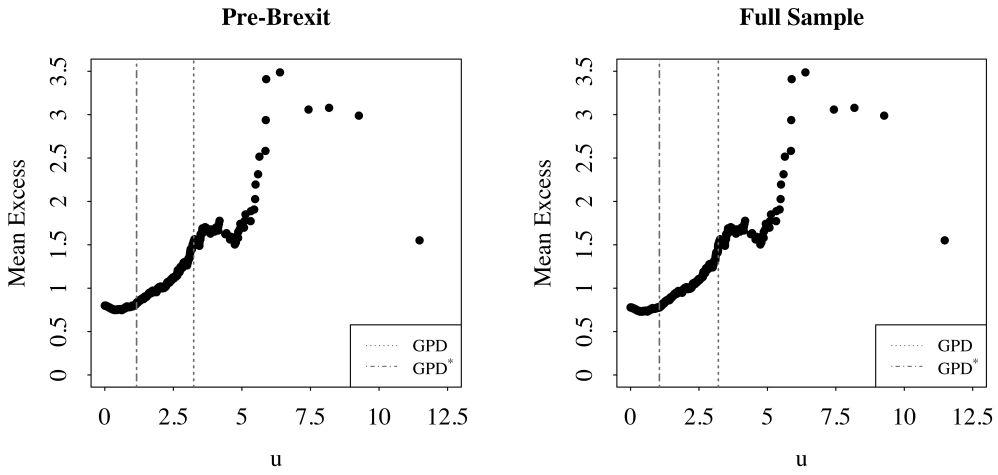
Figure 4 shows the mean excess plot for the full sample as well as the pre-Brexit sample. The mean excess function of an exponential distribution is constant, while that of a Pareto distribution tends to infinity for high values of $u$ (Embrechts et al., 1997). Since the extremes of an exponential distribution converge into a Gumbel distribution we expect the mean excess function of a distribution with $\xi = 0$ to be flat or increasing with a decaying speed. For both samples, the mean excesses are increasing with $u$, a sign that $\xi > 0$ is to be expected.

Examining the 0.975 quantile cdf plots we see that the GEV models using the block maxima method are overestimating the frequency of the losses in comparison to the empirical cdf for values below 7.6 (Figure 5). The GPD models using deterministic threshold selection appear to overestimate the frequency of the losses to an even greater extent. The GPD$^*$ models using a cross-validation threshold selection seem to fit the empirical cdf best. There is no visible difference in the model fit between the pre-Brexit and full sample models.

The estimates for the GEV models are evident in Table 15. We can see that the tail index estimate indeed, is less than 0.25. The $A^{*2}$ statistic for both sample periods rejects the null hypothesis that the return losses follow a GEV distribution. The estimates appear to be very close for the two considered sample periods.

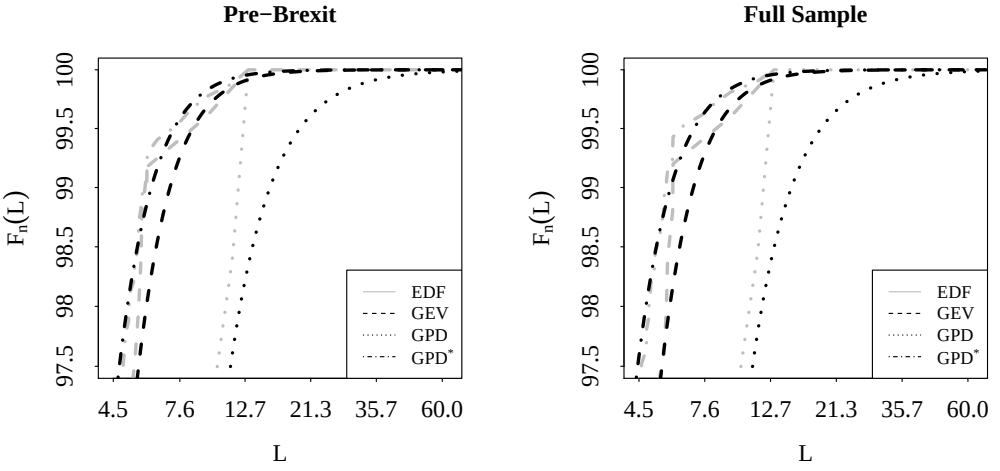
Table 16 shows the estimates for the GPD models using deterministic threshold selection. None of the $A^{*2}$ statistics rejects a GPD fit of the FTSE 100 extreme losses. The tail index estimates of the models with deterministic threshold selection are considerably higher than those of the GEV models. There is not a large difference in estimates between the sample periods.

The GPD$^*$ models tail index estimates are close to those of the GEV models (Table 17). In comparison, the $A^{*2}$ statistic is lower for the GPD$^*$ models, indicating a lack of evidence of differences between the fitted distribution and the empirical distribution. Just as for the previously considered models, the GPD$^*$ model estimates do not show any large differences between the sample periods.




(a) Empirical MEF for the pre-Brexit sample. (b) Empirical MEF for the full sample.

Figure 4: Empirical MEFs with selected deterministic threshold $u_0$ (dotted line) and cross-validation threshold $u_0^*$ (dash-dotted line). The mean excesses and threshold $u$ are given in percentage units.




(a) Fitted models to the pre-Brexit return losses. (b) Fitted models to the full sample return losses.

Figure 5: Empirical distribution functions (solid), fitted GEV (dashed), GPD with deterministic threshold $u_0$ selection (dotted) and cross-validation threshold $u_0^*$ selection (dash-dotted). The GPD model using threshold $u_0^*$ is denoted GPD$^*$.

Table 15: GEV models.

| | $\mu$ | $\beta$ | $\xi$ | $A^{*2}$ | Sig. | $n$ |
|---|---|---|---|---|---|---|
| Pre-Brexit | 1.646 | 0.691 | 0.2124 | 13.883 | .000 | 243 |
| Full Sample | 1.635 | 0.674 | 0.2164 | 16.619 | .000 | 250 |

*Note:* Models are fitted to the pre-Brexit sample (1984-01-03 – 2016-06-23) as well as the full sample (1984-01-03 – 2019-05-31) using ML estimation. The block size is 16 for the pre-Brexit model and 17 for the full sample model. Sample size $n$ is the number of block maxima used when fitting the model. The values of $\mu$ and $\beta$ are given in percentage units.

Table 16: GPD models with deterministic threshold selection.

| | $u_0$ | $\beta$ | $\xi$ | $A^{*2}$ | Sig. | $n$ |
|---|---|---|---|---|---|---|
| Pre-Brexit | 3.24 | 1.099 | 0.3391 | 0.377 | .409 | 62 |
| Full Sample | 3.21 | 1.072 | 0.3343 | 0.382 | .400 | 65 |

*Note:* Models are fitted to the pre-Brexit sample (1984-01-03 – 2016-06-23) as well as the full sample (1984-01-03 – 2019-05-31) using the Hill estimator. The threshold is denoted $u_0$. Significance levels of the $A^{*2}$ statistics are given for each model. The sample size $n$ is the number of exceedances used when fitting the model. The values of $u_0$ and $\beta$ are given in percentage units.

Table 17: GPD models with cross-validation threshold selection.

| | $u_0$ | $\beta$ | $\xi$ | $A^{*2}$ | Sig. | $n$ |
|---|---|---|---|---|---|---|
| Pre-Brexit | 1.13 | 0.660 | 0.1924 | 0.219 | .838 | 885 |
| Full Sample | 1.08 | 0.620 | 0.2043 | 0.222 | .830 | 1008 |

*Note:* Models are fitted to the pre-Brexit sample (1984-01-03 – 2016-06-23) as well as the full sample (1984-01-03 – 2019-05-31) using the ML estimator. The threshold is denoted $u_0$. Significance levels of the $A^{*2}$ statistics are given for each model. The sample size $n$, is the number of exceedances used when fitting the model. The values of $u_0$ and $\beta$ are given in percentage units.

### 2.4.5 Risk Models of the FTSE 100 losses

We proceed our analysis to the risk of the FTSE 100 return losses. A summary of estimates for the pre-Brexit and full sample period is given in Table 18 and 19. The tail index estimate is larger than zero for all models but as was previously mentioned the fitted GEVD density is significantly different from the empirical one (see Table 15). The tail index estimates of the GPD model significantly differs from the GEV and GPD$^*$ model estimates for both sample periods.

None of the VaR and ES interval estimates supports a significant difference in risk between the two sample periods. The GEV model VaR and ES estimates are significantly different from the empirical, GPD and GPD$^*$ ones. This highlights the difficulties in choosing the correct block size to obtain accurate tail loss risk estimates. We can not find evidence of a significant difference in the GPD and GPD$^*$ model VaR and ES estimates.

As a concluding remark, there is evidence of moderately heavy tails in the FTSE 100 return losses. Overall, the GEV model using block maxima failed to capture the tail risk in both sample periods. This likely stems from using a too small block size to adequately capture risk far out in the tail. No evidence was found of a significant difference in modelling tail risk using a GPD model with the Hill estimator and deterministic threshold selection in comparison to the ML estimator with cross-validation threshold selection method. Both models appear to capture the tail behaviour of the FTSE 100 log-returns well. Next, we will describe how to model risk in a multivariate setting.

Table 18: Summary of estimates, pre-Brexit sample.

| | $\xi$ | $\alpha$ | $\text{VaR}_{.99}$ | $\text{ES}_{.975}$ |
|---|---|---|---|---|
| GEV | 0.2124 | 4.708 | 1.085 | 1.204 |
| | [ 0.112, 0.313 ] | | [ 1.009, 1.160 ] | [ 1.009, 1.400 ] |
| GPD | 0.3391 | 2.949 | 3.799 | 4.213 |
| | [ 0.255, 0.424 ] | | [ 3.447, 4.150 ] | [ 3.695, 4.730 ] |
| GPD$^*$ | 0.1924 | 5.198 | 3.949 | 4.187 |
| | [ 0.115, 0.269 ] | | [ 3.650, 4.248 ] | [ 3.796, 4.578 ] |
| Empirical | | | 3.875 | 4.155 |
| | | | [ 3.453, 4.741 ] | [ 3.655, 4.681 ] |

*Note:* Estimates of VaR and ES are given in percentage units. Intervals at a 0.95 level for $\xi$, $\text{VaR}_{0.99}$ and $\text{ES}_{0.975}$ are given in square brackets below each estimate. Tail index $\xi$ is fitted with asymptotic normal intervals. Intervals for model VaR and ES are performed using the delta method. Intervals for the empirical sample are fitted using m-out-of-n percentile bootstrap with $B = 1000$ resamples and a resample size of $m = 2827$ for the full sample and $m = 2591$ for the pre-Brexit sample.

Table 19: Summary of estimates, full sample.

| | $\xi$ | $\alpha$ | $\text{VaR}_{.99}$ | $\text{ES}_{.975}$ |
|---|---|---|---|---|
| GEV | 0.2164 | 4.621 | 1.071 | 1.189 |
| | [ 0.118, 0.315 ] | | [ 0.999, 1.143 ] | [ 0.999, 1.378 ] |
| GPD | 0.3343 | 2.991 | 3.669 | 4.091 |
| | [ 0.253, 0.416 ] | | [ 3.374, 4.024 ] | [ 3.619, 4.562 ] |
| GPD* | 0.2043 | 4.895 | 3.856 | 4.102 |
| | [ 0.130, 0.278 ] | | [ 3.571, 4.142 ] | [ 3.723, 4.481 ] |
| Empirical | | | 3.689 | 4.064 |
| | | | [ 3.271, 4.339 ] | [ 3.609, 4.567 ] |

*Note:* Estimates of VaR and ES are given in percentage units. Intervals at a 0.95 level for $\xi$, $\text{VaR}_{0.99}$ and $\text{ES}_{0.975}$ are given in square brackets below each estimate. Tail index $\xi$ is fitted with asymptotic normal intervals. Intervals for model VaR and ES are performed using the delta method. Intervals for the empirical sample are fitted using m-out-of-n percentile bootstrap with $B = 1000$ resamples and a resample size of $m = 33\,333$.

## 3. Multivariate Risk Models

Multivariate risk modelling is a concept that is still evolving. While copulas are a mainstay of multivariate risk theory, graphical models are emerging as a way to overcome some of the difficulties as dimensions scale up. The next section presents some of the main theories on the subjects as well as more recent developments.

### 3.1 Multivariate Extreme Value Theory

#### 3.1.1 Copulas

Every joint distribution function of a pair of random variables contains information on their marginal behaviours as well as dependence structure. Copulas supply a simple way to model these dependence structures.

**Definition 3.1.** A copula is a multivariate distribution function $C : [0,1]^d \to [0,1]$ with standard uniform marginals, i.e. $C(1,\dots,1,u_i,1,\dots,1) = u_i, \forall i \in \{1,\dots,d\}, u_i \in [0,1]$.

The appeal of copulas is illustrated in the traditional results of Sklar (1959). The following theorem shows that all multivariate cdf's can be expressed as copulas which may be constructed using univariate marginals.

**Theorem 7** (Sklar's theorem, 1959)**.** *Let $F$ be a joint cdf with marginals $F_1,\dots,F_d$. Then there exists a copula $C$ such that,*

$$F(x_1,\dots,x_d) = C(F_1(x_1),\dots,F_d(x_d)), \quad \forall x_i \in \mathbb{R}, \quad i = 1,\dots,d. \tag{15}$$

*If the marginals are continuous, then C is unique. Conversely, if C is a copula and* $(F_i, 1 \leq i \leq d)$ *are univariate d.f., then F defined above is a multivariate df with marginals* $F_1, \ldots, F_d$.

We proceed to define the density function of a copula if existing.

**Definition 3.2.** The density function $c$ of a copula $C$ is defined as

$$c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}.$$

Consequently, a bivariate joint density is the product of the marginal densities and the density $c$ of the copula,

$$f(x_1, x_2) = c(F(x_1), F(x_2)) f_1(x_1) f_2(x_2).$$

Two examples of elliptical copulas commonly used in financial models are defined next.

**Definition 3.3.** A bivariate Gaussian copula with parameter $\rho \in (-1, 1)$ is defined as,

$$C_\rho^{Ga}(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\} dxdy,$$

where $\Phi$ denotes the standard Gaussian cdf.

**Definition 3.4.** A bivariate $t$-copula with parameter $\rho \in (-1, 1)$ and degrees of freedom $\nu \geq 2$ is defined as,

$$C_\rho^t(u, v) = \frac{\Gamma(\nu/2 + 1)}{\Gamma(\nu/2)\nu\pi\sqrt{1-\rho^2}} \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right)^{-(\nu+2)/2} dxdy,$$

where $t_\nu^{-1}$ denotes the inverse of the Student t cdf.

Another family of copulas is the Archimedean copula. Two examples from the Archimedean copula family are defined next, the Gumbel copula, being an extreme value copula and the Clayton copula.

**Definition 3.5.** A bivariate Gumbel copula is defined as,

$$C_\rho^{Gu}(u, v) = \exp\left\{-\left((-\log u)^\rho + (-\log v)^\rho\right)^{1/\rho}\right\}, \quad \rho \geq 1.$$

**Definition 3.6.** A bivariate Clayton copula is defined as,

$$C_\rho^{Cl}(u, v) = (u^{-\rho} + v^{-\rho} - 1)^{-1/\rho}, \quad \rho > 0,$$

These copulas are capable of handling independence, where $C_\rho(u, v) = uv$, as well as perfect dependence, $C_\rho = \min(u, v)$. In the Gumbel case, when $\rho = 1$, independence is present. Perfect dependence corresponds to a $\rho$ approaching zero. For the Clayton copula, as $\rho$ goes to +0, $u$ and $v$ are independent. Perfect dependence is associated with a $\rho$ approaching infinity in the Clayton copula.

Although being widely used in bivariate models, copulas can be computationally expensive in high dimensional settings. Furthermore, estimating a $d \times d$ covariance matrix $\Sigma$ as the number of assets increases to thousands can be challenging. Since financial risk management is concerned with current risks, using data from many years ago may provide a false image of today's risk. An interesting way to overcome these shortcomings is to utilise a graphical model with a sparse inverse covariance matrix. In the next chapter, the theoretical foundation for such models is presented.

### 3.1.2 Probabilistic Graphical Models

Let $X = (X_1, \ldots, X_p)$ be a random vector with distribution $N(\mu, \Sigma)$, precision matrix $\Omega = \Sigma^{-1}$ and sample covariance matrix $\hat{\Sigma} = S$. We denote a graph $\mathcal{G}$ consisting of a set of nodes $\mathcal{V}$ and edges $\mathcal{E}$. If the nodes are labelled distinct from each other the graph is called a labelled graph. While each node is associated with a random variable $X_i$, the edges represent the probabilistic interactions between these variables. For each ordered pair $(i, j)$ there is an edge between $X_i$ and $X_j$ if $(i, j) \in \mathcal{E}$. Furthermore, we use the notation $X_{\setminus\{i,j\}}$ to denote the variables contained in $X$ that are neither $X_i$ or $X_j$. Figure 6 visualises three types of graphical models. The set of nodes in Figure 6 a, is written $\mathcal{V} = \{A, B, C\}$ while the set of edges can be written $\mathcal{E} = \{e_1, e_2\}$.

Conditional independence is a main concept in graph theory. An edge is absent only if $X_i$ is conditionally independent of $X_j$ given the other variables $X_{\setminus\{i,j\}}$. We write this as $X_i \perp\!\!\!\perp X_j | X_{\setminus\{i,j\}}$ and note that this can only be true if $p(X_i, X_j | X_{\setminus\{i,j\}}) = p(X_i, | X_{\setminus\{i,j\}}) p(X_j, | X_{\setminus\{i,j\}})$. An important result is that, if we can write the joint probability distribution $p(X_i, X_j, X_{\setminus\{i,j\}})$ in terms of some functions $f$ and $g$, conditional independence exists.

**Definition 3.7.** (Factorisation criterion)

$$X_i \perp\!\!\!\perp X_j \mid X_{\setminus\{i,j\}} \iff p(X_i, X_j, X_{\setminus\{i,j\}}) = f(X_i, X_{\setminus\{i,j\}}) g(X_j, X_{\setminus\{i,j\}}), \tag{16}$$

for some functions $f$ and $g$, and for all $X_{\setminus\{i,j\}}$ with $p(X_{\setminus\{i,j\}}) > 0$.

In a directed graph, each edge has a direction represented by an arrow (Figure 6, a). A directed cyclic graph has a structure where the dependencies are chained together in a loop (Figure 6, b). The complex dependence structure in directed cyclic graphs makes this model problematic to use in practice. We say that a graph is acyclic if there is no way to move from a node $e_1, \ldots, e_k$ at a directed path such that $e_k = e_1$. A Bayesian network is an example of a directed acyclic graph.

Undirected graphs are useful for describing interactions between variables, where no natural directionality exists. The edges $\mathcal{E}$ in an undirected graph has no direction (Figure 6, c). A crucial feature of undirected graphical models is that, if the variables modelled possess the Markov property, they can be represented as a Markov random field. To determine whether a distribution satisfies the Markov property we first need to introduce the notion of the clique.

A set of nodes where all nodes are neighbours of each other's is referred to as a clique (Murphy, 2012). As an example, the nodes A and B in Figure 6 c, possess the clique property. If we would include node A, B and C in a clique, the clique property is lost as the nodes A and C are not neighbours. A clique where the number of neighbours cannot be made larger without losing the clique property is called a maximal clique. We let $c$ denote a clique with potential function $\Psi_c(X_c)$ where the set of maximal cliques is $\mathcal{C}$. The potential function is a concept in physics and relates to energy rather than probability (Murphy, 2012).

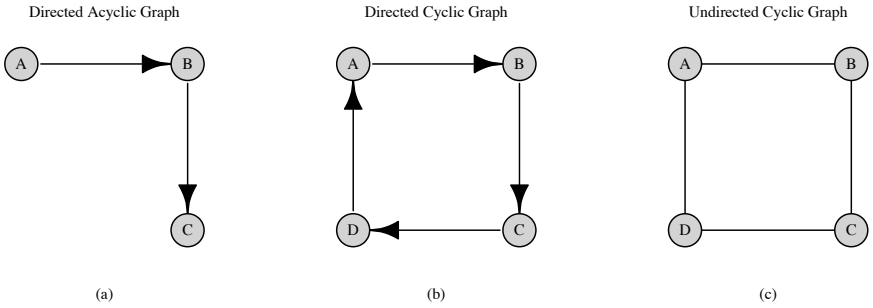



Figure 6: Different types of probabilistic graphical models. (a) A sample directed acyclic graph. (b) A sample directed cyclic graph. (c) A sample undirected cyclic graph.

The Hammersley-Clifford theorem states that if a positive distribution $p$ factorises over an undirected graph $\mathcal{G}$ this distribution satisfies the Markov property. It turns out that being able to express the joint density in terms of a number of functions rather than knowing all marginals' and their relationships have many advantages. Mainly, it allows us to express a high-dimensional distribution in terms of a smaller number of factors.

**Theorem 8.** *(Hammersley and Clifford, 1971) A positive distribution $p(X) > 0$ satisfies the CI properties of an undirected graph $\mathcal{G}$ if and only if p can be represented as a product of factors, one per maximal clique,*

$$p(X) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \Psi_c(X_c),$$

*where $\mathcal{C}$ is the set of all the (maximal) cliques of $\mathcal{G}$, and $Z$ is the normalising constant given by*

$$Z \stackrel{\Delta}{=} \sum_{X} \prod_{c \in \mathcal{C}} \Psi_c(X_c).$$

An example of the Hammersley-Clifford theorem in practice is the Gaussian Markov random field. For multivariate Gaussian distributions, conditional independence such as defined in (17) means that the precision matrix $\Omega_{ij} = 0$.

**Theorem 9.** *Let $X$ be a normal distributed r.v. with mean $\mu$ and precision matrix $\Omega > 0$. Then for $i \neq j$,*

$$X_i \perp\!\!\!\perp X_j \mid X_{\setminus\{i,j\}} \iff \Omega_{ij} = 0. \tag{17}$$

**Definition 3.8.** A random vector $X = (X_1, \ldots, X_p)^\top \in \mathbb{R}^p$ is called a Gaussian Markov random field with respect to a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\mu$ and precision matrix $\Omega > 0$ if and only if its density has the form

$$p_X(x) = (2\pi)^{-p/2} \mid \Omega \mid^{1/2} \exp\left\{-\frac{1}{2}(X-\mu)^\top \Omega (X-\mu)\right\} \tag{18}$$

and

$$\Omega_{ij} \neq 0 \quad \iff \quad \{i,j\} \in \mathcal{E} \quad \text{for all} \quad i \neq j.$$

*Proof.* [Theorem 9] (Rue and Held, 2005) We partition $X$ as $(X_i, X_j, X_{\setminus\{i,j\}})$ and then use the multivariate factorisation criterion on $p_X(X_i, X_j, X_{\setminus\{i,j\}})$. We fix $i \neq j$ and assume $\mu = 0$ without loss of generality. From (18) we get

$$\begin{aligned} p_X(X_i, X_j, X_{\setminus\{i,j\}}) &\propto \exp\left(-\frac{1}{2}\sum_{k,l} X_k \Omega_{kl} X_l\right) \\ &\propto \exp\left(-\frac{1}{2} X_i X_j (\Omega_{ij} + \Omega_{ji}) - \frac{1}{2}\sum_{\{k,l\}\neq\{i,j\}} X_k \Omega_{kl} X_l\right). \end{aligned}$$

Only the first term involve $X_i X_j$ if and only if $\Omega_{ij} \neq 0$. We see from (16) in Definition 3.7 that

$$p_X(X_i, X_j, X_{\setminus\{i,j\}}) = f(X_i, X_{\setminus\{i,j\}}) g(X_j, X_{\setminus\{i,j\}})$$

for some functions $f$ and $g$, if and only if $\Omega_{ij} = 0$. □

The estimation of graphical models can be a challenging task. When the number of parameters $p$, is considerably smaller than the number of observations $n$, maximum likelihood estimation of $\Sigma$ is a viable alternative. However, when $p > n$ the estimate $\hat{\Sigma}$ is not positive definite such that the joint density is undefined. A remedy is to use lasso methods and estimate $\Sigma$ by minimising

$$-\ell(\Omega) + \lambda \sum_{j\neq k} |\, \Omega_{jk} \,| \tag{19}$$

where $\lambda$ is the regularisation parameter,

$$\ell(\Omega) = \frac{1}{2}(\log |\, \Omega \,| - \mathrm{tr}(\Omega S) - p\log(2\pi))$$

the log-likelihood and $S$ the sample covariance matrix. While many estimation methods of $\Omega$ have been proposed, an efficient estimator is the glasso algorithm (Friedman et al., 2008). The glasso algorithm is a block coordinate descent algorithm, meaning that it successively minimises (19) along the coordinate directions until convergence.

### 3.1.3 The Nonparanormal

This section connects the preceding ones, giving an example of how copulas and graphical models relate to each other. The nonparanormal distribution was introduced by Liu et al. (2009) as a means to estimate undirected graphs in high dimensional settings. It can be shown that the nonparanormal under some restrictions is equivalent to a semi-parametric Gaussian copula but first we define the nonparanormal distribution.

**Definition 3.9.** (Liu et al. (2009)) If there exist functions $\{f_j\}_{j=1}^p$ for a random vector $X = (X_1, \ldots, X_p)^\top$ such that $Z = f(X) \sim N(\mu, \Sigma)$, where $f(X) = (f_1(X_1), \ldots, f_p(X_p))$, then $X$ has a nonparanormal distribution

$$X \sim NPN(\mu, \Sigma, f).$$

If each function $f_j$ is monotone and differentiable the joint density can be expressed such as

$$p_X(x) = \frac{1}{(2\pi)^{p/2} \mid \Sigma \mid^{1/2}} \exp\left\{-\frac{1}{2}(f(x)-\mu)^\top \Sigma^{-1}(f(x)-\mu)\right\} \prod_{j=1}^{p} \mid f_j'(x_j) \mid . \tag{20}$$

To see how the non-paranormal relates to copulas, consider the Lemma given next.

**Lemma 1.** *(Liu et al. (2009)) When the functions $f_j$ are monotone and differentiable the nonparanormal distribution $NPN(\mu, \Sigma, f)$ is a Gaussian copula.*

*Proof.* (Liu et al. (2009)) As was shown in Theorem 7 (Sklar, 1959) any joint distribution can be expressed in terms of a copula $C$ such as

$$F(x_1, \ldots, x_p) = C\{F_1(x_1), \ldots, F_p(x_p)\}.$$

In the nonparanormal case

$$F(x_1, \ldots, x_p) = \Phi_{\mu,\Sigma}(\Phi^{-1}(F_1(x_1)), \ldots, \Phi^{-1}(F_p(x_p)))$$

where $\Phi_{\mu,\Sigma}$ is the multivariate cdf and $\Phi$ is the univariate standard Gaussian cdf. The copula corresponding to the nonparanormal distribution becomes,

$$C(u_1, \ldots, u_p) = \Phi_{\mu,\Sigma}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)).$$

We can see that this is a Gaussian copula with parameters $\mu$ and $\Sigma$. Consequently, if each function $f_j$ is differentiable the density of $X$ has the form of (20). □

The density in (20) is identifiable under the condition that the mean and variance of $f_j$ equal that of $X_j$ such as,

$$\mu_j = \mathbb{E}(Z_j) = \mathbb{E}(X_j) \text{ and } \sigma_j^2 = \Sigma_{jj} = \mathbb{V}(Z_j) = \mathbb{V}(X_j). \tag{21}$$

We can then write the marginal distribution function $F_j(x)$ of $X_j$

$$F_j(x) = \mathbb{P}(X_j \leq x) = \mathbb{P}(Z_j \leq f_j(x)) = \Phi\left(\frac{f_j(x) - \mu_j}{\sigma_j}\right),$$

where the form of $f_j$ then can be derived as

$$f_j(x) = \mu_j + \sigma_j \Phi^{-1}(F_j(x)). \tag{22}$$

**Lemma 2.** *(Liu et al. (2009)) If $X \sim NPN(\mu, \Sigma, f)$ is nonparanormal and each $f_j$ is differentiable, then $X_i \perp\!\!\!\perp X_j \mid X_{\setminus\{i,j\}}$ if and only if $\Omega_{ij} = 0$, where $\Omega = \Sigma^{-1}$.*

*Proof.* (Liu et al. (2009)) The density in (20) factors with respect to the graph of $\Omega$ and according to Theorem 8 therefore obey the global Markov property of the graph. □

**Lemma 3.** *(Liu et al. (2009)) Define*

$$h_j(x) = \Phi^{-1}(F_j(x))$$

*and let $\Lambda$ be the covariance matrix of $h(x)$. Then $X_j \perp\!\!\!\perp X_k \mid X_{\setminus\{j,k\}}$ if and only if $\Lambda^{-1}_{jk} = 0$.*

*Proof.* (Liu et al. (2009)) Writing the covariance matrix as

$$\Sigma_{jk} = \mathbb{C}(Z_j, Z_k) = \sigma_j \sigma_k \mathbb{C}(h_j(X_j), h_k(X_k)),$$

we can express the covariance matrix such as $\Sigma = D\Lambda D$ and $\Sigma^{-1} = D^{-1}\Lambda^{-1}D^{-1}$, with $D$ being the diagonal matrix with diagonal elements $\sigma_1, \sigma_2, \ldots, \sigma_p$. Hence, we have established that the zero pattern of $\Lambda^{-1}$ is equal to that of $\Sigma^{-1}$. □

As a consequence of Lemma 3, we do not need to estimate the means or standard deviations to estimate the graph.

## 3.2 Methodology

Next, copula models and probabilistic graphical models are described in Section 3.2.1 and 3.2.2. In Section 3.2.3 we present the multivariate risk model while the evaluation methods are summarised in Section 3.2.4.

### 3.2.1 Copula Models

Consider a realisation of i.i.d. random vectors $\mathbf{x}_i, i \in \{1, \ldots, n\}$ of the random vectors $\mathbf{X}_i, i \in \{1, \ldots, n\}$ from a joint distribution $H$, with marginals $F_j$ and copula $C$. Assuming known marginals, $u_{ij} = F_j(x_{ij}), i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\}$ is a sample from $C$. However, in most applications marginals are not known. While parametric and non-parametric methods to estimate the marginals exist, a non-parametric approach using pseudo-observations is most commonly used. We follow the same procedure, estimating the pseudo-observations such as

$$\hat{u}_{ij} = \frac{n}{n+1} \hat{F}_{n,j}(x_{ij}) = \frac{r_{ij}}{n+1}, \tag{23}$$

where $\hat{F}_{n,j}$ is the EDF for the $j$ : th margin and $r_{ij}$ the rank of $x_{ij} \in \{1, \ldots, n\}$. We use a maximum-likelihood estimator for the copula parameters as this estimator has shown good properties even in high dimensional settings regardless if the marginals are known (Hofert et al., 2013). For details on the estimation, see Nagler (2019).

Kendall's tau is a measure of the strength of association in large values among two random variables. The population parameter is defined as

$$\tau = \mathbb{E}[\operatorname{sgn}((X_1 - X_1')(X_2 - X_2'))],$$

where $(X_1', X_2')^\top$ is an independent pair of random vectors with the same distribution as $(X_1, X_2)$. The corresponding sample estimate for the random sample $\mathbf{U}_i = (U_{i1}, U_{i2})^\top$ can be found as

$$\hat{\tau}_n = \binom{n}{2}^{-1} \sum_{1 \le i \le j \le n} \operatorname{sgn}[(U_{i1} - U_{j1})(U_{i2} - U_{j2})]. \tag{24}$$

To get an estimate of the dependence specific of the tails we calculate the coefficient of upper tail dependence as

$$\begin{aligned}\lambda_u(X_1, X_2) &= \lim_{\alpha \to 1} \mathbb{P}[X_2 > \text{VaR}_\alpha(X_2) \mid X_1 > \text{VaR}_\alpha(X_1)] \\ &= \lim_{\alpha \to 1} \frac{1 - 2\alpha + C(\alpha, \alpha)}{1 - \alpha},\end{aligned} \tag{25}$$

and the lower tail dependence

$$\begin{aligned}\lambda_l(X_1, X_2) &= \lim_{\alpha \to 0} \mathbb{P}[X_2 \leq \text{VaR}_\alpha(X_2) \mid X_1 \leq \text{VaR}_\alpha(X_1)] \\ &= \lim_{\alpha \to 0} \frac{C(\alpha, \alpha)}{\alpha}.\end{aligned} \tag{26}$$

The empirical tail dependence coefficient is now described. The method of Schmidt and Stadtmüller (2006) is utilised in our analysis (see also Drees and Huang, 1998). A software implementation for *r* is available using the *tdc* function in the *FRAPO* package (Pfaff, 2016). Assume that $n$ two-dimensional random vectors $(X_1^{(1)}, X_2^{(1)}), \ldots, (X_1^{(n)}, X_2^{(n)})$ are i.i.d. with distribution function $F$ and marginal distribution functions $G$, $H$, and copula $C$. We define the empirical copula $C_n$ as

$$C_n(u, v) = F_n(G_n^{-1}(u), H_n^{-1}(v)), \quad (u, v)^\top \in [0, 1]^2,$$

where $F_n$, $G_n$, $H_n$ is the empirical distribution functions corresponding to $F$, $G$, $H$ and

$$F_n(x_1, x_2) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}_{\{X_1^{(j)} \leq x_1, X_2^{(j)} \leq x_2\}}.$$

Let $R_{n_1}^{(j)}$ and $R_{n_2}^{(j)}$ denote the rank of $X_1^{(j)}$ and $X_2^{(j)}$, $j = 1, \ldots, n$, respectively. The empirical upper tail coefficient is then estimated as

$$\hat{\lambda}_{u,n}(x_1, x_2) := \frac{n}{k} \bar{C}_n \left( \frac{k x_1}{n}, \frac{k x_2}{n} \right) \approx \frac{1}{k} \sum_{j=1}^{n} \mathbb{I}_{\{R_{n_1}^{(j)} > n - k x_1, R_{n_2}^{(j)} > n - k x_2\}}. \tag{27}$$

and the lower empirical tail coefficient is estimated as

$$\hat{\lambda}_{l,n}(x_1, x_2) := \frac{n}{k} C_n \left( \frac{k x_1}{n}, \frac{k x_2}{n} \right) \approx \frac{1}{k} \sum_{j=1}^{n} \mathbb{I}_{\{R_{n_1}^{(j)} \leq k x_1, R_{n_2}^{(j)} \leq k x_2\}}. \tag{28}$$

We choose a value on the threshold as $k = \sqrt{n}$.

### 3.2.2 Probabilistic Graphical Models

We follow the method of Liu et al. (2009) to estimate the undirected graphs. For the software implementation in *r* to estimate and visualise these graphs see Jiang et al. (2019). Let $X^{(1)}, \ldots, X^{(n)}$ be a sample of size $n$ with $X^{(i)} = (X_1^{(i)}, \ldots, X_p^{(i)})^\top \in \mathbb{R}^p$. We define the estimator of $h_j, j \in \{1, \ldots, p\}$,

$$\hat{h}_j(x) = \Phi^{-1}(\tilde{F}_j(x))$$

where $\tilde{F}_j$ is an estimator of $F_j$. To estimate $\tilde{F}_j$, the truncated Winsorized estimator is used,

$$\tilde{F}_j(x) = \begin{cases} \delta_n & \text{if } \hat{F}_j(x) < \delta_n \\ \hat{F}_j(x) & \text{if } \delta_n \leq \hat{F}_j(x) \leq 1 - \delta_n \\ (1 - \delta_n) & \text{if } \hat{F}_j(x) > 1 - \delta_n, \end{cases}$$

where the choice of the truncation parameter $\delta_n$ results in a variance-bias trade-off. Liu et al. (2009) suggest that a good convergence rate in the estimate of a high dimensional precision matrix is achieved with a truncation parameter of

$$\delta_n = \frac{1}{4n^{1/4}\sqrt{\pi \log n}}.$$

The transformation function estimate is then given by,

$$\tilde{f}_j(x) = \hat{\mu}_j + \hat{\sigma}_j \tilde{h}_j(x)$$

where

$$\tilde{h}_j(x) = \Phi^{-1}(\tilde{F}_j(x)),$$

and the sample mean is $\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^n X_j^{(i)}$, while the sample standard deviation is $\hat{\sigma}_j = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_j^{(i)} - \hat{\mu}_j)^2}$.

We can then proceed to define the sample covariance matrix $S_n(\tilde{f})$ and sample mean $\mu_n(\tilde{f})$ of $\tilde{f}(X^{(1)}), \ldots, \tilde{f}(X^{(n)})$,

$$\begin{aligned} S_n(\tilde{f}) &= \frac{1}{n}\sum_{i=1}^n (\tilde{f}(X^{(i)}) - \mu_n(\tilde{f}))(\tilde{f}(X^{(i)}) - \mu_n(\tilde{f}))^\top \\ \mu_n(\tilde{f}) &= \frac{1}{n}\sum_{i=1}^n \tilde{f}(X^{(i)}). \end{aligned}$$

The maximum likelihood estimator of the precision matrix $\Omega$ can be found as $\hat{\Omega}_n^{\text{ML}} = S_n(\tilde{f})^{-1}$. The $\ell_1$-regularised estimator is the solution to the optimisation problem

$$\hat{\Omega}_n = \arg\min_{\Omega} \left\{ \text{tr}\left(\Omega S_n(\tilde{f})\right) - \log|\,\Omega\,| + \lambda \,\|\,\Omega\,\|_1 \right\} \tag{29}$$

with $\lambda$ being the regularisation parameter and $\|\,\Omega\,\|_1 = \sum_{j \neq i} |\,\Omega_{jk}\,|$. Finally, the graph is estimated such as $\hat{\mathcal{G}}_n = \{(j,k) : \hat{\Omega}_{jk} \neq 0\}$.

### 3.2.3 Risk Model

To quantify the risk within the FTSE 100 index from the estimated graphical model we perform dimension reduction by an eigenvalue decomposition of the covariance matrix. Bivariate $t$-copulas are then estimated between the nonparanormally transformed monthly block maxima of the return losses for each stock and the principal components. Finally, the coefficient of tail dependence is estimated for each stock. We visualise the result using the *igraph* package (Csárdi, 2020). To extract the main cluster the *clusters* and *induced.subgraph* commands are utilised. The *layout.kamada.kawai* command is applied to the resulting graph. For an introduction to network inference see Vialaneix (2015). The stocks are labelled according to their tickers and sector classification is made according to the Industry Classification Benchmark standard (FTSE Russell, 2019).

Procedure to estimate the extreme tail risk within the FTSE 100 index.

---

1. Extract the block maxima $\boldsymbol{L}_{BM}$ from the return losses of the stocks in the index.
2. Obtain the transformed block maxima $\boldsymbol{L}_{npn}$ according to Section 3.2.2.
3. Fit the undirected graphical model $\hat{\mathcal{G}}_{npn}$ to the transformed block maxima $\boldsymbol{L}_{npn}$ using glasso.
4. Extract the covariance matrix as $\hat{\Sigma}_{npn} = \hat{\Omega}_{npn}^{-1}$.
5. Perform an eigenvalue decomposition of the estimated covariance structure $\Sigma_{npn}$ in the previous step and obtain the matrix of eigenvectors $\boldsymbol{v}$.
6. The components are now calculated as $\boldsymbol{Y} = \boldsymbol{L}_{npn} \cdot \boldsymbol{v}$.
7. Fit a bivariate Student t copula between each vector of the transformed losses $\boldsymbol{L}_{npn}$ and the vectors in $\boldsymbol{Y}$.
8. Estimate the upper coefficient of tail dependence for each copula in the previous step.
9. Visualise the graph scaling the size of the nodes according to the upper coefficient of tail dependence.

---

#### 3.2.4 MODEL EVALUATION

To compare the fit of bivariate copulas, the Bayesian information criterion (BIC) is used (see Schwarz, 1978). We also compare the tail coefficients of the fitted copula with the empirical ones.

When the dimension is large in relation to the sample size, BIC has shown to perform poorly. Model evaluation for high-dimensional graphical models, therefore, requires different methods. Several suggestions to account for this shortcoming have been made. These include the modified rotation information criterion (RIC), the stability approach for regularisation selection method (StARS) and the extended Bayesian information criterion (EBIC) (Lysen, 2009; Liu et al., 2010; Foygel and Drton, 2010). We evaluate the fit using all of these criteria and compare the impact on the sparsity of the precision matrix.

Sparsity can be measured in many ways, we use the following definition,

$$Q = \frac{\sum \mathbb{I}_{\{\Omega_{ij}=0, i\neq j\}}}{\sum \mathbb{I}_{\{\Omega_{ij}, i\neq j\}}}.$$

Note that the standard measure used in the *huge* package (see Jiang et al., 2019) is counting the non-zero off-diagonal elements while our definition is the sum of the number of zero elements in comparison to the total number of off-diagonal elements.

### 3.3 Simulation Study II

Before proceeding to a real data set, we perform some simulations to get an idea of the behaviour of some common copulas. We can easily see the symmetric appearance of the elliptical copulas from the normal and Student t distributions (Figure 7). These distributions do not allow for an asymmetry in the tail dependencies where the upper and lower tails differ. The Archimedean copulas allow for different dependence structures in the upper and lower tails. This is clear judging by the ranked scatter of the Gumbel and Clayton copulas. The Gumbel distribution has a strong upper tail dependence while the Clayton copula shows an even stronger lower tail dependence. Since a common assumption in financial risk theory is that the lower tail is heavier than the upper tail, a Clayton copula should be the best choice. We continue our analysis with a case study of the dependence between the FTSE 100 and S&P 500 indexes to see if this is true.

### 3.4 Case Study II: The Dependence Between the FTSE 100 and S&P 500 Indexes

#### 3.4.1 FTSE 100 AND S&P 500 DATA

A description of the data used in this study is presented in Table 20. To gain insights on the tail dependences between the FTSE 100 and S&P 500 indexes, log-returns are calculated for further analysis using copulas. In the final graphical analysis, 946 daily log-returns for 95 of the 100 stocks in the FTSE 100 index are calculated during the period 2015-09-23 to 2019-05-30. Five stocks were excluded because of missing data. Then, monthly block maxima are calculated resulting in 45 observations for each component.

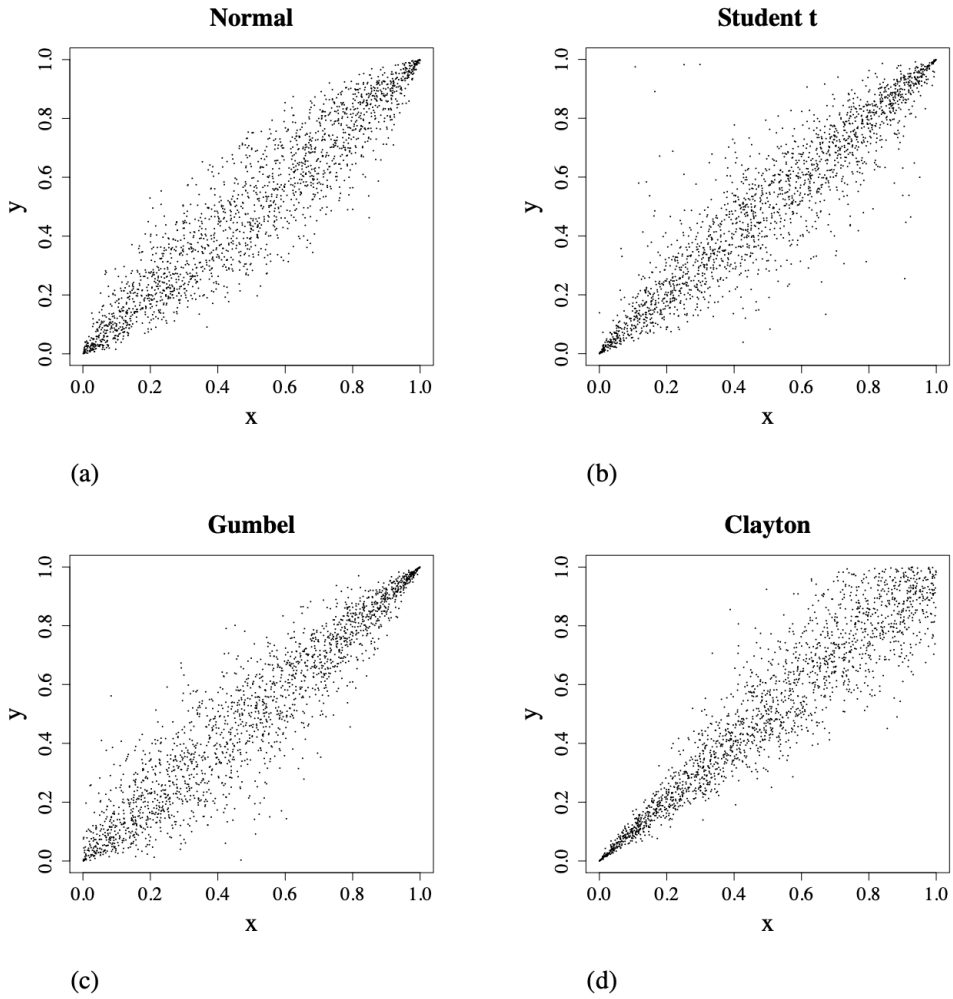



Figure 7: Two thousand simulated bivariate observations from the (a) Normal, (b) Student t, (c) Gumbel and (d) Clayton copulas. Parameters are chosen such that Kendall's tau is 0.8.

Table 20: Description of the FTSE 100 and S&P 500 stock index data.

| Data | Type | $n$ | |
|---|---|---|---|
| | | pre-Brexit | full sample |
| $S$ | Closing Value | 8285 | 9134 |
| $R_{\Delta t}$ | Log-Return | 8284 | 9133 |

*Note:* The table shows the number of data points for the series $S$ of closing values for FTSE 100 and S&P 500 log-returns $R$. The number of observations is given for the period before the Brexit referendum (1984-01-03 – 2016-06-23) as well as the full series (1984-01-03 – 2019–05–31).

### 3.4.2 DEPENDENCE BETWEEN THE BRITISH AND US STOCK MARKETS

To assess the dependence structures between the British and US markets before and after Brexit we estimate Pearson correlation and Kendall's tau for the pre-Brexit and full sample period. Both correlation measures are slightly lower during the full series compared to the pre-Brexit period (Table 21). The estimated linear dependencies using a Clayton copula is much higher during the pre-Brexit period and lower during the full sample period compared to the empirical estimate. The Gumbel copula is closer to the empirical estimate than the other copulas. Examining the non-linear dependencies we can tell that the Clayton copula still deviates more from the empirical estimate. Again, the Gumbel copula seems to coincide with the empirical estimate more closely than the other copulas. However, these correlation measures only consider the strength of dependence and not its structure. We proceed to examine the ranked scatter plots of the empirical pseudo-observations and simulations based on fitted copulas.

Table 21: Bivariate correlation between the marginals of simulated copulas as well as the empirical pseudo observations of the FTSE 100 and S&P 500 indexes.

| | pre-Brexit | | full sample | |
|---|---|---|---|---|
| | Pearson | Kendall's Tau | Pearson | Kendall's Tau |
| Normal | 0.423 | 0.290 | 0.435 | 0.298 |
| Student t | 0.418 | 0.297 | 0.409 | 0.290 |
| Gumbel | 0.409 | 0.284 | 0.414 | 0.286 |
| Clayton | 0.591 | 0.420 | 0.378 | 0.259 |
| Empirical | 0.407 | 0.288 | 0.406 | 0.286 |

*Note:* We first calculate the empirical pseudo observations of the full log-return series of FTSE 100 and S&P 500 stock indexes according to (23). Then, normal, Student t, Gumbel and Clayton copulas are fitted to these pseudo observations. The parameter estimates are used to simulate new data from the mentioned copulas. Finally, Pearson and Kendall's tau correlation coefficients are calculated, where Kendall's tau is estimated as in (24).

### 3.4.3 Tail Dependence Between the British and US Stock Markets

As we study the ranked scatter plot of the empirical and simulated copulas some properties become clear (Figure 8). There is no apparent difference in the dependence structure before and after Brexit. Furthermore, the normal and Gumbel copulas do not capture the lower dependence of the British and US stock markets well. Instead, the Student t and Clayton copulas appear to reflect the empirical copula more closely. We see a slightly higher lower dependence structure in the Clayton copula compared to the empirical copula. The $t$-copula more closely matches the empirical copulas in both tails of the distribution. The Gumbel copula seems to capture the upper dependence structure fairly close to the empirical distribution. In the next chapter, we present the model fit and tail dependence estimates.

### 3.4.4 Modelling the Tail Dependence Between the British and US Stock Markets

As shown in section 3.2.1 the tail dependence coefficients are closely related to value at risk. Knowing the relation between two financial indexes can provide more information on the index of interest. As an example of an application, given new information in an open market, one can assess the risk in a closed market before it opens. Since extreme outcomes rarely occur, we can also utilise more information by considering several indexes at once.

Table 22 show the estimates of the fitted copulas as well as the tail dependence coefficients for the empirical distribution. We can see that the lower empirical tail dependence coefficient is slightly larger pre-Brexit (0.29) than for the full sample (0.27). This indicates that the dependence in the return losses during the aftermath of the Brexit referendum was slightly lower. The $t$-copula appears to capture the lower tail dependences well for both sample periods. The Clayton copula that in theory should capture the lower tail dependence more accurately yield estimates that are much higher than the empirical estimates. Considering the upper tail of the distribution, the Gumbel copula most closely reflects the empirical estimate. The best overall fit is achieved using a $t$-copula according to the BIC evaluation measure. As a concluding remark, the Student t copula does not seem to be able to differentiate between the dependence structure of the pre-Brexit and full sample period.

A shortcoming of copulas is that the estimates in high dimensional models tend to be computationally expensive and in some cases inaccurate. Next, some ideas on how to extend multivariate risk to higher dimensions using graphical models are presented.

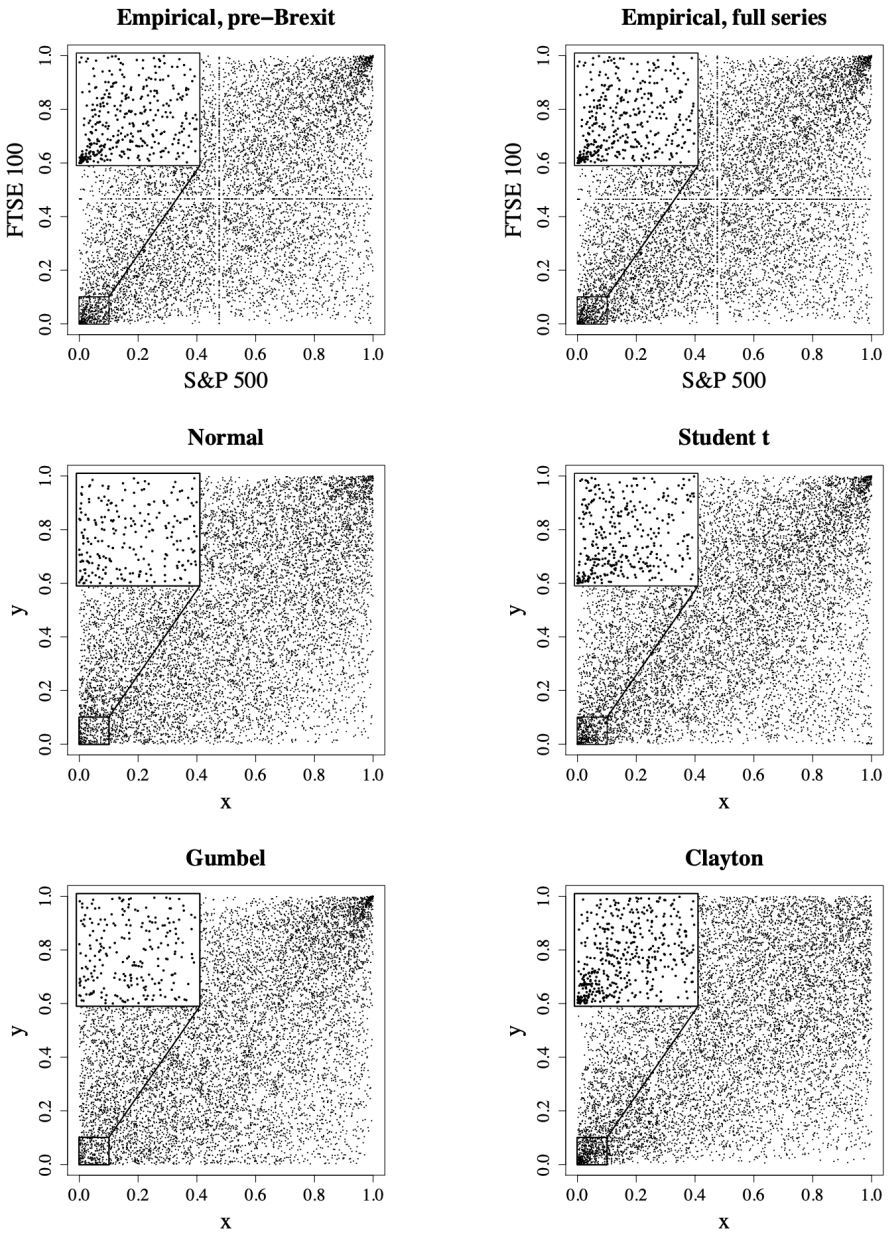



Figure 8: Standardised ranks for FTSE 100 and S&P 500 log-returns as well as simulated normal, Student t, Gumbel and Clayton copulas with the lower ranks, $\{y, x\} \in [0, 0.1]$ magnified. Copulas are simulated based on estimates fitted to the empirical pseudo-observations for the full sample period. The number of observations for the empirical copulas amounts to 8284 for the pre-Brexit period and 9133 for the full series and the simulations.

Table 22: Estimates and tail dependence coefficients for the fitted copulas and empirical observations.

| | pre-Brexit | | | | | full sample | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\rho}$ | $\nu$ | $\lambda_l$ | $\lambda_u$ | BIC | $\hat{\rho}$ | $\nu$ | $\lambda_l$ | $\lambda_u$ | BIC |
| Normal | 0.445 | | 0.000 | 0.000 | -1809.832 | 0.444 | | 0.000 | 0.000 | -1993.426 |
| Student t | 0.436 | 3.35 | 0.256 | 0.256 | -2377.183 | 0.438 | 3.36 | 0.256 | 0.256 | -2621.591 |
| Gumbel | 1.401 | | 0.000 | 0.360 | -1915.640 | 1.402 | | 0.000 | 0.361 | -2116.459 |
| Clayton | 0.677 | | 0.359 | 0.000 | -1736.844 | 0.674 | | 0.357 | 0.000 | -1901.032 |
| Empirical | | | 0.286 | 0.385 | | | | 0.274 | 0.358 | |

*Note:* We first calculate the empirical pseudo observations of the pre-Brexit and full log-return series of FTSE 100 and S&P 500 stock indexes according to (23). Then the normal, Student t, Gumbel and Clayton copulas are fitted to the empirical pseudo observations. The upper tail dependence coefficient, $\lambda_u$ is estimated according to (25) and the lower tail dependence coefficient $\lambda_l$ as in (26). For comparison, the upper empirical tail dependence coefficient is estimated according to (27) and the lower as in (28). The first copula parameter is estimated for all four distributions while the Student t copula has a second parameter $\nu$, which denotes the estimated degrees of freedom.

## 3.5 Case Study III: The Dependence Structure within the FTSE 100 index

### 3.5.1 The Precision Matrix of the FTSE 100 Index

Figure 9 shows a graphical lasso representation of the covariance structure of 95 of the stocks in the FTSE 100 index. Evaluating the models according to the RIC, the best fit was achieved using a smoothing parameter of approximately $\lambda = 0.53$. Using the StARS criterium with a threshold of 0.1 resulted in an optimal penalty parameter of $\lambda = 0.47$. The EBIC criterion indicates that choosing the smoothing parameter to $\lambda = 1$ gives the best model for all values of gamma. For this data set, it seems that the EBIC criterion is too conservative, resulting in a completely sparse graph structure. A reasonable conclusion is that setting lambda to approximately $\lambda = 0.5$ result in a good trade-off between sparsity and complexity.

We can see that increasing the imposed structure results in a sparser dependence structure (Figure 10). A sparser dependence structure results in a decreasing likelihood. Consequently, there is a trade-off between sparsity and model accuracy. However, it is challenging to judge the quality of the estimated graphs as the true dependence structure is not known. Although simulation studies can give some insights, real data often behave differently. Different criteria to assess the optimal model complexity gives different answers.

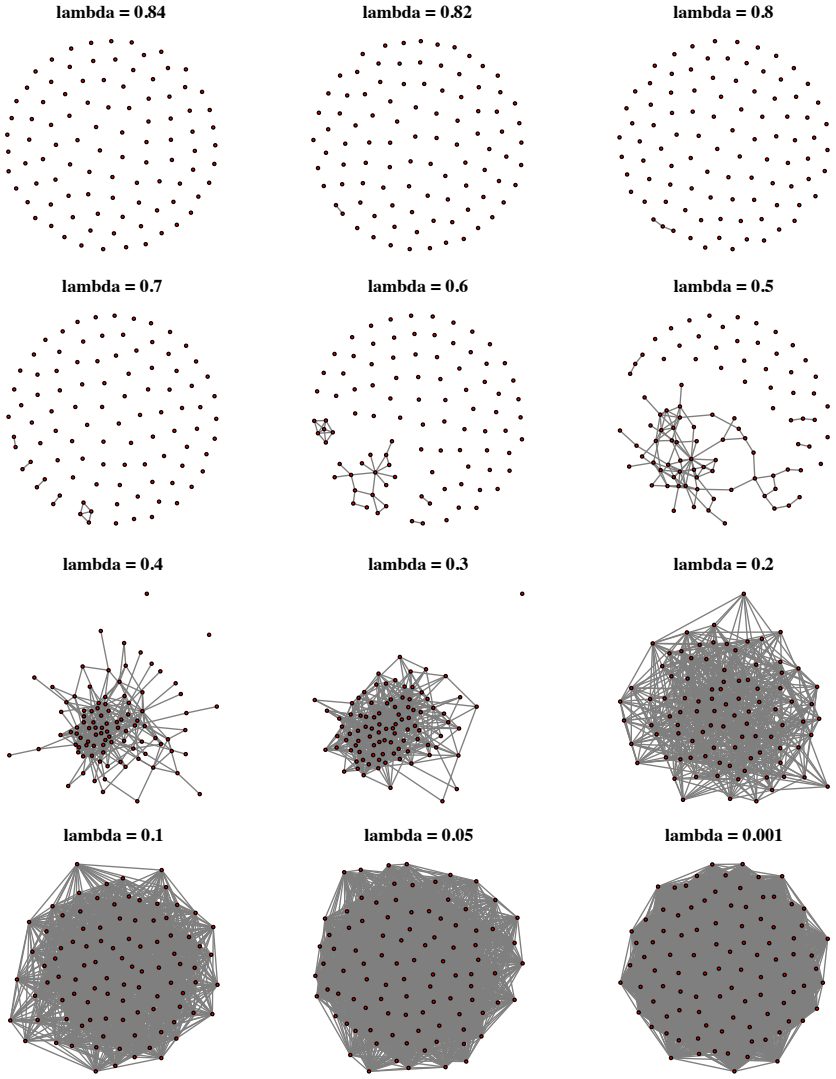



Figure 9: Graphical representation of the dependence structure between the FTSE 100 components extreme losses using 45 monthly block maxima spanning from 2015-09-23 to 2019-05-30. Log-returns for 95 of the FTSE 100 stocks were fitted using an $\ell_1$-regularised nonparanormal graphical model for 12 different values of the smoothing parameter $\lambda$.

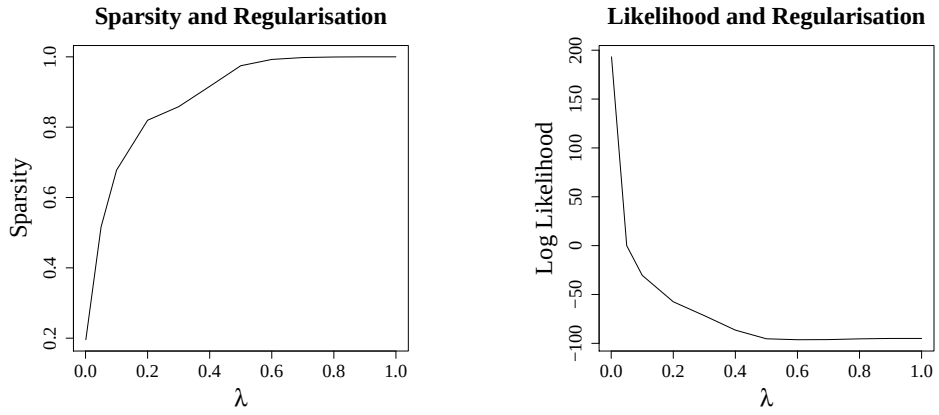



Figure 10: Sparsity $Q$, and log-likelihood as a function of $\lambda$ for 12 nonparanormal graphical models.

### 3.5.2 THE TAIL DEPENDENCE WITHIN THE FTSE 100 INDEX

Figure 11 visualises the inferred undirected graphical network using a penalty of $\lambda = 0.5$ by industry. We can see that the most connected cluster is composed of financials, basic materials, industrials and consumer goods/services. The healthcare companies are conditionally dependent on the financial and utilities sectors and the telecommunications and utilities sectors are dependent on the main cluster only through two companies, GLEN and GSK.

To gain insights on the actual risk in the FTSE 100 index we need to quantify the strength of the dependencies. For this purpose, an eigenvalue decomposition was performed on the covariance structure estimated in the previous section with $\lambda = 0.5$. For each stock, the nonparanormally transformed block maximums and the principal components were fitted to $t$-copulas. We then estimated the upper coefficient of tail dependence to quantify how strong the extreme tail risk is for each of the FTSE 100 stocks. While there are conditional dependencies between many of the stocks in the graphical model in Figure 11 the strength of these dependencies are not equal.

Figure 12 shows the graphical model with the upper coefficient of tail dependence between the individual stocks and the four first principal components. The node size is proportional to the size of the upper tail dependence coefficient with an absence of a node indicating a coefficient equal to zero. Only a few of the dependencies are strong enough to be considered a substantial risk. Most of the financial stocks that are highly connected have low tail dependence coefficients. The third principal component appears to capture the strongest tail dependencies. For estimates, see Appendix C. It is clear that the linear combination of extreme losses in the direction of the largest variance is not the most informative. The higher order of the components seems to capture the more extreme dependencies of the network.

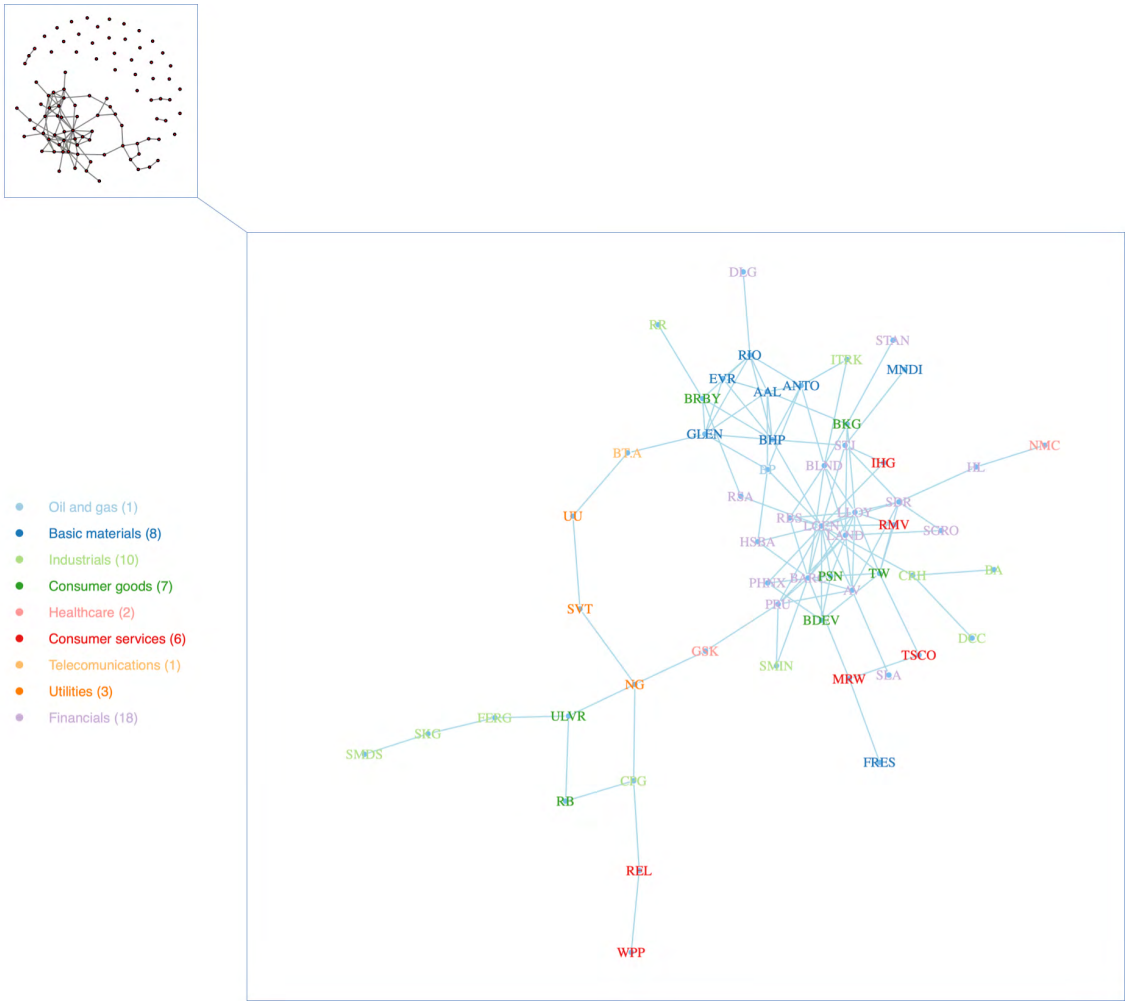



Figure 11: Inferred network of the dependence structure between 56 of the FTSE 100 stocks extreme losses by industry using 45 monthly block maxima spanning from 2015-09-23 to 2019-05-30. Nodes are labelled with the stocks tickers and color-coded according to the Industry Classification Benchmark standard. Log-returns for the FTSE 100 stocks were fitted using an $\ell_1$-regularised nonparanormal graphical model with smoothing parameter $\lambda = 0.5$.

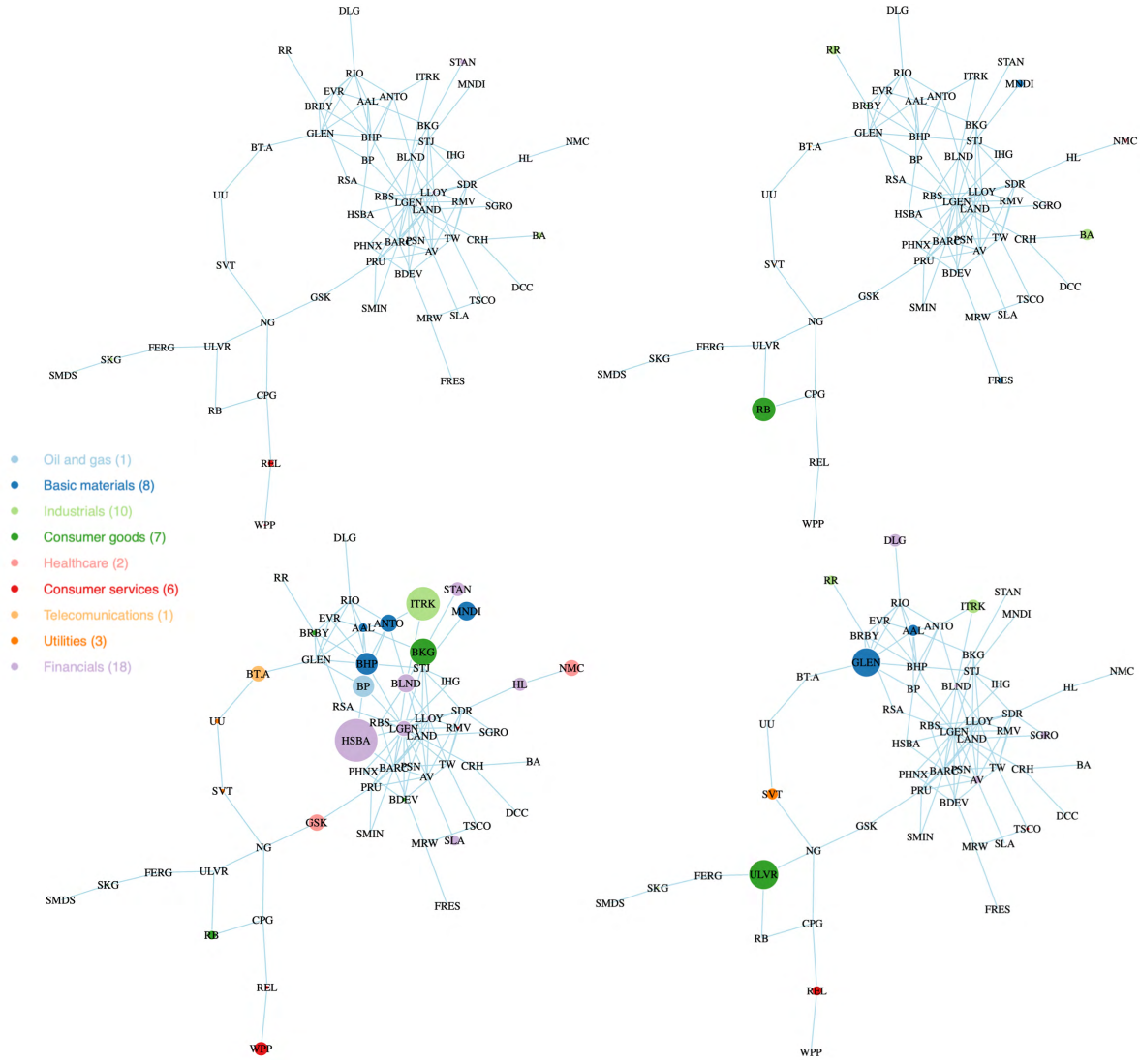



Figure 12: Graphical representation of the tail dependence in the FTSE 100 extreme losses. An eigenvalue decomposition was performed after which bivariate $t$-copulas were fitted between each stock and the principal components of the covariance structure in Figure 11. The upper row shows component 1 and 2 and the lower row shows component 3 and 4 from left to right. The node size is adjusted for the strength of the upper coefficient of tail dependence.

## 4. Conclusions

In this study, we have reviewed some methods to model multivariate risk. First, we compared some univariate risk models using simulated data and stock returns of the FTSE 100 index.

Using simulated data we found that the selection of block size and threshold requires a great deal of care in order for the models to capture the true underlying tail index. Overall, the extreme value models appeared to be capable of modelling risk, as measured by VaR and ES well, even when the tail index estimates had the wrong sign. Further studies are needed before any definitive conclusions can be given on these contradictory results.

Performing a case study where the British stock market risk pre-Brexit and post-Brexit were compared, we found no substantial difference, although the risk was slightly lower after the Brexit referendum. This highlights the difficulties in modelling rare events in short time windows even in the univariate case. No significant difference in the methods was detected when using real stock market return data.

Secondly, a study of bivariate risk using copulas was conducted using simulated data and stock returns of the FTSE 100 and S&P 500 indexes. The $t$-copula was most capable of modelling the lower tail dependence while the Gumbel extreme value copula fared better in the upper tail. The empirical tail dependencies were slightly lower post-Brexit in both tails. However, the estimated copulas were not able to capture these differences.

Finally, a multivariate analysis of the tail dependencies of the components in the FTSE 100 index was performed. Although the glasso model is a common choice for estimating sparse undirected graphical models, normality assumptions make it problematic for highly skewed data. A non-paranormal transformation was performed on the block maxima of the FTSE 100 index to impose normality. The estimated high-dimensional graphical model ($p > n$) showed that the financial sector is overrepresented in the network both in terms of the number of edges and nodes.

The severity of the dependencies of the extreme losses within the index was then quantified using a combination of copulas, PCA and undirected graphical models. We found that the extreme risk, as measured by the coefficient of tail dependence, was only associated with a few of the stocks in the network. Furthermore, the first component was not as informative as the succeeding three in capturing the extreme tail dependencies of the FTSE 100 index. Further research may focus on adding a temporal dimension to be able to quantify risk in both space and time.

## References

Abad, P., Benito, S., and López, C. (2014). A Comprehensive Review of Value at Risk Methodologies. *The Spanish Review of Financial Economics*, 12(1):15–32.

Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*. Springer, New York.

Anderson, T. W. and Darling, D. A. (1952). Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics*, 23(2):193–212.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Athreya, K. B. (1987). Bootstrap of the Mean in the Infinite Variance Case. *Ann. Statist.*, 15(2):724–731.

Balkema, A. A. and de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability*, 2:792–804.

Basel Committee on Banking Supervision (2004). *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Bank for International Settlements, Basel, Switzerland. https://www.bis.org/publ/bcbs107.pdf.

Basel Committee on Banking Supervision (2010). *Basel III: A global regulatory framework for more resilient banks and banking systems*. Bank for International Settlements, Basel, Switzerland. https://www.bis.org/publ/bcbs189_dec2010.pdf.

Basel Committee on Banking Supervision (2013). *Basel Committee on Banking Supervision. Consultative Document: Fundamental Review of the Trading Book: A Revised Market Risk Framework*. Bank for International Settlements. http://www.bis.org/publ/bcbs265.pdf.

Bickel, P., Gotze, F., and van Zwet, W. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica*, 7:1–31.

Bickel, P. J. and Freedman, D. A. (1981). Some Asymptotic Theory for the Bootstrap. *Ann. Statist.*, 9(6):1196–1217.

Blum, P. and Dacorogna, M. M. (2002). Extreme Moves in Daily Foreign Exchange Rates and Risk Limit Setting. In *Swiss Society of Investment Professionals Conference*.

Bräutigam, M., Dacorogna, M., and Kratz, M. (2018). Predicting Risk with Risk Measures: An Empirical Study. Working paper or preprint. https://hal-essec.archives-ouvertes.fr/hal-01791026.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Cramér, H. (1930). On the mathematical theory of risk. In *Försäkringsaktiebolaget Skandia 1855-1930 II*. Centraltryckeriet, Stockholm.

Cramér, H. (1946). *Mathematical methods of statistics*. Princeton University Press, Princeton.

Csárdi, G. (2020). *igraph: Network Analysis and Visualization*. R package version 1.2.4.1. https://CRAN.R-project.org/package=igraph.

Davison, A. C. and Smith, R. L. (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425.

Diethelm Wuertz, Tobias Setz, Y. C. (2017). *fExtremes: Rmetrics - Modelling Extreme Events in Finance*. R package version 3042.82. https://cran.r-project.org/web/packages/fExtremes/index.html.

Drees, H. and Huang, X. (1998). Best Attainable Rates of Convergence for Estimators of the Stable Tail Dependence Function. *Journal of Multivariate Analysis*, 64(1):25 – 46. http://www.sciencedirect.com/science/article/pii/S0047259X97917085.

Durrett, R. (2005). *Probability: Theory and Examples*. Thomson, Brooks/Cole, 3rd edition.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*, volume 38. Society for Industrial and Applied Mathematics.

Embrechts, P. A., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

Emmer, S., Kratz, M., and Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, 18(2). https://www.risk.net/journal-of-risk/2434913/what-is-the-best-risk-measure-in-practice-a-comparison-of-standard-measures.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of Statistics*, 43(1):276–298.

Ferreira, A., de Haan, L., and Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution. *Statistics*, 37(5):401 – 434.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180—-190.

Foygel, R. and Drton, M. (2010). Extended Bayesian Information Criteria for Gaussian Graphical Models. In Lafferty, J. D., Williams, C. K. I., Shawe-Taylor, J., Zemel, R. S., and Culotta, A., editors, *Advances in Neural Information Processing Systems 23*, pages 604–612. Curran Associates, Inc. http://papers.nips.cc/paper/4087-extended-bayesian-information-criteria-for-gaussian-graphical-models.pdf.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la societe Polonaise de Mathematique*, 6:93–116.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

FTSE Russell (2019). *Industry Classification Benchmark*. http://www.ftserussell.com/data/industry-classification-benchmark-icb. Accessed: 2020-05-05.

Gilleland, E. (2019). *distillery: Method Functions for Confidence Intervals and to Distill Information from an Object*. R package version 1.0-6, https://cran.r-project.org/web/packages/distillery/index.html.

Gnedenko, B. (1943). Sur la distribution du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3):423–453.

Hammersley, J. M. and Clifford, P. (1971). Markov fields on finite graphs and lattices. *Unpublished manuscript*, 46.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, 3(5):1163–1174. http://www.jstor.org/stable/2958370.

Hofert, M., Mächler, M., and McNeil, A. J. (2013). Archimedean copulas in high dimensions: Estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique*, 154(1):25–63.

Jenkinson, A. F. (1955). The Frequency Distribution of the Annual Maximum (or Minimum) of Meteorological Elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171.

Jiang, H., Fei, X., Liu, H., Roeder, K., Lafferty, J., Wasserman, L., Li, X., and Zhao, T. (2019). *huge: High-Dimensional Undirected Graph Estimation*. R package version 1.3.4. https://CRAN.R-project.org/package=huge.

Kratz, M. (2019). Introduction to Extreme Value Theory: Applications to Risk Analysis and Management. In *2017 MATRIX Annals*, pages 591–636. Springer, Cham. https://doi.org/10.1007/978-3-030-04161-8.

Laio, F. (2004). Cramer–von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research*, 40(9). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004WR003204.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer-Vlg, New York.

Liu, H., Lafferty, J., and Wasserman, L. (2009). The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs. *Journal of Machine Learning Research*, 10(10):2295–2328. http://jmlr.csail.mit.edu/papers/volume10/liu09a/liu09a.pdf.

Liu, H., Roeder, K., and Wasserman, L. (2010). Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models. In Lafferty, J. D., Williams, C. K. I., Shawe-Taylor, J., Zemel, R. S., and Culotta, A., editors, *Advances in Neural Information Processing Systems 23*, pages 1432–1440. Curran Associates, Inc. http://papers.nips.cc/paper/3966-stability-approach-to-regularization-selection-stars-for-high-dimensional-graphical-models.pdf.

Longerstaey, J. and Spencer, M. (1996). RiskMetrics$^{TM}$ — Technical Document. Technical Report 4$^{th}$ edition, Morgan Guaranty Trust Company of New York, New York. https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

Lysen, S. (2009). Permuted Inclusion Criterion: A Variable Selection Technique. *Publicly Accessible Penn Dissertations*, page 28. http://repository.upenn.edu/edissertations/28.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1):77–91.

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

Mersmann, O., Trautmann, H., Steuer, D., and Bornkamp, B. (2018). *truncnorm: Truncated Normal Distribution Continuous Distribution Functions*. R package version 1.0-8.

Millard, S. P. (2018). *EnvStats: An R Package for Environmental Statistics*. R package version 2.3.1. https://CRAN.R-project.org/package=EnvStats.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, Massachusetts.

Nagler, T. (2019). *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.3.0. https://CRAN.R-project.org/package=VineCopula.

Northrop, P. J. and Attalides, N. (2019). *threshr: Threshold Selection and Uncertainty for Extreme Value Analysis*. R package version 1.0.1. https://CRAN.R-project.org/package=threshr.

Northrop, P. J., Attalides, N., and Jonathan, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(1):93–120. https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12159.

Pfaff, B. (2016). *FRAPO: Financial Risk Modelling and Portfolio Optimisation with R*. R package version 0.4-1. https://CRAN.R-project.org/package=FRAPO.

Pickands, J. (1967). Maxima of Stationary Gaussian Processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 7(3):190–223.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1):119–131.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rappoport, P. (1993). A new approach: Average Shortfall. *JP Morgan Fixed Income Research Technical Document*.

Rozeff, M. S. and Kinney Jr, W. R. (1976). Capital market seasonality: The case of stock returns. *Journal of Financial Economics*, 3(4):379–402.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC, New York.

Saeb, A. (2018). *gnFit: Goodness of Fit Test for Continuous Distribution Functions*. R package version 0.2.0.

Samuelson, P. A. (1965). Rational Theory of Warrant Pricing. *Industrial Management Review*, 6(2):13–31.

Schmidt, R. and Stadtmüller, U. (2006). Non-Parametric Estimation of Tail Dependence. *Scandinavian Journal of Statistics*, 33(2):307–335. http://www.jstor.org/stable/4616927.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464. https://doi.org/10.1214/aos/1176344136.

Sklar, A. (1959). Fonctions de Répartition à N Dimensions Et Leurs Marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Stephens, M. A. (1986). Tests Based on EDF Statistics. In D'Agostino, R. B. and Stephens, M. A., editors, *Goodness-of-Fit-Techniques*. Marcel Dekker, Inc., New York, NY, USA.

Swanepoel, J. W. (1986). A note on proving that the (modified) bootstrap works. *Communications in Statistics - Theory and Methods*, 15(11):3193–3203.

Vialaneix, N. (2015). *An introduction to network inference and mining*. WikiStat. http://www.nathalievialaneix.eu/teaching/network.html.

von Mises, R. (1936). La distribution de la plus grande de $n$ valeurs. *Rev. Math. Union Interbalcanique*, 1:141–160.

## Appendix A.

This section is intended to guide the reader through the mathematical foundations of Theorem 3 and 4 by adding some details often left out in standard textbooks.

### A.1

A remarkable property of a standard Gaussian process $X = \{X_1, \dots, X_n\}$ with probability distribution $\mathbb{P}$ and the support $X \in \mathbb{R}$, is that its maximum $M_n$ asymptotically tends to a finite normalising constant $a_n$. We start this section by presenting a proof of Theorem 3 in Section 2.1.1.

**Theorem 3** (Durrett, 2005) *The ratio of the maximum of a standard Gaussian process and its normalising constant tends to one such as*

$$M_n a_n \underset{n\to\infty}{\to} 1 \text{ a.s.}$$

*where* $a_n = (2\log n)^{-1/2}$.

*Proof.* (Durrett, 2005) It can be shown that,

$$(6x)^{-1}e^{-x^2/2} \le \mathbb{P}(X_i \ge x) \le (2x)^{-1}e^{-x^2/2},\ x > 2.$$

Letting $\varepsilon > 0$, then

$$\sum_{n\ge} \mathbb{P}\left(\frac{M_n}{\sqrt{2\log n}} \ge \sqrt{1+\varepsilon}\right) = \sum_{n\ge 3} \mathbb{P}\left(M_n \ge \sqrt{2(1+\varepsilon)\log n}\right)$$

$$\le \sum_{n\ge 3} \frac{1}{2\sqrt{2(1+\varepsilon)\log n}} \exp\left(-2(1+\varepsilon)(\log n)/2\right) \le \sum_{n\ge 3} n^{-(1+\varepsilon)} < \infty,$$

where $\log n > 1$ when $n \le 3$

Using the Borel-Cantelli I Lemma,

$$\mathbb{P}\left(\frac{M_n}{\sqrt{2\log n}} \ge \sqrt{1+\varepsilon} \text{ i.o.}\right) = 0,$$

so that $\mathbb{P}(\limsup_n M_n/\sqrt{2\log n} > \sqrt{1+\varepsilon}) = 0$.

Hence, $\limsup_n M_n/\sqrt{2\log n} \le \sqrt{1+\varepsilon}$ a.s., and letting $\varepsilon \downarrow 0$ shows that $\limsup_n M_n/\sqrt{2\log n} \le 1$ a.s.

Now we consider the opposite direction.

$$\sum_{n\ge 2} \mathbb{P}\left(\frac{M_n}{\sqrt{2\log n}} \ge 1\right) = \sum_{n\ge 2} \mathbb{P}\left(M_n \ge \sqrt{2\log n}\right) \ge \sum_{n\ge 2} \frac{1}{6\sqrt{2\log n}} \exp\left(-(2\log n)/2\right)$$

$$= \frac{1}{6\sqrt{2}} \sum_{n\geq 2} \frac{1}{n\sqrt{\log n}} \geq \frac{1}{6\sqrt{2}} \int_2^\infty \frac{1}{m\sqrt{\log m}}\, dm = \infty.$$

Using Borel-Cantelli II lemma suggest that,

$$\mathbb{P}\left( \frac{X_n}{\sqrt{2\log n}} \geq 1 \text{ i.o.} \right) = 1,$$

such that $\limsup_n (M_n/\sqrt{2\log n}) \geq 1$ a.s. Since $\mathbb{P}\left( \sqrt{1-\varepsilon} < \frac{M_n}{\sqrt{2\log n}} < \sqrt{1+\varepsilon} \text{ i.o.} \right) = 1$, as $\varepsilon \to 0$ we have proven that $M_n/\sqrt{2\log n} \to 1$ a.s. □

### A.2

Let $\{X_i\}$ be an i.i.d. standard normal sequence of r.v.'s with maximum $M_n = \max(X_1, \ldots, X_n)$ and a sequence of real numbers $\{m_n\}$ such that $n(1 - F(m_n)) \to \zeta$ as $n \to \infty$ for $0 \leq \zeta \leq \infty$. Furthermore, we let $\Phi$ denote the standard normal distribution and $\phi$ the standard normal density. The convergence probabilities in Definition 2.1 can be reformulated from $\mathbb{P}({a_n}^{-1}(M_n - b_n) \leq x)$ to $\mathbb{P}(M_n \leq m_n)$ where $m_n(x) = a_n x + b_n$. This example is helpful in showing the intuition of the asymptotic distribution of extremes and can be extended to more complex sequences. We now give two helping lemmas after which a proof of Theorem 4 in Section 2.1.1 is presented.

**Lemma 4** (Adler and Taylor, 2007)**.** *In the limit as $x$ tends to infinity $\bar{\Phi}(m_n)$ is bounded such as*

$$\left( \frac{1}{m_n} - \frac{1}{m_n^3} \right) \phi(m_n) < \bar{\Phi}(m_n) < \frac{1}{m_n} \phi(m_n). \tag{30}$$

*Proof.* The upper bound in eq. (30) is due to the following inequality,

$$\int_{m_n}^{-\infty} e^{-u^2/2} du \leq \int_{m_n}^{-\infty} \frac{u}{m_n} e^{-u^2/2} du = \left[ \frac{d^{-1}}{du^{-1}} \left( \frac{u}{m_n} e^{-u^2/2} \right) \right]_{m_n}^{-\infty}$$
$$= \left[ -m_n^{-1} e^{-u^2/2} \right]_{m_n}^{-\infty} = \left[ \lim_{u \to -\infty} \left( -m_n^{-1} e^{-u^2/2} \right) - \left( -m_n^{-1} e^{-{m_n}^2/2} \right) \right]$$
$$= m_n^{-1} e^{-{m_n}^2/2}.$$

Similarly, one can derive the lower bound by substituting $u \mapsto m_n + v/m_n$ such as,

$$\int_0^{-\infty} e^{-u^2/2} du = \int_0^\infty e^{-({m_n}^2 + 2v + v^2/{m_n}^2)/2} {m_n}^{-1} dv$$
$$= {m_n}^{-1} e^{-{m_n}^2/2} \int_0^\infty e^{-(2v + v^2/{m_n}^2)/2} dv$$
$$= {m_n}^{-1} e^{-{m_n}^2/2} \int_0^\infty e^{-v} e^{-(v^2/{m_n}^2)/2} dv$$

$$\geq {m_n}^{-1} e^{-{m_n}^2/2} \int_0^\infty e^{-v}(1 - v^2/(2{m_n}^2))dv$$

by using the property that $\forall z \geq 0,\ e^{-z} > 1 - z$ letting $z = v^2/2{m_n}^2$. We complete the final step by noting that the remaining integral can be written as a gamma function,

$$\int_0^\infty e^{-v}(1 - v^2/(2{m_n}^2))dv = \int_0^\infty e^{-v}dv - \int_0^\infty e^{-v}v^2/(2{m_n}^2)dv$$

$$= 1 - \frac{1}{2{m_n}^2} \underbrace{\int_0^\infty e^{-v}v^2 dv}_{\Gamma(3)=2} = 1 - \frac{1}{{m_n}^2},$$

which by using the previous expression gives us,

$${m_n}^{-1} e^{-{m_n}^2/2} \left(1 - \frac{1}{{m_n}^2}\right) = \left(\frac{1}{m_n} - \frac{1}{{m_n}^3}\right) e^{-{m_n}^2/2}.$$

□

**Lemma 5.** *The ratio of* $\bar{\Phi}(m_n)$ *and* $\phi(m_n)/m_n$ *tends to one as* $m_n$ *goes to infinity.*

$$\lim_{m_n \to \infty} \frac{\bar{\Phi}(m_n)}{\phi(m_n)/m_n} = 1 \tag{31}$$

*Proof.* Let $f(m_n) = 1 - \Phi(m_n)$ and $g(m_n) = \phi(m_n)/m_n$. Using L'Hôpital's rule we know that

$$\lim_{m_n \to \infty} \frac{f(m_n)}{g(m_n)} = \lim_{m_n \to \infty} \frac{f'(m_n)}{g'(m_n)}$$

$$\frac{d}{dm_n}\left(\frac{1 - \Phi(m_n)}{\phi(m_n)/m_n}\right) = (-1)\frac{(1 - \Phi(m_n)) + m_n(-\phi(m_n))}{m_n\phi(m_n)}$$

$$= (-1)\frac{1 - \Phi(m_n) - m_n\phi(m_n)}{m_n\phi(m_n)} = \frac{\Phi(m_n) - 1}{m_n\phi(m_n)} + \frac{m_n\phi(m_n)}{m_n\phi(m_n)}$$

$$= 1 + \frac{\Phi(m_n) - 1}{m_n\phi(m_n)}$$

Now, the ratio of the derivative for the denominator and numerator in the second term can be written as

$$\frac{\phi(m_n)}{\phi(m_n) - m_n^2\phi(m_n)} = \frac{\phi(m_n)}{\phi(m_n)(1 - m_n^2)} = \frac{1}{1 - m_n^2}.$$

We can clearly see that this ratio tends to zero, and consequently the ratio of $\bar{\Phi}(m_n)/(\phi(m_n)/m_n)$ tends to one, or $\bar{\Phi}(m_n) \sim \phi(m_n)/m_n$. □

**Theorem 4** (Leadbetter et al., 1983; Cramér, 1946) *If $X$ is an i.i.d. (standard) normal sequence of r.v.'s then the asymptotic distribution of $M_n = \max(X_1, \dots, X_n)$ is of type $H_{3,0}$. Specifically*

$$\mathbb{P}(a_n{}^{-1}(M_n - b_n) \leq x) \to \exp(e^{-x}),$$

*where*

$$a_n = (2\log n)^{-1/2}$$

*and*

$$b_n = (2\log n)^{1/2} - \frac{1}{2}(2\log n)^{-1/2}(\log\log n + \log 4\pi).$$

*Proof.* (Leadbetter et al., 1983; Cramér, 1946)
Let $m_n$ be a transformation of $x$ such that $m_n = a_n x + b_n$. The asymptotic development for $\phi(m_n) = \int_{m_n}^{\infty} e^{-\frac{t^2}{2}}\,dt$ is given by $\phi(m_n) \sim e^{-\frac{m_n{}^2}{2}}(\frac{1}{m_n} - \frac{1}{m_n{}^3} + \frac{3}{m_n{}^5} \dots)$. Moving the first term out of the brackets and letting the rest be $\mathcal{O}\left(-\frac{1}{m_n{}^2}\right) = \mathcal{O}\left(\frac{1}{m_n{}^2}\right)$ we have that $\phi(m_n) \sim \frac{1}{m_n} e^{-\frac{m_n{}^2}{2}}\left(1 + \mathcal{O}\left(\frac{1}{m_n{}^2}\right)\right)$ such that

$$\frac{\sqrt{2\pi}\zeta(m_n)}{n} = \frac{1}{m_n} e^{-\frac{m_n{}^2}{2}}\left(1 + \mathcal{O}\left(\frac{1}{m_n{}^2}\right)\right).$$

The probability of obtaining a sample value larger than $m_n$ is

$$\zeta(m_n) = n\,(1 - \Phi(m_n)) = n\bar{\Phi}(m_n) = n(2\pi)^{-1/2} \int_{m_n}^{\infty} e^{-\frac{t^2}{2}}\,dt. \tag{32}$$

Letting $\zeta(m_n) = e^{-x}$ in (32) we get that $\bar{\Phi}(m_n) = (1/n)e^{-x}$.

We now use the upper bound of the Mill's ratio derived in Lemma 4 where $\bar{\Phi}(m_n) \sim \phi(m_n)/m_n$ letting

$$\frac{\bar{\Phi}(m_n)}{\phi(m_n)/m_n} \sim (1/n)e^{-x} m_n/\phi(m_n),$$

As was shown in Lemma 5 this ratio tends to one,

$$(1/n)e^{-x} m_n/\phi(m_n) \underset{n\to\infty}{\to} 1. \tag{33}$$

Taking the logarithm of the expression in eq. (33) we see that $-\log n - x + \log m_n - \log\phi(m_n) \to \log 1$ or equivalently,

$$-\log n - x + \log m_n + \frac{1}{2}\log 2\pi + \frac{m_n{}^2}{2} \underset{n\to\infty}{\to} 0. \tag{34}$$

Dividing each side by $\log n$ we get the expression,

$$-1 - \frac{x}{\log n} + \frac{\log m_n}{\log n} + \frac{\log 2\pi}{2\log n} + \frac{m_n{}^2}{2\log n} \underset{n\to\infty}{\to} 0.$$

We note the property that $({m_n}^2/2\log n) \to 1$ as $n$ tends to infinity shown in Theorem 3. Consequently, as $n$ tends to infinity, $2\log m_n - \log 2 - \log\log n \to 0$, since $(\log {m_n}^2/2\log n) = 2\log m_n - \log 2 - \log\log n \to \log 1 = 0$. It follows that,

$$\log m_n = \frac{1}{2}(\log 2 + \log\log n) + \mathcal{O}(1). \tag{35}$$

Exchanging $\log m_n$ in eq. 34 with the expression in eq. (35) we get that,

$$\frac{{m_n}^2}{2} = x + \log n - \frac{1}{2}\log 4\pi - \frac{1}{2}\log\log n + \mathcal{O}(1),$$

$${m_n}^2 = 2\log n\left(1 - \frac{x - \frac{1}{2}\log 4\pi - \frac{1}{2}\log\log n}{\log n} + \mathcal{O}\left(\frac{1}{\log n}\right)\right),$$

$$m_n = a_n x + b_n + \mathcal{O}\left((\log n)^{-1/2}\right) = a_n x + b_n + \mathcal{O}(a_n).$$

Consequently, if a bounded $\zeta$ exists such as described in Theorem 2 we have that, as $n$ tends to infinity $\mathbb{P}(M_n \le m_n) \to \exp(-e^{-x})$. Then, $\mathbb{P}(M_n \le a_n x + b_n + \mathcal{O}(a_n)) \to \exp(-e^{-x})$ as $n$ tends to infinity or equivalently,

$$\mathbb{P}(a_n^{-1}(M_n - b_n) + \mathcal{O}(1) \le x) \underset{n\to\infty}{\to} \exp(-e^{-x}).$$

□

## Appendix B.

R SCRIPT – 'EXTREME RISK'

### Description

Calculates extreme value parameters and risk estimates from the GEV and GP distributions with confidence intervals in a single function. Two choices of confidence intervals are available, asymptotic normal intervals using the delta method for the risk estimates and percentile bootstrap intervals. The parameters are estimated using either maximum likelihood, probability-weighted moment methods or hills for the GPD. This script was created using the packages *fExtremes* for the extreme value model fit and *distillery* for bootstrap intervals (Diethelm Wuertz, 2017; Gilleland, 2019). The *r* code is available on GitHub (see URL).

### Usage

```
extreme.risk(data, block, threshold, distr, type, method, estimate,
p, CI, alpha, N, m)
```

### Arguments

| | |
|---|---|
| `data` | Numerical vector or a one column data frame of positive values. |
| `block` | Block size, integer. |
| `threshold` | Threshold for the GPD model, numeric. |
| `distr` | The distribution to model. Either `"gev"` for the generalised extreme value distribution or `"gpd"` for the generalised pareto distribution. |
| `type` | The type of parameter estimation, either maximum likelihood estimation `"mle"`, or the probability weighted moment method `"pwm"`. |
| `method` | The method to use for the GP model, the hill estimator (`method = "hills"`) and standard fit using maximum likelihood of probability weighted moment is available (`method = "standard"`). |
| `estimate` | Can be any of the following, `beta`, `xi`, `VaR`, `ES` and `mu` if `distr = "gev"`. |
| `p` | The level at which VaR or ES is computed. |
| `CI` | The type of interval for estimates of the tail index and risk measures. Percentile bootstrap intervals (`CI = "bootstrap"`) and asymptotic normal intervals, (`CI = "delta"`) are supported. |
| `alpha` | The significance level for the confidence intervals. |
| `N` | Number of bootstrap resamples to make. |
| `m` | Resample size for each bootstrap sample. |

### Value

Returns a scalar when the estimate is `beta` or `mu` and a vector of the estimate and lower and upper bounds when the estimate is `xi`, `VaR` or `ES`. The estimates of `VaR` and `ES` are scaled to percentage units for use with financial return data.

### URL

https://github.com/ErikOldehed/extreme.risk

## Appendix C.

Table 23: Coefficient of tail dependence of bivariate $t$-copulas of the individual stocks and index PCs.

| Ticker | PCA1 | PCA2 | PCA3 | PCA4 |
|---|---|---|---|---|
| AAL | 0.00068 | 0.00000 | 0.07674 | 0.09291 |
| ANTO | 0.00000 | 0.00000 | 0.14274 | 0.00267 |
| AV | 0.00000 | 0.00020 | 0.00120 | 0.05999 |
| BA | 0.05543 | 0.09464 | 0.00000 | 0.00856 |
| BARC | 0.00025 | 0.00005 | 0.00584 | 0.00000 |
| BDEV | 0.00000 | 0.00000 | 0.03841 | 0.00000 |
| BKG | 0.00000 | 0.00024 | 0.21810 | 0.00000 |
| BHP | 0.00000 | 0.00000 | 0.17848 | 0.00000 |
| BP | 0.03130 | 0.00000 | 0.17787 | 0.00009 |
| BLND | 0.00056 | 0.00233 | 0.15073 | 0.03831 |
| BT.A | 0.00000 | 0.02421 | 0.12931 | 0.00000 |
| BRBY | 0.00775 | 0.02625 | 0.04732 | 0.00000 |
| CPG | 0.00000 | 0.00000 | 0.00000 | 0.00038 |
| CRH | 0.00000 | 0.00000 | 0.00001 | 0.00000 |
| DCC | 0.00000 | 0.00000 | 0.00001 | 0.00002 |
| DLG | 0.00000 | 0.00927 | 0.00000 | 0.10512 |
| EVR | 0.00000 | 0.00000 | 0.00003 | 0.00000 |
| FERG | 0.00000 | 0.01750 | 0.00009 | 0.00001 |
| FRES | 0.00000 | 0.04771 | 0.00002 | 0.00008 |
| GSK | 0.00000 | 0.00000 | 0.13536 | 0.00338 |
| GLEN | 0.00000 | 0.00000 | 0.00010 | 0.22830 |
| HL | 0.00480 | 0.00000 | 0.11264 | 0.00000 |
| HSBA | 0.00001 | 0.00609 | 0.34247 | 0.00001 |
| IHG | 0.00976 | 0.00229 | 0.00007 | 0.00007 |
| ITRK | 0.00000 | 0.00000 | 0.27078 | 0.11131 |
| LAND | 0.00150 | 0.00015 | 0.00161 | 0.00244 |
| LGEN | 0.01040 | 0.00000 | 0.12511 | 0.00001 |
| LLOY | 0.03549 | 0.00000 | 0.04479 | 0.00000 |
| MNDI | 0.00000 | 0.06603 | 0.15367 | 0.00144 |
| MRW | 0.00000 | 0.00000 | 0.00008 | 0.00707 |
| NG | 0.00000 | 0.00000 | 0.00000 | 0.00014 |
| NMC | 0.00000 | 0.04414 | 0.13260 | 0.00001 |
| PSN | 0.00000 | 0.00110 | 0.00000 | 0.00000 |
| PHNX | 0.00000 | 0.00000 | 0.00000 | 0.00001 |
| PRU | 0.00000 | 0.01895 | 0.00598 | 0.00004 |
| RB | 0.00000 | 0.19134 | 0.07490 | 0.00004 |
| REL | 0.05267 | 0.00593 | 0.03498 | 0.07984 |
| RIO | 0.00000 | 0.01759 | 0.00000 | 0.00173 |
| RMV | 0.00000 | 0.00000 | 0.02541 | 0.00000 |
| RR | 0.00564 | 0.08088 | 0.01580 | 0.08060 |
| RBS | 0.02549 | 0.00000 | 0.00108 | 0.00000 |
| RSA | 0.00000 | 0.00000 | 0.00000 | 0.00002 |
| SDR | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| SGRO | 0.00000 | 0.02918 | 0.00005 | 0.06267 |
| SVT | 0.00000 | 0.01561 | 0.04038 | 0.09392 |
| SMDS | 0.00000 | 0.00059 | 0.00000 | 0.00002 |
| SMIN | 0.00451 | 0.00343 | 0.00056 | 0.00000 |
| SKG | 0.03645 | 0.00000 | 0.02572 | 0.02520 |
| STAN | 0.04593 | 0.00000 | 0.11848 | 0.00000 |
| SLA | 0.00000 | 0.01767 | 0.08550 | 0.00000 |
| STJ | 0.00000 | 0.00000 | 0.01012 | 0.02057 |
| TW | 0.00000 | 0.00002 | 0.00000 | 0.00000 |
| TSCO | 0.00000 | 0.00000 | 0.00000 | 0.03093 |
| ULVR | 0.00000 | 0.00000 | 0.00007 | 0.23484 |
| UU | 0.00000 | 0.00000 | 0.04553 | 0.00023 |
| WPP | 0.02132 | 0.00000 | 0.11400 | 0.00013 |

Table 24: Student t copula parameter estimates between the transformed block maxima and PCAs of the covariance matrix estimated with a nonparanormal undirected graphical model using $\lambda = 0.5$.

| Ticker | PCA1 | PCA2 | PCA3 | PCA4 |
|---|---|---|---|---|
| AAL | -0.888 | -0.426 | 0.217 | -0.218 |
| ANTO | -0.857 | -0.503 | 0.267 | -0.201 |
| AV | -0.561 | -0.810 | 0.418 | -0.051 |
| BA | -0.127 | -0.134 | -0.131 | 0.120 |
| BARC | -0.487 | -0.798 | 0.559 | -0.421 |
| BDEV | -0.442 | -0.604 | -0.494 | -0.464 |
| BKG | -0.582 | -0.667 | 0.297 | -0.579 |
| BHP | -0.931 | -0.431 | 0.069 | -0.067 |
| BP | -0.661 | -0.511 | 0.136 | 0.211 |
| BLND | -0.587 | -0.656 | 0.517 | -0.466 |
| BT.A | -0.516 | -0.264 | 0.160 | -0.082 |
| BRBY | -0.615 | -0.413 | -0.015 | -0.119 |
| CPG | -0.454 | -0.295 | -0.061 | 0.322 |
| CRH | -0.355 | -0.563 | 0.019 | -0.034 |
| DCC | -0.359 | -0.412 | 0.032 | 0.104 |
| DLG | -0.555 | -0.378 | -0.066 | -0.065 |
| EVR | -0.798 | -0.452 | 0.135 | -0.065 |
| FERG | -0.220 | -0.459 | 0.207 | 0.050 |
| FRES | -0.544 | -0.556 | 0.085 | -0.283 |
| GSK | -0.301 | -0.450 | 0.109 | 0.319 |
| GLEN | -0.785 | -0.326 | 0.218 | 0.137 |
| HL | -0.370 | -0.482 | 0.098 | -0.204 |
| HSBA | -0.562 | -0.563 | 0.407 | -0.198 |
| IHG | -0.484 | -0.498 | 0.196 | 0.190 |
| ITRK | -0.485 | -0.297 | 0.402 | 0.022 |
| LAND | -0.484 | -0.658 | 0.444 | -0.293 |
| LGEN | -0.697 | -0.914 | 0.117 | 0.023 |
| LLOY | -0.420 | -0.783 | 0.135 | -0.632 |
| MNDI | -0.456 | -0.444 | 0.169 | 0.003 |
| MRW | -0.525 | -0.598 | 0.202 | -0.159 |
| NG | -0.299 | -0.397 | -0.077 | 0.244 |
| NMC | -0.139 | -0.170 | 0.129 | 0.079 |
| PSN | -0.483 | -0.753 | -0.438 | -0.386 |
| PHNX | -0.448 | -0.685 | -0.210 | 0.055 |
| PRU | -0.490 | -0.692 | 0.561 | 0.152 |
| RB | -0.060 | 0.062 | 0.109 | 0.150 |
| REL | -0.316 | -0.439 | -0.186 | 0.135 |
| RIO | -0.522 | -0.621 | -0.065 | -0.225 |
| RMV | -0.883 | -0.367 | 0.137 | -0.098 |
| RR | -0.442 | -0.323 | 0.283 | -0.248 |
| RBS | -0.502 | -0.682 | 0.409 | -0.564 |
| RSA | -0.389 | -0.482 | 0.004 | 0.110 |
| SDR | -0.491 | -0.740 | -0.029 | -0.056 |
| SGRO | -0.384 | -0.495 | 0.170 | -0.083 |
| SVT | -0.185 | -0.159 | -0.087 | 0.230 |
| SMDS | -0.277 | -0.381 | -0.045 | 0.086 |
| SMIN | -0.530 | -0.506 | 0.353 | -0.075 |
| SKG | -0.292 | -0.339 | 0.102 | -0.054 |
| STAN | -0.567 | -0.460 | 0.291 | -0.268 |
| SLA | -0.409 | -0.646 | 0.157 | -0.117 |
| STJ | -0.648 | -0.681 | 0.074 | 0.205 |
| TW | -0.449 | -0.773 | -0.385 | -0.465 |
| TSCO | -0.281 | -0.457 | -0.171 | -0.189 |
| ULVR | -0.280 | -0.351 | 0.191 | 0.254 |
| UU | -0.208 | -0.182 | -0.016 | 0.282 |
| WPP | -0.073 | -0.141 | -0.082 | 0.236 |

Table 25: Student t copula degrees of freedom estimates between the transformed block maxima and PCAs of the covariance matrix estimated with a nonparanormal undirected graphical model using $\lambda = 0.5$.

| | PCA1 | PCA2 | PCA3 | PCA4 |
|---|---|---|---|---|
| AAL | 3.37 | 30.00 | 5.77 | 2.43 |
| ANTO | 30.00 | 30.00 | 4.16 | 9.28 |
| AV | 30.00 | 5.28 | 30.00 | 4.19 |
| BA | 3.86 | 2.77 | 30.00 | 11.39 |
| BARC | 9.29 | 6.54 | 30.00 | 30.00 |
| BDEV | 30.00 | 30.00 | 2.53 | 30.00 |
| BKG | 30.00 | 6.93 | 2.95 | 30.00 |
| BHP | 13.14 | 30.00 | 2.33 | 30.00 |
| BP | 2.00 | 15.28 | 2.65 | 30.00 |
| BLND | 6.97 | 4.67 | 6.94 | 2.66 |
| BT.A | 30.00 | 4.52 | 3.67 | 30.00 |
| BRBY | 3.71 | 3.45 | 5.00 | 30.00 |
| CPG | 30.00 | 30.00 | 30.00 | 30.00 |
| CRH | 30.00 | 17.78 | 30.00 | 30.00 |
| DCC | 30.00 | 30.00 | 30.00 | 30.00 |
| DLG | 30.00 | 5.23 | 30.00 | 2.89 |
| EVR | 30.00 | 30.00 | 30.00 | 30.00 |
| FERG | 30.00 | 3.73 | 30.00 | 30.00 |
| FRES | 30.00 | 2.00 | 30.00 | 14.59 |
| GSK | 30.00 | 30.00 | 3.23 | 20.12 |
| GLEN | 11.23 | 30.00 | 30.00 | 2.00 |
| HL | 6.32 | 30.00 | 3.64 | 30.00 |
| HSBA | 12.51 | 4.36 | 2.00 | 20.79 |
| IHG | 4.35 | 6.16 | 30.00 | 30.00 |
| ITRK | 30.00 | 30.00 | 2.86 | 3.21 |
| LAND | 6.87 | 7.56 | 30.00 | 8.28 |
| LGEN | 2.87 | 30.00 | 3.49 | 30.00 |
| LLOY | 2.99 | 30.00 | 6.57 | 30.00 |
| MNDI | 30.00 | 2.04 | 3.24 | 14.19 |
| MRW | 30.00 | 30.00 | 30.00 | 7.83 |
| NG | 30.00 | 30.00 | 30.00 | 30.00 |
| NMC | 30.00 | 4.05 | 3.40 | 30.00 |
| PSN | 30.00 | 4.49 | 30.00 | 30.00 |
| PHNX | 30.00 | 30.00 | 30.00 | 30.00 |
| PRU | 30.00 | 2.33 | 30.00 | 30.00 |
| RB | 30.00 | 2.13 | 4.83 | 30.00 |
| REL | 2.92 | 5.38 | 4.39 | 4.87 |
| RIO | 30.00 | 2.79 | 30.00 | 9.81 |
| RMV | 30.00 | 30.00 | 8.30 | 30.00 |
| RR | 5.43 | 2.23 | 12.60 | 2.54 |
| RBS | 3.00 | 30.00 | 30.00 | 30.00 |
| RSA | 30.00 | 30.00 | 30.00 | 30.00 |
| SDR | 30.00 | 30.00 | 30.00 | 30.00 |
| SGRO | 30.00 | 2.86 | 30.00 | 3.89 |
| SVT | 30.00 | 6.19 | 4.80 | 5.23 |
| SMDS | 30.00 | 9.52 | 30.00 | 30.00 |
| SMIN | 4.99 | 5.55 | 30.00 | 30.00 |
| SKG | 3.64 | 30.00 | 7.81 | 6.14 |
| STAN | 2.00 | 30.00 | 5.03 | 30.00 |
| SLA | 30.00 | 2.65 | 4.86 | 30.00 |
| STJ | 22.81 | 30.00 | 10.09 | 10.05 |
| TW | 30.00 | 8.00 | 30.00 | 30.00 |
| TSCO | 30.00 | 30.00 | 30.00 | 4.60 |
| ULVR | 30.00 | 30.00 | 30.00 | 2.46 |
| UU | 30.00 | 30.00 | 5.09 | 30.00 |
| WPP | 6.34 | 30.00 | 2.64 | 30.00 |