EXPLORING THE EMPIRICAL APPLICATION OF MULTIVARIATE PEAK OVER THRESHOLD MODELS TO RISK MANAGEMENT

BACHELOR THESIS | INTERNATIONAL BACHELOR ECONOMETRICS AND OPERATIONS RESEARCH

Author: Felipe Benítez Schlapp
Student number: 445134
Supervisor: Jochem Oorschot, MSc.
Second Assessor: Dr. Rutger-Jan Lange

Date: 02/07/2020

Abstract

The promising performance of univariate peak over threshold models in risk management applications has been well documented in the literature, for example by Bilandi and Kudła (2016) and Marinelli et al. (2007). However, the same cannot be said for multivariate peak over threshold models. In order to contribute to this gap in the literature, this study explores the usage of the multivariate Generalized Pareto Distribution (GPD) model proposed in Kiriliouk et al. (2019) and the Hill method applied to a polar transformation of the multivariate data, as proposed in Resnick (2007), in the estimation of Value-at-Risk (VaR) and Expected Shortfall (ES). These risk measures are estimated for the daily losses of an equally weighted portfolio consisting of Amazon, Microsoft, J.P. Morgan and Johnson & Johnson, over the 2000-2020 period. The VaR estimates are evaluated using the Duration test for independence proposed in Christoffersen and Pelletier (2004), as well as the Dynamic Quantile test proposed in Engle and Manganelli (2004). The ES estimates are evaluated using the test proposed in McNeil and Frey (2000). The conclusions show that both methods seem to significantly overestimate VaR and ES. Hence, the application of these methods should be explored further, especially in combination with GARCH models which can account for conditional heteroskedasticity.

ERASMUS UNIVERSITEIT ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

Contents

1 Introduction

The fact that returns of financial assets exhibit heavy-tails is a well established stylized-fact that can be traced to Mandelbrot (1963). Heavy-tails refers to the fact that the statistical properties of returns can be thought of as being driven by large periodic shocks. This is opposite to conventional statistics which is focused on processes driven by averaging effects. As such, heavy-tailed distributions possess their own field of study, Extreme Value Theory (EVT), which offers an extensive literature documenting the properties and estimation methods available for such processes. Nowadays, EVT plays an important role in the field of portfolio risk management, which is focused on finding appropriate criteria to evaluate and handle the risks stemming from portfolios of financial assets. Its importance can be attributed to the fact that risk can be understood as the portfolios' exposure to the threat of large losses, and in evaluating large values one falls in the realm of EVT. However, this perception of risk and the application of EVT to portfolio risk management has only recently become applied in practice, as noted in Embrechts et al. (1999b). In the beginnings of the portfolio management field, which can be traced to Markowitz (1952), variance was initially introduced as the primary measure of risk. However, since then, many like Price et al. (1982) have uncovered flaws behind the usage of variance as a risk measure. The most trivial flaw is that variance considers not only losses, but also gains as sources of risk, which does not align with current definitions of risk.

Following the view that risk is only related to the downside potential of the portfolio, a plethora of risk measures have been developed to gauge the level of risk in a portfolio. Amongst the most popular ones lies Value-at-Risk (VaR), whose development is widely credited to Till Guldimann during his time at J.P. Morgan in the late 1980's. This measure is essentially a quantile of the distribution of portfolio losses, corresponding to a high level of confidence e.g. 0.95 or 0.99, as denoted in (9). Even though it provides a perception of the maximum loss that could be attained at a given level of confidence, VaR does not provide any information regarding how much a portfolio manager can expect to lose if VaR is exceeded. Expected Shortfall (ES) was proposed by Artzner et al. (1999) as a solution to this problem. ES is an estimate of the loss that can be expected given that VaR has been exceeded, as denoted in (10). It becomes apparent from both of these definitions that the crucial step in obtaining good VaR and ES estimates is to obtain a good estimate for the distribution of portfolio losses. In particular, as our interest lies in high quantiles of the distribution, it is crucial to have a good understanding of the tail of the distribution.

Given that there exist numerous techniques in EVT for estimating the tail of a distribution, an exploratory study comparing the relative performance of various methods can help practitioners looking to navigate the vast literature on this topic. This study seeks to build upon the work done by Bilandi and Kudła (2016), Harmantzis et al. (2006) and Marinelli et al. (2007) in evaluating the performance of univariate EVT methods for the estimation of financial risk measures, by carrying out a similar analysis on a multivariate level. In particular, this study will explore the performance of the Hill method applied on a polar transformation of multivariate data, as proposed in Resnick (2007), against the novel multivariate Generalized Pareto Distribution (GPD) estimation method proposed in Kiriliouk et al. (2019). Even though both methods fall within the semi-parametric Peak-over-Threshold class of models, they differ in their underlying assumptions as will be shown in section 2. Hence, it becomes interesting to evaluate if these differences result in significantly different VaR and ES estimates, and the effects that such differences may have for portfolio managers in practice.

The remainder of this study is structured as follows. In section 2, a theoretical background is provided to introduce the reader to the concept of Peak-over-Threshold estimation and the theory underpinning the two semi-parametric methods to be compared. Section 3 presents the data to be used and explores the properties of their tails. Section 4 provides an outline of the methods used to carry out the study and the results are summarized in Section 5. Lastly, Section 6 contains the conclusions of the study.

2 Theoretical Background

2.1 Extreme Value Theory Models

Extreme Value Theory (EVT) is concerned with understanding the behaviour of maxima, in the same way that conventional statistics is focused on the study of averages and moments. One could argue that EVT starts where conventional statistics stops. This is exemplified by the fact that one of the most general classes of distributions studied in EVT, i.e. heavy-tailed distributions, do not exhibit finite moment generating functions. Formally, Rolski et al. (2009) state that a distribution $F(x) = \mathbb{P}(X < x)$ exhibits heavy-tails if its moment generating function $M_X(t)$ is infinite for all $t > 0$. In particular, this implies that its tails (only the right tail considered in (1)) decay slower than that of any exponential distribution, as:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} dF(x) = \infty \implies \lim_{x\to\infty} \frac{1 - F(x)}{e^{-tx}} = \infty \, , \forall \, t \tag{1}$$

There exist two major strands of models in EVT used to study the tail behaviour of distributions: Block Maxima (BM) and Peak Over Threshold (POT). BM involes splitting the data into n subsamples, finding the maximum in each sample i.e. $M_1, \ldots, M_n$, and subsequently modelling the behaviour of these maxima through a Generalized Extreme Value Distribution (GEV). The properties of this method are derived from the theorems outlined in Fisher and Tippett (1928) and Gnedenko (1943), which are considered the fundamental building blocks of EVT. In particular, they show that the distribution of any properly normalized collection of maxima, such as that obtained through the BM approach, will converge to one of three distributions. These distributions, which are mentioned in decreasing order of tail mass as shown in Figure 1, are the Fréchet, Gumbel and Weibull distributions.

POT on the other hand utilizes the entire sample without subsampling, and is based on the premise that observations exceeding some sufficiently large threshold μ can be considered to be sufficiently tail. Sufficiently close is more formally referred to as the intermediate region close to the tail, and it corresponds to the collection of the largest order statistics in a sample. Note that the standing assumption is that one is never able to observe tail observations, as noted in Resnick (2007). The observations exceeding μ are referred to as exceedances, and the difference between an exceedance and μ is referred to as an excess. The behavior of these excesses can then modelled by a Generalized Pareto Distribution (GPD) as shown by the Pickands III et al. (1975) and Balkema and De Haan (1974). Similarly to GEV, GPD is able to capture all three types of tails. As a matter of fact, these distributions are asymptotically consistent as shown in De Haan and Ferreira (2007). Note that the type of tail is given by the shape parameter ξ for both the GEV and the GPD. In particular, $\xi > 0$ leads to a Fréchet distribution, $\xi < 0$ to a Weibull distribution and $\xi = 0$ to a Gumbel distribution. Nonetheless, this study only considers POT models, as they have been shown to outperform BM models in studies such as Marinelli et al. (2007).

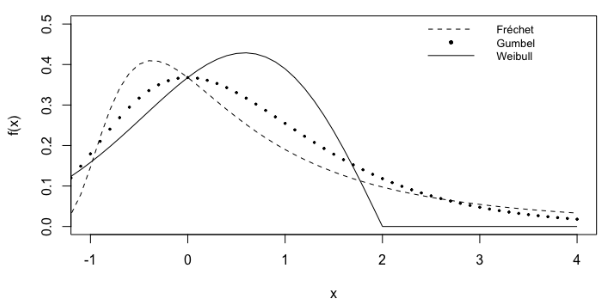

Figure 1: Extreme value densities with equal location and scale.

2.2 Maximum Domain of Attraction

As noted in the Fisher and Tippett (1928) and Gnedenko (1943) theorems, all tails can be shown to be of the extreme value distributions, that is Fréchet, Gumbel or Weibull. This concept is made rigorous through the definition of the maximum domain of attraction (MDA). According to De Haan and Ferreira (2007), a random variable $X \in \mathbb{R} : X \sim F(x)$ is said to be in the MDA of a multivariate non-degenerate[1] extreme value distribution $G(x)$, i.e. $F \in \mathcal{D}(G)$, if it satisfies Von Mises (1936) conditions. That is, if there exist sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that for n i.i.d. copies of X, i.e. $X_1, \dots, X_n$, it holds that:

$$\mathbb{P}\left(\frac{\max(X_1, \dots, X_n) - b_n}{a_n} \leq x\right) \to G(x) \tag{2}$$

Intuitively, this implies that as $n \to \infty$ the distribution of appropriately scaled sample maxima converges to one of the three above-mentioned extreme value distribution.

2.3 Fat-tailed distributions

The notion of fat-tailed distributions is formalized through the concept of regular variation. According to Danielsson (2011), a random variable $X \sim F(x)$ is said to be regularly varying at ∞ with tail index α if:

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = \lim_{t\to\infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha}, \quad \forall x > 0, \quad \alpha > 0 \tag{3}$$

where $\bar{F}(x) = 1 - F(x)$. This is denoted by Resnick (2007) as $F(x) \in RV_\alpha$. If $\alpha = 0$ the function is called slowly varying and denoted $L(x)$. Note that if $F(x) \in RV_\alpha$ then $L(x) = F(x)/x^{-\alpha}$, which implies that $F(x) = x^{-\alpha}L(x)$ for any regulary α-varying function. This representation is more commonly used in the literature. As shown in De Haan and Ferreira (2007), $F(x) \in RV_\alpha$ iff $F(x) \in \mathcal{D}(G_\alpha)$ where G denotes an extreme value distribution with $\alpha = 1/\xi > 0$. This implies that regular variation is equivalent to being in the MDA of a Fréchet distribution.

2.4 Multivariate Generalized Pareto Distribution

In order to make the step from univariate to multivariate POT models, consider $\mathbf{X} \in \mathbb{R}^d : \mathbf{X} = (X^{(1)}, \dots, X^{(d)})'$. Then, according to Rootzén et al. (2018a) if $\mathbf{X}$ is in the MDA of a nondegenerate multivariate max-stable[2] distribution with normalizing sequences $\mathbf{a}_n \in (0, \infty)^d$ and $\mathbf{b}_n \in \mathbb{R}^d$ it follows that:

$$\max\left\{\frac{\mathbf{X} - \mathbf{b}_n}{\mathbf{a}_n}, \eta\right\} \Big| \mathbf{X} \nleq \mathbf{b}_n \xrightarrow{d} \mathbf{Y} \text{ as } n \to \infty \tag{4}$$

where $\mathbf{Y}$ follows a multivariate GP distribution G and η denotes a vector of lower endpoints. The latter is only relevant when dealing with mass on lower-dimensional subspaces according to Kiriliouk et al. (2019). In essence, if μ is a sufficiently high threshold across all margins

[1] Nondegeneracy implies that the density is not condensed at a single point.

[2] Max-stability implies that the distribution of the maximum of n independent copies of the random variable is invariant up to a parameter specification.

$X^{(1)}, \ldots, X^{(d)}$ then $\mathbf{X} - \mu | \mathbf{X} \nleq \mu$ can be approximated by a multivariate GP distribution with scale $\boldsymbol{\sigma}$ and shape $\boldsymbol{\xi}$ parameters equal to the collection of estimated scale and shape parameters of the margins. Moreover, the marginal distributions and the dependence structure must also be estimated. Below a brief outline of the threshold stability property which will later be used to fit the dependence structure is provided. For a more comprehensive overview of this distribution and its properties see Rootzén et al. (2018a).

Threshold stability Consider a multivariate random variable $\mathbf{X} \in \mathbb{R}^d : \mathbf{X} \sim G_{\boldsymbol{\xi},\boldsymbol{\sigma}}$, where G is a multivariate GPD with $\mu = 0$. Then, for $\mathbf{w} \geq \mathbf{0}$ where $G_{\boldsymbol{\xi},\boldsymbol{\sigma}}(\mathbf{w}) < 1$ and $\boldsymbol{\sigma} + \boldsymbol{\xi}\mathbf{w} > \mathbf{0}$, threshold stability implies that:

$$\mathbf{X} - \mathbf{w} \mid \mathbf{X} \nleq \mathbf{w} \sim G_{\boldsymbol{\xi},\boldsymbol{\sigma}+\boldsymbol{\xi}\mathbf{w}} \tag{5}$$

Intuitively, this implies that the conditional distribution of the excesses remains multivariate generalized pareto when moving further into the tail (i.e. when increasing the threshold). In particular, the shape parameter remains the same, whilst the scale parameter changes from $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma} + \boldsymbol{\xi}\mathbf{w}$. Kiriliouk et al. (2019) proposes a test for the appropriateness of a multivariate generalized pareto distribution model using a special case of this property. Consider the threshold $\mathbf{w} = \mathbf{w}_t = \frac{\boldsymbol{\sigma}(t^{\boldsymbol{\xi}}-1)}{\boldsymbol{\xi}}$, which corresponds to the $1 - \frac{1}{t}$ quantile of $\mathbb{P}(\mathbf{X} > \mathbf{x} | \mathbf{X} > 0)$, such that threshold exceedances are Poisson(t) distributed. Then, for any set $A \subset \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x} \nleq 0\}$ it holds that for $t \geq 1$:

$$\mathbb{P}(\mathbf{X} \in \mathbf{w}_t + t^{\boldsymbol{\xi}} A) = \frac{\mathbb{P}(\mathbf{X} \in A)}{t} \tag{6}$$

where $\mathbf{w}_t + t^{\boldsymbol{\xi}} A = \{\mathbf{w}_t + t^{\boldsymbol{\xi}}\mathbf{x} : \mathbf{x} \in A\}$.

2.5 Hill Method

Lets first consider the Hill method in a univariate setting and then make the step towards the multivariate case. The univariate Hill method requires the presence of fat-tails, hence, we assume that the underlying distribution is in the MDA of the Fréchet. As previously mentioned, this is equivalent to assuming that $\bar{F}(x)$ tends to a Pareto distribution. Under this assumption, the Hill method estimates the tail index $\alpha = 1/\xi > 0$ via the Hill estimator proposed in Hill (1975). This estimator corresponds to the maximum likelihood estimator of a Pareto distribution's tail index and is given by:

$$\hat{\alpha}_{k,n} = \sum_{i=1}^{k} \frac{k}{\log X_{(i)} - \log X_{(k+1)}} \tag{7}$$

where $X_{(i)}$ denotes the i^{th} order statistic of a finite sample of observations $X_1, \ldots, X_n$. As this is a finite sample POT estimator, an additional assumption regarding the start of the Pareto behaviour of the tail is needed. This is the role of k within this estimator, i.e. to denote the number of highest order statistics which are considered to be in the intermediate region close to the tail. The choice of an appropriate k is a crucial step in using the Hill estimator due to the high sensitivity of the estimator towards this parameter, as noted by Nguyen and Samorodnitsky (2012) and De Haan et al. (2016). In particular, there exists a bias-variance trade-off, as large values of k imply using more observations which reduces the variance of the estimates, but increases the bias as more observations farther away from the tail are included in the estimation. Conversely, small values of k reduce the bias as the highest order statistics are the closest to the tail, but using a smaller amount of observations leads to a higher variance in the estimates.

Similarly to the univariate case, in order to model a collection of fat-tailed random variables $\mathbf{X} = (X^{(1)}, \ldots, X^{(d)})' \in \mathbb{R}^d$ via the Hill method it is necessary to assume multivariate regular variation (MRV). As noted in Resnick (2007), MRV implies that the joint distribution of $\mathbf{X}$ is in the maximum domain of attraction of the Fréchet distribution. Hence, in this setting we require the joint distribution to exhibit a fat-tail. In particular, because we are now in a multivariate context, it is practical to make use of the polar transformation $(R, \mathbf{S}) = (||\mathbf{X}||_1, \frac{\mathbf{X}}{||\mathbf{X}||_1})$. Note that here the unit-norm $||\mathbf{X}||_1 = \sum_{i=1}^{d} |X^{(i)}|$ is used for simplification, but Resnick (2007) mentions that any norm can be used. This transformation allows for the decomposition of $\mathbf{X}$ into the radial part R which contains information on the joint behaviour of $X^{(1)}, \ldots, X^{(d)}$, and the angular part $\mathbf{S}$ which provides information on the dependence structure of $X^{(1)}, \ldots, X^{(d)}$. Using this transformation, Resnick (2007) states that $\mathbf{X}$ is MRV if there exists a well-defined angular measure Ψ for the dependence structure in the limit, and if R is univariate regularly varying with an appropriate normalizing function $b(t) : \lim_{t\to\infty} b(t) = \infty$. This is summarized in equation (8), where $\xrightarrow{v}$ denotes vague convergence in measure and $c > 0$. For more information on vague convergence in measures see Kallenberg (1983).

$$t\mathbb{P}\left[\left(\frac{R}{b(t)}, \mathbf{S}\right) \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times \Psi \tag{8}$$

As shown by Basrak et al. (2002) under the assumption of equal tail indices across each $X^{(j)}$ for $j \in \{1, \ldots, d\}$, multivariate regular variation of $\mathbf{X}$ with tail index α implies univariate regular variation with tail index α for any linear combination of the components of $\mathbf{X}$. In a financial context, this implies univariate regular variation with tail index α for any portfolio $\mathbf{w}'\mathbf{X}$ where $\mathbf{w} \in \mathbb{R}^d$.

2.6 Value-at-Risk and Expected Shortfall

The risk measures that will be used to explore the benefits of the above-mentioned POT methods, are Value-at-Risk (VaR) and Expected Shortfall (ES). Consider a random variable $X \sim F(x)$ which denotes the loss of a portfolio. $VaR_X(p)$ corresponds to the $(1-p)^{th}$ quantile of $F(x)$ as stated in (9), whilst ES corresponds to the expected value of $X|X > VaR_X(p)$ as shown in (10).

$$\mathbb{P}(X \leq VaR_X(p)) = 1 - p \tag{9}$$

$$ES_X(p) = \mathbb{E}(X|X > VaR(p)) \tag{10}$$

Even though VaR is the most popular risk measure in practice, a lot of research has been done outlining its pitfalls. In particular, many such as Acerbi and Tasche (2002) have reached the conclusion that VaR is not a subadditive risk measure. As mentioned in Frittelli and Gianin (2002), subadditivity refers to the convexity of a risk measure, and it intuitively determines if the measure is capable of accounting for the well-known benefits of diversification. The lack of subadditivity thus implies that the risk from holding two assets in a portfolio is not smaller or equal than the risk of holding the two assets independently. However, recent research by Daníelsson et al. (2013) has shown that VaR is subadditive for high quantiles of fat-tailed distributions, i.e. if regular variation is satisfied. However, regular variation requires the random variable to be in the MDA of the Fréchet distribution, which clearly does not apply to all distributions. Moreover, the universe of distributions for which VaR is subadditive is constrained even further by Mainik and Rüschendorf (2010), who argues that VaR can only be subadditive for finite mean models, which imposes $\alpha > 1$ as an additional requirement. Hence, this motivates to also consider ES which satisfies subadditivity for any type of tail. For more information regarding the properties of risk measures see Frittelli and Gianin (2002).

3 Data

This study considers daily prices from Amazon (AMZN), Microsoft (MSFT), J.P. Morgan (JPM) and Johnson & Johnson (JNJ) from 03/01/2000 to 22/05/2020, retrieved from Yahoo Finance. These companies were chosen because they operate in seemingly different industries, (consumer discretionary, information technology, financials and health care, respectively) which could resemble the portfolio choice of an average diversification-seeking investor. Additionally, this timeframe was chosen because it encompasses various high-risk periods, where the probability of observing values close to the tail is maximized. In particular, this period includes the 2000's dotcom bubble, the 2008-2009 subprime mortgage crisis and the 2020 COVID-19 crisis.

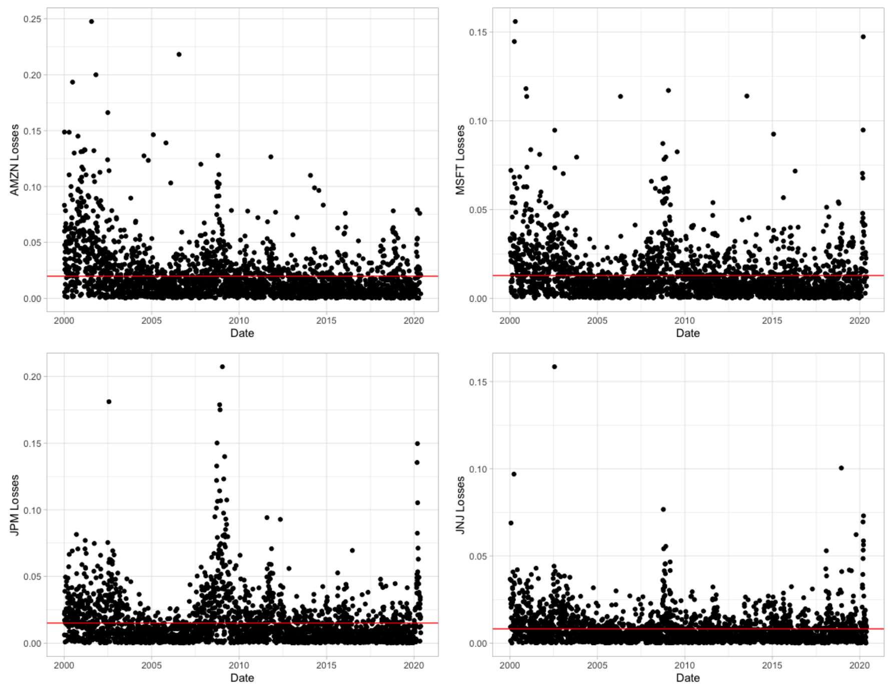

Figure 2: Daily losses of AMZN, MSFT, JPM and JNJ where the red line denotes the mean loss over the full period.

As it can be observed in Figure (2), both AMZN and MSFT exhibit their highest losses around the dotcom bubble. It can also be observed that all firms exhibit significantly larger than average losses during the 2008 subprime mortgage crisis. Lastly, even though all firms exhibit larger than average losses during the COVID-19 crisis, JPM and MSFT exhibit significantly higher relative losses than AMZN and JNJ. Hence, we can observe that even though these firms operate in seemingly unrelated industries, their losses seem to exhibit some degree of dependence during periods of high risk. Additionally, we can clearly observe that some periods exhibit significantly higher volatility than others, which is formally referred to as volatility clustering. This phenomenon could have a negative effect on the performance of the methods considered in this study, as they are unconditional models. Hence, unlike GARCH models, they fail to account for conditional heteroskedasticity (i.e. volatility clustering).

3.1 Preliminary Analyses

Even though it has been well-documented that financial returns exhibit heavy-tails, it is sensible to evaluate the appropriateness of this assumption previous to applying EVT methods. In particular, several visualization techniques which evaluate necessary (but not sufficient) conditions for heavy-tailed behaviour are considered in the remainder of this section.

Firstly, as noted in Embrechts et al. (1999a) QQ-plots of univariate losses against the exponential distribution can offer evidence of fat-tails in case of a concave plot, or thin tails in case of a convex plot. As can be observed in the QQ-plots for the full sample in Figure (3), all assets seem to exhibit relatively concave plots especially towards the tail of the plot (i.e. highest values of x), which supports the assumption of fat-tailed behaviour. This conclusion is supported by Zipf and MS plots which can be found in Appendix A.

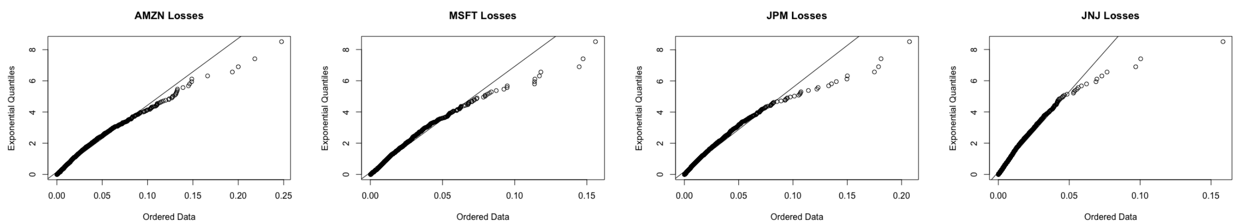

Figure 3: QQ-plots against the exponential distribution for the losses of AMZN, MSFT, JPM and JNJ for the period 03/01/2000 – 22/05/2020.

4 Methodology

Before providing an outline of how the models will be fitted and evaluated, it should be noted that only the first half of the sample, i.e. $\{0,\dots,\tau\}$, is used to derive the model specification. Note that observation 0 corresponds to the observation of 03/01/2000, and τ to that of 15/03/2010. The second half, i.e. $\{\tau,\dots,T\}$ where T corresponds to the observation of 21/05/2020, will be used for the evaluation of the estimated risk measures. Moreover, the parameter estimation is repeated on a daily basis throughout the evaluation period, using an expanding window. The choice of an expanding window is motivated by the fact that observations considered to be in the intermediate region close to the tail will be quite scarce, and using a moving window risks having a window with no observations belonging to it. Lastly, following Kiriliouk et al. (2019) the models are fit on the relative losses of the data. For the j^{th} asset, the relative loss at time t is given by $X_t^{(j)} = \frac{P_{t-1}^{(j)} - P_t^{(j)}}{P_{t-1}^{(j)}}$ where $P_{t-1}^{(j)}$ denotes the closing price of the asset at time (i.e. day) t. Note that relative losses are only defined for the period $\{1,\dots,\tau\}$, as the loss for $t = 0$ is undefined.

4.1 Fitting the models

4.1.1 Multivariate Generalized Pareto Distribution

Following Kiriliouk et al. (2019), the multivariate GP density will be fit to the data using the generators **U** and **T**, outlined in Rootzén et al. (2018b). These generators are simply means of fitting data to a multivariate GP distribution under different sets of assumptions. In particular,

U makes stricter assumptions on the data, which are less likely to hold but can be benefitial. For more information on the generators themselves see Rootzén et al. (2018b). Both of these methods require the data to be in standard form, that is, they can only be applied on a random variable $\mathbf{X}_0 \in \mathbb{R}^d$ with parameters $\boldsymbol{\sigma} = \mathbf{1} \in \mathbb{R}^d$ and $\boldsymbol{\gamma} = \mathbf{0} \in \mathbb{R}^d$. To this end, the empirical probability integral transform is applied to transform the relative losses into uniform scale. Subsequently, the empirically uniform-scaled data is transformed into the standardized Pareto scale using the transformation outlined in Kiriliouk et al. (2019): $X_0 = \frac{1}{1-U} \sim \text{Pareto}(x_m = 1, \alpha = 1)$ for U a uniformly distributed random variable. Note that Kiriliouk et al. (2019) also uses the standard Pareto scale in their applications even though they derive the theory based on the standard Exponential distribution. However, given that the standard Pareto scale is simply the exponential of the standard Exponential scale, the results are argued to be equivalent. Lastly, the multivariate GP density for the observed scale given in (12), is derived using the density for the standardized scale $h(\cdot; \mathbf{1}, \mathbf{0})$ and the data after being transformed back into the observed scale using (11).

$$\mathbf{X} \stackrel{\text{d}}{=} \frac{\boldsymbol{\sigma}(e^{\boldsymbol{\gamma}\mathbf{X}_0} - 1)}{\boldsymbol{\gamma}} \tag{11}$$

$$h(\boldsymbol{x}; \boldsymbol{\sigma}, \boldsymbol{\gamma}) = h\left(\frac{1}{\boldsymbol{\gamma}} \log(1 + \boldsymbol{\gamma}\boldsymbol{x}/\boldsymbol{\sigma}); \mathbf{1}, \mathbf{0}\right) \prod_{j=1}^{d} \frac{1}{\sigma_j + \gamma_j x_j} \tag{12}$$

where h is a density from **U** or **T**.

In order to fit (12) to the data, censored maximum likelihood is applied. This is done because only those observations considered to be in the intermediate region close to the tail are relevant for the estimation. Following Kiriliouk et al. (2019), we determine the threshold delimiting the start of the intermediate region close to the tail by using the stability property of multivariate GP distributions, together with the small-sample equivalent of the following measure of asymptotic dependence:

$$\chi_{1,\ldots,d}(q) = \frac{\mathbb{P}(F_1(x^{(1)}) > q, \ldots, F_d(x^{(d)}) > q)}{1-q}, \; q \in (0,1) \tag{13}$$

where F_j denotes the marginal loss distribution and $x^{(j)}$ the observed loss for asset $j \in \{1, \ldots, d\}$. The threshold stability property of multivariate GP distributions implies that $\chi_{1,\ldots,d}(q)$ should be constant for sufficiently high q. Hence, choosing $\boldsymbol{u} = (F_1^{-1}(q^*), \ldots, F_d^{-1}(q^*))$ where $q^* = inf\{0 < \tilde{q} < 1 : \chi_{1,\ldots,d}(q) = \chi, \; \forall q > \tilde{q}\}$ should provide an adequate threshold for the dependence structure. The appropriateness of using this as a threshold for the collection of marginals is evaluated by means of Chi-plots outlined below. After choosing an appropriate threshold $\boldsymbol{u}$ which denotes the start of the intermediate region close to the tail, the following likelihood is maximized over the fitting period $\{1,\dots,\tau\}$ to estimate the parameters of the multivariate GP distribution:

$$L(\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\gamma}) = \prod_{t=1}^{\tau} h\left(\boldsymbol{x}_t^{(D\setminus C_t)} - \boldsymbol{u}^{(D\setminus C_t)};\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\gamma}\right)$$

where $D = \{1,\dots,d\}$, $C_t = \{d \in D : x_t^{(d)} < u^{(d)}\}$ denotes the collection of assets censored at time t, $\boldsymbol{x}_t^{(D\setminus C_t)}$ and $\boldsymbol{u}^{(D\setminus C_t)}$ denote the collections of uncensored assets and thresholds for time t, and θ denotes the parameters assumed for the generator.

As it was mentioned in section 2, the multivariate GP distribution is able to capture all three types of tails. In order to find the best fitting tail, different classes of multivariate GP densities are evaluated, namely, Gumbel T, Gumbel U, Weibull, Reverse Exponential T and Reverse Exponential U. The one with the highest Akaike information criterium (AIC) is chosen as the appropriate model. For more information on these densities see Kiriliouk et al. (2019). In order to evaluate the appropriateness of more parsimonious models, likelihood ratio tests are used within the chosen class. For example the presence of a common shape parameter across all assets is evaluated to reduce the estimation from 4 shape parameters to 1. Lastly, in order to evaluate the goodness of fit, Kiriliouk et al. (2019) proposed examining QQ-plots for the marginal univariate GP distributions, and plots with estimated values of $\chi_{ij}(q)$ s.t. $i \neq j \in 1,\dots,d$ for $q \in [0.5, 1)$.

Lastly, having estimated the multivariate GP distribution $\mathrm{G}(\boldsymbol{\sigma},\boldsymbol{\xi})$ for a collection of translated assets losses $\mathbf{X} - \boldsymbol{u} \in \mathbb{R}^d$, the distribution for any portfolio consisting of these assets can be easily derived under the assumption of a common shape parameter $\xi \in \mathbb{R}$. This follows from the fact that, for a collection of weights $\mathbf{w} = (w^{(1)},\dots,w^{(d)})'$ the portfolio losses are given by $Y = \sum_{j=1}^{d} w_j \left(X^{(j)} - u^{(j)}\right)$. In particular, Kiriliouk et al. (2019) states that:

$$Y \middle| Y > 0 \sim \mathrm{GP}\left(\sum_{j=1}^{d} w^{(j)}\sigma^{(j)}, \xi\right) \tag{14}$$

where $w^{(j)} = 1/d$, $\forall j \in \{1,\dots,d\}$ as we consider an equally weighted portfolio. Hence, the shape parameter ξ remains the same, whilst the scale parameter becomes a weighted sums of the scale parameters of the marginals.

4.1.2 Hill Method

As mentioned in section 2.5, in order to apply the Hill method in a multivariate setting, i.e. to the losses $\mathbf{X} = (X^{(1)},\dots,X^{(d)})' \in \mathbb{R}^d$, it is necessary to assume multivariate regular variation (MRV). An additional relevant assumption to make in order to model the equally weighted (EW) portfolio of assets, is that $X^{(1)}, \dots, X^{(d)}$ exhibit the same tail index $\alpha > 0$. Mainik and Rüschendorf (2010) mention that the opposite case, i.e. allowing for different tail indices, can be considered trivial, as the asset with the lowest tail index will tend to dominate the rest. This assumption will be evaluated by eyeballing the bias-variance trade-off in the Hill plots of the assets, as done in De Haan et al. (2016). Hill plots correspond to plots of the Hill estimator for different threshold levels.

Even though we assume that all assets $X^{(1)}, \dots, X^{(d)}$ exhibit the same tail index, this is unlikely to be exact in a finite sample setting. Hence, in order to incorporate all assets in the estimation of the common α and make the results for the portfolio α more realistic, we first apply the polar transformation $(R, \mathbf{S}) = (||\mathbf{X}||_1, \frac{\mathbf{X}}{||\mathbf{X}||_1})$ and estimate the tail index of R. This incorporates the information of all assets into the estimation of the tail index. As mentioned in section 2, under these assumptions it suffices to estimate α for R to parametrize the tail of any portfolio constructed using $X^{(1)}, \dots, X^{(d)}$. The Hill estimator for the radial parts is given by:

$$\hat{\alpha}_k := \sum_{i=1}^{k} \frac{k}{\log R_{(i)} - \log R_{(k+1)}} \tag{15}$$

where $R_{(i)}$ denotes the i^{th} order statistic of our finite sample of estimation observations, i.e. $R_1, \dots, R_\tau$, and k denotes the threshold after which the data is assumed to belong to the region close to the tail.

4.2 Estimation of Extreme Risk Measures

4.2.1 Multivariate Generalized Pareto Distribution

In order to estimate VaR and ES for the multivariate GP method, the estimators outlined in Kiriliouk et al. (2019) are used. In particular, the estimators are given by plugging the sample estimates of the distribution's parameters $\hat{\xi}$ and $\hat{\sigma}$ and the empirical distribution value of $\bar{F}(u)$[3] into:

$$VaR_t^{GPD}(p) = \frac{1}{\xi} \sum_{j=1}^{d} w^{(j)} \sigma^{(j)} \left(\left(\frac{1-p}{\bar{F}(u)} \right)^{\xi} - 1 \right) + \sum_{j=1}^{d} w^{(j)} u^{(j)} \tag{16}$$

$$ES^{GPD}(p) = VaR^{GPD}(p) + \frac{\sum_{j=1}^{d} w^{(j)} \sigma^{(j)} + \xi \left(VaR^{GPD}(p) - \sum_{j=1}^{d} w^{(j)} u^{(j)} \right)}{1-\xi} \tag{17}$$

For the derivation of these estimators see Appendix B.

[3]This is given by $\hat{\bar{F}}(u) = \frac{C_T}{T}$, where C_T denotes the number of observations exceeding u in the sample, and T the total sample size.

4.2.2 Hill Method

Given that the portfolio losses are univariate regularly varying with index α for the Hill method, it suffices to focus on the univariate expression derived by Danielsson (2011) to estimate VaR as shown in (18). Moreover, according to Norton et al. (2019) it suffices to estimate the mean of the Pareto conditional excess distribution, to derive ES for the Hill method as shown in (19).

$$VaR^{Hill}(p) = u\left(\frac{\bar{F}_(u)}{p}\right)^{\frac{1}{\alpha}} \tag{18}$$

$$ES^{Hill}(p) = \mathbb{E}(Y|Y > VaR(p)) = \frac{\alpha VaR^{Hill}(p)}{\alpha - 1} \tag{19}$$

Hence, $VaR^{Hill}(p)$ and $ES^{Hill}(p)$ can be estimated by pluggin the estimated tail index $\hat{\alpha}$ and the empirical distribution value of $\bar{F}(u)$ into the expression above. The derivation of these estimators can be found in Appendix B.

4.3 Evaluation of Extreme Risk Measures

4.3.1 Evaluation of VaR Estimates

In order to evaluate the VaR estimates Berkowitz et al. (2011) found that the Dynamic Quantile (DQ) test for conditional coverage proposed in Engle and Manganelli (2004) as well as duration tests, such as the independence test proposed in christoffersen2004backtesting, offer significantly more power[4] than conventional tests such as those proposed in christoffersen1998evaluating and Kupiec (1995). Given that the models employed in this study are unconditional, meaning that the conditional heteroskedasticity is not being accounted for, independence tests will most likely reject the null hypothesis of correct VaR specification. However, for illustrative purposes both of the above-mentioned tests will be used to evaluate the goodness of the VaR estimates.

The duration test for independence proposed by Christoffersen and Pelletier (2004) is based on the duration of time between two VaR exceedances, i.e. $D_i = t_i - t_{i-1}$ where t_i denotes the time of the i^{th} violation. Under the null hypothesis of correct model specification, this duration should be memoryless (i.e. independent of previous durations) and have a mean of $1/(1-p)$ days. This can be tested via the Weibull distribution, which is memoryless if $b = 1$. This test corresponds to a likelihood ratio test with $H_0 : b = 1$, and it is performed using the implementation in the package

[4]Power refers to the probability of rejecting the null hypothesis when it is indeed false.

Ghalanos (2020). The log-likelihood is given by (20).

$$\begin{aligned}\ln L(D;\Theta) =& C_\tau \ln \bar{F}(D_\tau) + (1-C_1)\ln f(D_1;\Theta) + \sum_{i=\tau+1}^{T-1} \ln\big(f(D_i;\Theta)\big) \\ &+ C_T \ln \bar{F}(D_T) + (1-C_T)\ln f(D_T;\Theta)\end{aligned} \tag{20}$$

where $f(\cdot)$ and $\bar{F}$ correspond to the density and survival function of the Weibull distribution, respectively. Moreover, C_i indicates if a duration is censored. Hence, $C_\tau = 0$ if the observation at time τ (which corresponds to the start of the estimation period) exceeds its VaR, otherwise the first duration is left censored and $C_\tau = 1$. Similarly, $C_T = 0$ if the observation at time T (which corresponds to the end of the estimation period) exceeds its VaR, otherwise the last duration is right censored and $C_T = 1$. For censored observations the density is not informative, as we do not know the exact duration. However, given that we know that for a censored observation the duration was at least D_τ or D_T days, their contribution to the likelihood is given by the corresponding survival function value.

The DQ test proposed by Engle and Manganelli (2004) performs a linear regression of the Hit variable, $\text{Hit}_t(p) = \mathbb{1}(y_t > VaR(p)) - p$, on a constant, lagged Hits and other past information suspected of affecting the Hits. Under conditional coverage (i.e. both independence and unconditional coverage[5]), $\text{Hit}_t(p)$ would be uncorrelated with its lags and have an expected value of zero. Hence, we can test for conditional coverage by applying a Wald test (i.e. joint linear restrictions) to test the validity of a zero mean and lack of autocorrelation with respect to any lags or historical information. This test is performed using the implementation in the package Ardia et al. (2019) and its test statistic for $H_0 : \boldsymbol{\beta} = \mathbf{0}$ is given by:

$$DQ = \hat{\boldsymbol{\beta}}'(\text{Var}(\hat{\boldsymbol{\beta}}))^{-1}\hat{\boldsymbol{\beta}} \xrightarrow{d} \chi^2_g, \text{ as } T \to \infty$$

where $\hat{\boldsymbol{\beta}}$ denotes the OLS estimate of $\boldsymbol{\beta}$, $\text{Var}(\hat{\boldsymbol{\beta}})$ denotes the estimated OLS variance of $\hat{\boldsymbol{\beta}}$ and g denotes the number of regressors considered. Note that the convergence in distribution requires additional assumptions specified in Engle and Manganelli (2004).

4.3.2 Evaluation of ES Estimates

The evaluation of ES estimates is a disputed topic in the literature. On the one hand, there exist many methods for evaluating ES estimates in the literature, such as the one proposed in McNeil

[5] Unconditional coverage implies that at the p^{th} level of confidence, we should observe approximately $1 - p$ exceedances.

and Frey (2000). This method is based on the premise that the ES should not be significantly different from the average exceedance given that VaR has been exceeded. The test consists of a one-sided t-test as specified in (21), which is used to evaluate if the average difference between the estimated ES ($\hat{ES}_t$ for $\{\tau,\ldots,T\}$) and the average exceedance value $\bar{e}$[6], denoted by $\bar{d} = (T-\tau)^{-1}\sum_{t=\tau}^{T}(\hat{ES}_t - \bar{e})$, is significantly larger than zero. This is equivalent to testing if the expected shortfall is being systematically underestimated.

$$t = \frac{\bar{d}}{\sqrt{\text{Var}(\bar{d})/(T-\tau)}} \xrightarrow{d} \text{Student-t}_{(T-\tau-1)} \tag{21}$$

Note that $\text{Var}(\bar{d})$ denotes the sample variance of $\bar{d}$.

On the other hand, Gneiting (2011) showed that ES is not an elicitable measure, and used this to argue that it is not possible to evaluate ES estimates. Nonetheless, Fissler et al. (2015) showed that ES and VaR are jointly elicitable, meaning that joint tests could be a feasible alternative. In light of the discrepancies in the literature, this study is inclined to perform the abovementioned test under the following caveat. Given that VaR is usually computed for high levels of confidence, one would require an immense amount of observations to obtain a sufficiently large sample by which the Law of Large Numbers could be used to argued that the sample average is close to a true mean. Hence, the validity of the conclusions remains highly questionable. The test is implemented using the package Ghalanos (2020).

5 Results

5.1 Estimation and Evaluation of Semiparametric EVT Models

5.1.1 Multivariate Generalized Pareto Model

As mentioned in section 4 the first step in fitting the multivariate GP model is to determine an appropriate density. The results for the AIC's corresponding to each of the densities considered are summarized in Table (1). It can be observed that similarly to Kiriliouk et al. (2019), we find the Gumbel T density to offer the best fit to our data.

Table 1: AIC statistics for each Multivariate Generalized Pareto density considered.

	Gumbel T	Gumbel U	Gaussian	Reverse Exponential T	Reverse Exponential U
AIC	1835.137	1851.926	2037.270	1848.692	1850.985

[6] $\bar{e} = \sum_{t=\tau}^{T} \mathbb{1}(L_t > VaR_t(p))(L_t - VaR_t(p))$.

Next, we evaluate different model specifications of the Gumbel T density in search of a more parsimonious model specification. The resulting maximized log-likelihoods are summarized in Table (2). Likelihood ratio tests lead us to observe that M3 and M4 are significantly better models than M1 at the 5% level. Moreover, M2 is not significantly better than M1 at the 5% level with a p-value of 0.059. Given that M3 and M4 are not found to be significantly different even at the 10% level, we decide to proceed with M4 as it offers the most parsimonious model.

Table 2: Log-likelihoods for different model specifications of the chosen Gumbel T density

Model	Parameters	Maximized log-likelihood
M1	$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_3$	- 1821.1
M2	$\alpha_1, \alpha_2, \alpha_3, \alpha_4$	- 1824.9
M3	$\alpha, \beta_1, \beta_2, \beta_3$	- 1827.4
M4	α	- 1827.5

Figure (4) below contains plots for the fits of the marginal GP distributions for each of the four assets. We can observe that in general the fits tend to improve towards the tail, with the exception of AMZN, which seems to drift away from the 45-degree line towards its tail. However, these plots suggests that the model exhibits an overall appropriate fit at the marginal level.

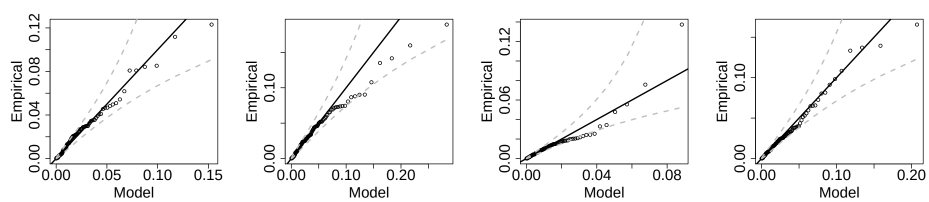

Figure 4: Marginal GPD QQ-plots. From left to right the plots correspond to MSFT, AMZN, JNJ and JPM.

When looking at the fit of the dependence structure, the results are not as promising as for the marginals. Figure (5) shows that based on the model selection, the value of the $\hat{\chi}(q)$ statistic assumed to be constant for sufficiently high q falls almost outside of its 95% confidence interval, which can be seen as a sign of misspecification. Additionally, this value is very close to 0 which indicates a weak dependence between the assets. This contradicts the initial observations from section 3, where the assets seemed to offer some dependence over periods of high risk.

We evaluate this concern further by looking at bivariate Chi-plots for each pair of assets in the portfolio, which are summarized in Figure (6). This is done to evaluate if the dependence structure between each pair of assets is being correctly incorporated into the model. We can immediately observe that the dependence structure between AMZN and JPM is not being correctly taken into

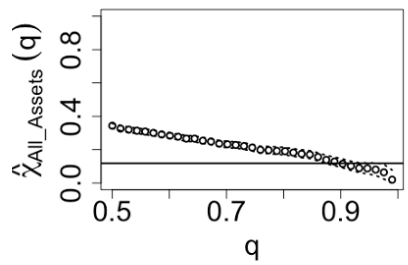

Figure 5: Chi plot for all assets

account, as the estimated $\hat{\chi}(q)$ value falls outside of its 95% confidence interval. Additionally, the estimated $\hat{\chi}(q)$ for the dependence between AMZN and JNJ falls very closely to the upper bound of the 95% confidence interval, which also raises concerns about misspecification. The rest of the plots can be found in Figure 12 in Appendix A, and their estimated $\hat{\chi}(q)$ seems to behave appropriately.

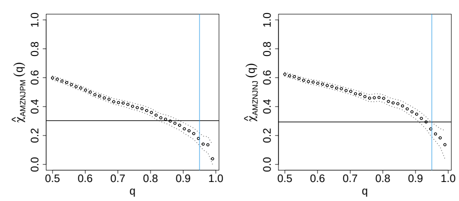

Figure 6: Bivariate Chi plots with fitted for AMZN and JPM, AMZN and JNJ

5.1.2 Hill Method Model

In this section we evaluate the validity of the assumptions underlying the Hill method. Starting from the left in the top row Figure (7) contains Hill plots for each asset, followed by Hill plots for the Radial parts and the equally weighted portfolio. In order to choose an appropriate cutoff point k, we evaluate the stability of the Hill-plot based on the bias-variance trade-off of the Hill estimator. In particular, the most stability seems to occur at $k \approx 128$ which corresponds to roughly 5% of the sample. For this choice of cutoff point, $\alpha = 2.9$ seems to be an appropriate estimate across all assets. Hence, the assumption of an equal tail index does not seem implausible. Moreover, this estimate also seems appropriate for both the Radial parts and the equally weighted portfolio at the chosen cutoff point k. Hence, These results are supportive of the assumption of equal tail indices across all assets. We proceed by considering the top 5% order statistics as the intermediate region close to the tail and by assuming equal tail indices.

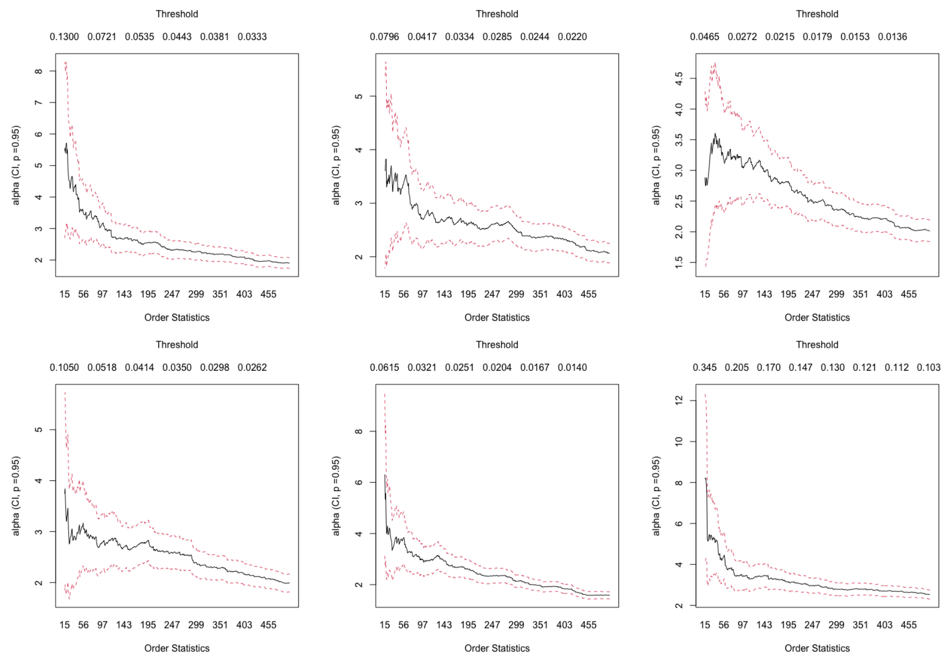

Figure 7: Hill plots for each asset, Radial parts and Equally Weighted portfolios. Order starting from the top left corner: AMZN, MSFT, JNJ, JPM, Radials, Portfolio.

5.2 Estimation and Evaluation of Extreme Risk Estimates

5.2.1 VaR Estimates

Figure (8) below summarizes the estimation results for VaR^{GPD} and VaR^{Hill} at the 99% and 95% confidence levels, in the left and right columns respectively. We can observe that the plots corresponding to the multivariate GPD method heavily overestimate VaR both at the 99% and 95% confidence levels, as no exceedances seem to have occurred even though we would expect 1% and 5% of observations to exceed the corresponding VaR's if correctly specified. The estimates obtained through the Hill method seem to offer a significantly more realistic view of the VaR, as some exceedances are indeed observed. The number of expected and actual exceedances are summarized in the first 2 columns of Table (3). The p-values corresponding to the tests outlined in the last 2 columns of Table (3). In particular, all test results yield a strong rejection of the hypotheses for correct specification, except for the duration test for independence (IND), which pertains to the assumption of independent exceedances. Clearly this result is also irrelevant as there are no exceedances for multivariate GP VaR estimates, thus the assumption of independence cannot be appropriately evaluated.

A reason for why both methods overestimate VaR, is that none of them are able to account for the conditional heteroskedasticity present in the data, as highlighted in section 3. This could be solved by combining these methods with a GARCH model, which accounts for this

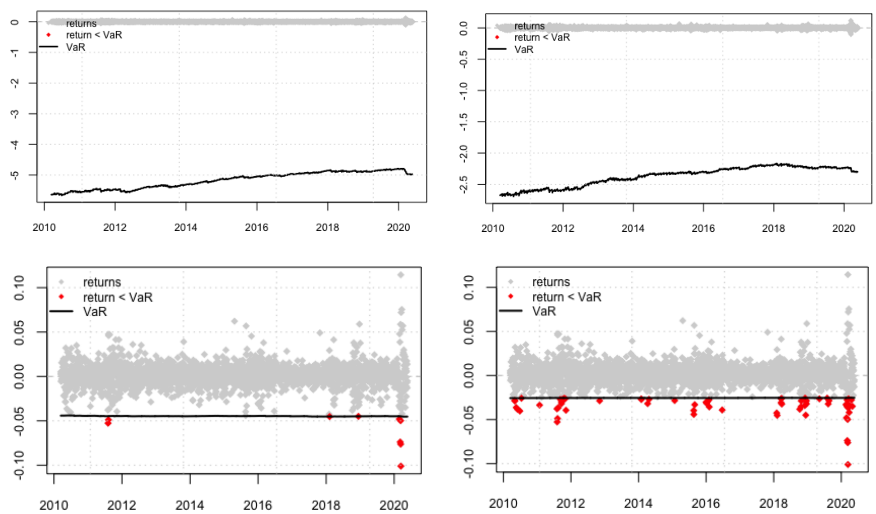

Figure 8: Plots of VaR estimates and returns for the backtesting period 16/03/2010 - 22/05/2020. Top row contains VaR^{GPD} estimates and bottom row VaR^{Hill} estimates. The left column contains 99% and the right 95% confidence level estimates.

Table 3: Number of VaR exceedances and P-values of tests.

Estimate	Expected Exceedances	Actual Exceedances	DQ P-value	IND P-value
$Var^{GPD}(0.01)$	25	0	0	0.317
$Var^{GPD}(0.05)$	128	0	0	0.317
$Var^{Hill}(0.01)$	25	10	0	0
$Var^{Hill}(0.05)$	128	72	0	0

property. Additionally, the multivariate GPD model could be argued to offer a significantly worse performance because there seems to be little to no dependence amongst the assets, as noted in Figure (5). The lack of dependence implies that the event of an extreme loss in one of the assets provides little information regarding the behavior of other assets. Thus, it could be argued that this uncertainty stemming from the lack of dependence can be a cause for overestimation.

5.2.2 ES Estimates

The 99% and 95% confidence ES estimates for each method can be observed in Figure (9). Based on the test of McNeil and Frey (2000), the ES estimates obtained from the multivariate GP method cannot be evaluated as no exceedances are empirically observed. For the Hill method estimates, the test can be carried out as there are exceedances both at the 99% and 95% confidence levels, as shown in Table (3). In particular, the test is unable to reject $H_0 : \bar{d} = 0$ for both sets of estimates, as it attains p-values of 0.98 and 0.94 for the 0.95% and 0.99% confidence ES estimates, respectively. Nonetheless, these results cannot be taken as clear indications of correct specification due to the concerns raised in Section 4.

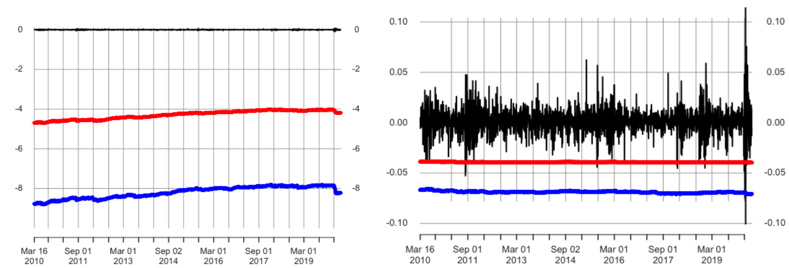

Figure 9: Plots of ES estimates for multivariate GP (left), multivariate Hill method (right) and returns of the equally weighted portfolio of assets, for the backtesting period 16/03/2010 - 22/05/2020. 95% confidence estimates are plotted in red, whilst 99% confidence estimates are plotted in blue.

6 Conclusions

The goal of this study was to explore the empirical application of multivariate peak over threshold models to risk management. To this end, the multivariate Generalized Pareto (GP) and Hill method were fitted on the losses of an equally weighted portfolio consisting of Amazon, Microsoft, J.P. Morgan and Johnson and Johnson over the period 03/01/2000 – 14/03/2010. These models were used to evaluate daily VaR and ES estimates for the portfolio over the period 15/03/2010 – 22/05/2020 using an expanding window.

Even though preliminary analyses motivated the use of extreme value theory techniques in the data, the results obtained for both methods were far from promising, particularly for the multivariate GP method. The results point towards the conclusion that both methods tend to overestimate both VaR as the null hypothesis of correct specification was reject by both the Dynamic Quantile and the Duration test for independence. With regards to the ES estimates, those of the multivariate GP method could not be evaluated with McNeil's test as there were no exceedances. However, those of the Hill method could be evaluated and led to the conclusion that the specification could not be deemed incorrect. Nonetheless, there has been plenty of discussion in the literature regarding the validity of ES tests, and this result should not be taken a strong argument in favour of the Hill method's performance.

A possible reason for these drastic negative results, is that the heteroskedasticity (or volatility clustering) present in the data is not being appropriately incorporated in any of the two models. Hence, further research exploring applications of these techniques combined with GARCH models could provide interesting results for practitioners. Moreover, the multivariate GPD model could be argued to offer a significantly worse performance than the Hill method model due to the lack of dependence amongst the assets.

References

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503.

Ardia, D., Boudt, K., and Catania, L. (2019). Generalized autoregressive score models in R: The GAS package. *Journal of Statistical Software*, 88(6):1–28.

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

Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, pages 792–804.

Basrak, B., Davis, R. A., and Mikosch, T. (2002). A characterization of multivariate regular variation. *Annals of Applied Probability*, pages 908–920.

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, 57(12):2213–2227.

Bilandi, M. M. and Kudła, J. (2016). Comparing the precision of different methods of estimating var with a focus on evt. *ACRN Oxford Journal of Finance and Risk Perspectives*, 5(1):109–147.

Christoffersen, P. and Pelletier, D. (2004). Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, 2(1):84–108.

Cirillo, P. (2013). Are your data really pareto distributed? *Physica A: Statistical Mechanics and its Applications*, 392(23):5947–5962.

Cirillo, P. and Taleb, N. N. (2016). On the statistical properties and tail risk of violent conflicts. *Physica A: Statistical Mechanics and its Applications*, 452:29–45.

Danielsson, J. (2011). *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab*, volume 588. John Wiley & Sons.

Daníelsson, J., Jorgensen, B. N., Samorodnitsky, G., Sarma, M., and de Vries, C. G. (2013). Fat tails, var and subadditivity. *Journal of econometrics*, 172(2):283–291.

De Haan, L. and Ferreira, A. (2007). *Extreme value theory: an introduction*. Springer Science & Business Media.

De Haan, L., Mercadier, C., and Zhou, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics*, 20(2):321–354.

Embrechts, P., Kluppelberg, C., and Mikosch, T. (1999a). Modelling extremal events. *British actuarial journal*, 5(2):465–465.

Embrechts, P., Resnick, S. I., and Samorodnitsky, G. (1999b). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2):30–41.

Engle, R. F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge University Press.

Fissler, T., Ziegel, J. F., and Gneiting, T. (2015). Expected shortfall is jointly elicitable with value at risk-implications for backtesting. *arXiv preprint arXiv:1507.00244*.

Frittelli, M. and Gianin, E. R. (2002). Putting order in risk measures. *Journal of Banking & Finance*, 26(7):1473–1486.

Ghalanos, A. (2020). *rugarch: Univariate GARCH models.* R package version 1.4-2.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Harmantzis, F. C., Miao, L., and Chien, Y. (2006). Empirical study of value-at-risk and expected shortfall models with heavy tails. *The journal of risk finance*.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, pages 1163–1174.

Kallenberg, O. (1983). Random measures, 3rd and 4th editions.

Kiriliouk, A., Rootzén, H., Segers, J., and Wadsworth, J. L. (2019). Peaks over thresholds modeling with multivariate generalized pareto distributions. *Technometrics*, 61(1):123–135.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The J. of Derivatives*, 3(2).

Mainik, G. and Rüschendorf, L. (2010). On optimal portfolio diversification with respect to extreme risks. *Finance and Stochastics*, 14(4):593–623.

Mandelbrot, B. (1963). New methods in statistical economics. *Journal of political economy*, 71(5):421–440.

Marinelli, C., d'Addona, S., and Rachev, S. T. (2007). A comparison of some univariate models for value-at-risk and expected shortfall. *International Journal of Theoretical and Applied Finance*, 10(06):1043–1075.

Markowitz, H. (1952). The utility of wealth. *Journal of political Economy*, 60(2):151–158.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3-4):271–300.

Nguyen, T. and Samorodnitsky, G. (2012). Tail inference: where does the tail begin? *Extremes*, 15(4):437–461.

Norton, M., Khokhlov, V., and Uryasev, S. (2019). Calculating cvar and bpoe for common probability distributions with application to portfolio optimization and density estimation. *Annals of Operations Research*, pages 1–35.

Pickands III, J. et al. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1):119–131.

Price, K., Price, B., and Nantell, T. J. (1982). Variance and lower partial moment measures of systematic risk: some analytical and empirical results. *The Journal of Finance*, 37(3):843–855.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

Rolski, T., Schmidli, H., Schmidt, V., and Teugels, J. L. (2009). *Stochastic processes for insurance and finance*, volume 505. John Wiley & Sons.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018a). Multivariate generalized pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018b). Multivariate peaks over thresholds models. *Extremes*, 21(1):115–145.

Von Mises, R. (1936). La distribution de la plus grande de n valuers. *Rev. math. Union interbalcanique*, 1:141–160.

Appendices

A Additional Preliminary Analyses

Zipf Plots As noted by Cirillo (2013) Zipf plots (also known as log-log plots) exploit the functional form of the Pareto distribution to evaluate the presence of a fat-tail. In particular, for a Pareto distribution the survival function is given by $\bar{F}(x) = 1 - F(x) = \left(\frac{x}{x_l}\right)^{-\alpha}$, where α denotes the tail index and x_l corresponds to the lower bound of the support. Hence, taking logs on both sides of the previous expression yields: $\ln \bar{F}(x) = \alpha \ln x_l - \alpha \ln x$, where $\alpha \ln x_l$ is constant for a given choice of α and x_l. Hence, in the presence of a Pareto tail ($\alpha > 0$) the plot should resemble a downward sloping line for large values of x. The Zipf plot for each of the assets considered over the full sample can be observed in Figure (10). All plots seem except AMZN's seem to offer a relatively linear behaviour for large values of x, which would support the presence of a fat-tail. AMZN appears to show quite some curvature even for large values of x, which questions the presence of a fat-tail.

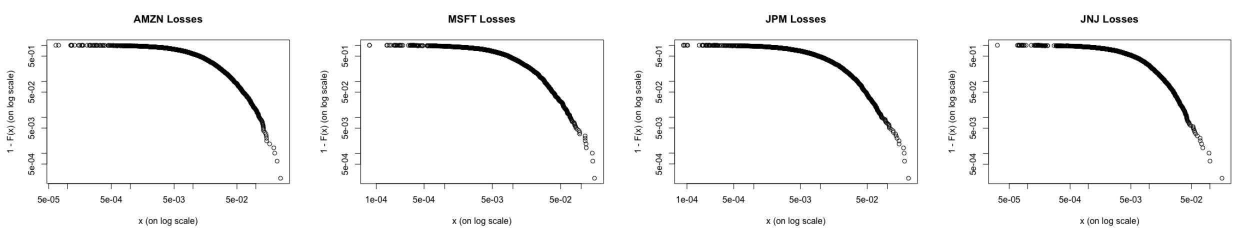

Figure 10: Zipf plot for the losses of AMZN, MSFT, JPM and JNJ for the period 03/01/2000 – 22/05/2020.

MS Plots Maximum-to-Sum (MS) plots will be used to evaluate which moments of the data could be considered to be well-defined. In particular, given that the moments of variables with Gumbel or Weibull tails are all theoretically well-defined as noted in Resnick (2007), finding evidence of undefined moments can be used to argue in favor of a Fréchet tail (i.e. fat-tail). As explained in Cirillo and Taleb (2016) an MS plot for the p^{th} moment consist of a plot of $R_n(p) = \frac{M_n(p)}{S_n(p)}$ for increasing values of the sample size $n \geq 0$. Note that $S_n(p) = \sum_{i=1}^{n} |x_i|^p$ and $M_n(p) = \max(|x_1|^p, \ldots, |x_n|^p)$ denote the partial sum and maximum, respectively. The above-mentioned result regarding the well-definedness of the p^{th} moment follows from the fact that if $\mathbb{E}(X^p) < \infty$ then $R_n(p) \xrightarrow{p} 0$ as $n \to \infty$. Figure (11) contains MS plots for $p = 1, \ldots, 4$ for each asset of the assets considered, over the fitting period. The plots for $p = 1$ and $p = 2$ lead to the conclusion that the first and second moments for every asset seem to be well-defined, as they appear to converge to zero as n increases. Moreover, the plots for $p = 3$ show that the third moment seems to be well-defined for AMZN, MSFT and JNJ, but not for JPM. Lastly, the plots for

$p = 4$ do not seem to converge to zero for any of the assets, which implies that the assumption of well-defined fourth moments does not seem plausible for any asset. Hence, these plots are also supportive of the presence of fat-tails.

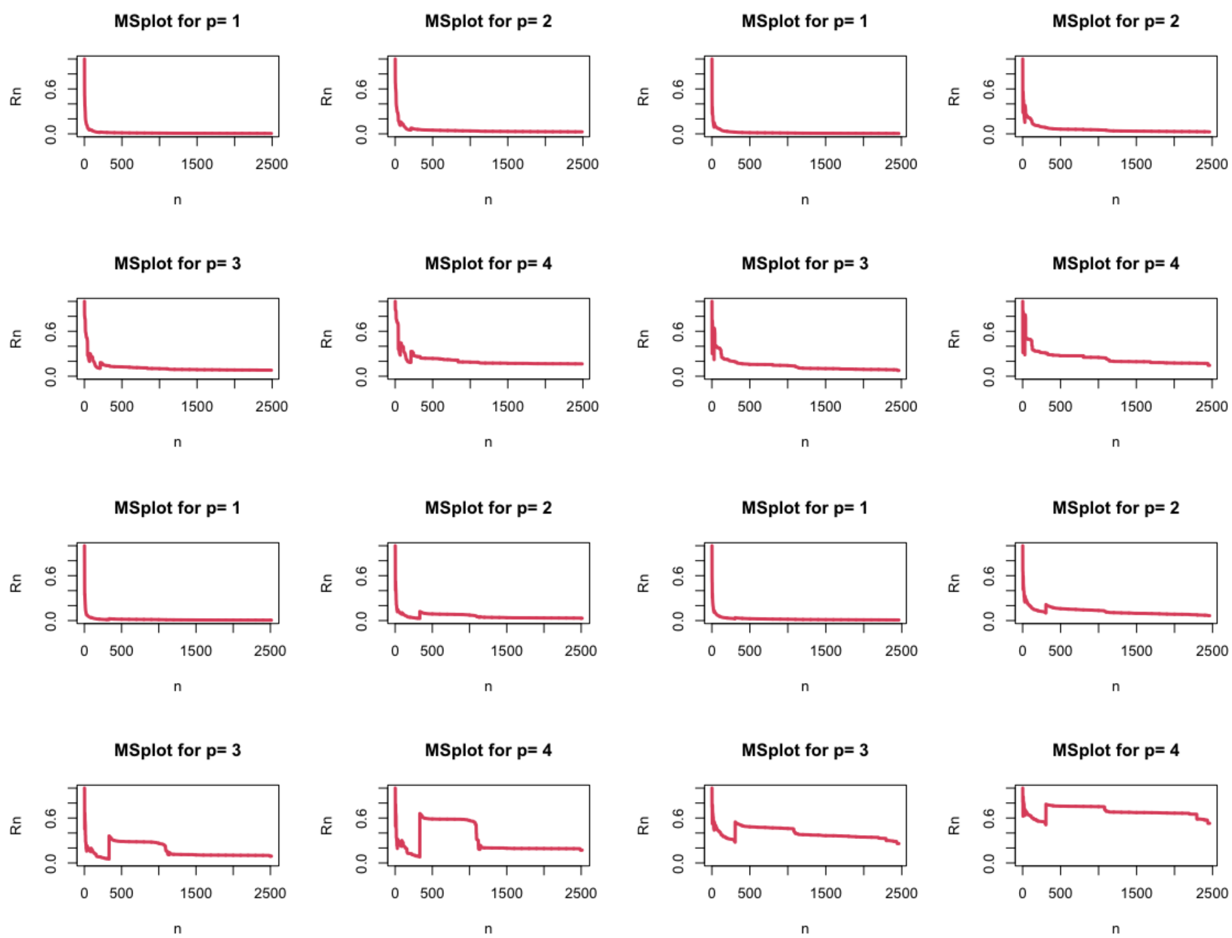

Figure 11: MS plots for $p = 1, \ldots, 4$ of the losses of AMZN (upper left quadrant), MSFT (upper right quadrant), JPM (lower left quadrant) and JNJ (lower right quadrant), for the period 03/01/2000 – 22/05/2020.

B Additional Derivations

VaR and ES estimators for multivariate GPD model The survival function $\bar{F}(y) = 1 - F(y)$ of the portfolio losses Y_t at a point $t \in \{\tau, \ldots, T\}$ in time, can be expressed as:

$$\begin{aligned}
\bar{F}(y) &= \mathbb{P}(Y_t > y) \\
&= \mathbb{P}(Y_t > y | Y_t \nleq u)\mathbb{P}(Y_t \nleq u), \text{ for } y \nleq u \\
&= \bar{F}_u(y)\bar{F}(u)
\end{aligned}$$

where $\bar{F}_u(y)$ denotes the multivariate GP distribution described in (14). Then, we can derive the corresponding VaR using the definition provided in (9) and solving for the quantile y as follows:

$$1-p=\bar{F}(u)\left(1+\frac{y-\sum_{j=1}^{d}w^{(j)}u^{(j)}}{\alpha\sum_{j=1}^{d}w^{(j)}\sigma_j}\right)^{\alpha}$$

$$VaR_t^{GPD}(p)=\alpha\sum_{j=1}^{d}w^{(j)}\sigma^{(j)}\left(\left(\frac{1-p}{\bar{F}(u)}\right)^{\frac{1}{\alpha}}-1\right)+\sum_{j=1}^{d}w^{(j)}u^{(j)} \tag{22}$$

In order to derive an estimator for $ES_t^{GPD}(p)$, one can make use of the threshold stability property of GP distributions as outlined in section 2. This allows us to see the distribution of Y_t conditional on exceeding $VaR_t^{GPD}(p)$ as the distribution of Y_t conditional on exceeding some threshold and then simply estimating the mean of the resulting distribution. Noting that the mean of a $GP(\mu,\sigma,\xi)$ distribution is given by $\mu+\frac{\sigma}{1-\xi}$ if $\xi<1$ (i.e. if the mean of the distribution is defined), we can express $ES_X(p)$ using the definition (10) as:

$$\begin{aligned} ES(p) &= \mathbb{E}(Y|Y>VaR(p)) && (23)\\ &= \mathbb{E}(VaR(p)+Y-VaR(p)|Y>VaR(p)) && (24)\\ &= VaR(p)+\mathbb{E}(Y-VaR(p)|Y>VaR(p)) && (25)\end{aligned}$$

By threshold stability of GP distributions it follows that:

$$Y-VaR(p)|Y>VaR(p)\sim GP\left(\sum_{j=1}^{d}w_j\sigma_j+\gamma(VaR(p)-\sum_{j=1}^{d}w_ju_j),\gamma\right)$$

Hence, we conclude that:

$$ES^{GPD}(p)=VaR^{GPD}(p)+\frac{\sum_{j=1}^{d}w_j\sigma_j+\gamma\left(VaR^{GPD}(p)-\sum_{j=1}^{d}w_ju_j\right)}{1-\gamma} \tag{26}$$

Hence, it follows that both $VaR_t^{GPD}(p)$ and $ES_t^{GPD}(p)$ can be estimated by plugging the sample estimates of the distribution's parameters and the empirical distribution, into (22) and (26) respectively.

VaR and ES estimators for Hill method In order to derive the VaR estimate for the Hill method, it suffices to focus on the univariate estimator derived by Danielsson (2011). Namely, note that we can express the probability of exceeding $VaR(p)$ conditional on being in the tail, i.e. $Y > u$ as:

$$\mathbb{P}(Y > VaR(p)|Y > u) = \frac{\mathbb{P}(Y > VaR(p))}{\mathbb{P}(Y > u)} = \frac{p}{\bar{F}(u)} \tag{27}$$

$$\mathbb{P}(Y > VaR(p)|Y > u) = \left(\frac{VaR(p)}{u}\right)^{-\frac{1}{\gamma}} = \left(\frac{VaR(p)}{u}\right)^{-\alpha} \tag{28}$$

where (27) follows from the definition of conditional probability and (28) follows from the assumption of a fat-tail, i.e. Pareto-distributed tail. Note that (28) is only true in the limit, but assuming a strict equality allows for the estimator to be derived. Equating these two expressions and solving for $VaR(p)$ yields:

$$VaR^{Hill}(p) = u\left(\frac{\bar{F}_(u)}{p}\right)^{\frac{1}{\alpha}} \tag{29}$$

To find an ES estimator for the Hill method we follow a similar approach to that used in Kiriliouk et al. (2019) to derive the ES of the multivariate GP method. In particular, note that due to the MRV assumption, we assume that the excessess are Pareto distributed. Similar to the GP distribution, the Pareto distribution also exhibits threshold stability. Hence, it suffices to estimate the mean of the conditional excess distribution to estimate ES. Given that the mean of a Pareto-distributed random variable Y is given by $\mathbb{E}(Y|Y > x_m) = \frac{\alpha x_m}{\alpha - 1}$ for $\alpha > 1$, it follows[7]:

$$ES^{Hill}(p) = \mathbb{E}(Y|Y > VaR(p)) \tag{30}$$

$$= \frac{\alpha VaR^{Hill}(p)}{\alpha - 1} \tag{31}$$

Hence, we can estimate $VaR^{Hill}(p)$ and $ES^{Hill}(p)$ using the estimated tail index $\hat{\alpha}$ and the empirical distribution value of $\bar{F}(u)$.

[7]Note that Norton et al. (2019) offers a formal proof of this result.

C Additional Results

Additional Bivariate Chi Plots These plots consist of the bivariate Chi plots with fitted values not included in Figure (6).

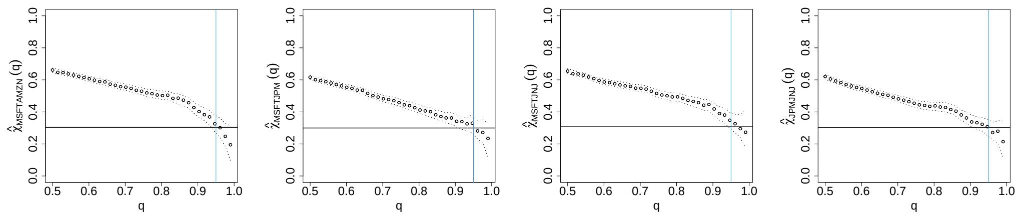

Figure 12: Bivariate Chi plots with fitted