TESIS DOCTORAL

AÑO 2023

QUANTIFICATION OF MARKET RISK IN THE CONTEXT OF CONDITIONAL EXTREME VALUE THEORY.

Autora: MªÁngeles Navarro Cervantes

PROGRAMA DE DOCTORADO EN ECONOMÍA Y EMPRESA

Directora: Dra. Sonia Benito Muela
Codirectora: Dra. Carmen López Martín

"The key message is that EVT cannot do magic but it can do a whole lot better than empirical curve-fitting and guesswork. My answer to the skeptics is that if people aren't given well-founded methods like EVT, they'll just use dubious ones instead"

Jonathan Tawn.Prof. of Statistics. Lancaster University.

"There is always going to be an element of doubt, as one is extrapolating into areas one doesn't know about. But what EVT is doing is making the best use of whatever data you have about extreme phenomena"

Richard Smiths, Prof. of Biostatistics. Gillings School of Global Public Health.

Agradecimientos

En primer lugar, quiero expresar mi gratitud a mis directoras de Tesis. Vuestra ayuda, apoyo y motivación han sido esenciales para mí durante este proceso. Sobre todo, gracias también por guiarme en momentos difíciles y por vuestra paciencia.

A Sonia, desde que me dirigiste el TFM hasta ahora, gracias por tu confianza. A Carmen, gracias por querer formar parte de este proyecto y codirigir la Tesis, y por tener siempre palabras de ánimo.

Agradecimiento especial a Javi, por estar siempre ahí, por tratar de entender sinceramente de qué iba todo esto, por tu apoyo incondicional, por creer en mí cuando a veces yo dudaba. Gracias por acompañarme siempre y por permitirme que te acompañe en tantas aventuras. Sobre todo la de este último año, con Martín.

A él también le hago una mención especial en esta dedicatoria. Tu sonrisa y tus primeras palabras han sido para mí un impulso de ánimo e ilusión en esta última etapa, aunque el doctorado en maternidad no ha hecho más que empezar.

A mi familia, gracias siempre por vuestros ánimos. Al final, llegar hasta aquí se ha debido también a vuestro esfuerzo.

A los que hoy no puedo agradecer personalmente su apoyo pero que los llevo siempre en mi recuerdo.

Y en definitiva, mi más sincero agradecimiento a todas aquellas personas, que con sus consejos o con su interés, han formado parte de este viaje y han hecho posible que hoy pueda presentar esta Tesis Doctoral con orgullo y satisfacción.

Por último, espero humildemente que mi aportación con este trabajo pueda ser de utilidad a la comunidad científica y que pueda ayudar a mejorar la comprensión y el conocimiento en este campo de estudio.

Contents

List of Figures

CHAPTER II

CHAPTER III

CHAPTER V

List of Tables

CHAPTER II

CHAPTER III

CHAPTER IV

List of abbreviations

AD	Anderson-Darling test
APARCH	Asymmetric Power ARCH Model
ARCH	Autoregressive Conditional Heteroskedasticity Model
BCBS	Basel Committee on Bank Supervision
BD	Bayer and Dimitriadis Expected Shortfall backtest
BMM	Block Maxima Method
BTC	Back-Testing Criterion
CDF	Cumulative Distribution Function
CvM	Cramer-von Mises test
CVaR	Conditional Value at Risk
DCR	Daily Capital Requirement
DQ	Dynamic Quantile Test
EGARCH	Exponential GARCH
ES	Expected Shortfall
EVT	Extreme Value Theory
FABL	Abad-Benito-López Firm Function
FLF	Sarma Firm Loss Function
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GEVD	Generalised Extreme Value Distribution
GPD	Generalised Pareto Distribution
KS	Kolmogorov-Smirnov Test
LRcc	Christoffersen's Likelihood ratio statistic for conditional coverage test
LRind	Christoffersen's Likelihood ratio statistic for the independence of violations

LRuc	Kupiec likelihood ratio statistic for the independence of violations
MDA	Maximum Domain of Attraction
MEF	Mean Excess Function
MEP	Mean Excess Plot
MF	McNeil and Frey Expected Shortfall backtest
MRLP	Mean Residual Life Plot
NZ	Nolde and Ziegel Expected Shortfall backtest
POT	Peaks Over Threshold Model
RC	Righi and Ceretta Shortfall backtest
RFL	Regulator Function López
RMSE	Root Mean Error
SD	Standard deviation
SMEF	Sample Mean of the Excesses Function
VaR	Value at Risk

Chapter I. Introduction.

1. Summary and objectives.

This Thesis presents an empirical analysis that employs the Extreme Value Theory (EVT) to examine different aspects related to quantifying market risk for an asset portfolio. We aim to explore the applicability of EVT in financial risk management and evaluate its effectiveness in measuring market risk through different approaches.

Although the theoretical framework as well as the specific objectives and results of each Chapter that make up this Thesis will be expanded upon later in this introduction, the fundamental objectives of this research can be summarised in the following points:

(i) Assessing to what extent the performance of the volatility model estimation under fat-tailed and skewed distributions let to improve market risk estimation in the framework of the Conditional Extreme Value Theory (Chapter II).

(ii) Examining how the choice of threshold affects the Generalize Pareto Distribution (GPD) quantiles and market risk measures (such as Value at Risk and Expected Shortfall) using the Peaks over Threshold approach (Chapter III)

(iii) Analyzing the sensitivity of Generalized Extreme Value Distribution (GEVD) parameters and market risk quantification to the choice of different block sizes with the Block Maxima approach (Chapter IV).

Thus, this research seeks to contribute to the existing literature on market risk by providing new insights and empirical evidence through the application of the Extreme Value Theory (EVT). Specifically, we focus on the methodology proposed by McNeil and Frey (2000) for estimating the market risk that combines the unconditional Extreme Value Theory with models of volatility, i.e, the Conditional Extreme Value Theory.

First, it is necessary to provide an overview of what Extreme Value Theory entails. This Theory is considered one of the most important statistical disciplines for estimating the likelihood of uncommon events based on observed outliers.

More formally, EVT focuses on the limiting distribution of the extreme values observed over a long period, which is independent of the distribution of the values themselves. In other words, EVT relates to the asymptotic behavior of extreme observations of a random variable.

Although EVT is well-established in many sciences such as engineering, insurance, and meteorology among others (see e.g., Embrechts et al., 1999; Reiss and Thomas, 2007), its application in the financial sector has gained more relevance in recent years. It has demonstrated its effectiveness in a variety of fields, ranging from the analysis of material strength to the estimation of the likelihood of natural disasters such as earthquakes and floods.

Specifically, regarding risk management, extreme value modeling has emerged recently as an important tool to quantify large financial losses from different sources of risk (see Cruz, 2002; Moscadelli, 2004; Fontnouvelle et al., 2007; Ergün and Jun, 2010; Žiković and Aktan, 2009; and Abad and Benito, 2013).

In order to contextualize our research, it is important to understand that risk is a key concept in financial decision-making that relates to the uncertainty surrounding future investment outcomes. Financial risk is the risk associated with financial transactions and investments. It encompasses a wide range of risks such as market, credit, liquidity, operational, and legal risks. Particularly, market risk is the risk of loss due to changes in the value of financial assets and is of particular concern to investors and financial institutions. This kind of risk is an inherent characteristic of investments due to the unpredictable nature of market fluctuations. By accurately assessing the nature and magnitude of financial risk, investors can make informed decisions to optimize their portfolios.

Thus, measuring financial risk is an essential component of effective risk management. Various measures of risk are used to quantify the potential loss of an investment, including Value at Risk (VaR) and Expected Shortfall (ES). These measures provide information on the potential size of losses that an investment could incur, allowing investors to assess the level of risk associated with different investment strategies.

Chapter I. Introduction.

Jorion (2001) defines the VaR measure as the maximum expected loss over a specified time horizon, under normal market conditions, at a given confidence level. The ES is defined as the average of all losses that are greater than or equal to VaR, i.e., the average loss in the worst α % cases. In contrast to the VaR measure, ES is a coherent risk measure, and it does not present tail risk[1].

Although to date, the VaR measure has been the most used for quantifying market risk, ES is gaining ground in part due to the change in the regulation set by the Basel Committee on Banking Supervision (BCBS). Under the new regulation, financial institutions must calculate the market risk capital requirements' risk based on the ES measure, replacing the VaR measure (BCBS, 2012, 2013, 2017).

To estimate the VaR, various methodologies have been developed. One of the most commonly used approaches by financial institutions due to its simplicity is the parametric method. This method assumes that financial returns are characterized by a known distribution. Thus, the VaR of a portfolio at the $1-\alpha$ confidence level is calculated by multiplying the conditional standard deviation of the portfolio returns by the α quantile of the distribution. Typically, the normal distribution is assumed for this purpose by financial institutions, being this assumption the major limitation of the method.

Empirical studies have demonstrated that financial returns do not adhere to a normal distribution. Most of the time, the skewness coefficient is negative and statistically significant, which indicates that the financial return distribution is skewed toward the left. Moreover, the empirical distribution of financial returns has been found to exhibit a considerable amount of excess kurtosis (fat tails and peakedness), as reported by Bollerslev (1987). Consequently, the actual losses tend to be much larger than what would be predicted by a normal distribution.

Among the semi-parametric methods to estimate VaR, the EVT has been proven to be one of the most successful in VaR estimation (see Abad et al., 2014). This method, well-established in many sciences such as engineering, insurance, meteorology and recently in the financial field, focuses on limiting the distribution of extreme returns

[1] VaR is not a coherent market risk measure as it violates the subadditivity condition, which may discourage diversification (see Artzner et al., 1999).

observed over a long period, which is independent of the distribution of the returns themselves.

Thus, the first goal of this Thesis (Chapter II) is to analyze if in the framework of the method based on the Conditional Extreme Value Theory, the estimation of the volatility model under a fat tail and skewness distribution contributes to improving the results in VaR estimation.

For this purpose, two models to estimate volatility are used. The one proposed by Creal et al. (2008, 2011) and Harvey and Chakravarty (2008), called Beta-t-EGARCH, integrates three characteristics of financial returns as are volatility clustering, leverage effect and long memory. And the model proposed by Harvey and Sucarrat (2013) called Beta-skewness-t-EGARCH, which extends the Beta-t-EGARCH model, and also allows us to capture features such as skewness and fat-tailed distributions inherent in financial returns. We compared the VaR estimates obtained with both models, calculated one day ahead at a 1% probability.

The study has been done for six assets bellowing to the telecommunication sector: ADP, Amazon, Cerner, Apple, Microsoft and Telefónica. The analysis period runs from January 1st, 2008 to the end of December 2013.

The results obtained seem to indicate that the heavy tail and skewed distribution outperform the symmetric distribution both in terms of accuracy VaR estimations and in terms of the firm's loss function but also regarding requirement capital.

Continuing with Extreme Value Theory, the first step is to identify which observations are considered outliers. Two approaches are available for this purpose. The first approach involves dividing the sample of observations into blocks of equal size and selecting the maximum observation in each block (known as the Block Maxima Method). The other alternative is to set a threshold and select all observations above this threshold (Peaks over Threshold).

Within the Peaks over Threshold (POT) approach, the extreme values above a high threshold are modeled using a generalized Pareto distribution (GPD). The main difficulty of this approach lies in the selection of the threshold, as different thresholds may provide different results. In the context of risk management, it is interesting to know to what extent the selection of the threshold impacts risk estimation.

Chapter I. Introduction.

From a theoretical point of view, threshold selection is a critical issue in the framework of the POT approach. This method establishes that excess returns over the threshold follow a GPD when the number of observations *n* tends to infinity The choice of threshold must be a balance between bias and variation. A threshold being too low is likely to violate the asymptotic basis of the model which leads to bias. However, a threshold being too high will generate few excesses leading to an increase in the variance of the estimators (Davison and Smith, 1990; Coles, 2001; MacDonald et al.,2011; Papalexiou et al.,2013; Wyncoll and Gouldby, 2013).

That is why, within the framework of the POT method, different techniques have been developed recently for the selection of a suitable threshold, although none of them have been proven to provide better results than others. In Chapter III we review the State of the Art of methodologies used for this purpose.

Our study is in accordance with Langousis et al. (2016) who remarked that "*the variety of existing methods for the threshold chosen, the fundamental differences in their theoretical underpinnings, and their relative performance when dealing with different types of data, make threshold detection an open question that can be addressed solely on the basis of a specific application*". With this in mind, the second purpose of the Thesis (Chapter III) has been to address whether there is an optimal threshold or, on the contrary, there is a range of thresholds that may be suitable for quantifying market risk. Iriondo (2017) offers preliminary evidence in favor of this last hypothesis. Following this author, we will analyze the impact of the threshold choice on the two risk measures: VaR and ES.

As far as our knowledge goes, there have been no studies conducted on this particular subject in the financial field. To cover this gap, we perform an empirical analysis with a double aim. First, to examine the sensitivity of the GPD quantiles to the threshold choice and second, to study the sensitivity of the market risk measures to this choice.

We analyze in detail the case of the S&P500 daily return from January 3rd, 2000, through December 30th, 2021, and later the study has been extended to a set of 14 assets from alternative markets: 7 stock indexes (CAC40, DAX30, FTSE100, HangSeng, IBEX35, Merval and Nikkei), four commodities (Copper, Gold, Crude Oil Brent and Silver) and three rates exchange (£ /€, $/€ and ¥/€). To apply the POT method, different thresholds have been considered from 80th to 99th. The daily forecasting is obtained one day ahead at the 95% and 99% confidence levels according to the Basel Committee on Banking Supervision's standards. The accuracy of the risk estimates has been evaluated through backtesting.

In our analysis, we first observe that the parameter estimations are highly sensitive to the selected threshold for estimating GPD, consistent with previous literature. However, we find that the quantiles of the GPD remain stable despite changes in the threshold, especially for high quantiles (95th, 96th, 97th, 98th and 99th), which are relevant for risk estimation. Secondly, for a large set of thresholds (from the 80th percentile to the 96th percentile), the VaR and ES estimates are practically equivalent. Finally, we use the ES (99%) estimations to calculate the market risk capital requirements, and our results reveal that there is a range of thresholds that produce the same outcome, with some differences observed at the higher percentiles.

Hence, the results suggest that when estimating market risk, researchers and practitioners should not place excessive emphasis on selecting a specific threshold, as a broad range of choices results in similar risk estimates. While the quantification of risk is not primarily dependent on threshold selection, minor differences are noticeable for certain thresholds. This may be of interest to financial institutions that may choose the threshold that minimizes the market risk capital requirement.

In contrast with the POT method, the Block Maxima Method (BMM) is based on the idea of dividing the dataset into *m* blocks of size *n* and then fitting the Generalized Extreme Value (GEV) distribution to the maximum *m*-block data series. Selecting the block size may be not trivial. The fit of the GEV distribution will be inaccurate if the block size is too small, leading to biased estimates, while a block size too large will lead to a smaller number of extreme observations and consequently a higher variance (Coles et al., 2001).

Thus, the aim of Chapter IV is two-fold. Firstly, it focuses on exploring the sensitivity of the parameters of the GEV distribution for different block sizes. We particularly examine the changes in the shape parameter, which determines the weight of the tail in the distribution. Additionally, we aim to investigate the impact of block size selection on the quantification of market risk. To achieve this goal, we assess the sensitivity of risk measures such as VaR and ES to changes in block sizes. The objective is to identify if there exists an optimal block size that leads to precise risk estimates.

Our analysis is performed for daily data on the S&P500 stock index from January 3rd, 2000 to December 31st, 2019, and the results are later extended to a broad set of assets to verify the findings (the same used in Chapter III). For calculating risk, two confidence levels have been used: 97.5% and 99%. To apply BMM, nine different block sizes have been considered: 5, 10, 21, 31, 42, 63, 126, 189 and 252 observations.

Our research indicates that the accuracy of the risk estimates through the risk measures mentioned above depends clearly on the block length chosen. Therefore, the block size may be a critical aspect that should not be chosen ad hoc. This leads us to think that, in line with the literature reviewed, BMM may not be the most reliable method for estimating market risk if the block size is selected arbitrarily. The use of this method could require further research on optimal block size selection techniques.

Lastly, we conduct a comparison of VaR estimations between the POT and BM methods. The results of this comparison indicate, as expected, that the POT method is consistent across different threshold choices, while the BM method is highly dependent on the selected block size to achieve a level of exceptions close to the theoretical one.

Finally, the Thesis ends with some concluding remarks shown in the Conclusions Section.

2. Methodology.

The analysis has been performed using software tools such as R and Matlab mostly for the calculation of volatility models, market risk and backtesting. The most relevant R packages for the development of the study are referenced in the bibliography Section. For the graphics, plots and tables mainly spreadsheet tools such as Excel have been used. The descriptive statistics have been calculated through SPSS. To optimize data processing when it required analyzing a large amount of data in R, it has been used Amazon Elastic Compute Cloud (EC2).

The formulation and algorithms for the calculation of VaR and ES risk measures for POT and BMM approaches are detailed in the Appendix at the end of this Thesis.

The databases consulted to obtain the data series used in this Thesis were mainly Thomson Reuters-Eikon and Yahoo Finance.

Chapter II*

Analyzing the role of the skewed distributions in the framework of conditional extreme value theory

Abstract

In this paper, we analyze the role of the heavy tail and skewed distribution in market risk estimation (Value at Risk (VaR)). In particular, we are interested in knowing if, in the framework of the conditional Extreme Value Theory, the estimation of the volatility model below the heavy tail and skewed distribution contributes to improving the VaR estimation with respect to those obtained from a symmetric distribution. The study has been done for six individual assets bellowing to the telecommunication sector: ADP, Amazon, Cerner, Apple, Microsoft and Telefonica. The analysis period runs from January 1st, 2008 to the end of December 2013. Although the evidence found is a little bit weak the results obtained seem to indicate that the heavy tail and skewed distribution outperform the symmetric distribution both in terms of accuracy VaR estimations and in terms of the firm's loss function. Furthermore, the market risk capital requirement fixed on the base of the VaR estimations is also lowest under a skewed distribution.

Keywords: Extreme Value Theory, Value at Risk, asymmetric distributions, volatility, GARCH models.

*Published in the journal *International Business Research*; Vol. 10, No. 11; 2017. ISSN 1913-9004 E-ISSN 1913-9012. Canadian Center of Science and Education. Online Published: October 13, 2017. DOI:10.5539/ibr.v10n11p88

Chapter II. Analyzing the role of the skewed distributions in the framework of conditional extreme value theory.

1. Introduction

A context of risk is one in which we do not know with certainty the consequences associated with a decision. The only thing that we know is the possible outcomes associated with it and the likelihood of achieving such results. In the financial field, the notion of risk implies that we know the various yields can potentially get to invest and also know the probability of achieving such results. This allows us to estimate the average expected yield and the possible diversion 'above' or 'below' the average value, that is, the risk. The most popular and traditional risk measure is volatility (variance). The traditional financial theory defines risk as the dispersion of returns due to movements in financial variables.

Another way of measuring risk, which is the most commonly used at present, is to evaluate the losses that may occur when the price of the asset that makes up the portfolio goes down. This is what Value at Risk (VaR) does. The Value at Risk of a portfolio indicates the maximum amount that an investor may lose over a given time horizon and with a given probability. In this case, the concept of risk is associated with the danger of losses.

According to Jorion (2001), "VaR measure is defined as the worst expected loss over a given horizon under normal market conditions at a given level of confidence". For the estimation of the VaR measure, different methodologies have been developed which can be grouped into three groups: (i) parametric methods (as Riskmetrics); (ii) non-parametric methods (as historical simulation) and (iii) semiparametric method (as Extreme Value Theory, filter historical simulation and CaViar Method)[1].

Among all of them, one of the most used by the financial institution is the parametric method. This method assumes that financial returns follow a known distribution. Thus, below this method, the VaR of a portfolio at $1-\alpha$ confidence level is calculated as the product of the conditional standard deviation of the return portfolio by the α (k_α) quantile of the assumed distribution[2]. Mostly the assumed distribution is the normal one. Below this assumption k_α is the α quantile of the standard normal.

[1] See Abad et al. (2014) for a review of those methodologies.

[2] This is true when we assume that the portfolio return mean is zero.

The major drawbacks of this method are related to the normal distribution assumption for financial returns. Empirical evidence shows that financial returns do not follow a normal distribution. The skewness coefficient is in most cases negative and statistically significant, implying that the financial return distribution is skewed to the left. This result is not in accord with the properties of a normal distribution, which is symmetric. Also, the empirical distribution of financial return has been documented to exhibit significantly excessive kurtosis (fat tails and peakedness) (see Bollerslev, 1987). Consequently, the size of the actual losses is much higher than that predicted by a normal distribution.

Taking this into account, the research in the framework of the parametric method has focused on investigating other density functions that capture the skew and kurtosis of financial returns. In this line, Abad et al (2016), Chen et al (2012), Polansky and Stoja (2010), Bali and Theodossiou (2008), Zhang and Cheng (2005), Ausin and Galeano (2007), etc., show that in the context of parametric method assuming fat tail and skewness distributions improve the performance of this model in VaR estimation.

As we will see later, below the Conditional Extreme Value Theory (which is one of the most successful methods in estimating VaR) the VaR of a portfolio at $(1-\alpha)\%$ confidence level is calculated as the product of the conditional standard deviation of the return portfolio by the α (k_α) quantile of Generalized Pareto distribution. Traditionally, the conditional standard deviation of the return portfolio is estimated by assuming a symmetric distribution for the financial return. So, in the same line that the aforementioned paragraphs in this paper we analyze if in the framework of the method based on the Conditional Extreme Value Theory, the estimation of the volatility model below a fat tail and skewness distribution contributes to improving the results in VaR estimation.

The rest of the paper is as follows. In Section 2 we present the volatility models used in the empirical application. In Section 3 we describe the Value at Risk measure and the methodology used in this paper to calculate this measure and the backtesting techniques. The results of the empirical application are presented in Section 4. Section 5 includes the main conclusions.

2. Volatility models

According to Bollerslev, Engle and Nelson (1994), financial returns are characterized by a variance changing over time, alternating periods of low volatility followed by high volatility. In other words, as noted by Mandelbrot (1963), “large changes tend to be followed by large changes, of either sign and small changes tend to be followed by small changes.” This effect, known as *volatility clustering*, indicates the presence of conditional heteroscedasticity in the return series and the need to model the behavior of the conditional variance. Furthermore, financial returns are subject to leverage effect (Black, 1976). This means that volatility tends to be higher after negative returns, this is typically attributed to leverage (hence the name).

To capture cluster in volatility Engle (1982) proposed Autoregressive Conditional Heteroscedasticity (ARCH), which featured a variance that does not remain fixed but rather varies throughout a period. Bollerslev (1986) extended the ARCH model into the Generalized ARCH (GARCH). The GARCH model captures volatility clustering but does not capture the leverage effect. In this model, the positive surprises have the same effect on volatility as negative surprises. To capture the leverage effect have been developed asymmetric GARCH models for instance the APARCH model or the EGARCH model proposed by Nelson (1991).

On the other hand, it is well known that long memory property is observed in the volatility of the financial series, which cannot be captured by traditional GARCH models, so it is necessary to model a long-term component. To capture this characteristic Engle and Lee (1999) proposed a model in which the variance is decomposed into long and short-term components. The main role of the short-term component is to capture the temporary increase in the variance after an impact on the price. While the persistence or long-term memory indicates how long the shock takes to be absorbed.

Creal et al. (2008, 2011) and Harvey and Chakravarty (2008) proposed a model that integrates three characteristics aforementioned: cluster in volatility, leverage effect and long memory. The model proposed by these authors is known as Beta-t-EGARCH model. In this model, the short-term component responds in the same way as in the traditional GARCH. Nevertheless, the long-term component is not sensitive to extreme observations, as it does in the standard GARCH model. That is, in the standard GARCH, the existence of an outlier in yields has a persistent effect on volatility, which increases the variance. In contrast, the Beta-t-EGARCH model corrects volatility in case of the existence of an outlier, absorbing the effect and returning to previous levels of volatility.

Formally let $r_1, r_2, \dots, r_n$ be a sequence of independent and identically distributed random variables representing financial returns. Assume that $\{r_t\}$ follows a stochastic process:

$$r_t = exp(\lambda_{t|t-1})(\varepsilon_t), \varepsilon_t \sim t(0, \sigma_\varepsilon, \nu) \qquad \nu > 2 \qquad (1)$$

Where the conditional error ε_t is distributed as a t-student with zero mean, unconditional standard deviation σ_ε and degrees of freedom parameter ν. $\lambda_{t|t-1}$ is the conditional scale or volatility, that does not need to be equal to the conditional standard deviation. The conditional standard deviation is obtained as $\sigma_\varepsilon \; exp\,(\lambda_{t|t-1})$[3].

In the Beta-t-EGARCH model, $\lambda_{t|t-1}$ in (1) is defined as:

$$\lambda_{t|t-1} = \delta + \lambda^{+}_{1,t|t-1} + \lambda^{+}_{2,t|t-1} \qquad (2)$$

Where the long-term component is $\lambda^{+}_{1,t|t-1} = \varphi_1 \lambda^{+}_{1,t-1|t-2} + k_1 u_{t-1}$ and the short-term component is $\lambda^{+}_{2,t|t-1} = \varphi_2 \lambda^{+}_{2,t-1|t-2} + k_2 u_{t-1}$. The leverage effect may be introduced into the model using the sign of the observations. Thus, the short component with leverage effect is as follows:

$$\lambda^{+}_{2,t|t-1} = \varphi_2 \lambda^{+}_{2,t-1|t-2} + k_2 u_{t-1} + k^*(-r_{t-1})(u_{t-1} + 1)$$

Where u_t is the score conditional which is given by:

$$u_t = \frac{(\nu+1)(r_t)^2}{\nu exp(2\lambda_{t|t-1}) + (r_t)^2} - 1 \;, \qquad -1 \le u_t \le \nu, \nu > 0$$

[3] The specification for the standard GARCH model is as follows: $r_t = \sigma_t \varepsilon_t$, $\varepsilon_t \sim$IID $(0, \sigma_\varepsilon^2)$, where σ_t^2 is the scale o volatility, which is modeled $\sigma_t^2 = \delta + \varphi_1 \sigma_{t-1}^2 + k_1 \varepsilon_{t-1}^2$ with $\sigma_\varepsilon^2 = 1$.

Taking the signs of minus, r_{t-1} means that the parameter of k^* is normally non-negative for stock returns as in the GARCH model, the long-term component $\lambda^{+}_{1,t|\mathrm{t}-1}$, has φ_1 close to one or even set equal to one, while the short-term component $\lambda^{+}_{2,t|\mathrm{t}-1}$, will typically have a higher k combined with a lower φ. The model is not identifiable if the $\varphi_1 = \varphi_2$. Imposing the constraint $0 < \varphi_2 < \varphi_1 < 1$ ensures identifiability and stationarity.

Finally, the empirical literature has shown that financial returns also exhibit skewness and fat-tailed distributions. To capture these features Harvey and Sucarrat (2013) extended the Beta-t-EGARCH model, combining the skewness of the conditional distribution with a leverage effect in the dynamic of the scale. This model is known as Beta-skewness-t-EGARCH.

Skewness is introduced into the Beta-t-EGARCH model using the method proposed by Fernandez and Steel (1998) (see Harvey and Sucarrat (2013) for more details of this method). Thus, in Equation (1), they assume that the conditional error ε_t is distributed as a skewed t-student with mean μ_ε, scale σ_ε, degrees of freedom parameter v and skewness parameter γ[4].

$$r_t = exp(\lambda_{t|\mathrm{t}-1})(\varepsilon_t - \mu_\varepsilon), \varepsilon_t \sim st(\mu_\varepsilon, \sigma_\varepsilon, v, \gamma) \qquad v > 2,\ \gamma\, \epsilon (0, \infty) \qquad (3)$$

Where $\lambda_{t|\mathrm{t}-1}$ is given by:

$$\lambda_{t|\mathrm{t}-1} = \delta + \lambda^{+}_{1,t|\mathrm{t}-1} + \lambda^{+}_{2,t|\mathrm{t}-1} \qquad (4)$$

The long-term component is $\lambda^{+}_{1,t|\mathrm{t}-1} = \varphi_1 \lambda^{+}_{1,\mathrm{t}-1|\mathrm{t}-2} + k_1 u_{t-1}$ and the short-term component is $\lambda^{+}_{2,t|\mathrm{t}-1} = \varphi_2 \lambda^{+}_{1,\mathrm{t}-1|\mathrm{t}-2} + k_2 u_{t-1} + k^*(-(r_{t-1}))(u_{t-1} + 1)$. Again, only the short-term component has a leverage effect. In this model, the conditional score is given by:

[4] The conditional error ε_t is an uncentred (i.e., mean not necessarily equal to zero) skewed t variable with v degrees of freedom, skewness parameter γ. A centered and symmetric t-distribution variable with mean zero is obtained when $\gamma = 1$, in which $\mu_\varepsilon = 0$, whereas a left-skewed (right-skewed) t-variable is obtained when $\gamma < 1$, $(\gamma > 1)$. More details on the distribution can be found in Harvey and Sucarrat (2013) and Sucarrat (2013).

$$u_t = \frac{(\nu+1)(r_t+\mu_\varepsilon exp(\lambda_{t|t-1}))(r_t)}{\nu\gamma^{2(r_t+\mu_\varepsilon exp(\lambda_{t|t-1}))}exp(2\lambda_{t|t-1})+(r_t+\mu_\varepsilon exp(\lambda_{t|t-1}))^2}$$

3. Risk Measure methodology

Let $r_1, r_2, \ldots \ldots r_n$ be identically distributed independent random variables representing the financial returns. Use $F(r)$ to denote the cumulative distribution function $F(r) = Pr\ (r_t < r|\Omega_{t-1})$ conditioned to the information set, Ω_{t-1}, available at time $t-1$. Assume that $\{r_t\}$ follows the stochastic process:

$$r_t = \mu + \sigma_t\varepsilon_t \qquad \varepsilon_t \sim iid(0,1) \tag{5}$$

Where μ represents the mean of returns; $\sigma_t^2 = E(\varepsilon_t^2|\Omega_{t-1})$ and ε_t has the conditional distribution function $G(\varepsilon)$, $G(\varepsilon) = Pr\ (\varepsilon_t < \ \varepsilon\ |\Omega_{t-1})$.

The Value at Risk with a given level of confidence $1-\alpha$, denoted by $VaR\ (\alpha)$, is defined as the α quantile of the probability distribution of financial returns.

$$F(VaR(\alpha)) = Pr\big(r_t < VaR(\alpha)\big) = \alpha \tag{6}$$

There are two ways to estimate this quantile: (1) inverting the distribution function of financial returns, $F(r)$, or (2) inverting the distribution function of innovations, $G(\varepsilon)$. The latter case will also require estimating the standard deviation of returns.

$$VaR(\alpha) = F^{-1}(\alpha) = \ \mu + \sigma_t G^{-1}(\alpha) \tag{7}$$

Thus, the VaR estimation involves the specifications of the distribution function of financial returns, $F(r)$, or the distribution function of innovations, $G(\varepsilon)$, along with the standard deviation of returns, σ_t.

The Historical simulation method, Monte Carlo simulation and Unconditional approach based on the Extreme Value Theory focus on the estimation of $F(r)$, while Parametric Method and the Conditional Extreme Value Theory estimate $G(\varepsilon)$.

Below, we will describe the Conditional Extreme Value Theory, which has been used in this study to compute VaR.

3.1 Conditional Extreme Value Theory

The Extreme Value Theory (EVT) approach focuses on the limiting distribution of extreme returns observed over a long period, which is essentially independent of the distribution of the returns themselves.

There are two methods based on EVT: a) Block Maxima Model (BMM) proposed by McNeil (1998) and b) Peaks Over Threshold (POT). This second model is generally considered to be the most useful for practical applications, due to the more efficient use of the data on extreme values.

Within the POT models framework, we can distinguish two types of analysis: (i) the Semi-Parametric Model built around the Hill estimator and (ii) and the fully parametric models based on the Generalized Pareto distribution (GPD). The latter method is commonly used in practice.

Below, the fundamental theory of this approach is described, considering both unconditional and conditional versions.

Let $r_1, r_2, \ldots \ldots r_n$ be a random sequence of observations representing the financial returns, given a threshold denoted by u, we will be interested in excess losses over the threshold[5] denoted by $y_1, y_2, y_3, \ldots y_{N_u}$, where $y_i = r_i - u$ and N_u are the number of sample data greater than u.

Then, for instance, if the threshold is equal to 1.5 %, we are left with all returns lower to 1.5 %.

The Extreme Value Theory assumes that the distribution of excess losses above the threshold is a Generalized Pareto distribution given for the following expression:

$$G_{k,\xi}(y) = 1 - [1 + \frac{k}{\xi} y]^{-1/k} \tag{8}$$

Where k y ξ are the parameters of the distribution[6]

It can be shown that under this assumption, the α percentile of the distribution, as the VaR can be estimated as:

[5] The most common is to use the 10% percentile as the threshold level.

[6] These parameters can be estimated by Maximum likelihood.

$$VaR(\alpha) = G_{k,\xi}^{-1}(\alpha) = u + \frac{\xi}{k}\left[\left[\frac{n}{N_u}(1-\alpha)\right]^{-\xi} - 1\right] \quad r < u \tag{9}$$

Where n represents the number of sample data.

The extreme value method described in the preceding paragraphs does not consider the level of volatility. This method is known as *Unconditional Extreme Value Theory*.

Since financial returns are variables that are characterized by heteroscedasticity, McNeil and Frey (2000) proposed a new methodology for estimating VaR that combines Extreme Value Theory with models of volatility, called *Conditional Extreme Value Theory*.

According to this theory, the VaR of a portfolio at a confidence level of $1-\alpha$ can be calculated as:

$$VaR(\alpha)_t = \mu + \sigma_t G_{k,\xi}^{-1}(\alpha) \tag{10}$$

Where σ_t^2 represents the conditional standard deviation of the financial returns and $G_{k,\xi}^{-1}$ is the α quantile of the GPD.

To estimate the conditional standard deviation of the returns in this paper we use the beta-t-EGARCH model and the beta-skewness-t-EGARCH models presented in Section 2.

3.2 Backtesting

To the adequacy of the VaR estimates we use two alternative approaches: statistical tests that evaluate the accuracy of the estimates and loss functions.

To test the accuracy of the VaR estimates, we use some standard tests: (i) Unconditional Coverage test (Kupiec, 1995), (ii) Conditional Coverage test (Christoffersen, 1998) and (iii) Dynamic Quantile (Engle and Manganelli, 2004).

To implement these tests, we must first define an exception indicator. This indicator is calculated as follows:

$$I_{t+1}\begin{cases}1 & si \quad r_{t+1} < VaR(\alpha)\\ 0 & si \quad r_{t+1} > VaR(\alpha)\end{cases} \tag{11}$$

Kupiec (1995) shows that if we assume that the probability of getting an exception is constant, then the number of exceptions $x = \sum I_{t+1}$ follows a binomial distribution $B(N, \alpha)$, where N is the number of observations. An accurate VaR measure should produce Unconditional Coverage ($\hat{\alpha} = \frac{\sum I_{t+1}}{N}$) equal to α percent.

The Unconditional Coverage test has a null hypothesis $\hat{\alpha} = \alpha$, with a likelihood ratio statistic:

$$LR_{uc} = 2[log(\hat{\alpha}^x (1-\hat{\alpha})^{N-x}) - log\,(\alpha(1-\alpha)^{N-x})] \qquad (12)$$

Which follows an asymptotic $\chi^2(1)$ distribution.

The Conditional Coverage test, developed by Christoffersen (1998), jointly examines whether the percentage of exceptions is statistically equal to the one expected $(\hat{\alpha} = \alpha)$ and the serial independence of the exception indicator.

The likelihood ratio statistic of this test is given by $LR_{cc} = LR_{uc} + LR_{ind}$ which is asymptotically distributed as $\chi^2(2)$ and the LR_{ind} statistic is the likelihood ratio statistic for the hypothesis of serial independence against first-order Markov dependence.

A similar test for the significance of the departure of $\hat{\alpha}$ from α is the backtesting criterion statistic (BTC):

$$Z = (N\hat{\alpha} - Na) / \sqrt{Na\,(1-a)} \qquad (13)$$

which follows an asymptotic N(0,1) distribution.

Finally, the Dynamic Quantile test proposed by Engle and Manganelli (2004) examines if the exception indicator is uncorrelated with any variable that belongs to the information set Ω_{t-1} available when the VaR is calculated. This is a Wald test of the hypothesis that all slopes are zero in the regression model:

$$I_t = \beta_0 + \sum_{i=1}^{p} \beta_i I_{t-i} + \sum_{j=1}^{q} \mu_j X_{t-j} \qquad (14)$$

where X_{t-j} are explanatory variables contained in Ω_{t-1} .

This statistic introduced as explanatory variables lags of VaR. Under H_0 the exception indicator cannot be explained by the level of VaR, i.e. $VaR(\alpha)$ is usually an explanatory variable to test if the probability of an exception depends on the level of the VaR.

Additionally, we evaluate the magnitude of the losses experienced. For this purpose, we have considered two loss functions: the regulator's loss function and the firm's loss function.

Lopez (1998, 1999) proposed a loss function that reflects the utility function of a regulator. This function assigns a quadratic specification when the observed portfolio losses exceed the VaR estimate. Thus, we penalize only when an exception occurs according to the following specification:

$$RLF_t : \begin{cases} (VaR_t - r_t)^2 \; if \;\; r_t < VaR(\alpha) \\ 0 \qquad\qquad\qquad\quad otherwise \end{cases} \tag{15}$$

This loss function gives higher scores when failures take place and it considers the magnitude of these failures. In addition, the quadratic term ensures that large failures are penalized more than small failures.

This could be optimal from the perspective of the regulator, but not from the point of view of the firm since a model that generates a VaR too high leads the firm to incur high capital costs.

Taking this into account, Sarma et al. (2003) define the firm's loss function as follows:

$$FLF_t : \begin{cases} (VaR_t - r_t)^2 \; if \;\; r_t < VaR(\alpha) \\ -\beta VaR_t \qquad\qquad\quad otherwise \end{cases} \tag{16}$$

where β is the opportunity cost of capital.

This function penalizes the cases in which no exceptions are multiplying the VaR by a factor β. As Abad et al. (2014) indicate, this product does not capture the opportunity cost of capital precisely. Therefore, in line with Sarma, they propose a new firm's loss function that is expressed as follows:

$$FABL_t : \begin{cases} (VaR_t - r_t)^2 \; if \;\; r_t < VaR(\alpha) \\ \beta\,(r_t - VaR_t) \quad\;\; otherwise \end{cases} \tag{17}$$

As can be determined in this function, the exceptions are penalized as usual in the literature, following the instructions of the regulator. When there are no exceptions, the loss function penalizes the difference between the VaR and returns weighted by a factor that represents an interest rate. This product is exactly the opportunity cost of the capital, ie, the excess capital held by the firm.

Finally, we evaluate the VaR estimate on the bases of daily capital requirement (see McAleer et al., 2013). These authors adapt to daily terms the function used by the financial institutions for calculating market risk capital requirement in a 10 days horizon (Basel II). The daily capital requirement at time t can be calculated as follow (BCBS, 1996; 2006):

$$DCR_t = sup\{-k \times \overline{VaR}_{60}, -VaR_{t-1}\} \tag{18}$$

where DCR_t represents the daily market capital requirement at time t, which is the higher of $k \times \overline{VaR}_{60}$ and $-VaR_{t-1}$; $\overline{VaR}_{60}$ is the mean VaR over the previous 60 working days; $(3 \leq k \leq 4)$ is the Basel II violation penalty (see Table 1).

4. Empirical Application.

4.1 Data analysis.

The data set consists of the daily returns of six companies from 01/04/2000 to 31/12/2013 (approximately 3520 observations). The companies are ADP, Amazon, Apple, Cerner, Microsoft and Telefonica. The computation of the returns (r_t) is based on the formula $r_t = ln(P_t) - ln(P_{t-1})$ where P_t is the price at time t. The evolution of daily prices and yields is represented in Figure 1.

Table 2 contains descriptive statistics of the return series. For each company, the unconditional mean is very close to zero. The highest unconditional standard deviations are 3.68 (Amazon) and 2.74 (Cerner) and the lowest is 1.66 (ADP). For the rest series of returns, the standard deviation moves between 2.08 and 2.73. The skewness statistic is negative for four of the series considered. This means that in most cases the distribution of those returns is skewed to the left. For all the series, the excess kurtosis statistics are above 3, implying that the distributions of those returns have much thicker tails than the normal distribution.

In Figure 1 we observe that the range of the fluctuation of the returns changes over time and they evolve according to the idea of a cluster in volatility (Mandelbrot, 1963). To capture this and other characteristics of the return as the leverage effect (Black, 1976) we use the beta-t-EGARCH model, proposed by Creal et al. (2008, 2011) and Harvey and Chakravarty (2008), and the beta-skewness-t-EGARCH proposed by and Harvey and Sucarrat (2013). In Table 3 we present the coefficient estimations of both models for each asset in the whole period.

All the parameters estimated are positive and statistically significant. Just in the case of Amazon k_2 is not significant. In addition, the estimations of φ_2 and φ_1 satisfy the identifiability and stationary conditions ($\varphi_2 < \varphi_1 \leq 1$). The parameter k^*, which captures the "leverage effect", is positive in all cases, indicating that volatility tends to be higher after negative returns. The value of this parameter moves around 0.015 (Microsoft) and 0.034 (Apple). To last, in the case of the beta-skewness-t-EGARCH model, the parameter γ is inferior to one (γ <1) in the case of ADP, Cerner, Apple and Telefónica. This means that the distributions of those assets are skewed to the left. For the case of Amazon and Microsoft, the distribution is skewed to the right (γ >1).

4.2 VaR applications

In this Section, the beta-t-EGARCH model and the beta-skewness-t-EGARCH model presented in Section 2 are used to calculate the VaR one day ahead at 1% probability and then these estimations are compared. The analysis period ran from January 1st, 2008, to December 31st, 2013. The comparison of the VaR estimates has been conducted in terms of evaluating the accuracy of the VaR estimates and the loss function. In Figure 2 we present the returns and the VaR estimates obtained from both volatility models for all assets considered. As we can observe, the risk assumed by the companies varies along the sample being especially high in 2008-2009. From the naked eye, it seems that there are no significant differences between the VaR estimation from both models.

To evaluate the accuracy of the VaR estimates, several standard tests are used. The results of these tests are presented in Table 4. For each index, it is presented the number and the percentage of exceptions obtained with each volatility model considered. The percentages of exceptions are marked in bold. Below the percentages, the statistics used to test the accuracy of the VaR estimates are presented. These statistics are as follows: (i) the Unconditional Coverage test (LRuc); (ii) Backtesting Criterion; (iii) statistics for serial independence (LRind); (iv) the Conditional Coverage test (LRcc) and (v) the Dynamic Quantile test (DQ). When the null hypothesis that "the VaR estimate is accurate" has not been rejected by any test, we shaded the region of the figure that represents the number and percentage of the exceptions.

The percentage of exceptions goes below 1% in almost all cases indicating that both models overestimate risk. Just in the case of Apple, the risk is underestimated. However, it is worth noting that the number and the percentage of exceptions are close to the expected one. The accuracy tests used to test formally the performance of the volatility models in terms of VaR corroborate this hypothesis. These results indicate that both models provide accurate VaR estimates in all cases.

Additionally, intending to detect differences between both models (beta-t-EGARCH and beta-skewness-t-EGARCH) we follow Gerlach et al. (2011) and focus on analyzing the ratio VRate/α and some statistics of it. This ratio is calculated as the quotient of percentage exception by the value of α, which is 1%. The beta-t-EGARCH model provides a VRate/α close to one for three assets (ADP, Cerner and Telefónica) the same as the beta-skewness-t-EGARCH (ADP, Cerner and Apple) (Table 5). Table 6 displays summary statistics for VRate/α for each model across the 6 assets. The Std(1) column shows the standard deviation from the expected ratio of 1 (not mean sample), while the 1st column counts the assets where the model ranked had VRate/α closest to 1. According to these statistics, the beta-skewness-t-EGARCH model provides better results as the mean of the ratio is closer to one and the std(1) is lower than those provided by the beta-t-EGARCH model.

Thus, although the evidence is weak the results indicate that in terms of accuracy, the beta-skewness-t- EGARCH model may outperform the beta-t-EGARCH which has been estimated under a symmetric distribution.

Another way to compare the VaR estimates that is often used in the VaR literature is through a loss function. The loss function measures the magnitude of the losses experienced. A model that minimizes the total loss is preferred to other models. For this purpose, we have considered two loss functions: the regulator's loss function proposed by Lopez (1998, 1999) and the firm's loss function (Abad et al. (2015)). The results of these loss functions are presented in Table 7[7].

According to the regulator´s loss function, there is no model superior to others. The beta-t-EGARCH model provides the lowest losses for ADP, Amazon and Telefónica while the beta-skewness-t-EGARCH model provides the lowest losses for Apple, Cerner and Microsoft. Thus, from the point of view of the regulator, both models seem to be equivalent. However, according to the firm´s loss function, which takes the opportunity cost of capital into account, the beta-skewness-t-EGARCH model outperforms the beta-t-EGARCH model by providing the lowest losses for all assets considered. Although in daily terms these differences are reduced, in annual and monetary terms these differences are higher. For instance, for a portfolio value of 10 million dollars, the differences in annual capital opportunity cost provide for both models to move around 3500 dollars (ADP) and 92250 dollars (Telefónica). Moreover, for a portfolio value of 100 million dollars, these differences move around 350000 and 952500 dollars[8]. These data reflect that although the differences in daily capital opportunity cost are small, in annual and monetary terms become significant.

To last, we compare VaR estimates in terms of daily capital requirement (DCR) which have been calculated according to Equation (18). The average of the DCR moves around 10% and 20% depending on the asset (Table 8). For almost all assets considered the Beta-t-skewness-EGARCH model provides the lowest daily capital requirement. The difference between these models moves around 0.01%-0.24% depending on the asset.

As a resume, we can conclude that in the framework of the Conditional Extreme Value Theory considering skewness t-student distributions for the returns may contribute to improving the accuracy of VaR estimations with respect to the symmetric student-t distribution. Furthermore, the results obtained by the loss function indicate that this kind of distribution may be preferred by financial companies, as they provide the opportunity

[7] In order to calculate the firm's loss function, we proxy the price of capital with the interest rate of the Eurosystem monetary policy operations of the European Central Bank.

[8] The annual capital opportunity cost is calculated by multiplying the average of the daily opportunity cost by 250 days. And the average of the daily opportunity cost is calculated as the data included in Table 7 (Panel b) divide by 100 times the portfolio value.

capital cost lowest. In addition, for banks, using a skewed student-t distribution enables them to maintain the lowest possible market risk capital requirements.

5. Conclusions

It is well documented in the literature that the financial return distribution is characterized as being skewed and exhibits an important excess of kurtosis. Thus, assuming a normal distribution for VaR estimation may take us to underestimate risk. Taking this into account, the research in the framework of a parametric method for VaR estimation has focused on investigating other density functions that capture the skew and kurtosis of financial returns. In this vein, recent papers show that in the context of this method assuming a fat tail and skewness distributions improve the results.

In the same line, we evaluate the role of the heavy tail and skewed distribution in VaR estimation in the framework of the Conditional Extreme Value Theory. Below the conditional EVT, the Value at Risk of a portfolio at $(1-\alpha)\%$ confidence level is calculated as the product of the conditional standard deviation of the return portfolio by the α (k_α) quantile of Generalized Pareto distribution. Traditionally, the conditional standard deviation of the return portfolio is estimated by assuming a symmetric distribution for the financial return, such as normal or student-t distribution. Thus in this paper, we analyze if, in the framework of this method, the estimation of the volatility model below a fat tail and skewness distribution contributes to improving the results in VaR estimation.

The study has been done for six individual assets bellowing to the telecommunication sector: ADP, Amazon, Cerner, Apple, Microsoft and Telefónica. The analysis period runs from January 1^{st,} 2008 to the end of December 2013. Although the evidence found is a little bit weak the results obtained seem to indicate that the heavy tail and skewed distribution outperform the symmetric distribution both in terms of accuracy VaR estimations as in terms of the firm’s loss function and capital requirement.

Table 1. Basel Accord Penalty Zones

Zone	Number of exceptions	k
Green	0 to 4	3
Yellow	5	3.4
	6	3.5
	7	3.65
	8	3.75
	9	3.85
Red	10 or more	4

Note: The number of exceptions is given for 250 trading days

Table 2. Descriptive Statistics

Assets	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque Bera
ADP	0.013	0.023	11.179	-26.892	1.665	-1.148*	26.678*	83004
						(0.041)	(0.083)	(0.001)
Amazon	0.042	-0.004	29.618	-28.457	3.681	0.450*	13.170*	15289
						(0.041)	(0.083)	(0.001)
Apple	0.066	0.077	12.239	-21.999	2.733	-0.446*	8.041*	3844
						(0.041)	(0.083)	(0.001)
Cerner	0.066	0.063	22.063	-19.159	2.746	0.441*	11.995*	11981
						(0.041)	(0.083)	(0.001)
Microsoft	-0.035	0.000	16.455	-18.623	2.081	-0.497*	12.480*	13327
						(0.041)	(0.083)	(0.001)
Telefónica	0.029	0.000	20.096	-21.279	2.474	-0.011*	8.483*	4410
						(0.041)	(0.083)	(0.001)

Note: This table presents the descriptive statistics of the daily percentage returns of ADP, Amazon, Apple, Cerner, Microsoft and Telefónica. The sample period is from January 3rd, 2000 to December 31st, 2013. The returns are calculated as $Rt=100.\ln(P_t/P_{t-1})$. The standard error of the skewness and kurtosis coefficients are calculated as $\sqrt{6/n}$ and $\sqrt{12/n}$ respectively. The JB statistic is distributed as the Chi-square with two degrees of freedom; in brackets is their p-value. * denotes significance at 1% level.

Table 3. Estimation of the parameters

		δ (se)	φ1 (se)	φ2 (se)	k1 (se)	k2 (se)	κ* (se)	df (se)	γ (se)	LogL (BIC)
ADP (T=1511)	beta-t-EGARCH	0.759 (0.253)	1.000 (0.000)	0.986 (0.006)	0.012 (0.006)	0.018 (0.007)	0.024 (0,004)	5.352 (0.458)	--	-5981.5 (3.416)
	beta-skewness-t-EGARCH	0.357 (0.143)	0.988 (0,009)	0.997 (0,001)	0.024 (0,011)	0.008 (0,011)	0.022 (0,003)	5.386 (0,464)	0.953 (0,020)	-5982.1 (3,418)
Amazon (T=1511)	beta-t-EGARCH	1.587 (0.234)	1.000 (0.000)	0.922 (0.051)	0.022 (0.006)	0.011 (0.009)	0.024 (0.007)	4.179 (0.269)	--	-8522.3 (4.860)
	beta-skewness-t-EGARCH	1.592 (0.236)	1.000 (0.000)	0.899 (0.052)	0.023 (0.005)	0.010 (0.009)	0.023 (0.007)	4.202 (0.270)	1.065 (0.023)	-8518.17 (4.860)
Apple (T=1511)	beta-t-EGARCH	1.435 (0,197)	1.000 (0.000)	0.882 (0.033)	0.015 (0.004)	0.025 (0.008)	0.034 (0.006)	6.136 (0.567)	--	-8013.3 (4.571)
	beta-skewness-t-EGARCH	1.432 (0.196)	1.000 (0.000)	0.885 (0.032)	0.015 (0.004)	0.026 (0.008)	0.034 (0.006)	6.115 (0.0566)	0.978 (0.022)	-8012.9 (4.573)
Cerner (T=1511)	beta-t-EGARCH	1.462 (0.221)	1.000 (0.000)	0.964 (0.012)	0.014 (0.004)	0.027 (0.008)	0.023 (0.004)	4.742 (0.354)	--	-7700.91 (4.393)
	beta-skewness-t-EGARCH	1.465 (0.221)	1.000 (0.000)	0.964 (0.012)	0.014 (0.004)	0.027 (0.008)	0.024 (0.005)	4.726 (0.354)	0.989 (0.021)	-7700.8 (4.395)
Microsoft (T=1511)	beta-t-EGARCH	0.826 (0.304)	1.000 (0.001)	0.971 (0.013)	0.019 (0.005)	0.030 (0.008)	0.015 (0.004)	4.780 (0.350)	--	-6675.8 (3.810)
	beta-skewness-t-EGARCH	0.319 (0.121)	0.943 (0.023)	0.996 (0.002)	0.032 (0.008)	0.020 (0.006)	0.015 (0.004)	4.830 (0.356)	1.004 (0.022)	-6672.8 (3.811)
Telefónica (T=1511)	beta-t-EGARCH	1.185 (0.192)	1.000 (0.000)	0.972 (0.008)	0.007 (0.003)	0.036 (0.006)	0.020 (0.004)	8.034 (0.927)	--	-7716.9 (4.402)
	beta-skewness-t-EGARCH	1.193 (0.191)	1.000 (0.000)	0.973 (0.008)	0.007 (0.003)	0.037 (0.006)	0.021 (0.004)	8.084 (0.945)	0.973 (0.022)	-7716.2 (4.404)

Note: The table reports the parameter estimates of the beta-t-EGARCH model (Equation 2) and the beta-skewness-t-EGARCH model (Equation 4). (se) denote the standard deviation (in parentheses). Log-L is the maximum likelihood value and BIC is the Bayesian Information Criterion. κ* is the parameter that captures the "leverage effect". γ is the parameter that captures skewness in distribution. γ <1, (γ >1) denotes skewed to the left (right).

(i) beta-t-EGARCH model

$$r_t = exp(\lambda_{t|t-1})(\varepsilon_t), \varepsilon_t \sim t(0, \sigma_\varepsilon, \nu) \qquad \nu > 2$$

$$\lambda_{t|t-1} = \delta + \lambda^{+}_{1,t|t-1} + \lambda^{+}_{2,t|t-1}$$

$$\lambda^{+}_{1,t|t-1} = \varphi_1 \lambda^{+}_{1,t-1|t-2} + k_1 u_{t-1}$$

$$\lambda^{+}_{2,t|t-1} = \varphi_2 \lambda^{+}_{2,t-1|t-2} + k_2 u_{t-1} + k^{*}(-r_{t-1})(u_{t-1} + 1)$$

(ii) beta-skewed-t-EGARCH model

$$r_t = exp(\lambda_{t|t-1})(\varepsilon_t - \mu_\varepsilon), \varepsilon_t \sim st(\mu_\varepsilon, \sigma_\varepsilon, \nu, \gamma) \qquad \nu > 2,\ \gamma\, \epsilon (0, \infty)$$

$$\lambda_{t|t-1} = \delta + \lambda^{+}_{1,t|t-1} + \lambda^{+}_{2,t|t-1}$$

$$\lambda^{+}_{1,t|t-1} = \varphi_1 \lambda^{+}_{1,t-1|t-2} + k_1 u_{t-1}$$

$$\lambda^{+}_{2,t|t-1} = \varphi_2 \lambda^{+}_{1,t-1|t-2} + k_2 u_{t-1} + k^{*}(-(r_{t-1}))(u_{t-1} + 1)$$

Table 4. Accuracy test

	Beta-t-EGARCH	Beta- skewness-t-EGARCH
ADP		
Nº exceptions	11	11
% exceptions	**0.73**	**0.73**
LR uc	0.46	0.46
BTC	0.23	0.23
LR ind	0.79	0.79
LR cc	0.74	0.74
DQ	0.98	0.08
Amazon		
Nº exceptions	10	10
% exceptions	**0.66**	**0.66**
LR uc	0.35	0.35
BTC	0.17	0.17
LR ind	0.81	0.81
LR cc	0.63	0.63
DQ	0.99	0.98
Apple		
Nº exceptions	18	16
% exceptions	**1.19**	**1.06**
LR uc	0.63	0.88
BTC	0.30	0.39
LR ind	0.66	0.70
LR cc	0.81	0.92
DQ	0.53	0.98
Cerner		
Nº exceptions	12	12
% exceptions	**0.79**	**0.79**
LR uc	0.58	0.58
BTC	0.29	0.29
LR ind	0.77	0.77
LR cc	0.82	0.82
DQ	0.14	0.14
Microsoft		
Nº exceptions	12	12
% exceptions	**0.79**	**0.79**
LR uc	0.58	0.58
BTC	0.29	0.29
LR ind	0.77	0.77
LR cc	0.82	0.82
DQ	1.00	1.00
Telefónica		
Nº exceptions	14	17
% exceptions	**0.93**	**1.13**
LR uc	0.85	0.75
BTC	0.38	0.35
LR ind	0.74	0.68
LR cc	0.93	0.87
DQ	0.84	0.89

Note: VaR violation ratios of the daily returns (%) are boldfaced. The table reports the p-values of the following tests: (i) the unconditional coverage test (LRuc); (ii) the backtesting criterion (BTC); (iii) statistics for serial independence (LRind), (iv) the conditional coverage test (LRcc) and (v) Dynamic Quantile test (DQ). A p-value greater than 5% indicates that the forecasting ability of the VaR model is accurate. The shaded cells indicate that the null hypothesis that the VaR estimate is accurate is not rejected by any test.

Table 5. Ratio Vrate/alpha at alpha=1% for each VaR model

	ADP	Amazon	Apple	Cerner	Microsoft	Telefónica
Beta-t-EGARCH	0.73	0.66	1.19	0.79	0.79	0.93
Beta-skewness-t-EGARCH	0.73	0.66	1.06	0.79	0.79	1.13

Note: Shaded cells indicate closest to 1 in that index.

Table 6. Summary statistics for the ratio Vrate/alpha at alpha=1%

	Mean	Median	Std(1)	1st
Beta-t-EGARCH	0.82	0.79	0.25	3
Beta-skewness-t-EGARCH	0.84	0.79	0.23	3

Note: Shaded cells indicate the most favored in each column. Std (1) is the standard deviation in ratios from an expected value of 1. 1st indicates the number of markets where that model's VRate/α ratio ranked closest to 1.

Table 7. Loss functions

Panel (a): Lopez Loss Function (%)

	ADP	Amazon	Apple	Cerner	Microsoft	Telefónica
Beta-t-EGARCH	0.0721	0.5856	3.1905	0.8017	1.3473	0.7175
Beta-skewness-t-EGARCH	0.0793	0.6757	3.1416	0.8002	1.2492	0.7367

Panel (b): ABL Loss Function (%)

	ADP	Amazon	Apple	Cerner	Microsoft	Telefónica
Beta-t-EGARCH	0.0708	0.1312	0.10218	0.09079	0.09398	0.10219
Beta-skewness-t-EGARCH	0.07066	0.1301	0.102	0.09037	0.0938	0.09838

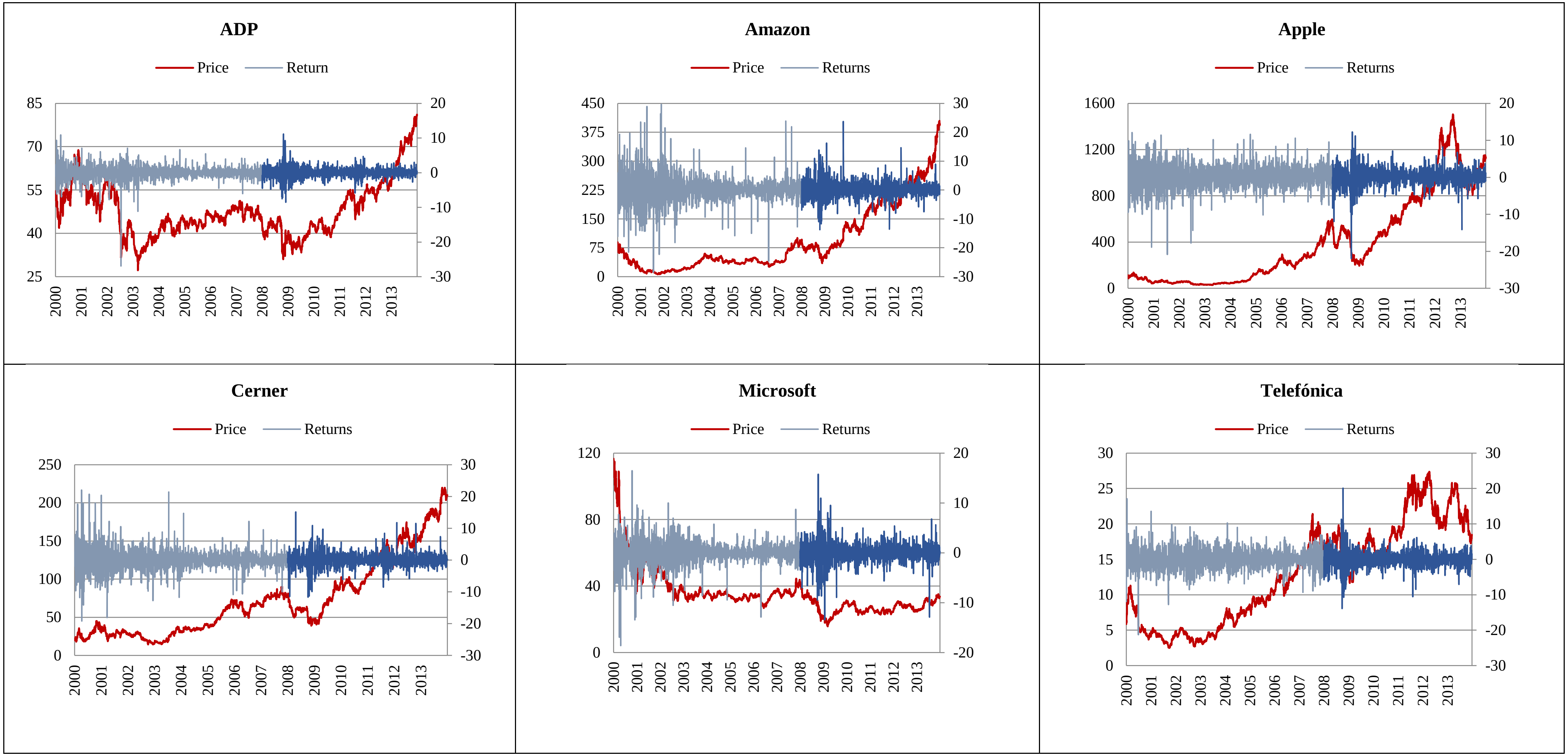
The Table reports the values of the different loss functions of each VaR model at 99% confidence levels. In both case Table shows the average of the losses. The shaded cells denote the minimum value for the different loss functions.

Table 8. Daily Requirement Capital

	ADP	Amazon	Apple	Cerner	Microsoft	Telefónica
Beta-t-EGARCH	11.70%	20.98%	16.78%	15.92%	15.10%	16.69%
Beta-skewness-t-EGARCH	11.82%	20.72%	16.57%	15.88%	15.11%	16.36%

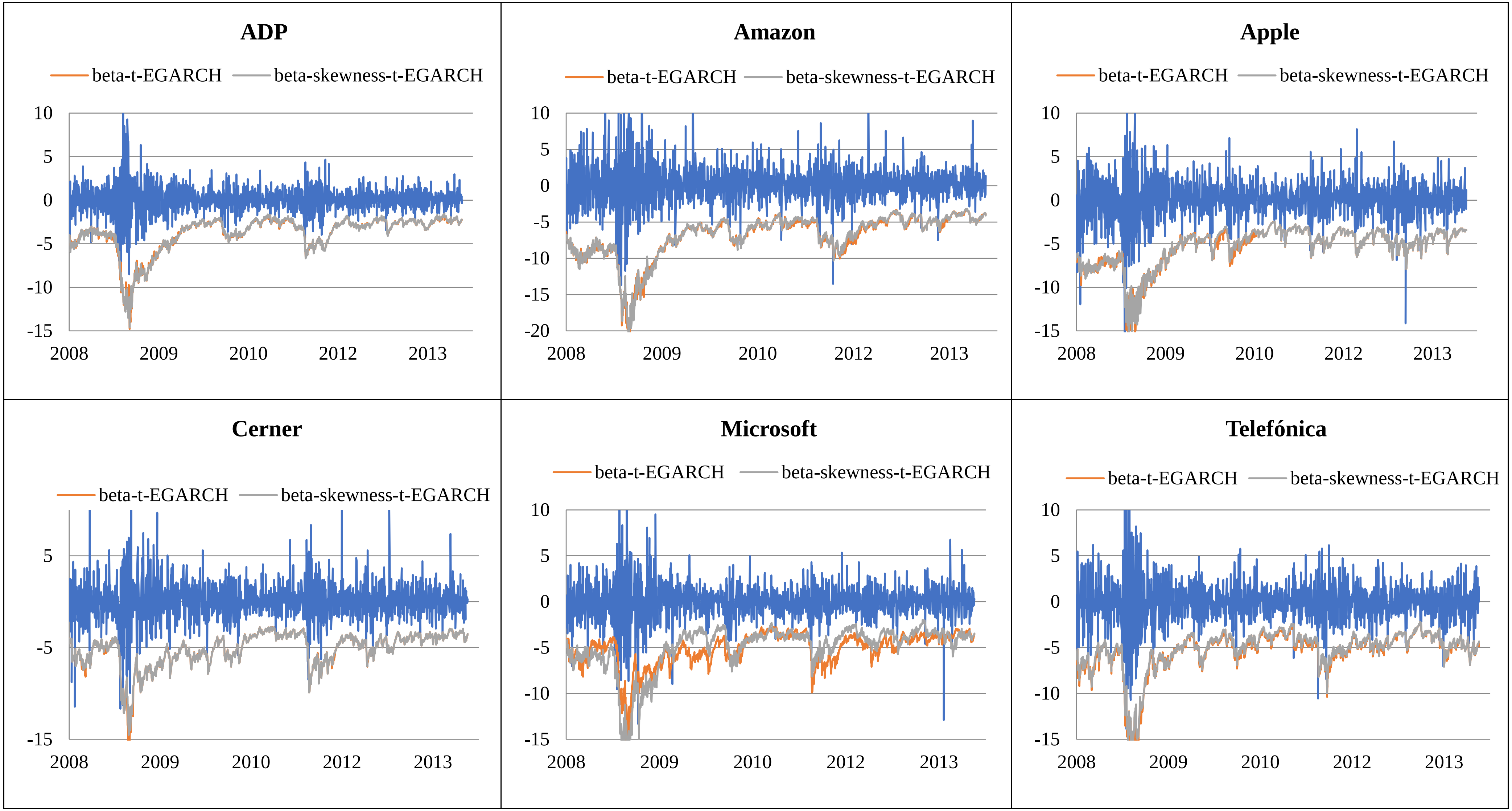
Note: The table reports the average daily capital requirement (DCR) obtained according to Equation (18). For each asset the shaded cells denote the model that provides the lowest average of the DCR.

Figure 1. Evolution of daily prices and yields.



Note: This figure illustrates the daily evolution of price (red) and returns (blue) of six assets (ADP, Amazon, Apple, Cerner, Microsoft and Telefonica) from January 3rd, 2000 to December 31st, 2013.

Figure 2. Returns and VaR estimations



Note: This figure illustrates the return (blue) and VaR estimations obtained from the beta-t-EGARCH model and the beta skewness-t- EGARCH model. The analysis period goes from January 1st 2008 to December 31st, 2013.

Chapter III*

Assessing the importance of the choice threshold in quantifying market risk under the POT approach (EVT)

Abstract

From a theoretical point of view, the selection of thresholds is a critical issue in the framework of the Peaks Over Threshold (POT) approach, which is why in the last decade numerous methodologies have been proposed for its selection. In this paper, we address this subject from an empirical point of view by assessing to what extent the selection of the threshold is decisive in quantifying the market risk. For measuring market risk, we use the Value at Risk (VaR) and the Expected Shortfall (ES) measures. The results obtained indicate that there is a large set of thresholds that provide similar Generalized Pareto Distribution (GPD) quantiles estimators and as a consequence similar market risk measures. Just only, for large thresholds, those corresponding to the 98^{th} and 99^{th} percentile of the GPD some differences are found. It means that the choice of threshold in the framework of the POT method may not be relevant in quantifying market risk when we use the VaR and ES measures for this task.

Keywords: Extreme Value Theory, Peaks over Threshold, Value at Risk, Expected Shortfall, Generalized Pareto Distribution.

Acknowledgments. The authors gratefully acknowledge the financial support granted by Ministerio de Ciencia e Innovación PID2020-113367RB-I00 and Young Talent Research Project UNED 2022 (076-044355 ENER-UK).

*Published in the journal *Risk Managemet*; Vol. 25, No. 6; 2023. ISSN 1460-3799 E-ISSN 1743-4637. Palgrave Macmillan-Springer Nature. Online Published: January 6, 2023. DOI: 10.1057/s41283-022-00106-w

*Also it was published as a working paper. The complete reference is Benito, S., López-Martín, C. and Navarro, M. (September 2018). Instituto Complutense de Análisis Económico (ICAE). WP nº 1820. ISSN: 2341-2356 Universidad Complutense de Madrid.

Chapter III. Assessing the importance of the choice threshold in quantifying market risk under the POT approach (EVT).

1. Introduction

Extreme value analysis has wide applications in many fields, such as civil engineering (Wu and Qiu, 2018), climatology (Davison and Smith, 1990; Kharin et al. 2013), seismology (Beirlant et al. 2018), hydrology (Katz et al. 2002, Carreau et al. 2017, Bader et al. 2018), insurance (Reiss and Thomas, 2007) and finance (Embrechts et al., 1997, Fontnouvelle et al., 2007 and Abad and Benito, 2013, among others). For instance, the extreme values of vehicle load play an important role in bridge design and risk assessment (Wu and Qiu, 2018). In seismology and climatology, extreme value analysis is used to study earthquakes (Beirlant et al. 2018) and extreme precipitation (Bader et al. 2018). In hydrology, extreme value analysis is an important tool for studying coastal flood risk (Haigh et al. 2010; McMillan et al. 2011). In the field of finance, extreme value modeling is important to quantify large financial losses from different sources of risk: operational, credit and market risk (see Cruz, 2002; Moscadelli, 2004; Fontnouvelle et al., 2007; Ergün and Jun, 2010; Žiković and Aktan, 2009; and Abad and Benito, 2013).

Traditionally, the study of the extreme values of fat tail distributions has been carried out through the Extreme Value Theory (EVT). EVT comprises mainly two approaches – the Block Maxima Method (BMM) and the Peaks over Threshold (POT) approach-. In the former, the data set is divided into blocks, and a Generalized Extreme Value (GEV) distribution is fitted to the sample of maximums or minimums extracted from these blocks. In the context of the POT method, a threshold is determined above which the excesses are fitted with the Generalized Pareto distribution (GPD) (Queensley et al. 2019).

Although the BMM and POT approaches should lead asymptotically to the same results, in practice the POT provides more suitable extreme quantile estimations due to the more efficient use of the data for the extreme values, see Cunnane (1973) and Madsen et al. (1997a). These authors show that the POT approach performs better than BMM, independently of the estimation method used. Similar results have been reported by Wang (1991), Madsen et al. (1997b).

From a theoretical point of view, threshold selection is a critical issue in the framework of the POT approach. The choice of threshold must be a balance between bias and variation. A threshold being too low is likely to violate the asymptotic basis of the model which leads to bias. However, a threshold being too high will generate few excesses leading to an increase in the variance of the estimators (Davison and Smith, 1990; Coles, 2001; MacDonald et al.,2011; Papalexiou et al.,2013; Wyncoll and Gouldby, 2013).

That is why, within the framework of the POT method, different methods have been developed for the selection of the suitable threshold. Those methods can be divided into two groups: (i) graphical approaches, based on a visual inspection of plots, such as the mean excess plot (Davison and Smith, 1990), stability parameters plot (Coles, 2001) and Hill plot (Drees et al., 2000) among others and (ii) numerical approaches (Ferreira et al. 2003; Thompson et al., 2009; Northrop and Coleman, 2014; Li et al., 2014, Wadsworth and Tawn, 2012, Naveau et al., 2016) which are more objective methods. Recently, new methods have been developed with the aim of automating some of the existing proposals, especially those based on visual data inspection, see for instance Wu and Qiu (2018), Bader et al. (2018), Caballero-Megido et al. (2017) and Queensley et al. (2019) among others.

The aforementioned papers focus on studying new methods for the selection of the optimal threshold, assuming that the estimates of the quantiles of the Generalized Pareto distribution are highly sensitive to the threshold choice in which case such efforts would be fully justified (Langousis et al., 2016). In the field of finance, the existing literature on this issue is quite scarce, especially in the area of market risk management. As far as we know in this field there are no studies on this subject. To cover this gap, we carry out an empirical analysis with a double aim. First, to analyze the sensibility of the GPD quantiles to the threshold choice and second, to study the sensitivity of the market risk measure to this choice. For measuring market risk, we use the Value at Risk (VaR)[1] and the Expected Shortfall (ES)[2] measures. To last, we calculate the market risk capital requirements and evaluate their sensitivity to the threshold choice.

---

[1] The VaR of a portfolio is defined as the worst expected loss over a given horizon under normal market conditions at a given level of confidence. Formally speaking, the VaR(α) of a portfolio at (1-α)% confidence level is the percentile α % of the return portfolio distribution.

[2] The ES is defined as the average of all losses that are greater than or equal to VaR, i.e., the average loss in the worst α % cases. In other words, this measure provides the expected value of an investment in the worst α % of the cases.

This study is in accordance with Langousis et al. (2016) who remarked that "*the variety of existing methods for threshold chosen, the fundamental differences in their theoretical underpinnings, and their relative performance when dealing with different types of data, make threshold detection an open question that can be addressed solely on the basis of a specific application*". With regard to this, we think, that the area of market risk management -where daily data and large samples are used- could give rise to different results from those obtained in other areas of science where the periodicity of the data is annual and consequently the size of the samples is reduced.

Thus, in this paper we analyze in detail the case of the S&P500 and later extend that study to a set of 14 assets from alternative markets: seven stock indexes (CAC40, DAX30, FTSE100, HangSeng, IBEX35, Merval and Nikkei), four commodities (Copper, Gold, Crude Oil Brent and Silver) and three rates exchange (£ /€, $/€ and ¥/€). The results obtained indicate that there is a large set of thresholds that provide similar GPD quantiles estimators and, as a consequence, similar market risk measures. Just only for large thresholds, those corresponding to the 98th and 99th percentile of the GPD some differences are found. It means that the choice of threshold in the framework of the POT method may not be relevant in quantifying market risk when we use the VaR and ES measures for this task. Regards to the market risk capital requirement, we find that these charges do not differ much among the thresholds. Nevertheless, if the objective of the financial institutions is to minimize these charges, they might be interested in the selection of a specific threshold.

The remainder of the paper is organized as follows. In Section 2, we show the methodology used for the study. In Section 3, we submit the data and the results obtained for the particular case of the S&P500 index. Section 4 displays a robustness analysis. The main conclusions are presented in Section 5.

2. Methodology

2.1 Extreme Value Theory

The Extreme Value Theory (EVT) studies the asymptotic behavior of extreme values of a random variable. This theory has wide applications in many fields, such as civil engineering (Wu and Qiu, 2018), climatology (Davison and Smith, 1990; Kharin et al. 2013), seismology (Beirlant et al. 2018), hydrology (Katz et al. 2002, Carreau et al.

2017, Bader et al. 2018), insurance (Reiss and Thomas, 2007) and finance (Embrechts et al., 1997), among others.

Within the EVT context, there are two approaches that study extreme events. The first one, based on the Generalized Extreme Value (GEV) distribution, models the distribution of the minimum or maximum realizations and it is known as the Block Maxima (Minima) Method (BMM). The second one is the Peaks Over Threshold (POT) approach based on the Generalized Pareto distribution (GPD) (Pickans, 1975) which models the exceedances over a particular threshold. In the next lines, we introduce these approaches.

2.1.1 Fisher –Tippett theorem

Suppose that $X_1, X_2, \dots, X_n$ is a sequence of independently and identically distributed random variables with a distribution function $F(x) = \Pr(X_t \leq x)$ and denote $M_n = \max(X_1, X_2, \dots, X_n)$ as a sample of maxima of this series, the distribution function (CDF) of M_n is represented as:

$$P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \prod_{i=1}^{n} F(x) = F^n(x) \tag{1}$$

Although $F(x)$ is unknown, Fisher and Tippet's theorem establishes an asymptotic approach for $F^n(x)$. This theorem establishes that given a sequence of $b_n > 0$, $a_n \in R$, the maximum normalized value $Z_n = \frac{M_n - a_n}{b_n}$ converges to a non-degenerated distribution H, and this distribution is the Generalized Extreme Value (GEV) distribution, $\lim_{n\to\infty} \Pr\left(\frac{M_n - a_n}{b_n} \leq x\right) \longrightarrow H(x)$.

The algebraic expression for such generalized distribution is as follows:

$$GEV_{\xi,\mu,\sigma}(x) = e^{-\left[1+\xi\frac{(x-\mu)}{\sigma}\right]^{-\frac{1}{\xi}}} \tag{2}$$

defined on $\left(1 + \frac{\xi(x-\mu)}{\sigma}\right) > 0$, where $\sigma > 0$ is the scale parameter, $-\infty < \mu < \infty$ is the location, and $-\infty < \xi < \infty$ is known as the shape parameter of the GEV distribution and characterizes the tail behavior of the distribution. The prior distribution is a generalization of the three types of distributions, depending on the value taken by ξ :

- Gumbel (ξ =0) type I family. It has light extremes, not heavy extremes

$$\Lambda(x) = e^{e^{-\frac{x-\mu}{\sigma}}} \qquad \forall x \in \Re$$

- Fréchet (ξ >0) type II family. This distribution is particularly useful for patterning financial returns as it has very heavy tails.

$$\Phi_{\xi,\mu,\sigma}(x) = \begin{cases} 0 & x \leq \mu \\ e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}} & x > \mu \end{cases}$$

- Weibull (ξ <0) type III family. This distribution is used when the extremes are lighter (softer) than those from the normal distribution, and thus, it is not particularly useful for applications related to financial yields (returns).

$$\Psi_{\xi,\mu,\sigma}(x) = \begin{cases} e^{-\left(-\frac{x-\mu}{\sigma}\right)^{\frac{-1}{\xi}}} & x \leq \mu \\ 1 & x > \mu \end{cases}$$

2.1.2 Peaks over Threshold Approach (POT)

In general, we are not only interested in the maxima of observations but also in the behavior of large observations which exceed a high threshold. One method of extracting extremes from a sample of observations, X_t, $t = 1, 2, \dots n$ with a distribution function $F(x) = \Pr(X_t \leq x)$ is to take the exceedances over a predetermined high threshold u. An exceedance of a threshold u occurs when $X_t > u$ for any t in $t = 1,2, \dots, n$. Thus, an excess over u is defined as $y = X_t - u$. This approach is known as POT.

Although the BMM and POT approaches should lead asymptotically to the same results, in practice the POT provides more suitable extreme quantile estimations due to the more efficient use of the data for the extreme values, see Cunnane (1973) and Madsen et al. (1997a). These studies show that the POT approach performs better than BMM, independently of the estimation method used. Similar results have been reported by Wang (1991), Madsen et al. (1997b), and Tanaka and Takara (2002) among others.

Let x_0 be the finite or infinite right endpoint of the distribution F. That is to say, $x_0 = \sup\{x \in R: F(x) < 1\} \leq \infty$. The distribution function of the excesses (y) over the threshold, u is given by $F_u(y) = P\big((X-u) \leq y \mid X > u\big)$ for $0 \leq x \leq x_0 - u$. Thus,

$F_u(y)$ is the probability that the value of X exceeds the threshold u by no more than an amount y, given that the threshold is exceeded. This probability can be written as:

$$F_u(y) = \frac{F(y+u) - F(u)}{1 - F(u)} \tag{3}$$

This distribution can be approximated by the Generalized Pareto distribution (GPD) which is usually expressed as a two-parameter distribution[3]:

$$G_{k.\xi}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-\frac{1}{\xi}} & if\ \xi \neq 0 \\ 1 - exp\left(-\frac{y}{\sigma}\right) & if\ \ \xi = 0 \end{cases} \tag{4}$$

where ξ and $\sigma > 0$ are the shape parameter and the scale parameter, respectively[4]. Note that if the distribution of M_n^* converges to a GEV distribution for block maxima with parameter ξ, then the distribution of exceedances over the threshold converges to the GPD with the same parameter ξ (Rodríguez G. 2017).

Using this approximation, the distribution function of X will be given by $F(x) = \big(1 - F(u)\big)\, F_u(y) + F(u)$. Replacing F_u(y) by GPD and $F(u)$ by its empirical estimator $(n - N_u)/n$, where n is the total number of observations and N_u the number of observations above the threshold u, we have

$$F(x) = 1 - \frac{N_u}{n}\left(1 + \frac{\xi}{\sigma}(x - u)\right)^{-\frac{1}{\xi}} \tag{5}$$

For a given probability $\alpha > F(u)$, the quantile α, which is denoted by q_α, is calculated by inverting the tail estimation formula to obtain

---

[3] The study of Jobst (2007) provides favorable evidence in favor of this approach.

[4] The traditional Extreme Value Theory (EVT) assumes that the data are stationarity. When stationarity is assumed, parameters that determine the distribution function (Generalized Pareto and Generalized Extreme Value distribution) are independent of time. However, in practice, it is often the case that stationarity assumptions (such as independence and identical distribution) for time series extremes are violated. If the process is non-stationary, the parameters of distributions are time-dependent, and the properties of the distribution vary with time. To capture the non-stationarity of extreme data, new approaches have been developed in the framework of the Extreme Value Theory. Some applications of these new approaches can be found in Cheng and AghaKouchak (2014), Cheng et al. (2014), Ruggiero et al. (2010), Chavez-Demoulin and Embrechts (2004) among others.

$$q_\alpha = u + \frac{\sigma}{\xi}\left(\left(\frac{n}{N_u}(1-\alpha)\right)^{-\xi} - 1\right) \quad (6)$$

The distributional choice is motivated by a theorem (Balkema and de Haan, 1974; Pickands, 1975) which states that, for a certain class of distributions, the GPD is the limiting distribution for the distribution of the excesses, as the threshold tends to the right endpoint:

$$\lim_{u \to x_0} sup\left|F_n(y) - GPD_{\xi,\sigma}(y)\right| = 0$$

This theorem is fulfilled if and only if F is in the maximum domain of attraction (MDA) of the Generalized Extreme Value distribution H_ξ, $(F \in MDA(H_\xi))$. It means that if, for a given distribution F, an appropriately normalized maximum sample converges to a non-degenerated distribution H_ξ, then this is equivalent to say H_ξ is the MDA for F for some value of ξ.

The class of distribution F for which the condition $F \in MDA(H_\xi)$ holds is large; essentially all commonly encountered continuous distributions show the kind of regular behavior for sample maximum described by Equation (1).

2.2 Threshold selection method

The approaches developed for selecting the suitable threshold can be divided into two groups: (i) subjective approaches based on graphical analysis, such as the mean excess plot, stability parameters plot and Hill plot among others and (ii) numerical approaches. In its turn, the latter can be divided into various categories: (a) non-parametric approach; (b) approaches based on goodness of fit test; (iii) mixture models; (iv) simple naïve methods; (v) computational approaches and (vi) other approaches. In the following lines, we describe briefly these methods (see Scarrot and McDonald, 2012 and Langousis et al., 2016 for a more detailed review of these methods).

2.2.1 Graphical approaches.

Due to its simplicity, the graphic method most commonly used for selecting threshold is the mean excess plot (MEP) also called mean residual life plot (MRLP) introduced by Davison and Smith (1990). This instrument is a graphical tool based on the sample means of the excesses function (SMEF), which is defined as:

$$SMEF(u) = \frac{\sum_{i}^{N_u}(r_i - u)_{\{r_i > u\}}}{N_u}$$

The sample means excess function is an estimate of the excess mean function (MEF), $e(u) = E[(X-u)|X>u]$. For the GPD, the excess mean function is given by a linear function in u[5]:

$$e(u) = \frac{\sigma}{1-\xi} + \frac{\xi}{1-\xi}u \qquad (7)$$

This finding means that for $0 < \xi < 1$ and $\sigma + u\xi > 0$, the mean excess plot should resemble a straight line with a positive slope. Empirical estimates of the sample mean excesses are typically plotted against a range of thresholds. Thus, the general rule for the choice of the optimal threshold will be to choose a value of u for which the resulting line has a positive slope. An application of this method can be found in Beirlant et al. (2004).

The main problem associated with the sample mean excess plot is subjectivity. As Queensley et al. (2019) remark, "*judging from where the graph is approximately linear using only the eyeball inspection approach, is a rather subjective choice so that different thresholds may be selected by different viewers of the plot*".

The second type of plot is the parameter stability plot (Coles, 2001) (shape and scale) created by fitting the GPD using a range of thresholds. This method involves plotting $\hat{\sigma}$ and $\hat{\xi}$ together with confidence intervals and selecting the value of u from which the estimates are no longer stable (see Coles, 2001). This type of plot may present some inconsistencies, showing different flat sections for different ranges of threshold (Scarrot and McDonald, 2012).

Other graphic approaches are these based on quantile plots and plots comparing the empirical cumulative distribution function and the cumulative GPD. According to quantile plots, the proper threshold is selected as the lowest threshold above which the plot shows a linear trend. When we compare the empirical with the theoretical distribution function the proper threshold is selected as the lowest threshold above which the

[5] If a distribution function is subexponential, the mean excess function tends to infinity, if it is an exponential distribution the mean excess function is a constant and for the normal distribution the mean excess function tends to zero.

differences between the empirical and the theoretical distribution function seem minimum. Hill plot, explored by Drees et al. (2000) can also be included in this group. The Hill plot plots the Hill estimator of the tail index for a set of thresholds. According to this plot, the optimal threshold is the lowest threshold at which the Hill estimator is stabilized. This tool suffers from many of the same benefits and drawbacks that the MEP, and has been referred to as the Hill horror plot by Resnick (1997).

2.2.2 Numerical approaches.

The approaches aforementioned are based on judgment (Caballero-Megido, 2017) so they can be rather subjective and require substantial expertise to interpret these diagnostics as a method of threshold selection (Davison and Smith, 1990; Coles, 2001; Solari and Losada, 2012a). To overcome these limitations some numerical approaches have been developed which lead to a more objective decision.

The numerical approaches are numerous and can be classified into different categories: (a) non-parametric approach; (b) approaches based on goodness of fit test; (c) simple naïve methods; (d) mixture models; (e) computational approaches and (e) other approaches. In the following lines, we resume each one of these categories.

(a) **Nonparametric methods** that are intended to locate the changing point between extreme and nonextreme regions of the data (see e.g., Gerstengarbe and Werner, 1989, 1991; Werner and Gerstengarbe, 1997; Domonkos and Piotrowicz, 1998; Lasch et al., 1999; Cebrián et al., 2003; Cebrián and Abaurrea, 2006; Karpouzos et al., 2010, among others).

(b) **Approaches based on the Goodness of fit test**, where the threshold is selected as the lowest level above which the GPD provides an adequate fit to the exceedances. To analyze the goodness of fit of the GPD, Kolmogorov-Smirnov test and Anderson-Darling test can be used. Applications of this method can be found in Davison and Smith (1990), Dupuis (1999), Choulakian and Stephens (2001), Northrop and Coleman (2014), Langousis et al. (2016) among others.

In this category, we also include the method based on the Root Mean Square Error (RMSE) proposed by Li et al.(2014). The RMSE measures the difference between analytical and observed CDFs of exceedances for different thresholds. The threshold with the lowest RMSE is considered the best one.

(c) **Simple naïve methods**. Given the general order statistic convergence properties, various rules of thumb have been derived from the literature. Simple fixed quantile rules, like the upper 10% rule of DuMouchel (1983). Ferreira et al. (2003) use the square root of the number of data (n) to specify the number of exceedances (N_u). Ho and Wan (2002) and Omran and McKenzie (1999) use the rule $N_u = \frac{n^{2/3}}{\log(\log(n))}$ proposed by Loretan and Philips (1994) to determine the optimal number of exceedances. Neftci (2000), followed by Bekiros and Georgoutsos (2005), proposes the estimation of the threshold as $1.176\ \sigma_0$, where σ_0 is the standard deviation of the sample. In other studies, these methods are classified as ad-hoc methods or rules of thumb.

(d) **Methods in the other category** are based on mixtures of a GPD for the tail and another distribution for the "bulk" joined at the threshold (e.g., MacDonald et al., 2011; Wadsworth and Tawn, 2012; Naveau et al., 2016). Treating the threshold as a parameter to estimate, these methods can account for the uncertainty from threshold selection in inferences. The major drawback of such models is their ad-hoc heuristic definitions, the asymptotic properties of which are still little understood. They have also not had time to be well used in practice and currently, there is no readily available software implementation to allow practitioners to gain wider experience (Scarrot and McDonald, 2012).

(e) **Computational approaches.** Other researchers have suggested using techniques that provide an optimal trade-off between bias and variance. This method involves using bootstrap simulations to numerically calculate the optimal threshold considering the trade-off between bias and variance. Applications of this method can be found in Danielsson et al. (2001), Drees et al. (2000), Ferreira et al. (2003), Hall (1990) and Bairlant et al. (2004). In general, the restrictive assumptions underlying these approaches hinder their wide applicability.

(f) **Other approaches.** Other approaches different from the aforementioned are proposed by Dupuis (2000), Thompson et al. (2009) and De Zea Bermudez et al. (2001). See Scarrot and McDonald (2012) for a detailed review of these methods.

Recently, new methods have been developed to automate some of the existing proposals, especially those based on visual data inspection, see for instance Wu and Qiu

(2018), Bader et al. (2018), Caballero-Megido et al. (2017) and Queensley et al. (2019), Schneider et al. (2021) among others. Wu and Qiu (2018) propose a method to select the suitable threshold based on multiple criteria decision analysis (MCDA). In MCDA, Chi-Square test, Kolmogorov Smirnov (K-S) test and Root Mean Square Error (RMSE) are combined as the test criteria and the weight of these criteria is calculated using the entropy method. Thus, the MCDA can integrate results obtained from the goodness-of-fit test under different criteria into a comprehensive one, which makes the selection more scientific and objective (Wang et al. 2009). Bader et al. (2018) develop an efficient technique to evaluate and apply the Anderson–Darling test to the sample of exceedances above a fixed threshold.

In order to automate threshold selection, this test is used in conjunction with a recently developed stopping rule that controls the false discovery rate in ordered hypothesis testing. Caballero-Megido et al. (2017) propose a new automated method that mimics the enduringly popular visual inspection method. The purpose of the automated graphic threshold selection (AGTS) method, in absence of a priori threshold value, is to guide in the choice of the threshold which requires judgment and expertise, making the process simple and approachable, whilst being reproducible and less subjective. Queensley et al. (2019) propose an alternative way of selecting the threshold where, instead of choosing individual thresholds in isolation and testing their fit, they make use of the bootstrap aggregate of these individual thresholds which are formulated in terms of quantiles. The method incorporates the visual technique and is aimed at reducing the subjectivity associated with solely using the eye inspection approach (EIA). Schneider et al. (2021) suggest a couple of automated methods for threshold selection. The first one consists in estimating and minimizing the integrated square error (ISE) between the exponential density and its parametric estimator employing the Hill estimator. This is based on the null hypothesis that the log-spacings between a sample of thresholds are indeed exponentially distributed. The error function that obtains is called the inverse Hill statistic (IHS). This method exhibits high fluctuations for small thresholds, which might make the automated selection of the minimum highly variable. To control this problem the authors propose a smooth IHS. The second method consists in look for a sample fraction of optimal thresholds that minimize the asymptotic mean squared error (AMSE) of the Hill estimator.

2.3 Risk measure

According to Jorion (2001), "VaR measure is defined as the worst expected loss over a given horizon under normal market conditions at a given level of confidence". Thus, VaR is a conditional quantile of the asset return loss distribution.

Let $X_1, X_2, \ldots, X_n$ be identically distributed independent random variables representing the financial returns. Using $F(x)$ to denote the cumulative distribution function, $F(x) = \Pr\,(X_t \leq x|\Omega_{t-1})$ conditioned to the information available at $t-1$ (Ω_{t-1}). Assume that $\{X_t\}$ follows the stochastic process given by

$$X_t = \mu_t + \tilde{\sigma}_t z_t \qquad z_t \sim iid(0{,}1) \tag{8}$$

where $\tilde{\sigma}_t^2 = E(z_t^2|\Omega_{t-1})$ and z_t has the conditional distribution function $G(z)$, $G(z) = \mathrm{P}(z_t < z|\Omega_{t-1})$. The VaR with a given probability $\alpha \in (0, 1)$, denoted by $VaR(\alpha)$, is defined as the α quantile of the probability distribution of financial returns $\mathrm{F}\big(\mathrm{VaR}_t(\alpha)\big) = \Pr\big(\mathrm{X_t} < \mathrm{VaR}_t(\alpha)\big) = \alpha$. In this paper, we use the POT approach to estimate the tail of the distribution of the standardized residuals and thus later estimate the risks measure. As the GPD is only defined for positive values, we multiply our data by (-1) and thus move the left tail to the right side. Therefore, the VaR of a portfolio at α% probability will be calculated as:

$$VaR_t(\alpha) = \mu_t + \tilde{\sigma}_t\, q_{1-\alpha} \tag{9}$$

where μ_t and $\tilde{\sigma}_t$ represent the conditional mean and the conditional standard deviation of the returns[6] and $q_{1-\alpha}$ is the quantile $(1-\alpha)$ of the GPD (Equation 6).

The Expected Shortfall (ES) with a given probability $\alpha \in (0, 1)$, denoted by $ES(\alpha)$, is defined as the average of all losses that are greater than or equal to VaR, i.e., the average loss in the worst α % cases:

$$ES_t(\alpha) = E[X|\, X \geq VaR(\alpha)] = \mu_t + \tilde{\sigma}_t E[z|\, z \geq q_{1-\alpha}\,] \tag{10}$$

[6] For estimating the volatility of the return, we use an APARCH model, which is given by the next expression: $\sigma_t^\delta = \alpha_0 + \alpha_1\,(|\varepsilon_{(t-1)}| - \gamma\varepsilon_{(t-1)})^\delta + \beta\sigma_{t-1}^\delta$, $\alpha_0, \beta, \delta > 0$, $\alpha_1 \geq 0$, $-1 < \gamma < 1$. In this model, the γ parameter captures the leverage effect (Black, 1976), which means that volatility tends to be higher after negative returns.

It can be demonstrated[7] that the mean of the excess distribution $F_{q_{1-\alpha}}(y)$ over the threshold $q_{1-\alpha}$ is given by:

$$E(z|z \geq q_{1-\alpha}) = \frac{q_{1-\alpha}}{1-\xi} + \frac{\sigma - \xi u}{1-\xi} \quad (11)$$

Replacing (11) in (10) we obtain the ES measure under the conditional EVT.

$$ES_t(\alpha) = E[X|\, X \geq VaR(\alpha)] = \mu_t + \tilde{\sigma}_t \left[\frac{q_{1-\alpha}}{1-\xi} + \frac{\sigma - \xi u}{1-\xi}\right] \quad (12)$$

2.4 Backtesting

2.4.1 Backtesting VaR

To evaluate the accuracy of the VaR estimates, several tests have been used. All of these tests are based on the indicator variable. We have an exception when $r_{t+1} < VaR_\alpha$; then, the exception indicator variable (I_{t+1}) is equal to one (zero in other cases).

To check the accuracy of the VaR estimates, we have used four standard tests: unconditional (LR_{uc}), independent and conditional coverage (LR_{ind} and LR_{cc}) and dynamic quantile (DQ) tests.

Kupiec (1995) shows that if we assume that the probability of obtaining an exception is constant, the number of exceptions $x = \sum I_{t+1}$ follows a binomial distribution $B(N, \alpha)$, where N represents the number of observations. An accurate measure VaR_α should produce an unconditional coverage $\left(\hat{\alpha} = \frac{\sum I_{t+1}}{N}\right)$ equal to α percent. The unconditional coverage test has a null hypothesis $\hat{\alpha} = \alpha$, with a likelihood ratio statistic:

$$LR_{uc} = 2[log(\hat{\alpha}^x(1-\hat{\alpha})^{N-x}) - log(\alpha^x(1-\alpha)^{N-x})] \quad (13)$$

which follows an asymptotic $\chi^2(1)$ distribution. The conditional coverage test, developed by Christoffersen (1998), jointly examines whether the percentage of exceptions is statistically equal to the one expected ($\hat{\alpha} = \alpha$) and the serial independence

[7] A more detailed theoretical development can be found in McNeil et al. (2005), Chapter 7.

of the exception indicator. The likelihood ratio statistic of this test is given by $LR_{cc} = LR_{uc} + LR_{ind}$, which is asymptotically distributed as $\chi^2(2)$, and the LR_{ind} statistic is the likelihood ratio statistic for the hypothesis of serial independence against first-order Markov dependence[8]. Finally, the DQ test proposed by Engle and Manganelli (2004) examines if the exception indicator is uncorrelated with any variable that belongs to the information set Ω_{t-1}, available when the VaR is calculated. This test is a Wald test of the hypothesis that all slopes are zero in the regression:

$$I_t = \beta_0 + \sum_{i=1}^{p} \beta_i I_{t-i} + \sum_{j=1}^{q} \mu_j X_{t-j} \tag{14}$$

where X_{t-j} are the explanatory variables contained in Ω_{t-1}. This statistic is introduced as five explanatory variable lags of VaR. Under the null hypothesis, the exception indicator cannot be explained by the level of VaR, i.e., $VaR(\alpha)$ is usually an explanatory variable to test if the probability of an exception depends on the level of the VaR.

2.4.2 Backtesting ES

In this paper, we use McNeil and Frey (2000) test for the conditional expected shortfall. This test is likely the most successful in the literature. These authors develop a test to verify that a model provides much better estimates of the conditional expected shortfall than any other. The authors are interested in the size of the discrepancy between the return r_{t+1} and the conditional Expected Shortfall forecast $ES_t(\alpha)$ in the event of quantile violation. The authors define the residuals as follows:

$$Y_{t+1} = \frac{r_{t+1} - ES_{t+1}(\alpha)}{\sigma_{t+1}} \tag{15}$$

[8] The LR$_{ind}$ statistic is $LR_{ind} = 2[\log L_A - \log L_0]$ and has an asymptotic $\chi^2(1)$ distribution. The likelihood function under the alternative hypothesis is $L_A = (1-\pi_{01})^{N_{00}}\, \pi_{01}^{N_{01}}\, (1-\pi_{11})^{N_{10}}\, \pi_{11}^{N_{11}}$, where N$_{ij}$ denotes the number of observations in state $_j$ after having been in state i in the previous period, $\pi_{01} = N_{01}/(N_{00}+N_{01})$ and $\pi_{11} = N_{11}/(N_{10}+N_{11})$.
The likelihood function under the null hypothesis ($\pi_{01} = \pi_{11} = \pi = (N_{11}+N_{01})/N$) is $L_0 = (1-\pi)^{N_{00}+N_{01}}\, \pi^{N_{01}+N_{11}}$.

Replacing Equation (8) and Equation (12) in Equation (15), we have the next expression:

$$y_{t+1} = z_{t+1} - E(z|z < q_\alpha) \tag{16}$$

It is clear that, under model (5), these residuals are *i.i.d*. and that, conditional on $\{r_{t+1} < VaR_{t+1}(\alpha)\}$ or equivalent $\{z_{t+1} < q_\alpha\}$, they have an expected value of zero. Suppose we again backtest on days in the set T. We can form empirical versions of these residuals on those specific days on which violations have occurred, i.e., days on which $\{r_{t+1} < VaR_{t+1}(\alpha)\}$. The authors call these residuals exceedances and denote them by $\{\hat{y}_{t+1}: t \in T.\ \ r_{t+1} < VaR_{t+1}(\alpha)\}$ where $\hat{y}_{t+1} = \frac{r_{t+1} - \widehat{ES}_{t+1}(\alpha)}{\hat{\sigma}_{t+1}}$ and $\widehat{ES}_{t+1}(\alpha)$ is an estimation of the conditional expected shortfall.

Under the null hypothesis, in which we correctly estimate the dynamic of the process μ_{t+1} and σ_{t+1} and the first moment of the truncated innovation distribution $E(z|z < q_\alpha)$, these residuals should behave such as an *i.i.d* sample with a mean of zero. Thus, for testing whether the estimates of the expected shortfall are correct, we must test if the sample mean of the residual is equal to zero against the alternative that the mean of y is negative. Indeed, given a sample $\{y_{t+1}\}$ of size N (where N is the number of violations in the T period), the sample mean $\bar{y}$ converges in distribution to standard normality, as N tends to ∞ by the central limit theorem. In other words, given mean μ_y and variance σ_y of population

$$\sqrt{N}\left(\frac{\bar{y} - \mu_y}{\sigma_y}\right) \to N(0,1) \tag{17}$$

By applying the central limit theorem, the statistics for testing the null hypothesis are given by

$$t = \frac{\bar{y}}{\frac{S_y}{\sqrt{N}}} \sim t_{N-1} \tag{18}$$

where $\bar{y}$ and S_y are the sample mean and the sample standard deviation, respectively, of the exceedance residuals.

2.5 Forecasting daily market risk capital charges

Basel II Accord required financial institutions to meet daily capital requirements based on VaR estimates (BCBS, 1996; 2006). The Basel II Accord specified that daily capital charges (DCC) must be set at the higher of the previous day's VaR or the average VaR over the last 60 business days, multiplied by a value between 3 and 4 depending on the number of violations (see Table 1) that occurred in the 250 days prior to the estimation of capital charges $DCC_t = sup\{-k \times \overline{VaR}_{60}, -VaR_{t-1}\}$.

Recently, the Basel Committee for Banking Supervision (BCBS) has promoted a change in international financial regulation. Under the new regulation based on the Basel solvency framework (BCBS, 2012, 2016, 2017, 2019), known as Basel III, financial institutions must calculate the market risk capital requirements based on the Expected Shortfall (ES) measure, replacing the Value at Risk (VaR) measure.

Following Chang et al. (2019), we evaluate the market risk capital requirement based on the ES measure, which is the market risk benchmark according to Basel III. Thus, the forecasting daily market risk capital requirement (DCR) at time *t* can be calculated as follow:

$$DCR_t = sup\{-k \times \overline{ES}_{60}, -ES_{t-1}\} \tag{19}$$

3. Case of study

3.1 Dataset overview

The data consist of the S&P500 stock index extracted from the Thomson-Reuters-Eikon database. The index is transformed into returns by taking the logarithmic differences of the closing daily price (in percentage). We use daily data for the period January 3rd, 2000, through December 30th, 2021. The sample size is 5534. Figure 1 shows the evolution of the daily index and returns of the S&P500. The index shows a sawtooth profile alternating periods with an upward slope with a period of sudden decreases.

In addition, we can observe that the range fluctuation of daily returns is not constant, which means that the variance of the returns changes over time. The volatility of the S&P500 was particularly high from 2008 to 2009, coinciding with the period known as the Global Financial Crisis, and in the first quarter of 2020, coinciding with the beginning of the COVID-19 pandemic. The basic descriptive statistics are provided in

Table 2. The unconditional mean daily return is very close to zero (0.021%) which is typical of daily returns. The skewness statistic is negative, implying that the distribution of daily returns is skewed to the left. The kurtosis coefficient shows that the distribution has much thicker tails than the normal distribution. Similarly, the Jarque-Bera statistic is statistically significant, rejecting the assumption of normality. All this evidence shows that the empirical distribution of daily returns cannot be fit by a normal distribution, as it exhibits a significant excess of kurtosis and asymmetry (fat tails and peakness).

Before continuing, we briefly summarize the steps performed in this study. For each of the thresholds selected, *first of all*, we evaluate the fit of the GPD, *second,* we analyzed the stability of the GPD parameters. In the *third* place, the sensitivity of the high quantiles of the GPD to the threshold choice is evaluated (Section 3.2). *Later*, we evaluate the sensitivity of the risk market measure to the threshold choice (Section 3.3). *Fifth*, we assess the accuracy of the estimated risk measures (Section 3.4). *Finally*, for each selected threshold we calculate the capital charges based on the ES measure (Section 3.5). The objective is to evaluate how sensitive the capital requirements are to the choice of the threshold.

3.2 Fitting the GPD

In this Section, we fit GPD to the data for a set of 20 thresholds. We aim to evaluate the sensitivity of the parameters and quantiles of the generalized Pareto distribution to changes in the threshold. The thresholds were chosen from a quantile range between the 80th to the 99th percentile at 1% increments. As the size of the sample is 5534 daily returns, the percentile 80th gives 1107 exceedances while the 99th percentile gives 56 exceedances. As the threshold increases by one unit, the number of exceedances decreases by 55 units. According to the theory, the distribution of the exceedances is defined as $F_n(X-u)$ for $X \geq u$ may be approximated by the GPD denoted by $G_{\xi,\sigma}(y)$. Thus, for each threshold, we fit a GPD and check that the sample of the excesses above the threshold follows a $G_{\xi,\sigma}(y)$. Examples of this fit can be seen in Figure 2 for the 5534 returns with thresholds set at 0.74% and 2.76%. These thresholds give 1107 and 56 exceedances respectively. Parameters are estimated by maximum likelihood and the resulting GPD curves are superimposed on the empirical estimate of the distribution function of the exceedances. As we can see, GPD seems to fit pretty well the exceedances samples. In concordance with this, the Kolmogorov-Smirnov test used to test if the sample of exceedances follows a GPD, can not be rejected in any case (see Table 3).

Let $u_1, u_2, \ldots, u_n$ be the set of thresholds selected ($n = 20$). For $j = 1, \ldots, n$, let $\hat{\xi}_{u_j}$ and $\hat{\sigma}_{u_j}$ be the estimators of the shape and scale parameters based on the exceedances over the threshold u_j. Figure 3 displays the estimation of ξ and σ, respectively, as a function of the threshold u.

We observe that as the threshold increases, the value of ξ increases. In the case of the scale parameter, the opposite occurs; as the threshold increases, the value of σ is reduced. As we expected, in both cases, the accuracy of the estimations decreases as the threshold increases.

The estimation of the shape parameter, which determines the weight of the tail in the distribution, is very sensitive to changes in the threshold. For instance, the value of ξ increases by 2233% when the threshold moves from the 80th percentile to the 90th percentile. From the 90th percentile to the 99th percentile, the increase is equal to 227%. The value of the scale parameter is also sensitive to changes in the threshold; however, in this case, the changes are not that striking.

Thus, in accordance with the literature, we find that the parameter estimations are very sensitive to the threshold selected for estimating GPD. But, what about the GPD quantiles? Do they depend on the threshold choice? To answer this question, we analyze the sensitivity of the high GPD quantiles to changes in the threshold. For this analysis, we just only focus on the high quantiles (95%, 96%, 97%, 98% and 99%) as they are the only relevant quantiles in quantifying market risk. The quantile of the GPD is calculated using Equation (6).

Figure 4 displays these quantiles as a function of the threshold u. What draws our attention is that the line representing the quantiles as a function of the threshold is completely flat, which means that the high quantile of the GPD does not depend on the threshold choice, at least in the range of threshold considered in this paper. This result is pretty striking. Table 4 displays the differences in α quantiles obtained for all thresholds considered against the threshold benchmark. Panel (a) displays these differences regard to the threshold corresponding to the percentile 90^{th,} which is in the middle point of the considered range vector and has been used successfully in many empirical papers in VaR estimate (see Abad and Benito, 2013; Benito et al., 2017).

Panel (b) displays these differences with regard to the threshold corresponding to percentile 97th, which is the optimal threshold according to the Mean Excess Plot (See Figure 5). In Panel (a) we observe that for a large set of thresholds, from a return corresponding to the 81st percentile to a return corresponding to the 97th percentile, the differences in quantile estimation do not exceed the 7 basis points. Even more, from a return corresponding to the 85th percentile to a return corresponding to the 94th percentile, the differences in quantile estimation {95th to 98th} do not exceed the 2 basis points. As regards Panel (b), although the differences in quantile estimation {95th to 98th} are larger compared with Panel (a), they do not exceed 8 basis points from a return corresponding to the 80th percentile to a return corresponding to the 99th percentile. Just only in the case of the 99th quantile, the differences are somewhat higher.

As the estimation of the market risk depends on the quantile of the GPD, this preliminary analysis may suggest that the choice of the threshold in the framework of the POT method may not be very relevant in quantifying market risk.

3.3 Sensitivity of the risk measures to changes in the threshold

We can say that the analysis presented in the previous section is in accordance with the literature; we observe that the estimates of the parameters that describe the Generalized Pareto distribution depend significantly on the threshold selected for the estimation. However, surprisingly the high quantiles of the GPD keep approximately constant. In this Section, we want to go a step further by assessing to what extent the selection of the threshold affects the quantification of financial risk. With this objective, a set of 20 thresholds has been selected. The parametric estimates corresponding to these thresholds were presented in the previous section.

To quantify the risk, we use VaR and ES measures, which were presented in Section 2.3. The expression for these measures is given by:

$$VaR_t(\alpha) = \mu_t + \tilde{\sigma}_t\, q_{1-\alpha} \qquad\qquad ES_t(\alpha) = \mu_t + \tilde{\sigma}_t \left[\frac{q_{1-\alpha}}{1-\xi} + \frac{\sigma+\xi u}{1-\xi}\right]$$

where μ_t is the conditional mean return that is assumed constant ($\mu_t = \mu$), $\tilde{\sigma}_t$ represents the conditional standard deviation of the return; $q_{1-\alpha}$ is the percentile $1-\alpha$ of the GPD and ξ and σ are the shape and scale parameters of the GPD. For the estimation of the conditional standard deviation of the returns, we use an APARCH model.

The sample period is divided into a learning sample from January 3rd, 2000, to December 30th, 2016, and a forecast sample from January 3rd, 2017, to the end of December 2021. For each day of the forecast period, we will generate estimations of VaR and ES measures. These forecasting measures are obtained one day ahead at the 95% and 99% confidence levels.

Table 5 presents the descriptive statistics of the differences between the market risk estimates obtained from the threshold corresponding to the 90th percentile and the market risk estimates obtained from the remaining selected thresholds. For VaR estimates at the 95% confidence level, from a return corresponding to the 80th percentile to a return corresponding to the 96th percentile, the mean of the differences does not exceed the 3 basis points with a standard deviation between 1 and 3 basis points. For the thresholds corresponding to the 97th and 99th percentiles the mean of the differences in the VaR estimate at a 95% confidence level, increases moving between 6 and 9 basis points.

The standard deviation of these differences also increases, moving between 4 and 28 basis points. For these thresholds, the minimum difference becomes 45 basis points (99th percentile), while the maximum difference becomes 229 basis points (99th percentile). For VaR estimates at the 99% confidence level, we find similar results. For a large set of thresholds (from the 82nd percentile to the 96th percentile), the mean and standard deviation of the differences are very small, not exceeding 5 basis points. Only in the case of the threshold corresponding to 99th percentiles, the differences are higher. In summary, we find that for a large set of thresholds (the return corresponding to the 80th percentile to the 96th percentile) the quantification of risk that we obtain from VaR measures is similar. The same conclusion can be drawn from the ES measure. At the 95% confidence level and from the 82nd percentile to the 98th percentile, the mean and standard deviation of the differences do not exceed 3 basis points. At the 99% confidence level, the differences are even smaller and do not exceed 1 basis point (from the 80th percentile to the 99th percentile).

Thus, we can conclude that in the selected range, the choice of threshold in the framework of the POT method may not be very relevant in quantifying market risk.

3.4 Analyzing the quality of the risk estimates

In this section, we are interested in analyzing the accuracy of the risk measures (VaR and ES) obtained from the conditional EVT. In addition, we will analyze if the quality of these measures depends on the threshold selected for applying EVT. Therefore, we will use the backtesting techniques presented in Section 2.4.

To evaluate the accuracy of the VaR estimates, we have used four standard tests: unconditional (LR_{uc}), independent (LR_{ind}), conditional coverage (LR_{cc}) and dynamic quantile (DQ) tests.

The results of these tests are presented in Table 6. In this table, we also present the number and the percentage of exceptions. The first thing that draws our attention when viewing Table 6 is that for a large set of thresholds (from the 82nd percentile to the 93rd percentile), the number of exceptions is very close to the expected one[9]. In the cases in which the number of exceptions differs from the theoretical one, the differences are very small. Thus, at the 95% confidence level, the percentage of exceptions ranges from 4.61% to 5.56%, corresponding to the 80th percentile and the 99th percentile. At the 99% confidence level, the percentage of exceptions ranges from 1.43% to 1.59%, also very similar to the expected one (1%). To test statistically whether the number of exceptions is equal to the theoretical one, we use the aforementioned test. We cannot reject the null hypothesis "that the VaR estimates are accurate" for any of the thresholds selected. To test whether the ES estimations are correct, we use the procedure proposed by McNeil and Frey's (2000) test. The results of these tests are displayed in Table 6. The null hypothesis, which states that the ES (95%) estimates are correct, is rejected for all thresholds at both 5% and 1% probability. However, the hypothesis that the ES (99%) estimates are correct is rejected at 5% for all thresholds, but not at 1%.

The results presented in this section indicate that the choice of threshold in the framework of the POT method may not be relevant in quantifying market risk when we use the VaR and ES measures for this task.

---

[9] For the forecasting period considered in this study, which has 1258 observations, the expected number of exceptions is 63 at a 95% confidence level and 13 at a 99% confidence level.

3.5 Analysing the sensitivity of forecasting daily capital charges to the selected threshold

In this section, we carry out an empirical application in which we evaluate the sensitivity of the daily market risk capital requirement (DCR) to the threshold choice. For this proposal, we follow Chang et al. (2019) and calculate the DCR according to Equation (19). Figure 6 shows the mean of the DRC calculated on the base of the ES measure at a 99% confidence level for each of the thresholds selected.

The visual inspection of this Figure suggests that there is a large set of thresholds (85^{th} to 95^{th}) that provide similar results, observing some differences in the lowest (80^{th} to 84^{th}) and highest (96^{th} to 99^{th}) thresholds. Table 7 which shows the mean, standard deviation and range of daily capital requirements, confirms these previous results. For a large set of thresholds (80^{th} to 95^{th} proxy) the differences in DRC are under 5 basis points. This implies that for investment portfolios worth 1 million euros, the differences do not exceed 50 thousand euros. However, for thresholds outside this range, the differences are somewhat greater.

4. Robustness Analysis

In the above section, we show that in the framework of the POT approach, there is a set of thresholds that provides a similar market risk estimate. It is due to the fact that the GPD quantiles are not sensitive to the threshold choice. Just only for the threshold $99^{th,}$ some differences are found. To corroborate the validity of this result, in this section, we extend the study to a set of 14 assets. In accordance with the performed study for the S&P500, for each of these assets, we select a set of 20 thresholds, from the 80^{th} percentile to 99^{th} percentile. For each threshold selected, first, we apply the conditional POT approach to analyze the sensitivity of the GPD quantiles to the threshold choice. Second, we obtain forecast VaR and ES measures 1 day ahead and analyze the differences among them for the set of thresholds selected. To last, we study the sensitivity of the market risk capital charges to the threshold choice.

Before analyzing the accuracy of the market risk measure, we evaluate the sensitivity of high quantiles from Generalized Pareto distribution to changes in the threshold. As in the case of S&P500, for this analysis, we just only focus on the high quantiles (95%, 96%, 97%, 98% and 99%) as they are the only relevant quantiles in quantifying market risk. The quantile of the GPD is calculated using Equation (6). Figure 7 displays the GPD quantiles as a function of the threshold for all assets considered. Again, what draws our attention is that the line representing the quantiles as a function of the threshold is relatively flat in the threshold range (80^{th} to 96^{th}). Just in the case of the high threshold, corresponding to the 97^{th} to 99^{th} percentiles, some differences are observed, especially for the threshold corresponding to the 99^{th} percentile. For this threshold, the differences are around 25 basis points becoming 50 basis points for some assets as the Merval index.

After checking that the GPD fits well the upper tail of the distribution of the assets for the set of thresholds considered, we calculate the market risk measure at the 95% and 99% confidence levels. For evaluating the accuracy of the VaR estimates, we use the standard tests that we presented in Section 2.4: LR_{uc}, LR_{ind}, LR_{cc} and DQ. For each asset, Table 8 displays the number of times that each of these tests is rejected for the 20 thresholds selected.

In the footnote in Table 8, we indicate the set of thresholds for which the null hypothesis is rejected. For instance, for CAC40 at a 95% confidence level, LR_{uc} test is rejected once for the threshold corresponding to the 99^{th} percentile. The results obtained for VaR are as follows. According to LR_{uc} tests, in 7 of the 15 considered assets, we do not find evidence against the null hypothesis that the “VaR(5%) estimate is accurate”. This result is independent of the selected threshold, although, for certain indexes, this hypothesis is rejected for some tests performed over the threshold corresponding to the 99^{th} percentile. The results found for VaR at the 99% confidence level are even more conclusive than those for VaR at the 95% confidence level. According to LR_{uc} tests, in 14 of the 15 considered assets, we do not find evidence against the null hypothesis that the “VaR (1%) estimate is accurate”. In certain cases, the accuracy tests provide evidence against the null hypothesis; however, in these cases, the rejection does not depend on the threshold selected. For instance, for the Cooper commodity, the DQ test rejects the null hypothesis for all thresholds. These results suggest that the quantification of the risk through the VaR measure does not depend on the threshold selected for this objective.

To test whether the ES estimations are correct, we use the procedure proposed by McNeil and Frey (2000) test. Overall, we do not find evidence against the null hypothesis that the average of the discrepancy measure is equal to zero indicating that all the thresholds provide correct ES estimations for both 95% and 99% confidence levels.

The results presented in this section corroborate those obtained for S&P500, indicating that the quantification of market risk through the VaR and ES measures does not depend on the threshold selected for applying the POT method.

To last, for all assets, we calculate the market risk capital requirement on the base of the ES at 99%. Table 9 shows the mean of these requirements. Again we find that in general for all assets there is a wide set of thresholds that give similar results. Just only the extreme thresholds provide capital requirements something different. If the aim of a financial institution is to minimize the market risk capital charges, the optimal threshold is the threshold corresponding to 90th percentile for six of the 14 assets considered. For three indexes (IBEX35, Merval and Nikkei) the highest thresholds (98th and 99th) are the best, while for the commodities, the thresholds that minimize market risk capital requirement are the lowest(80th).

5. Conclusions

The conditional Extreme Value Theory has been proven to be one of the most successful in estimating market risk. The implementation of this method in the framework of the POT model requires choosing a threshold return for fitting the Generalized Pareto distribution. Threshold choice involves balancing bias and variance. To determine the optimal threshold, several techniques have been proposed such as graphic methods, ad hoc methods, or methods based on goodness-of-fit contrasts. However, none of these techniques have been proven to provide better results than others.

In this paper, we ask if the threshold choice is relevant in measuring market risk. In other words, in this study, we assess to what extent the selection of the threshold is decisive in quantifying the market risk. To measure market risk, we have used the Value at Risk (VaR) and Expected Shortfall (ES) measures. The study has been done for the S&P500 index.

Previously, we analyse both, the sensitivity of the parameter estimates and GPD quantiles to the threshold choice. The results obtained are as follows. First, we find that following the literature, the parameter estimations are very sensitive to the selected threshold for estimating GPD. However, the quantiles of the GPD do not change much when the threshold changes, particularly for high quantiles (95th, 96th, 97th, 98th and 99th), which are relevant in risk estimation. Second, for a large set of thresholds (from the 80th percentile to the 96th percentile), the VaR estimations are practically equivalent. A similar finding occurs for the ES measure. In the last application, we calculate the market risk capital requirements on the basis of the ES(99%) estimations. The results reveal that there is a set of thresholds that provides the same results finding some differences for the higher percentiles.

The results obtained indicate that from the market risk management point of view, there is not an optimal threshold but that there is a set of optimal thresholds which provide similar market risk measures. Thus, we can conclude that in market risk estimation the researchers and practitioners should not focus excessively on the threshold choice, as a wide range of them produce the same risk estimates.

To corroborate these results, we have extended the S&P500 index study to a set of 14 assets (stock market indexes, commodities and exchange rates). The results obtained for these assets corroborate the results obtained for S&P500.

To last, although overall the quantification of the risk does not depend on the threshold choice, for a certain threshold some differences are found therefore, the financial institution may be interested in choosing the threshold that minimizes the market risk capital requirement.

Figure 1. S&P500

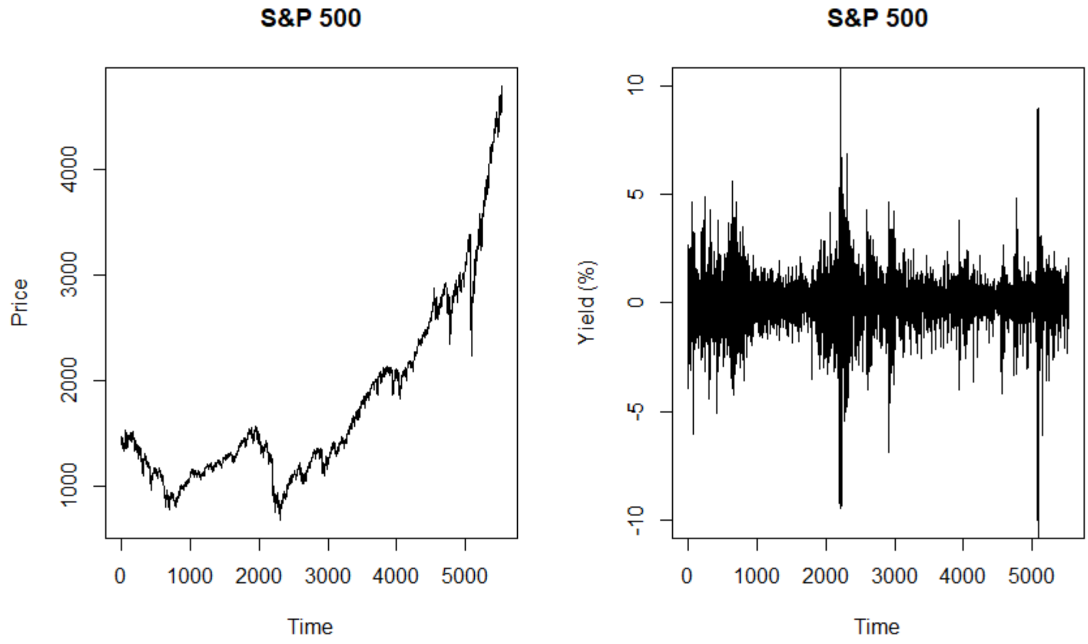
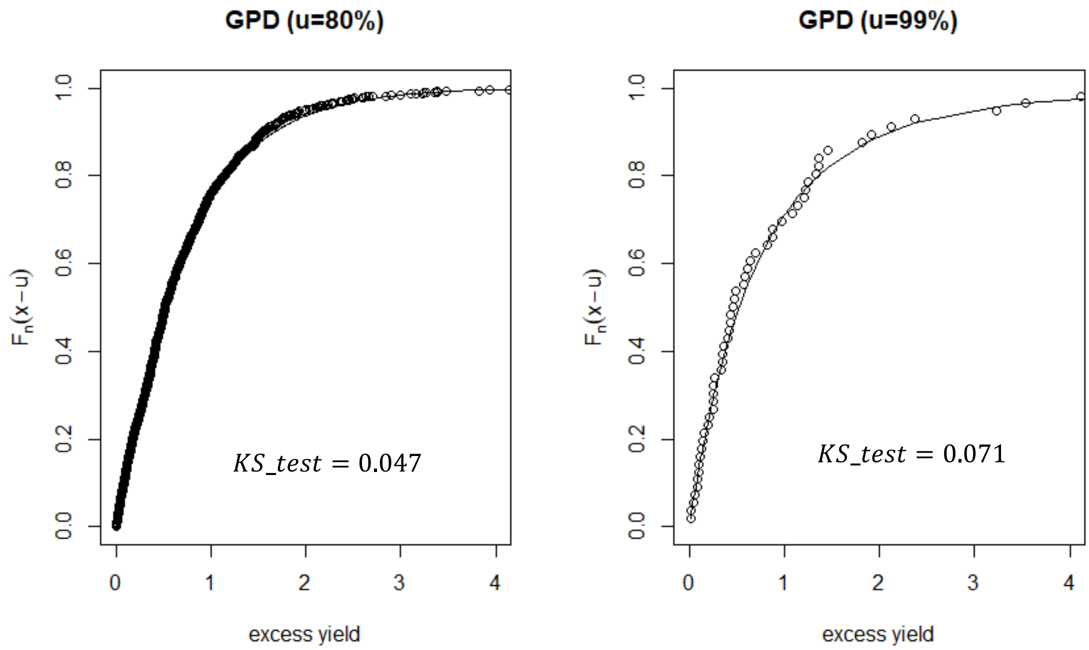


Figure 2. GPD



Note: In the left plot, GPD is fitted to 1107 exceedances over the threshold of 0.74%. In the right plot, GPD is fitted to 56 exceedances over the threshold of 2.76%.

Figure 3. Maximum Likelihood Estimations GPD parameters.

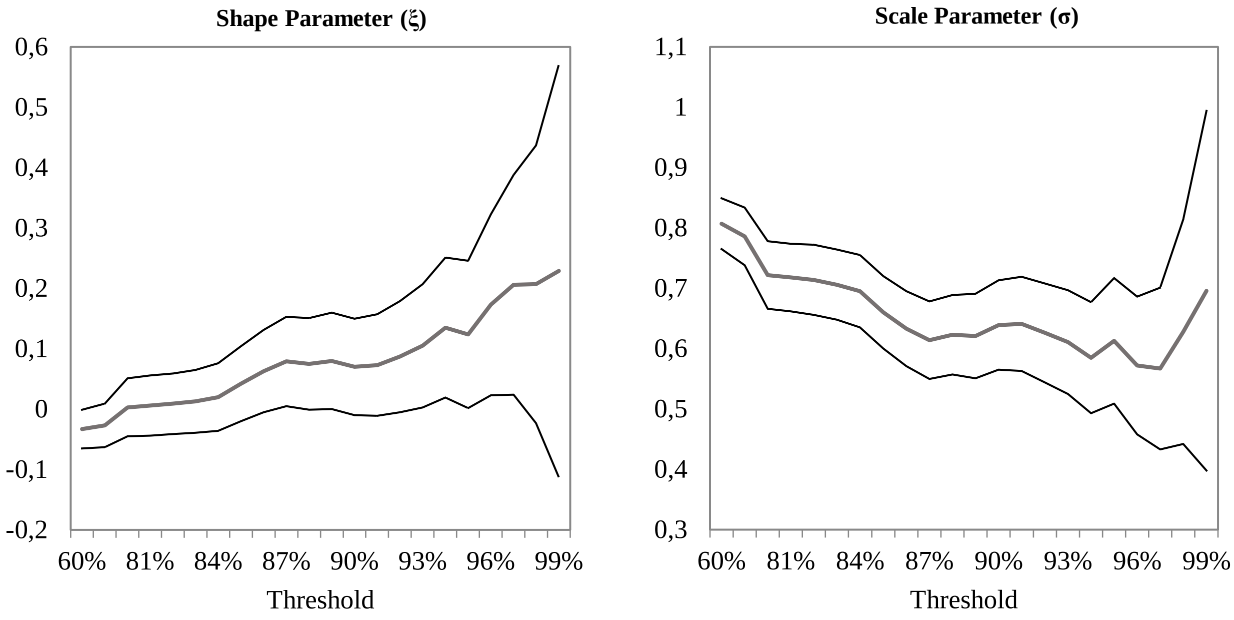
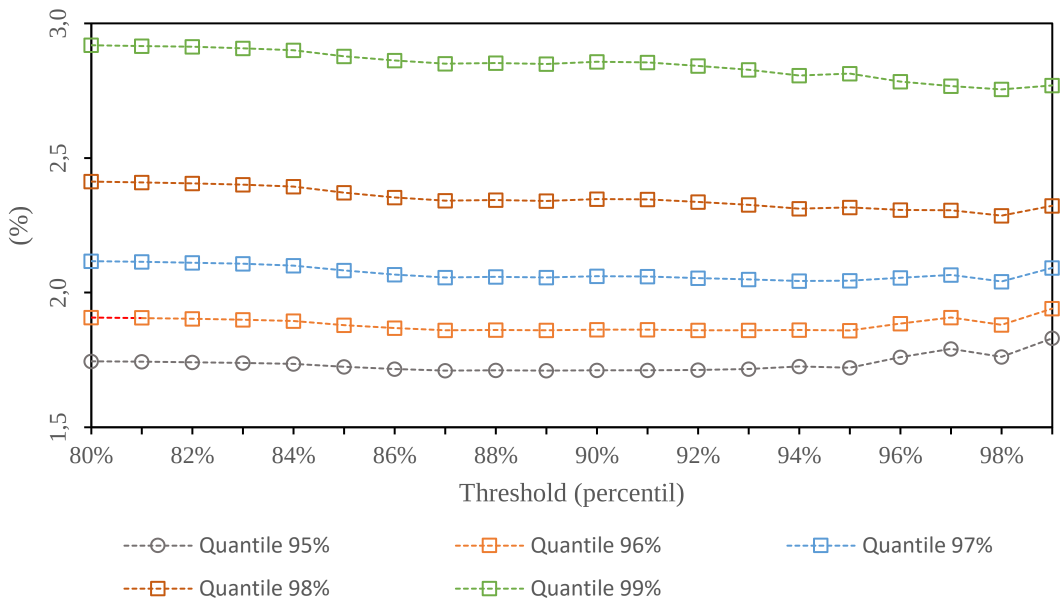


Figure 4. GPD quantiles S&P500



Note: GPD quantiles at 95%, 96% 97%, 98%, and 99% probability are displayed as a function of the thresholds.

Figure 5. Mean Excess Plot

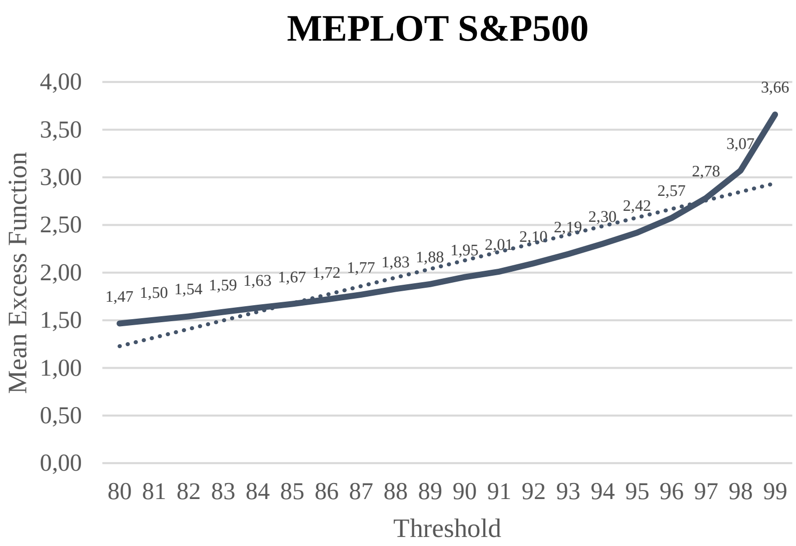


Figure 6. Mean Market Risk Capital Requirement calculated on the ES measure

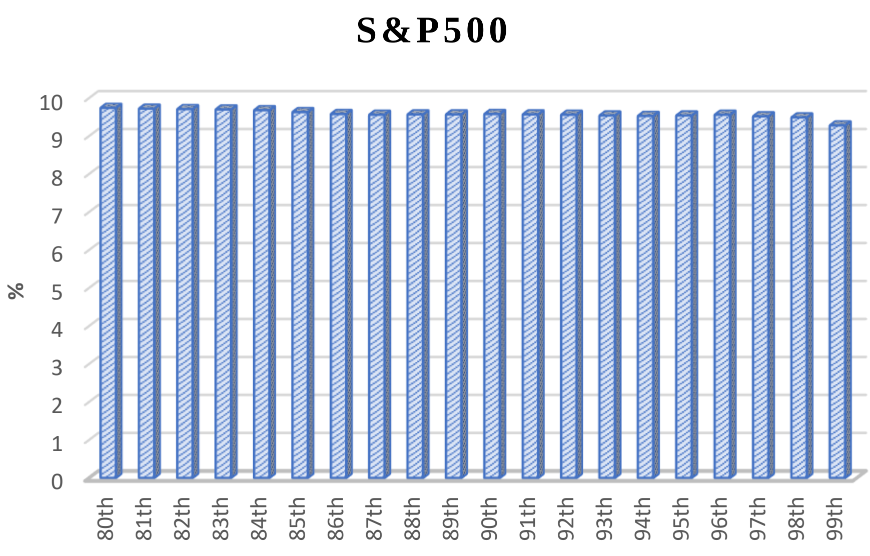


Figure 7. GPD quantiles all assets.

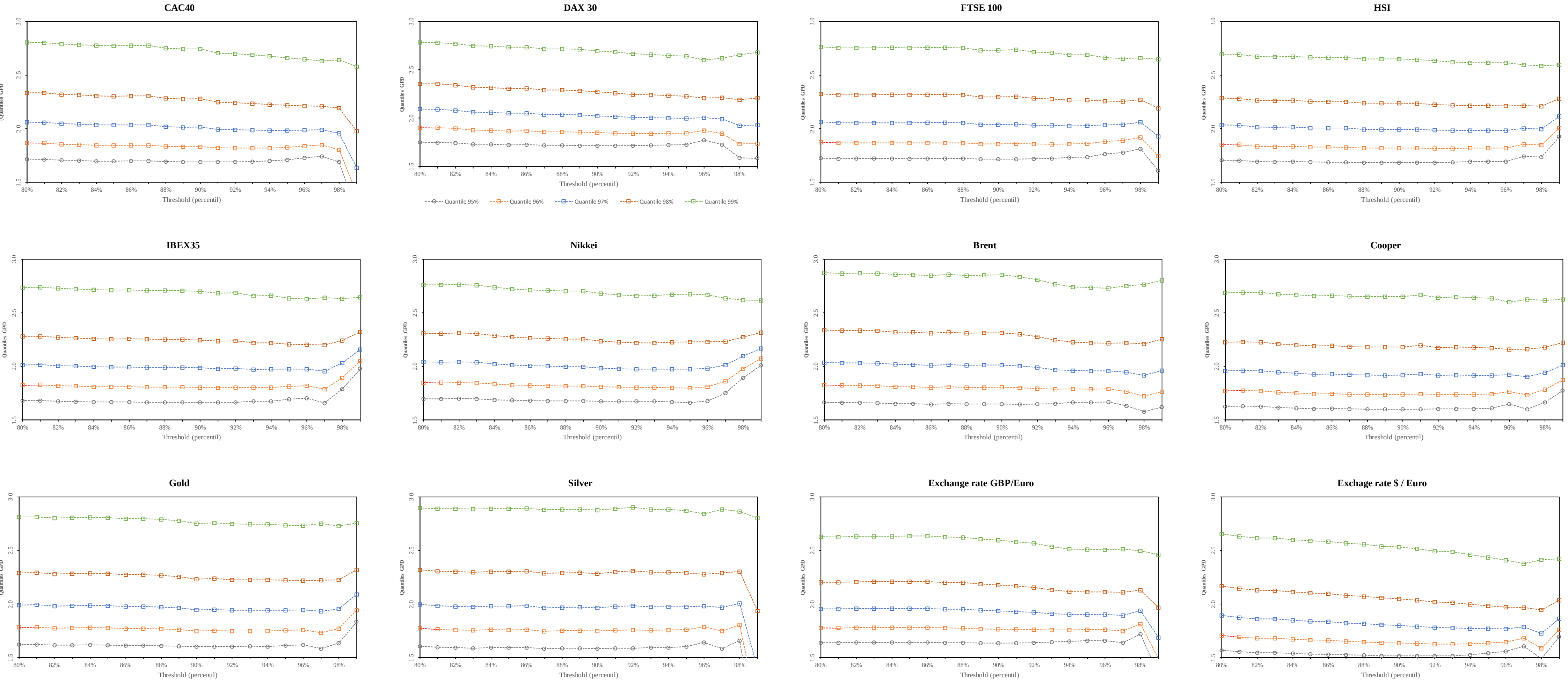


Table 1. Basel Accord Penalty Zones

Zone	Number of exceptions	k
Green	0 to 4	3
Yellow	5	3.4
	6	3.5
	7	3.65
	8	3.75
	9	3.85
Red	10 or more	4

Note: The number of exceptions is given for 250 trading days.

Table 2. Descriptive Statistics

	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque Bera
S&P 500	0.0214	0.0639	10.957	-12.765	1.2391	-0.401* (0.033)	11.067* (0.066)	28334 (0.000)

Note: This table presents the descriptive statistics of the daily returns of S&P500. The sample period is from January 3rd, 2000 to December 30th, 2021. The index return is calculated as R_t=100(ln(P_t)-ln(P_{t-1})) where P_t is the index level for period t. Standard errors of the skewness and excess kurtosis are calculated as $\sqrt{6/n}$ and $\sqrt{24/n}$ respectively. The JB statistic is distributed as the Chi-square with two degrees of freedom. (*) denotes significance at the 5% level.

Table 3. Maximum Likelihood Estimations (GPD)

Percentiles	Threshold return	Exceedances	ξ	σ	KS test	Critical Value
80th	0.74	1107	0.003 (0.024)	0.722 (0.028)	0.047	0.041*
81th	0.78	1052	0.006 (0.025)	0.718 (0.028)	0.048	0.042*
82th	0.82	996	0.009 (0.025)	0.714 (0.029)	0.050	0.043*
83th	0.87	941	0.013 (0.026)	0.706 (0.029)	0.051	0.044*
84th	0.92	886	0.020 (0.028)	0.695 (0.030)	0.048	0.046*
85th	0.98	830	0.042 (0.031)	0.660 (0.030)	0.033	0.047
86th	1.04	775	0.063 (0.034)	0.633 (0.031)	0.019	0.049
87th	1.10	720	0.079 (0.037)	0.614 (0.032)	0.023	0.051
88th	1.15	664	0.075 (0.038)	0.623 (0.033)	0.024	0.053
89th	1.20	609	0.080 (0.040)	0.621 (0.035)	0.026	0.055
90th	1.26	554	0.070 (0.040)	0.639 (0.037)	0.030	0.058
91th	1.32	498	0.073 (0.042)	0.641 (0.039)	0.034	0.061
92th	1.41	443	0.087 (0.046)	0.626 (0.041)	0.031	0.065
93th	1.51	388	0.105 (0.051)	0.611 (0.043)	0.028	0.070
94th	1.62	332	0.135 (0.058)	0.585 (0.046)	0.039	0.075
95th	1.72	277	0.124 (0.061)	0.613 (0.052)	0.046	0.082
96th	1.88	222	0.173 (0.075)	0.572 (0.057)	0.058	0.091
97th	2.07	166	0.206 (0.091)	0.567 (0.067)	0.072	0.106
98th	2.28	111	0.207 (0.115)	0.628 (0.093)	0.045	0.129
99th	2.76	56	0.229 (0.171)	0.696 (0.149)	0.071	0.182

Note: ξ: Shape parameter; σ: scale parameter. The standard deviation is given in parentheses. KS test is the Kolmogorov-Smirnov test. The critical value at a 95 % of confidence level is calculated as 1.36/√(n). In the cases denoted with (*), we can not reject the null hypothesis at 1%. The critical value at a 99% of confidence level is calculated as 1.63/√(n).

Table 4: Differences in quantiles

Panel (a)

	Differences in quantiles				
u	**95%**	**96%**	**97%**	**98%**	**99%**
80th	3	4	6	6	6
81st	3	4	5	6	6
82nd	3	4	5	6	6
83rd	3	4	5	5	5
84th	2	3	4	5	4
85th	1	2	2	2	2
86th	0	1	1	1	0
87th	0	0	-1	-1	-1
88th	0	0	0	0	0
89th	0	0	-1	-1	-1
90th	**0**	**0**	**0**	**0**	**0**
91st	0	0	0	0	0
92nd	0	0	0	0	0
93rd	0	0	-1	-1	-1
94th	0	0	-1	-2	-3
95th	1	0	-2	-4	-5
96th	1	0	-2	-3	-4
97th	5	2	-1	-4	-7
98th	8	4	0	-4	-9
99th	5	2	-2	-6	-10

Panel (b)

	Differences in quantiles				
u	**95%**	**96%**	**97%**	**98%**	**99%**
80th	-5	0	5	11	15
81st	-5	0	5	10	15
82nd	-5	0	5	10	15
83rd	-5	-1	4	9	14
84th	-6	-1	3	9	13
85th	-7	-3	2	7	11
86th	-7	-4	0	5	9
87th	-8	-5	-1	3	8
88th	-8	-5	-1	4	9
89th	-8	-5	-1	3	8
90th	-8	-4	0	4	9
91st	-8	-5	-1	4	9
92nd	-8	-5	-1	3	8
93rd	-7	-5	-2	2	6
94th	-7	-5	-2	1	4
95th	-7	-5	-2	1	5
96th	-3	-2	-1	0	2
97th	**0**	**0**	**0**	**0**	**0**
98th	-3	-3	-2	-2	-1
99th	4	3	3	2	0

Note: In the left plot the GPD quantiles at 95%, 96% 97%, 98%, and 99% probability are displayed as a function of the thresholds. The tables capture the differences in quantiles related to the 90th (Panel (a)) and 97th percentile (Panel (b)). We shaded the differences that oscillate between 3 and 4 basis points in light gray. Differences greater than 4 basis points are shaded in dark gray.

Table 5. Differences between VaR and ES estimates. Descriptive statistics. Optimal threshold 90%

	95% confidence level								99% confidence level							
Threshold (u)	VaR				ES				VaR				ES			
	Mean	S.D.	Max	Min	Mean	S.D.	Max	Min	Mean	S.D.	Max	Min	Mean	S.D.	Max	Min
80%	-3	2	-1	-23	-4	3	-1	-24	-5	4	-2	-33	1	1	8	-1
81%	-3	2	-1	-22	-3	2	-1	-23	-5	3	-1	-32	1	1	8	0
82%	-3	2	-1	-19	-3	2	-1	-20	-4	3	-1	-27	1	1	8	0
83%	-2	2	-1	-18	-3	2	-1	-18	-4	3	-1	-25	1	1	9	0
84%	-2	2	-1	-16	-2	2	-1	-15	-3	2	-1	-21	1	1	9	0
85%	-1	1	0	-8	-1	1	0	-5	-2	1	0	-6	1	1	8	0
86%	0	0	0	-2	0	0	2	-1	0	1	6	-1	0	0	3	0
87%	0	0	1	0	0	0	5	0	1	1	10	0	0	0	0	-2
88%	0	0	0	0	0	0	4	0	0	1	8	-1	0	0	0	-1
89%	0	0	2	0	0	1	7	-1	0	2	15	-2	0	0	1	-3
90%	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0	0	0	0	0	0	0	0
91%	0	0	1	0	0	0	4	0	1	1	9	-1	0	0	0	0
92%	0	0	0	0	0	1	6	0	1	1	13	-1	0	0	0	0
93%	-1	0	0	-3	1	1	8	0	3	2	21	1	0	0	1	-1
94%	-1	1	0	-7	1	1	10	0	4	3	31	1	0	0	1	-1
95%	0	0	2	-1	1	1	9	0	3	2	20	0	0	0	1	-2
96%	-3	3	0	-28	1	1	5	0	5	4	40	1	0	1	6	0
97%	-6	4	-2	-38	0	1	3	-2	6	5	45	2	1	1	7	0
98%	-8	4	0	-27	-1	2	14	-5	7	5	50	2	1	1	5	-1
99%	9	28	229	-45	7	16	140	-19	8	9	84	2	0	1	2	-3

Note: This table shows some descriptive statistics of the differences between the VaR and ES estimations obtained under the threshold u_j (j=1, …, 20) and the VaR and ES estimates obtained under the optimal threshold. The optimal threshold is given by the 90th percentile. We shaded in light gray the differences that oscillate between 3 and 4 basis points. Differences greater than 4 basis points are shaded in dark gray.

Table 6: Backtesting VaR and ES for S&P500 (2017-2021)

	95 % confidence level							99% confidence level						
	Exceptions		VaR				ES	Exceptions		VaR				ES
Threshold	Nº	%	LR_{uc}	LR_{ind}	LR_{cc}	DQ	MF	Nº	%	LR_{uc}	LR_{ind}	LR_{cc}	DQ	MF
80th	58	4.61	0.67	0.38	0.62	0.01	0.01	18	1.43	0.34	0.45	0.48	0.18	0.03
81st	58	4.61	0.67	0.38	0.62	0.01	0.01	18	1.43	0.34	0.45	0.48	0.18	0.03
82nd	58	4.61	0.67	0.38	0.62	0.01	0.01	18	1.43	0.34	0.45	0.48	0.18	0.03
83rd	58	4.61	0.67	0.38	0.62	0.01	0.01	18	1.43	0.34	0.45	0.48	0.18	0.03
84th	58	4.61	0.67	0.38	0.62	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03
85th	58	4.61	0.67	0.38	0.62	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03
86th	58	4.61	0.67	0.38	0.62	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03
87th	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.04
88th	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03
89th	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.04
90th	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03
91st	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03
92nd	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03
93rd	58	4.61	0.67	0.38	0.62	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03
94th	58	4.61	0.67	0.38	0.62	0.01	0.00	19	1.51	0.26	0.48	0.42	0.23	0.05
95th	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.04
96th	58	4.61	0.67	0.38	0.62	0.01	0.00	19	1.51	0.26	0.48	0.42	0.23	0.05
97th	56	4.45	0.55	0.54	0.69	0.01	0.00	19	1.51	0.26	0.48	0.42	0.23	0.05
98th	55	4.37	0.49	0.51	0.64	0.00	0.00	19	1.51	0.26	0.48	0.42	0.23	0.05
99th	70	5.56	0.55	0.32	0.51	0.00	0.07	20	1.59	0.20	0.52	0.36	0.28	0.08

Note: The table shows the p-value for the following statistics: the unconditional coverage test (LR_{uc}), statistics for serial independence (LR_{ind}), the Conditional Coverage test (LR_{cc}), the Dynamic Quantile test (DQ) and McNeil and Frey test (MF).

Table 7. Statistics of the daily capital requirement (ES 99%)

	Mean	S.D.	Range
80%	**11,09**	6,07	31,11
81%	**11,09**	6,07	31,12
82%	**11,09**	6,07	31,11
83%	**11,09**	6,07	31,11
84%	**11,09**	6,07	31,11
85%	**11,09**	6,07	31,13
86%	11,12	6,08	31,19
87%	11,14	6,09	31,23
88%	11,13	6,09	31,22
89%	11,14	6,10	31,26
90%	11,13	6,09	31,22
91%	11,13	6,09	31,23
92%	11,13	6,09	31,22
93%	11,13	6,09	31,23
94%	11,25	6,05	31,23
95%	11,14	6,10	31,24
96%	11,23	6,04	31,17
97%	11,22	6,02	31,16
98%	11,22	6,01	31,17
99%	11,45	6,50	33,96

Note: The table shows the mean of forecasting DCR for the S&P500. The shaded cells report the mean DCR with a difference of less than 5 basis points.

Table 8. Backtesting VaR and ES all assets.

	95% confidence level					99% confidence level				
	VaR				ES	VaR				ES
	LRuc	LRind	LRcc	DQ	MF	LRuc	LRind	LRcc	DQ	MF
CAC40	1[(1)]	0	0	1[(1)]	1[(3)]	0	0	0	0	0
DAX30	0	0	0	0	2[(4)]	0	0	0	0	0
FTSE100	0	0	0	0	19[(9)]	0	0	0	0	0
HANG SENG	0	0	0	1[(1)]	0	0	0	0	0	0
IBEX35	1[(1)]	0	1[(1)]	0	1[(1)]	0	0	0	0	0
MERVAL	0	0	0	0	0	0	0	0	0	0
NIKKEI	2[(2)]	0	1[(1)]	0	1[(1)]	0	0	0	0	0
S&P500	0	0	0	20[(10)]	15[(5)]	0	0	0	0	0
COPPER	0	0	0	0	0	0	0	0	20[(10)]	0
GOLD	1[(1)]	0	0	0	0	0	0	0	0	0
OIL BRENT	2[(2)]	0	1[(1)]	1[(1)]	4[(8)]	17[(6)]	0	4[(7)]	0	0
SILVER	1[(1)]	0	1[(1)]	0	1[(1)]	0	0	0	0	0
$/€	1[(1)]	0	1[(1)]	0	0	0	0	0	0	0
£/€	2[(2)]	0	2[(2)]	0	0	0	0	0	2[(2)]	0
¥/€	0	0	0	0	0	0	0	0	1[(1)]	0

Note: The table counts the number of rejections for the 20 thresholds considered. Reject for: (1) threshold corresponding to 99th percentile (u=99%); (2) thresholds corresponding to 98th and 99th percentiles; (3) threshold corresponding to 97th percentile; (4) thresholds corresponding to 96th and 97th percentiles; (5) thresholds corresponding to the interval [84th, 98th]; (6) thresholds 81st and the interval [84th, 99th]; (7) thresholds in the percentiles range [95th, 98th]; (8) thresholds corresponding to the interval [93rd, 96th]; (9) All thresholds except for that corresponding to 99th percentile; (10) All thresholds. For the case of ES backtesting, the acceptance level of the null hypothesis is set at 1%.

Table 9. Forecasting daily capital charges based on ES measure (99% confidence level)

Threshold	CAC40	DAX	FTSIE 100	HANG SENG	IBEX 35	MERVAL	NIKKEI	COPPER	GOLD	BRENT	SILVER	$/€	£/€	¥/€
80th	11.48	12.24	10.18	12.22	12.67	27.36	12.20	**14.82**	**9.76**	**26.27**	**20.95**	**4.57**	4.23	**5.88**
81th	11.46	12.23	10.18	12.22	12.67	27.35	12.20	**14.82**	**9.76**	26.44	21.03	**4.57**	4.23	5.89
82th	11.46	12.19	10.18	12.20	12.65	27.35	12.21	**14.82**	9.77	26.28	21.14	**4.57**	4.23	5.89
83th	11.46	12.20	10.18	12.59	12.65	27.34	12.19	14.87	9.77	26.29	21.23	4.58	4.23	5.89
84th	11.45	12.20	10.18	12.21	12.64	27.35	12.15	14.91	9.77	26.49	21.11	4.58	4.23	5.90
85th	11.46	12.19	10.18	12.59	12.64	27.34	12.13	14.95	9.77	26.50	21.14	4.59	4.23	5.89
86th	11.46	12.19	10.19	12.59	12.64	27.35	12.13	14.96	9.77	26.49	21.15	4.59	4.23	5.89
87th	11.46	12.17	10.19	12.59	12.63	27.31	12.12	14.97	9.77	26.50	21.27	4.60	4.23	5.89
88th	11.43	12.17	10.18	**12.58**	12.62	27.28	12.12	14.98	9.88	26.49	21.26	4.61	4.22	5.89
89th	11.43	**12.16**	**10.16**	**12.58**	12.61	27.28	12.11	14.98	9.89	26.50	21.23	4.63	**4.21**	**5.88**
90th	11.43	12.32	**10.16**	**12.58**	12.60	27.27	12.11	14.97	9.91	26.51	21.26	4.64	4.24	**5.88**
91th	**11.41**	12.31	10.17	**12.58**	12.58	27.21	12.11	14.96	9.92	26.76	21.19	4.67	4.23	**5.88**
92th	11.42	12.32	10.13	12.70	12.59	27.09	12.11	14.98	9.93	26.80	21.15	4.70	4.22	5.90
93th	11.42	12.32	10.12	12.69	12.54	27.05	12.11	14.98	9.94	27.14	21.69	4.71	4.22	5.93
94th	11.42	12.32	10.22	12.68	12.52	26.99	12.12	15.12	9.94	27.18	21.24	4.76	4.22	5.96
95th	11.42	12.35	10.22	12.70	12.49	26.96	12.14	15.12	9.93	27.45	21.72	4.88	4.24	6.05
96th	11.42	12.36	10.20	12.68	12.46	26.90	12.08	15.19	9.93	28.03	21.79	4.94	4.23	6.18
97th	11.61	12.40	10.19	12.63	12.50	26.88	11.99	15.16	9.95	27.65	21.73	5.10	4.23	6.24
98th	11.60	12.38	10.21	12.62	12.40	**26.85**	**11.92**	15.13	9.94	27.46	21.75	4.95	4.23	6.58
99th	11.62	12.45	10.20	12.61	**12.32**	27.26	11.94	15.15	9.90	27.37	22.06	5.62	4.27	7.40

Note: The table shows the mean of forecasting DCR for all the assets considered. We remark in bold the minimum mean daily capital requirement and shadow the cases in which the differences regarding the 90th percentile are less than 5 basis points.

Chapter IV

Assessing the selection of block size in the quantification of market risk under the Block Maxima approach (EVT)

Abstract

The conditional Extreme Value Theory (EVT) has been proven to be the most successful performing market risk estimation. This study focuses on the classical EVT, i.e Block Maxima Method (BMM). One key issue in implementing this approach requires the choice of block size for fitting the Generalized Extreme Value (GEV) distribution. The aim of the present study is two-fold. Firstly, we investigate the sensitivity of the parameters of the GEV distribution for different block lengths, especially we pay attention to the changes in the estimation of the shape parameter, which determines the weight of the tail in the distribution. Secondly, we also want to find out whether the selection of block size is significant in the quantification of market risk. For this latter purpose, we assess the sensitivity of risk measures such as Value at Risk and Expected Shortfall to changes in block sizes. With this in mind, we attempt to identify whether there is an optimal block size that leads to accurate risk estimates. The study has been done for a daily S&P500 stock index and later extended to a large set of assets to corroborate the results obtained. The findings indicate that the BBM does not provide satisfactory results in estimating market risk, as the results are highly sensitive to the block size selected.

Keywords: Extreme Value Theory, Block Maxima Method, Value at Risk, Expected Shortfall, Generalized Extreme Value Distribution.

Chapter IV. Assessing the selection of block size in the quantification of market risk under Block Maxima approach (EVT).

1. Introduction

Extreme Value Theory (EVT) has emerged as one of the most important statistical disciplines for predicting the probability of unusual events from observed outliers. More formally, EVT focuses on the limiting distribution of the extreme values observed over a long period, which is independent of the distribution of the values themselves. Extreme Value Theory is well-established for many sciences such as engineering, insurance, and meteorology among others (see e.g., Embrechts et al., 1999; Reiss and Thomas, 2007). In this sense, it has been proven to be very useful in applications as disparate as the analysis of the strength of materials or the probability of recurrence of natural disasters such as earthquakes or floods.

More recently, EVT has been gaining ground in the financial field. This increased popularity is due to the interest shown by the financial community in analyzing the impact of extreme variations, such as stock market crashes. Since the seminal work of Longin (1998), there have been several studies in the literature where the empirical applicability of EVT has been used to estimate extreme risks of the financial market (see, among others, Novales and Garcia-Jorcano, 2019; Mögel and Auer, 2018; Abad and Benito, 2013, Brooks et al., 2005). Also, an interesting discussion about the potential of extreme value theory in risk management is given in Diebold et al. (1998).

In the EVT context, there are two approaches. One of them, the Block Maxima Method (BMM), models directly the distribution of minimum or maximum realizations. The other one, Peaks over Threshold (POT), models the exceedances above a particular threshold.

Within the POT model, the extreme values above a high threshold are modeled using a generalized Pareto distribution (GPD). The main difficulty of this approach lies in the selection of the threshold, as different thresholds may provide different results. In the context of risk management, it is interesting to know to what extent the selection of the threshold impacts risk estimation. This issue has been largely studied and discussed in Chapter III[1] of the present Thesis.

[1] Chapter III has been published in the Journal *Risk Management*. The complete reference is Benito Muela, S., López-Martín, C. and Navarro Cervantes, M. Á. (2023). "Assessing the importance of the choice threshold in quantifying market risk under the POT approach (EVT)". Risk Management 25, 6. DOI: 10.1057/s41283-022-00106-w (Available online 6th January 2023).

In contrast, with the POT method, BMM is based on the idea of dividing the dataset into *m* blocks of size *n* and then fitting the Generalized Extreme Value distribution (GEV) to the maximum m-block data series. As in the case of the POT method, selecting the block size is not trivial. The fit of the GEV distribution will be inaccurate if the block size is too small, leading to biased estimates, while a block size too large will lead to a smaller number of extreme observations and consequently a higher variance (Coles et al., 2001).

When we examined the literature on the use of the BMM, we find few recent papers that propose techniques for the selection of optimal block size (Dkengne et al., 2020; Özari et al., 2019; Wang et al., 2016). But, in general, the choice is done without making any assumptions (Santinelli et. al., 2014; Singh et al., 2013) or according to a natural time division (Engeland et. al., 2004; Gilli et. al., 2006).

In the same spirit as Chapter III, in this study, we first investigate the sensitivity of the parameters of the GEV distribution for different block lengths, and second, we attempt to answer the question of whether block size selection is a determinant in the accurate estimation of market risk. Specifically, in this Chapter, we want to find out whether the block size choice is important in quantifying market risk and assess the sensitivity of the risk measures to changes in block sizes. As in the previous Chapter, the risk measures used in this analysis are Value at Risk (VaR) and Expected Shortfall (ES).

Finally, a third contribution emerges from this work. As in this study, we use similar data to those used in Chapter III we compare the performance of the BM method with the POT method in estimating market risk[2]. Regarding the performance of the BM method in comparison to the POT method, the literature is somewhat ambiguous in the area of finance. For instance, Flugentiusson (2012) indicates the BMM as inferior to the POT method. Similar results are presented by Marinelli et al. (2007), Caires (2009) and Coles (2001). This last author says that "modeling only block maxima is a wasteful approach to extreme value analysis if other data on extremes are available". The work of Jobst (2007) also favors the use of POT, giving less effectiveness to BMM. His justification is based on the fact that the instability of the parameters at high percentile levels is largely caused by the lack of sufficient empirical data, which permits only simple parametric models of asymptotic tail behavior. In this case, the POT method seems to be

[2] Initially we present an exhaustive analysis for the S&P500 and later we extend the work to a set of 14 assets, the same as in the previous Chapter.

the most useful to derive out-of-sample estimates of asymptotic tail behavior. Carvalhal and Mendes (2003), however, obtain satisfactory results for Asian financial markets. They show that the market risk estimations obtained over a one-month under the BM method, are all accurate. Szubzda (2019), compares VaR estimates obtained through POT and BMM taking two block sizes (monthly and bimonthly). VaR estimates based on the BM method make the estimation more conservative than those obtained through the POT method, providing more efficient results during crisis periods, but it may, in exchange, be more costly for financial institutions during calm periods. Also, Qian Yiping et. al. (2010) conducted empirical research on loss data of the operational risk from Chinese commercial banks and pointed out that the BMM could obtain a more accurate result with fewer data. Cunnane (1973) states that for $\xi = 0$ and maximum Likelihood estimators, the POT estimate for a high quantile is better only if the number of exceedances is larger than 1.65 times the number of blocks. Similar conclusions are found in Wang (1991) who shows that POT is as efficient as BMM for high quantiles, based on the Probability-weighted moments (PWM) estimator. Madsen et al. (1997) suggest that POT is preferable for $\xi > 0$ again, only with the number of exceedances larger than the number of blocks.

In this context, this study contributes to the existing literature by shedding some light on the performance of the Maximum Block model concerning the POT method in the area of risk management, and more specifically with regard to estimating market risk.

The remainder of the paper is organized as follows. In Section 2, we show the methodology used for the study. In Section 3, we submit the data and the results obtained for the particular case of the S&P500 index. Section 4 displays a robustness analysis. Section 5 shows the comparison between POT and BMM VaR results. The main conclusions are presented in Section 6.

2. Theoretical Framework

2.1 Extreme Value Theory

Extreme Value Theory (EVT) relates to the asymptotic behavior of extreme observations of a random variable. Many researchers have contributed to the theoretical discussion on EVT, see for instance, Embrechts et al. (1997), Reiss and Thomas (1997), and Coles (2001) among others. Even though EVT has previously found large applicability in climatology and hydrology, there also are some applications of this theory in finance literature in recent years, see for instance, Novales and Garcia-Jorcano, (2019), Mögel and Auer (2018), Louzis et al. (2012), Brooks et al. (2005) and Abad and Benito (2013).

Within the EVT context, two approaches let to capture the extreme value of the tails distribution: the first one is based on Generalized Extreme Value Distribution (GEVD), which models the distribution of minimum or maximum realizations. This approach is known as the Block Maxima (Minima) Method (BMM). The second one is the Peak Over Threshold (POT) approach, based on the Generalized Pareto Distribution (GPD) (Pickands 1975), which models the exceedances of a particular threshold. As it was said in the introduction, this Chapter, focused on the study of extremes in financial markets using BMM.

The Block Maxima Method consists in dividing the dataset into m blocks of size n. Let be $X_1m, \ldots, X_nm$ a series of independent and identically distributed random variables from a time interval m, then the maximum values can be defined as $M_m = max(X_1m, \ldots, X_nm)$. The CDF of M_m is represented as:

$$P(M_m \leq x) = P(X_1m \leq x, \ldots., X_nm \leq x) = \prod_{i=1}^{n} F(x) = F^n(x) \quad (1)$$

Since $F(x)$ is unknown, an approach is to look for an asymptotical distribution for a large value of maxima that can be estimated based on extreme data. This is analogous to approximating the distributions of sample means by the normal distribution based on the central limit theory.

Theorem 1. (Fisher and Tippet, 1928; Gnedenko, 1943).

Given $F^n(x) < 1, \forall\, x < \infty$, M_m converges in probability to 0 for any $x < x^+$, where $x^+$is the upper end-point of F, i.e, the smallest value of x such that $F(x) = 1$ or it converges to 1 for any $x \geq x^+$. Thus, to achieve a non-degenerate behavior limit when $n \longrightarrow \infty$, M_m has to be standardized.

Fisher and Tippet's Theorem states that given a sequence of $b_m > 0, a_m \in R,$ the maximum normalized value $M_m^* = \frac{M_m - a_m}{b_m}$ converges to a non-degenerated distribution H, being this distribution the Generalized Extreme Value distribution (GEVD),

$$\lim_{m \to \infty} Pr\left(\frac{M_m - a_m}{b_m} \leq x\right) \longrightarrow H(x). \tag{2}$$

2.1.1 Generalized Extreme Value Distribution.

The algebraic expression for such generalized distribution is as follows:

$$H_{\xi,\mu,\,\sigma}(x) = e^{-\left[1+\xi\frac{(x-\mu)}{\sigma}\right]^{-\frac{1}{\xi}}} \tag{3}$$

defined on $\left(1 + \frac{\xi(x-\mu)}{\sigma}\right) > 0$, where $\sigma > 0$ is the scale parameter, $-\infty < \mu < \infty$ is the mean, and $-\infty < \xi < \infty$ is known as the shape parameter of the GEV distribution and characterizes the tail behavior of the distribution, i.e. the limiting distribution of the normalized maximum. Feller (1971, p.279) proves that the shape parameter is invariant under time aggregation[3].

The prior distribution is a generalization of the three types of distributions, depending on the value taken by ξ:

[3] The traditional Extreme Value Theory (EVT) assumes that data is stationarity. When stationarity is assumed, parameters that determine the distribution function (Generalized Pareto and Generalized Extreme Value distributions) are independent of time. However, in practice, it is often the case that stationarity assumptions (such as independence and identical distribution) for financial time series are violated (because of clustering property). If the process is non-stationary, the parameters of distributions are time-dependent, and the properties of the distribution vary with time. To capture the non-stationarity of extreme data, new approaches have been developed in the framework of the extreme value theory. Some applications of these new approaches can be found in Cheng and AghaKouchak (2014), Cheng et al. (2014), Ruggiero et al. (2010), Chavez-Demoulin and Embrechts (2004) among others.

- Gumbel (ξ=0) type I family.

$$\Lambda(x) = e^{e^{-\frac{x-\mu}{\sigma}}} \qquad \forall x \in \Re \tag{4}$$

- Fréchet (ξ>0) type II family.

$$\Phi_{\xi,\mu,\sigma}(x) = \begin{cases} 0 & x \leq \mu \\ e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}} & x > \mu \end{cases} \tag{5}$$

Gumbel and Fréchet's distributions have infinite right endpoints. The main difference between them is in the decay of the tail.

- Weibull (ξ<0) type III family. This distribution is a short-tailed distribution with a finite right endpoint.

$$\Psi_{\xi,\mu,\sigma}(x) = \begin{cases} e^{-\left(-\frac{x-\mu}{\sigma}\right)^{\frac{-1}{\xi}}} & x \leq \mu \\ 1 & x > \mu \end{cases} \tag{6}$$

Note that the generalized expression of the distribution, $H_{\xi,\mu,\ \sigma}$, is continuous in the parameter ξ. Since the shape parameter is related to the decay of the tail of the distribution, an important characteristic is that if the distribution of M_m^* converges to a GEV distribution for block maxima with parameter ξ, then the distribution of exceedances over the threshold converges to the GPD with the same parameter ξ (Rodriguez G. 2017).

2.1.2 Maximum Domains of Attraction

If Equation (2) holds for some non-degenerate distribution function *H*, then *F* is said to be in the maximum domain of attraction of *H*, written $F \in MDA(H)$.

If the tail of *F* declines exponentially, then H_ξ is of the Gumbel type and ξ=0. When it comes to financial modeling, it is often erroneously assumed that the only interesting models are the power-tailed distributions of the Fréchet type. Nevertheless, the Gumbel type is also interesting because it includes many distributions with much heavier tails than normal (McNeil et al. 2005). In this case, some underlying distributions in the domain of attraction of H_ξ are normal, log-normal, exponential, gamma, and chi-squared.

If the tail of *F* decays by a power function, being the rate of decay $\alpha = 1/\xi$ also known as the tail index of the distribution, then H_ξ is of the Fréchet class and $\xi > 0$.

Distributions in its domain of attraction are called fat-tailed distributions (e.g., Pareto, Cauchy, Student-t, log gamma, F, and Burr) which are of particular interest in financial applications and the most studied in the context of EVT.

Finally, if the tail of F is finite then H_ξ is of the Weibull type and $\xi < 0$. Distributions in the domain of attraction are those with bounded support such as uniform and beta distributions. This is the least relevant class for market risk analysis since it refers to distributions with finite right endpoints. However, it still could be useful in financial modeling in the area of credit risk (McNeil et al., 2005).

2.2 Block Maxima Method and block size selection.

The Block Maxima approach has its origin in the field of hydrology, where the existence of seasonal periodicity of the observations suggests the application of this method. [4] [5] We also find its application in the field of finance in studies such as Bekiros et al. (2005) in which they compare the predictive ability of Value at Risk (VaR) estimates obtained from different techniques, including POT and BMM.

As discussed in the previous section, BMM is based on the idea of fitting the GEV distribution function to the maximum data series. The parameters of GEV distribution can be estimated using different techniques, including the Maximum Likelihood Estimation (MLE), which is the one we will use later in the empirical study[6]. In the implementation of MLE, it is assumed that the block size is sufficiently large so that regardless of whether the observations are dependent or not, the block maxima observations can be taken as independent[7].

The idea is therefore to divide the dataset into m blocks of size n. Denote the maxima observations per block as $X_1m, \ldots, X_nm$ and the density of the GEV distribution as $h_{\xi,\mu,\sigma}$, the algebraic expression for the log-likelihood is as follows:

[4] According to Ferreira and de Haan (2015) the Block Maxima approach may be preferable to POT when (i) the observations are not exactly independent and identically distributed (i.i.d.), for example when it exists a seasonal periodicity or there is a natural way of blocking the data; or (ii) when they have a short dependence that plays a role within blocks but not between them (Katz et al., 2002 and Madsen et al., 1997).

[5] BMM is more often and more effectively used in hydrology (Abad et al., 2013).

[6] These estimates are consistent and asymptotically efficient in the case of $\xi > -1/2$ as shown in Smith (1985).

[7] According to McNeil et al (2005) dependence leads generally to a slower convergence to the GEV distribution. In this case, may be advisable larger block sizes than are used in the iid case.

$$l(\xi,\mu,\sigma;X_1m,\ldots,X_nm) = \sum_{i=1}^{n} ln\, h_{\xi,\mu,\sigma} =$$

$$= -m\, ln\, \sigma - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{m} ln\left(1+\xi\,\frac{M_{mi}-\mu}{\sigma}\right) - \sum_{i=1}^{m} ln\left(1+\xi\,\frac{M_{mi}-\mu}{\sigma}\right)^{-1/\xi} \qquad (7)$$

which must be maximized subject to the constraints: $\sigma > 0$ and $1+\xi\,\frac{M_{mi}-\mu}{\sigma} > 0$.

In the selection of block size, a trade-off arises: the fitting to the GEV distribution will be inaccurate if the block size, n, is too small, leading to biased estimates, while a block size that is too large will lead to a smaller number of extreme observations per block and consequently a higher variance (Coles et al., 2001).

When we examined the literature on the use of the BMM, we did not find an agreed method that allows the selection of the block size. This choice is done either ad-hoc (Santinelli et. al., 2014; Singh et. al., 2013) or according to a natural time division (Engeland et. al.; 2004, Gilli et. al., 2006). Bystörm (2004) already defined the difficulty of making the right choice as the “optimal block size problem”.

Specifically, in the context of financial market data, we find some examples of block selection. Christoffersen et al. (1998) suggest a block size of 10 to 15 days to adequately model the S&P500 series. Longin (2000) proposes in his paper a block size of 21 days. McNeil et al. (2005) consider quarterly, semi-annual and annual blocks. Lindholm (2015) applies the maximum block method with block sizes $n = 21, 63, 126,$ and 252, but after an analysis only considers continuing with $n = 63$ (quarterly) to avoid the bias-variance problem. Allen et al. (2011) apply BMM on ASX and S&P500 returns with quarterly and annual block sizes. Carvalhal and Mendes (2003) apply blocks length of one month ($n = 21$), two months ($n = 42$), three months ($n = 63$), and six months ($n = 126$) to Asian stock markets.

In most papers, the selection of block size is an ad-hoc decision. However, some works search for the most appropriate block selection using different techniques. Tsay (2010) suggests choosing a block size by evaluating the model fit based on procedures such as parameter evaluation and distribution fit plots. Recent works such as Özari et. al., (2019); Wang et. al., (2016), and Dkengne et al., (2020) perform several methods that might be used for selecting the optimal block size. The first one puts the spotlight on the need to select an appropriate block size and proposes a simple computational method to specify the optimal block size. The second paper, applied to the engineering field, proposes a comprehensive evaluation based on multiple-criteria decision-making, dividing the data into blocks and analyzing the fit results graphically and quantitatively using goodness-of-fit tests of the GEV distribution function. The third paper proposes an automatic method to identify the block size. The proposed scheme is illustrated on two datasets, applied to engineering and meteorology fields.

Following the references aforementioned, in this study, the GEV distribution is fitted to the data of maximums that we get from 9 different block sizes corresponding to an approximate trading week (5), two weeks (10), a month (21), six weeks (31), two months (42), a quarter (63), a semester (126), nine months (189) and a year (252). Then we check if they are consistent with the theory. It is to say, we check if the maximum sample we get from these block sizes follows a GEV distribution. To do that, we use several procedures: (i) graphical methods such as QQ-plot and (ii) computational methods such as the Kolmogorov-Smirnov (K-S test), Anderson-Darling test, and Cramer-von Mises test.

The K-S test quantifies the distance between the theoretical and the empirical distribution function. The test statistic is as follows:

$$K = Max|F(x_i) - F_t(x_i)| \quad (8)$$

where $F(x_i)$ is the observed distribution function of a random sample of n observations and is the $F_t(x_i)$ is the theoretical distribution. The smaller K the better the goodness of fit. The null hypothesis states that there is no difference between the two distributions.

The Anderson-Darling test is a modification of the KS test and evaluates whether a set of data came from a specific distribution population. It is characterized by being more sensitive to the tails of the distribution, i.e. it assigns more weight to the tails than the K-S test and therefore has better power against fatter tails. The null hypothesis assumes that the data follow a certain distribution. The test statistic is as follows:

$$AD = -n\frac{1}{n}\sum_{i=1}^{n}(2i-1)[F(x_i) + ln\,(1 - F(x_{n-i+1}))] \quad (9)$$

where i is the ith sample, calculated when the data is sorted in ascending order and n is the sample size.

Finally, the Cramer-von-Mises test looks at the sum of squares of the differences and tests the null hypothesis that a sample comes from a pre-specified population distribution. The statistic test is as follows:

$$CVM = \frac{1}{12n} + \sum_{i=1}^{n} \left(F(x_i) - \frac{2i-1}{2n} \right)^2 \tag{10}$$

2.3 Risk measure.

The Basel Committee on Banking Supervision (BCBS) introduced the first framework for minimum capital requirements for market risk in January 1996. The main purpose was to ensure that financial institutions could hold minimum levels of capital that would reduce their exposure to market risk by absorbing potential losses arising from extreme movements in market prices in times of financial turbulence.

The subsequent global financial crisis (2008) exposed the weaknesses in the design of the market risk requirements, especially revealing the fact that it did not exist an adequate capitalization of the financial institutions to cover potential shortfalls.

In 2017 the Basel Committee on Banking Supervision published a reform with standards for minimum capital requirements for market risk, which were later revised in 2019. These revisions were expected to come into effect in 2022.

One of the aspects that have been the subject of attention and review in the regulatory framework concerns the internal models approach for measuring market risk. These internal models are based on risk measures such as Value at Risk (VaR) and Expected Shortfall (ES).

2.3.1 Value at Risk (VaR).

Among the market risk measures, the one most widely used by regulators to date is the Value at Risk measure (VaR) (Morgan, 1996). Jorion (2001) defined VaR as "the worst expected loss over a given horizon under normal market conditions at a given level of confidence". Thus, VaR is a conditional quantile of the asset return loss distribution.

More formally, let X_1, X_2, X_n be identically distributed independent random variables representing the financial returns. Using $F(x)$ to denote the cumulative distribution function, $F(x) = \Pr\,(X_t \leq x|\Omega_{t-1})$ conditioned to the information available at $t-1$ (Ω_{t-1}). Assume that $\{X_t\}$ follows the stochastic process given by

$$X_t = \mu_t + \tilde{\sigma}_t z_t \qquad z_t \sim iid(0,1) \tag{11}$$

where $\tilde{\sigma}_t^2 = E(z_t^2|\Omega_{t-1})$ and z_t has the conditional distribution function $\mathrm{G(z)}$, $\mathrm{G(z)} = \mathrm{P(z_t < z|\Omega_{t-1})}$.

The VaR with a given probability α ∈ (0, 1), denoted by $VaR(\alpha)$, is defined as the α quantile of the probability distribution of financial returns $F\big(VaR_t(\alpha)\big) = Pr\big(X_t < VaR_t(\alpha)\big)$ = α, i.e, at a given confidence level α, VaR_α can be defined as the α-*th* quantile of the distribution *F*:

$$VaR_\alpha = F^{-1} \tag{12}$$

where F^{-1} is the quantile function defined as the inverse of the distribution *F*. Assuming that the extreme observations follow a GEV distribution, then unconditional VaR is represented by the quantile function from the GEV, q_α, which is obtained by inverting such distribution. Substituting the parameters for the estimated parameters obtained from MLE and assuming a short position the expressions of q_α is as follows:

$$q_\alpha = \begin{cases} \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \{-nln(1-\alpha)\}^{-\hat{\xi}}\right], for\ \hat{\xi} \neq 0 \\ \hat{\mu} + \hat{\sigma}\{-nln(1-\alpha)\} \qquad\qquad ,for\ \hat{\xi} = 0 \end{cases} \tag{13}$$

where *n* is the block size from which the maximum observed returns are obtained and $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ are the location, scale, and shape estimated parameters.

McNeil and Frey (2000) proposed an approach that combines the EVT and GARCH models and that leads to a more accurate estimation of VaR in comparison with simple EVT methods, especially as GARCH models take into account the volatility clustering or conditional heteroscedasticity feature of financial returns. The application of this approach is most often found in the estimation of VaR using POT. However, it is also found in works such as Ghorbel and Trabelsi (2008) that consider a conditional GEV distribution and conclude that, although the conditional EVT is more accurate with POT than with the BMM, the conditional Block Maxima method for the CAC 40 index at high confidence level produce acceptable VaR forecasts.

This study extends McNeil and Frey's approach to the BMM and creates conditional VaR forecasts with the maxima following an autoregressive process with an APARCH conditional variance structure[8].

Therefore, for a one-day horizon, estimates of the dynamic VaR will be calculated as:

$$VaR_t(\alpha) = \mu_t + \sigma_t\, q_\alpha \tag{14}$$

where μ_t and σ_t[9] represent the conditional mean and the forecasted volatility of the returns respectively and q_α is the quantile function of the GEVD. (See *Appendix A*).

2.3.2 Expected Shortfall (ES)

An alternative developed by Artzner et al., (1997; 1999) that measures excess losses above VaR, i.e, that incorporates both the frequency and the size of extreme events, is the *Expected Shortfall* (ES) or Conditional Value at Risk.

The ES with a given probability $\alpha \in (0, 1)$, denoted by $ES(\alpha)$, is defined as the expected size of a loss that exceeds VaR_α, i.e., the average loss in the worst α % cases:

$$ES_t(\alpha) = E[X|\, X < VaR(\alpha)] = \mu_t + \sigma_t\, E[z|\, z < q_n\,] \tag{15}$$

This measure was proposed to satisfy some statistical properties of extreme losses that are violated by VaR, such as the principle of subadditivity, making ES a more consistent risk measure than VaR[10]. Nevertheless, it is not free from some shortcomings, such as its non-elicitable nature, which will make backtesting difficult to assess the accuracy of risk estimates.

---

[8] APARCH model is given by the expression: $\sigma_t^\delta = \alpha_0 + \alpha_1(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1})^\delta + \beta\sigma_{t-1}^\delta$, $\alpha_0, \beta, \delta > 0$, $\alpha_1 \geq 0, -1 < \gamma < 1$. In this model, the γ parameter captures the asymmetric effects on volatility (Black, 1976), if volatility tends to be higher after negative returns, then occurs the so-called leverage effect.

[9] Do not get confused this parameter with the scale parameter of GEV distribution (σ).

[10] In 2011, the Basel Committee on Banking Supervision acknowledged the failure of VaR to satisfy coherence (BCBS 2011b, pp. 17–20). In 2012, it proposed phasing out VaR (BCBS 2012, p. 20). In 2013, the Committee recognized numerous "weaknesses … in using Value-at-Risk (VaR) for determining regulatory capital requirements, including its inability to capture tail risk" (BCBS 2013, p. 3). It adopted the Expected Shortfall at a confidence level of 97.5 %, "a broadly similar level of risk capture as the 99th percentile VaR threshold" (BCBS 2014, pp. 14, 19)". Chen (2018).

We can define the algorithm of Expected Shortfall (ES) based on the Block Maxima Method as follows:

Let q_α be the α-*th* lower quantile for the logarithm returns, thus $q_\alpha = VaR(\alpha)$.

Let z_t be the standardized maximum returns and $f(z_t)$ be the density function of standardized maximum returns

$$ES(\alpha) = -\frac{1}{\alpha}\int_{-\infty}^{-q_n} z_t f(z_t) dz \tag{16}$$

If the standardized logarithm returns follow the extreme-value distribution then,

$$ES_\alpha(\alpha) = -\frac{\sigma_t}{n\alpha}\int_{-\infty}^{-q_n} z_t(1+\xi z_t)^{\frac{1}{\xi}-1} exp\left[-\frac{1}{n}(1+\xi z_t)^{1/\xi}\right] dz - \mu_t \tag{17}$$

where n is the block size, z_t are the standardized maximum returns, σ is the scale parameter of GEV distribution, ξ is the shape parameter and μ is the location. (See *Appendix B*).

2.4 Backtesting

2.4.1 Backtesting VaR

To evaluate the accuracy of the VaR estimates, we have used five standard tests: unconditional (LR_{uc}), independent and conditional coverage (LR_{ind} and LR_{cc}), Backtesting Criterion (BTC) and dynamic quantile (DQ) tests. All of these tests are based on an indicator variable. We have an exception when $r_{t+1} < VaR_\alpha$; then, the exception indicator variable (I_{t+1}) is equal to one otherwise its value is zero.

$$I_t = \begin{cases} 1 \; if \; r_t < VaR_t(\alpha) \\ 0 \; if \; r_t \geq VaR_t(\alpha) \end{cases} \tag{18}$$

Kupiec (1995) shows that if we assume that the probability of obtaining an exception is constant, the number of exceptions $x = \sum I_{t+1}$ follows a binomial distribution $B(N, \alpha)$, where N represents the number of observations. An accurate measure VaR_α should produce an unconditional coverage $\left(\hat{\alpha} = \frac{\sum I_{t+1}}{N}\right)$ equal to α percent. The unconditional coverage test has a null hypothesis $\hat{\alpha} = \alpha$, The rejection or non-rejection of the null hypothesis is verified through the maximum likelihood ratio:

$$LR_{uc} = 2[log(\hat{\alpha}^x(1-\hat{\alpha})^{N-x}) - log(\alpha^x(1-\alpha)^{N-x})] \tag{19}$$

which follows an asymptotic $\chi^2(1)$ distribution.

The LRcc test, developed by Christoffersen (1998), jointly examines whether the percentage of exceptions is statistically equal to the one expected ($\hat{\alpha} = \alpha$) and the serial independence of the exception indicator. The likelihood ratio statistic of this test is given by $LR_{cc} = LR_{uc} + LR_{ind}$, which is asymptotically distributed as $\chi^2(2)$, and the LR_{ind} statistic is the likelihood ratio statistic for the hypothesis of serial independence against first-order Markov dependence[11].

This test has some drawbacks: (i) the use of the Markov chain only allows the influence of past violations to be measured and does not allow for the influence of other exogenous variables; (ii) the way of verifying independence does not take into consideration other alternatives that take into account the independence of orders higher than one.

A similar test for the significance of the departure of $\hat{\alpha}$ from α is the backtesting criterion statistic (BTC):

$$Z = \left(N\hat{\alpha} - Na\right) / \sqrt{Na\left(1-a\right)} \tag{20}$$

which follows an asymptotic N(0,1) distribution.

Finally, the DQ test, proposed by Engle and Manganelli (2004), overcomes the two drawbacks of Christofersen's conditional coverage test and examines if the exception indicator is uncorrelated with any variable that belongs to the information set Ω_{t-1}, available when the VaR is calculated. This test is a Wald test of the hypothesis that all slopes are zero in the regression:

$$I_t = \beta_0 + \sum_{i=1}^{p} \beta_i I_{t-i} + \sum_{j=1}^{q} \mu_j X_{t-j} \tag{21}$$

where X_{t-j} are the explanatory variables contained in Ω_{t-1}. This statistic is introduced as five explanatory variable lags of VaR. Under the null hypothesis, the exception indicator

[11] The LRind statistic is $LR_{ind} = 2\left[\log L_A - \log L_0\right]$ and has an asymptotic $\chi^2(1)$ distribution. The likelihood function under the alternative hypothesis is $L_A = \left(1-\pi_{01}\right)^{N_{00}} \pi_{01}^{N_{01}} \left(1-\pi_{11}\right)^{N_{10}} \pi_{11}^{N_{11}}$, where N_{ij} denotes the number of observations in state j after having been in state i in the previous period, $\pi_{01} = N_{01}/(N_{00}+N_{01})$ and $\pi_{11} = N_{11}/(N_{10}+N_{11})$. The likelihood function under the null hypothesis ($\pi_{01} = \pi_{11} = \pi = (N_{11}+N_{01})/N$) is $L_0 = \left(1-\pi\right)^{N_{00}+N_{01}} \pi^{N_{01}+N_{11}}$.

cannot be explained by the level of VaR, i.e., $VaR(\alpha)$ is usually an explanatory variable to test if the probability of an exception depends on the level of the VaR.

2.4.2 Backtesting ES

The main issue with the ES is its difficulty to be backtested because it is not an elicitable measure (Gneiting, 2011; Heinrich, 2014; Weber 2006). A non-elicitable measure may present difficulties with robust estimation and backtesting. The elicitability is an important mathematical property for evaluating the interpretation of a forecast, especially in the context of risk measurement. An elicitable risk measure can be defined as the minimizer of a loss function or scoring function. However, recent papers have supported measures where ES is jointly elicitable with VaR (Fissler and Ziegel, 2016). Thus, in this paper, we use Righi and Ceretta (2013), McNeil and Frey (2000), Nolde and Ziegel (2017) and Bayer and Dimitriadis (2022). These last three tests are specified in the R package '*Esback*' (Bayer and Dimitriadis, 2020) which we will use in the development of the backtesting in this paper.

McNeil and Frey (2000) test is likely the most successful in the literature. These authors develop a test to verify that a model provides much better estimates of the conditional expected shortfall than any other. The authors are interested in the size of the discrepancy between the return r_{t+1} and the conditional expected shortfall forecast $ES_t(\alpha)$ in the event of quantile violation. The authors define the residuals as follows:

$$Y_{t+1} = \frac{r_{t+1} - ES_{t+1}(\alpha)}{\sigma_{t+1}} \tag{22}$$

Replacing Equation (11) and Equation (15) in Equation (22), we have the next expression:

$$y_{t+1} = z_{t+1} - E(z|z < q_\alpha) \tag{23}$$

These residuals are *i.i.d.* and conditional on $\{r_{t+1} < VaR_{t+1}(\alpha)\}$ or equivalent $\{z_{t+1} < q_\alpha\}$, they have an expected value of zero. Suppose we again backtest on days in the set T. We can form empirical versions of these residuals on those specific days on which violations have occurred, i.e., days on which $\{r_{t+1} < VaR_{t+1}(\alpha)\}$. The authors call these residuals exceedances and denote them by $\{\hat{y}_{t+1}: t \in T.\ \ r_{t+1} < VaR_{t+1}(\alpha)\}$

where $\hat{y}_{t+1} = \frac{r_{t+1} - \widehat{ES}_{t+1}(\alpha)}{\hat{\sigma}_{t+1}}$ and $\widehat{ES}_{t+1}(\alpha)$ is an estimation of the conditional expected shortfall.

Under the null hypothesis, in which we correctly estimate the dynamic of the process μ_{t+1} and σ_{t+1} and the first moment of the truncated innovation distribution $E(z|z < q_\alpha)$, these residuals should behave such as an *i.i.d* sample with a mean of zero. Thus, for testing whether the estimates of the Expected Shortfall are correct, we must test if the sample mean of the residual is equal to zero against the alternative that the mean of y is negative. Indeed, given a sample $\{y_{t+1}\}$ of size N (where N is the number of violations in the T period), the sample mean $\bar{y}$ converges in distribution to standard normality, as N tends to ∞ by the central limit theorem. In other words, given mean μ_y and variance σ_y of population

$$\sqrt{N}\left(\frac{\bar{y} - \mu_y}{\sigma_y}\right) \to N(0,1) \tag{24}$$

By applying the Central Limit Theorem, the statistics for testing the null hypothesis are given by

$$t = \frac{\bar{y}}{\frac{S_y}{\sqrt{N}}} \sim t_{N-1} \tag{25}$$

where $\bar{y}$ and S_y are the sample mean and the sample standard deviation, respectively, of the exceedance residuals.

We use Bootstrap to test the hypothesis, which does not assume the underlying distribution of the residuals.

Righi and Ceretta (2013) propose a test based on the dispersion of the truncated distribution of returns beyond VaR instead of the whole probability function. They do not limit the approach to the Gaussian case, unlike McNeil and Frey, allowing for other distribution functions. Righi & Ceretta's approach is based on the BT_t series which represents the standardized VaR violations with respect to the ES and the standard deviation truncated by VaR (SD_t).

$$BT_t = \begin{cases} \frac{r_t - ES_t}{SD_t} & if\ r_t < VaR_t(\alpha) \\ 0 & if\ r_t \geq VaR_t(\alpha) \end{cases} \tag{26}$$

$$SD_t = (\sigma^2 VaR[z_t | z_t < F^{-1}(\alpha)])^{1/2} \tag{27}$$

This test verifies whether the average observed deviation from ES is zero, $H_0: E[BT_t] = 0$, $H_1: E[BT_t] < 0$, i.e in case of a violation $r_t < VaR_t(\alpha)$, this test allows us to know how far the occurred loss is from the ES, computed in units of the dispersion measure. This test also requires simulation analysis to compute the p-value.

Nolde and Ziegel (2017) propose a two-stage backtest. The first stage is based on a traditional backtest. Following Fissler, Ziegel, and Gneiting (2015), traditional backtests are those with a null hypothesis of the type "The risk measurement procedure is correct" which is rejected if the risk measurement under consideration is making inaccurate predictions. Nevertheless, these kinds of tests present mainly two drawbacks: They are not suited to compare different risk estimation procedures and they may be insensitive when increasing the information set (Holzmann and Eulert 2014, Davis 2016). According to Nolde and Ziegel (2017), traditional backtesting can be formalized in the form of conditional calibration tests. In the case of ES, the conditional calibration test is related to the backtest for ES of McNeil and Frey (2000) based on exceedance residuals.

To avoid the above-mentioned limitations of traditional backtests, the authors suggest that regulators may additionally apply a comparative backtest, based on a consistent scoring function, in the second stage in cases where a traditional backtest is passed[12]. This requires a standard model against which the bank's internal model is to be tested. However, comparative backtests need an elicitable risk measure. As we highlighted at the beginning of this section, ES is a non-elicitable ES measure, but it turns out to be jointly elicitable with VaR.

To sum up, this two-stage procedure lets to test that conditional VaR and ES predictions are at least as large as their optimal conditional predictors and that the internal model predicts at least as well as the standard model.

Finally, Bayer and Dimitriadis (2022) introduce a regression-based backtest for ES forecast[13], which extends the classical Mincer and Zarnowitz (1969) test to ES. They propose to estimate a regression that models the conditional ES at level α as a linear

[12] Bank for International Settlements (2014) and Fissler, Ziegel and Gneiting (2015) have recently proposed to replace traditional backtests by comparative backtests based on strictly consistent scoring functions.

[13] We can find recent literature using a regression procedure to backtest ES in Bayer and Dimitriadis (2022), Barendse et. al., (2018) and Patton et al., (2019).

function $Es_\alpha = Y_t|\mathcal{F}_{t-1} = \gamma_1 + \gamma_2\hat{e}_t$, where Y_t represents the daily log-returns of a financial asset and it is the response variable, ES forecasts $\hat{e}_t$ is the explanatory variable and $\mathcal{F}_{t-1}$ is the given information set. For correctly specified ES forecasts, the intercept and slope parameters equal to zero and one are tested by using a Wald statistic.

Because of the non-elicitable character of ES, it is not feasible to estimate the regression parameters for the ES stand-alone. Thus, the authors propose to build a test based on a joint VaR and ES regression, by specifying an auxiliary quantile regression equation that allows for different specifications: *Auxiliary*, *Strict*, and *Intercept* Expected Shortfall (ESR) backtests (Bayer and Dimitriadis, 2022). First, using auxiliary VaR forecasts $\hat{v}_t$ as the explanatory variable in the quantile equation, but only test the ES specific parameters, they refer to this test as the *Auxiliary ESR* backtest. Second, using the ES forecasts $\hat{e}_t$ as the explanatory variable in both, the quantile and the ES equation and again only test on the ES pecific parameters. The authors refer to this test as the *Strict ESR* backtest. And finally, they introduce an intercept variant of the Strict ESR backtest by fixing the slope parameter in the regression to one, and by only estimating and testing the intercept term. They refer to this backtest as the *Intercept ESR* backtest.

3. Case of study

3.1 Dataset Overview

The data used throughout this study consist of daily prices of the S&P500 stock index extracted from the Yahoo Finance database. We use daily data for the period of January 3, 2000, through December 31, 2019. The index is transformed into returns by taking the logarithmic differences of the closing daily price in percentage. The number of data in the sample is 5029.

Figure 1 displays the evolution of the daily index and returns of the S&P 500 during the sample length. The plotted returns over time show heteroskedasticity and clustered volatility, i.e variance changes over time, alternating periods of low volatility followed by high volatility, being a distinctive feature of financial returns (Bollerslev et al.,1994). The volatility of the index was particularly high from 2008 to 2009 coinciding with the global financial crisis. Table 1 provides basic descriptive statistics of data.

The unconditional mean and median are about zero as expected and while the standard deviation is around 1.19. Evidence of non-normality can be found by looking at the negative skewness statistic and excess kurtosis (above 3) that imply a high peaked distribution, with a thicker tail that is skewed to the left, i.e the losses are somewhat larger than the profits. Similarly, the p-value for the Jarque-Bera test tells us that we can strongly reject the normality assumption.

All this evidence suggests that the empirical distribution of daily returns cannot be fit by a normal distribution.

3.2 Fitting the GEV distribution.

Before we apply the block maxima method we must first choose the block size, *n*. As we highlighted in Section 2.2, the main problem in selecting block size is the trade-off between bias and variance. In the aforementioned section, different references were extracted from the literature about the selection of the block in the field of financial returns. Following these references, in this section, the GEV distribution is fitted to the data of maximums we get from 9 different block sizes corresponding to an approximate trading week (5), two weeks (10), a month (21), six weeks (31), two months (42), a quarter (63), a semester (126), nine months (189) and a year (252) respectively.

After estimating the parameters of GEVD (μ,σ,ξ) using Maximum Likelihood[14], one can assess the model fit based on procedures such as parameter evaluation and distribution fit plots. Regarding this, various diagnostic plots are shown in Figure 2. In this Figure, we display for each block size the QQ-plot and a plot with the density function theoretical (GEVD) and the histogram.

As we can see, for each block size the corresponding density seems consistent with the histogram. In the QQ-plot, we can observe that the points are near-linear, although this precision is lost as we use a larger block size and therefore, we have a lower number of maximums.

Figure 3 displays the plots with the fitted GEV distribution and the empirical cumulative distribution. GEV distribution seems to fit quite well the distributions of maximums for each block size and in concordance with this, the null hypothesis for the Kolmogorov-Smirnov test cannot be rejected in any case (see Table 2).

[14] We find Maximum Likelihood estimation as the most appropriate technique according to Gaines and Denny (1993) and Leder et. al. (1998).

Similar tests but more sensitive to the extremes of the distribution, such as Anderson-Darling and Cramer-von Mises, have been applied and the results show that in the case of the weekly block, we reject the null hypothesis that the empirical estimate distribution function comes from the underlying theoretical distribution (see Table 2).[15]

All these graphical and test diagnostics lend support to the fitted GEVD model for blocks larger than 5. Therefore, we will not proceed with the quantification of risk measures for the weekly block.

Finally, we can check the sensitivity of the parameters of the GEVD for different block lengths. In Table 2 we observe that the estimation of the shape parameter, which determines the weight of the tail in the distribution, increases by 678% when the block size moves from 5 to 252.

As we can observe, the scale and location parameters are statistically significant at 5%. This is not the case for the shape parameter. The fact that the shape parameter is not statistically significant for any block size implies that ξ =0, which determines that the non-degenerate distribution function is of Gumbel type and not Frechet. According to McNeil et al. (2005), although power-tailed distributions, and therefore of the Frechet class, are interesting when working with financial returns, this class is not the only interesting one, since the Gumbel class also contains many distributions with heavier tails than normal.

In line with these results, we can see graphically in Figure 4 the plot for the estimated parameters as a function of the block size. We observe that there are different changes in the trend of the shape parameter, however, from block 42 onwards, its value tends to concentrate and stabilize around 0.10.

The accuracy of the estimations also decreases considerably as the block length increases, as we can see with the band of the confidence interval in the dashed line, implying considerable uncertainty about the value of ξ. According to the asymptotic distribution theory, there is less accuracy in the standard error of the parameter estimates as the block size increase, i.e, decrease the number of maxima (Coles, 2001).

We can highlight some aspects of the behavior of the estimated parameters. First, we can observe the stability of the tail index around a particular value since it is an

[15] The calculations were performed with the '*gnfit*' R package (Saeb, 2018).

intrinsic parameter of the process of returns (see Figure 5). Second, there is an increase in the location parameter, with values that go from 0.8082 to 3.2388. Third, it is known that the behavior of the scale parameter is not specified a priori as the distribution of extremes may contract or expand[16]. In this study, the scale parameter goes around 0.6 up to block 63 from which the parameter increases.

We can find similar results in Tsay (2005) who performs the estimation for different sample sizes (monthly, quarterly, weekly, and yearly) and concludes that the estimation of the scale and location parameters increase when the sample size increases. Also, he determines that the shape parameter is stable for extreme values when the sample size is around 63. Again the same conclusion is obtained in Longin (1996) about the stability of the tail index, especially for periods longer than a semester or block size 126. From this analysis, we can conclude that the estimation of the parameters of the GEV distribution is sensitive to changes in block sizes.

3.3 Market risk estimation.

In this section, we analyze the market risk estimation (Value at Risk and Expected Shortfall) that we get from the Block Maximum Method[17] and the sensitivity of these risk measures to changes in block sizes. These estimations are obtained by applying Equations (14) and (17).

To estimate the risk measures, we divided the sample period into a learning sample from January 3, 2000, to December 31, 2010, and a forecast sample from January 3, 2011, to December 31, 2019. The daily forecasting is obtained one day ahead at the 97.5% and 99% confidence levels according to the Basel Committee on Banking Supervision's standards. In particular, we report the forecasts based on a series of rolling time windows, instead of recursive estimation (see Elliott and Timmermann, 2013). The use of rolling time windows or non-overlapping estimation windows (in line with the current market risk regulations introduced by Basel III in 2014[18]) not only avoids the dependence on the

---

[16] Gumbel (1958).

[17] From losses previously multiplied by (-1) and therefore in the right tail (short position), we obtain the maxima per block and estimate the parameters to obtain the percentile. This percentile will be therefore of positive sign. We will multiply again by (-1) to move to the left tail of the losses.

[18] The Basel Committee on Banking Supervision (BSBC) was keen on the overlapping approach in the version of the Basel III Proposals in 2013, but not in the subsequent revised version in 2014.

underlying time series, which may lead to biased estimations but also prevents anomalous events will be repeated in the sample.

Figure 6 displays the VaR estimates we get from different block sizes at a 2.5% probability. At first glance, it can be seen that the VaR estimates with a block size equal to 10 underestimate risk while block sizes above 42 tend to overestimate risk. In the case of ES, the overestimation of risk is even more obvious (see Figure 7). Only block 10 seems to capture somewhat better the risk. We get similar results for VaR and ES estimates at a 1% probability.

The graphical analysis presented above suggests that the risk estimates obtained through VaR or ES measures are very sensitive to the selected block size. To corroborate this result, we compute some descriptive statistics of these estimations by block size (see Table 3). As can be seen, there are significant differences in the mean and variance of these estimates, depending on the block size. Overall, we find that the larger the block size, the higher the estimated mean loss and variance. For example, with a probability of 2.5%, the average loss goes from -0.43 with a block size of 10 to -3.63 with a block size of 252. The minimum of the VaR estimations is also very different, going from -1.52 with a block size of 10 to -13.75 with a block size of 252. Similar results are obtained for the ES at 2.5% and 1% probability.

Although it seems very clear that there are important differences in market risk estimations depending on the block size selected we check statistically if there is any significant difference between the means of these estimates to which we can use a one-way ANOVA test[19], see Table 4.

As the p-value is less than the significance level of 0.05, we can conclude that there are significant differences between the means of the estimates generated for all the block sizes.

To determine if the difference between means of specific pairs of block sizes is statistically significant, it is possible to perform multiple pairwise-comparison through a

[19] This test compares means between groups of data. In our case, these groups are the risk estimates obtained for each block size, which is the factor variable. In a one-way ANOVA test, the null hypothesis states that the differences between means equal zero, i.e, a significant p-value determines that some of the groups are different.

Tukey HSD test (Tukey Honest Significant Differences). The null hypothesis states that the difference between the means of a pair of block sizes equals zero.

Table 5 shows p-values for this pairwise test. Panel A displays the test for VaR estimates at 1% and 2.5% probability and Panel B does the same for ES estimates at 1% and 2.5% probability. As we can see in Panel A, any pair of block sizes provide similar mean estimates as the p-value is less than 0.05 and we reject the null hypothesis. Panel B shows the comparison for ES estimates. In this case, for some pairs of blocks, we accept the null hypothesis. Thus, at a 1% probability, we find that ES estimates are similar to blocks 31-63, 42-189 and 126-252. At a 2.5% probability, we find that there is not much difference in the ES estimates in pairs 21-31, 21-63, 31-63, 42-189 and 126-252.

To sum up, the analysis presented in this section suggests that the risk estimates obtained by the Maximum Block method, via VaR or ES, are highly sensitive to the selected block size, being very different from each other depending on the size chosen. In order to see the quality of these risk estimates, we will analyze the results using backtesting.

3.4. Backtesting.

In this section, we analyze the accuracy of the market risk estimations obtained from the BM method. We also analyze if the quality of these measures is related to the block size. The results are displayed in Table 6.

Column three of Table 6, reports the percentage of times that the S&P500 returns go below VaR at a 2.5% probability level. Consistent with what we observed in Figure 6, the percentage of exceptions obtained by the smallest block sizes, 10 and 21, is too high, indicating that these blocks underestimate the risk. On the contrary, the higher blocks size (126, 189 and 252) clearly overestimate risk, as they provide a reduced number of exceptions. Thus, at first glance, it seems that the lowest and highest block sizes considered provide inaccurate VaR estimation. Just only the intermediate block sizes, 31, 42 and 63, provide an exception percentage around to the theoretical one.

To test formally the accuracy of the VaR estimate we use several standard tests: unconditional (LR_{uc}), independent (LR_{ind}), conditional coverage (LR_{cc}), the backtesting criterion (BTC), and dynamic quantile (DQ) tests. We consider that a VaR measure is accurate if it passes at least 4 of the 5 tests[20]. According to this criterium, only block size 42 yields accurate VaR estimates at a 97.5% confidence level. The lower blocks such as 10 or 21 only pass one test. In contrast, higher blocks, 126, 189 and 252 only pass 2 or 3 out of the 5 tests considered, indicating that these estimations are inaccurate. On the other hand, at a 99% confidence level, only blocks 126 and 189 pass at least 4 out of 5 tests. The rest of the blocks considered just only pass one or two tests.

Regarding the Expected Shortfall measure, the results are even more discouraging. To test whether the ES estimations should be accepted we have used the procedure proposed by McNeil and Frey (2000), Righi and Ceretta (2013), Nolde and Ziegel (2007), and Bayer and Dimitriadis (2022)[21]. We consider that ES measure is accurate if it passes at least 3 of the 4 tests. No block passes 3 of the 4 proposed tests. At 97.5% confidence level, blocks 10, 21, 31, 42, and 126 pass just one of the 4 tests. The remaining blocks (63,189,252) pass 2 tests. At a 99% confidence level, the results are also very poor as most blocks pass just two of the 4 tests.

To sum up, the results obtained in this analysis are sensible to the confidence level, the block size, and the market risk measure considered VaR or ES. In the case of VaR, just only block size 42 provide accurate estimates at a 97.5% confidence level. However, at a 99% confidence level, block size 126 and 189 seems to be optimal. In the case of the ES measure any block size yield appropriate estimates.

Given the results presented and in line with the previously mentioned literature in Section 1, we can conclude that the BMM is not the most appropriate when it comes to market risk estimation under the EVT framework. The accuracy of the risk estimates seems to be very sensitive to the chosen block size. Thus, the block size may be a critical aspect that should not be chosen ad hoc. The use of this method could require further research on optimal block size selection techniques.

4. Robustness analysis.

To corroborate the validity of the results reported in the previous section -which are a few discouraging- we extend the study to a set of 14 assets: 7 stock market indexes

[20] Following the criteria in Campbell (2005), only sequences that satisfy unconditional coverage and independence properties can be described as evidence of an accurate VaR model.

[21] The calculations were performed with the '*esback*' package in R, Bayer and Dimitriadis (2020).

(CAC40, DAX30, FTSE100, Hang Seng, IBEX35, Merval, and Nikkei), 4 commodities (Copper, Gold, Crude Oil Brent, and Silver) and 3 exchange rates (₤ /€, $/€ and ¥/€).[22]

Table 7 shows the percentage of exceptions for each block size. Panel A displays the results obtained for a confidence level of 97.5% and Panel B displays the results at a 99% confidence level. In this table, blocks that pass at least 4 out of 5 tests are highlighted in light grey.

At a 97.5% confidence level (Panel A), block size 31 provides accurate VaR estimates in 9 out of the 14 assets considered while block size 42 provides accurate VaR estimates in 5 out of the 14 assets. As in the case of the S&P500, it seems that at this confidence level, the intermediate block sizes provide the best VaR estimates. Again, we find that the lowest block sizes, 10 and 21, provide very poor VaR estimations as they clearly underestimate risk. Along the same line, the highest blocks overestimate risk.

For a confidence level of 99% (Panel B), block 63 provides accurate VaR estimates in 7 out of 14 assets (5 of which are indexes). Block 126 passes the tests for 4 out of 14 assets and block 189 for 4 out of 14 assets. Taking into account the percentage of exceptions closest to the theoretical one, as well as the tests, the block size that seems to have the best performance for the calculation of estimates in the quantification of risk is 63.

This analysis helps us to corroborate that risk estimates are indeed sensitive to the choice of block size when applying the Block Maxima Method in the EVT framework. Very small block sizes, which result in too many observations and being included as extremes, lead to underestimates of risk, with very high exception rates. Conversely, large block sizes, such as the annual block, result in the inclusion of too few observations and therefore very low exception rates and an overestimated VaR.

Table 8 displays the results obtained for the backtesting of the ES measure. In this table, we report the number of tests that do not reject the null hypothesis (the average observed deviation from ES, in case of a violation r < VaR, equals zero). For the stock indexes, no block size provides good results as in the case of the S&P500 index. For the exchange rates and the commodities, some block sizes provide good results but again they are not consistent. For instance, at a 99% confidence level, block size 126 provides

[22] The daily data has been extracted from the Yahoo Finance database.

adequate ES estimates for 4 out of the 7 assets, and block size 31 does so in 2 out of the 7 cases. The rest of the blocks perform well just in one case or no one. At a 97.5% confidence, we find similar results although somewhat worse.

5. Comparison of POT and BMM VaR for market risk.

As we know, the main difference between the two methods lies in the way they handle extreme events. While the BMM groups data into fixed-length blocks and uses the maximum value within each block to estimate the parameters of the Generalized Extreme Value (GEV) distribution, in the case of the POT method, on the other hand, the extreme values above a particular threshold are modeled using a Generalized Pareto distribution (GPD). Both methods could have their advantages and limitations, and the choice between the two methods will depend on the specific characteristics of the data being analyzed. Regarding the performance of the BM method in comparison to the POT method in the area of finance, the literature is somewhat ambiguous (see for instance Szubzda and Chlebus, 2019).

With the aim of contrasting the performance of the Maximum Block model concerning the POT method in estimating market risk, we present in this section a comparison of the VaR estimates obtained under the POT method and the Block Maxima method for the set of assets analyzed in this study (7 indexes, 4 commodities and 3 exchanges rates).

As in the present study, to estimate the VaR measure under the POT method, we take the forecast period from January 3, 2011, to the end of December 2019. For each day of the forecast period, we will generate estimations of VaR. These forecasting measures are obtained one day ahead at the 97.5% and 99% confidence levels. Assuming that for a certain threshold u the distribution of observations above the threshold is the GPD, we get the quantile:

$$q_{1-\alpha} = u + \frac{\sigma}{\xi}\left(\left(\frac{n}{N_u}(1-\alpha)\right)^{-\xi} - 1\right)$$

where n is the total number of observations and N_u the number of observations above the threshold u and, σ and ξ are the scale and shape parameters, respectively, of the GPD. From this quantile we can get the VaR measure as follows:

$$VaR_t(\alpha) = \mu_t + \sigma_t \, q_{1-\alpha}$$

being μ_t and σ_t represent the conditional mean and the conditional standard deviation of the returns (calculated through an APARCH conditional variance structure) and $q_{1-\alpha}$ is the quantile $(1-\alpha)$ of the GPD.

Table 9 shows the percentage of exceptions obtained for the VaR estimates at a probability level of 2.5% for block sizes 10 to 252 and thresholds 87 to 94. Table 10 shows the same for a 1% probability level. By analyzing these Tables, we can draw the following conclusions.

In the case of BMM, the percentage of exceptions varies greatly depending on the block size chosen. Thus, for smaller blocks, we found a strong underestimation of the VaR measure. The opposite is true for larger blocks. Only in some cases (blocks 63 and 126 for a probability level of 1% and blocks 31 and 42 for a probability level of 2.5%), the exception percentage is close to the expected one.

This leads us to think that obtaining accurate market risk estimates through this method may depend on the selected block size and probability level, making necessary some computational method to determine the optimal block size for a given sample (see for instance Özari et al., 2019). Therefore the BMM may not be the most reliable approach to estimating market risk if an ad hoc block size is selected.

In the case of the POT method, its greater consistency can be observed. For most assets (with some exceptions such as the CAC40 and Merval) the percentage of exceptions obtained from the VaR estimates is close to the expected probability level regardless of the threshold chosen, i.e, no matter what threshold we choose, the VaR estimates could be accurate. The selection of the threshold in the POT method may not be so critical when estimating market risk (see Benito et al., 2023) and this makes it the most convenient method for estimating market risk under the EVT framework.

This comparison shows the superiority of the POT method over the BBM in estimating market risk.

6. Conclusions.

This study is framed within the Extreme Value Theory (EVT) and more specifically within the Block Maxima Method (BMM).

The Extreme Value Theory focuses on the limiting distribution of the extreme values observed over a long period, which is independent of the distribution of the values themselves. In the EVT context, there are two approaches. One of them is the BMM which models directly the distribution of minimum or maximum realizations. The other one, Peaks over Threshold (POT), models the exceedances above a particular threshold.

Extreme Value Theory has been widely used in many sciences such as engineering, insurance, meteorology and more recently finance. In this last area, the POT method has reaped very good results, especially in quantifying market risk. The performance of the BMM in this field has been scarcely studied. Thus, the study carried out in this paper contributes to three aspects. First, it analyses the performance of the BMM in estimating the market risk of a portfolio. As we have said before, few studies have analyzed this issue. Second, we assess to what extent the selection of the block size is significant in quantifying market risk. And finally, a third contribution emerges from this work. As in this study, we use similar datasets to those used in Chapter III[23] we can compare the performance of the BMM with the POT method in estimating market risk.

For this study daily data of the S&P500 returns have been used from January 3rd, 2000 to December 31st, 2019. To quantify market risk two measures have been considered: Value at Risk (VaR) and Expected Shortfall (ES**)**. For calculating risk, two confidence levels have been used: 97.5% and 99%. To apply BMM, nine different block sizes have been considered: 5, 10, 21, 31, 42, 63, 126, 189 and 252 observations.

The results obtained are as follows. First, we detect that the VaR estimations are highly sensitive to the block size selected for fitting GEVD. Both, the lowest block size and the higher block size provide inaccurate market risk estimations. Just only the intermediate block size 42 at 97.5% confidence level and 126 and 189 at 99% confidence level provide reasonable VaR estimations. The remaining blocks provide imprecise

[23] Chapter III has been published in the Journal *Risk Management.* The complete reference is Benito Muela, S., López-Martín, C. and Navarro Cervantes, M. Á. (2023). "Assessing the importance of the choice threshold in quantifying market risk under the POT approach (EVT)". Risk Management 25, 6. DOI: 10.1057/s41283-022-00106-w (Available online 6th January 2023).

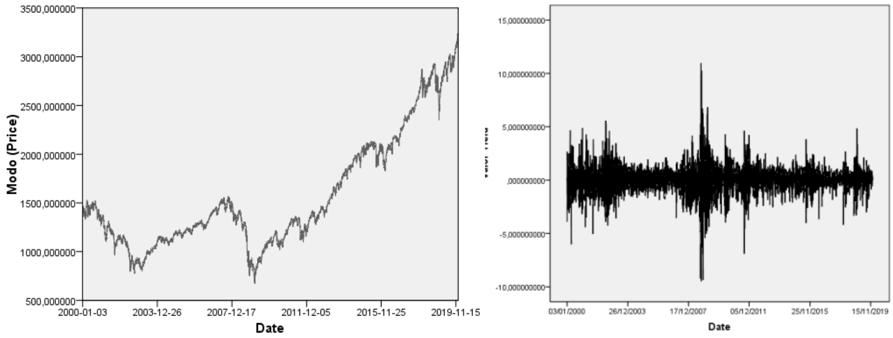
estimates of VaR. Second, in the case of Expected Shortfall, we found a strong risk overestimation. Any block size provides appropriate risk estimations.

To corroborate these results, we have extended the S&P 500 index study to a set of 14 assets (stock market indexes, commodities, and rate exchanges). The results go in the same direction, indicating that the Block Maxima Method does not provide appropriate market risk estimations and then fundamentally depends on the block size selected. Just only the intermediate block size seems to perform well although the results are not robust, in the sense that no one block size performs well for all assets considered.

Finally, a comparison of the VaR estimates obtained under the POT method and the BM method is carried out, through the percentages of exceptions. This comparison shows the consistency of the POT method regardless of the chosen threshold compared to the BM method, in which obtaining a level of exceptions close to the theoretical one depends on the chosen block size. This leads us to conclude the need to determine an optimal block size selection method in case of using this method in the estimation of market risk.

Furthermore, one aspect to bear in mind when considering the results obtained and which could be of interest to future research is the stationarity of the data. We have only considered the case of stationary and independently distributed random variables instead of non-iid variables which are supposed to hold in most financial markets because of clustering property.

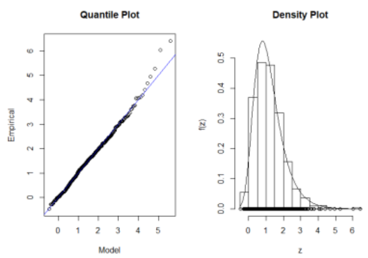
Figure 1. S&P500



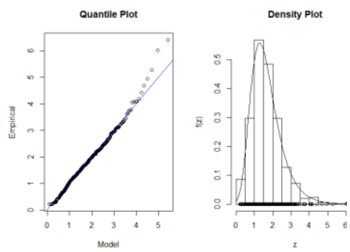
Note. Daily closing prices, percentage, and log-returns of the S&P 500 Index from January 03rd, 2000, through December 31st, 2019.

Figure. 2. Quantile and density plots

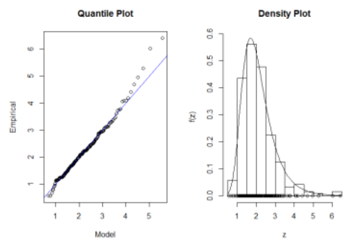
Block size=5



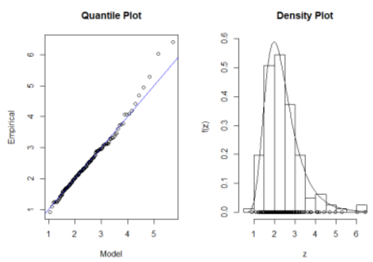
Block size=10



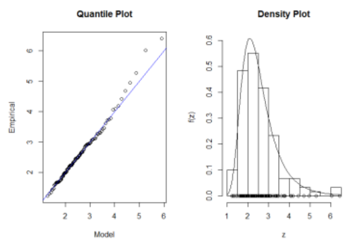
Block size=21



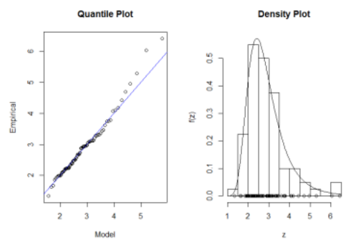
Block size=31



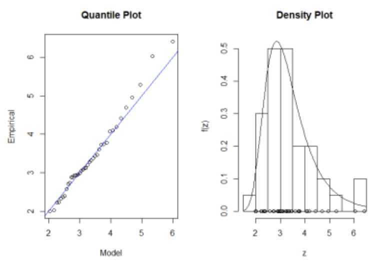
Block size=42



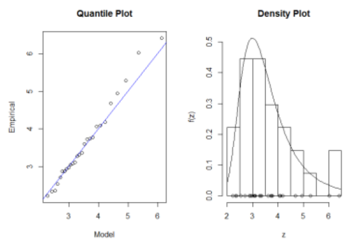
Block size=63



Block size=126



Block size=189



Block size=252

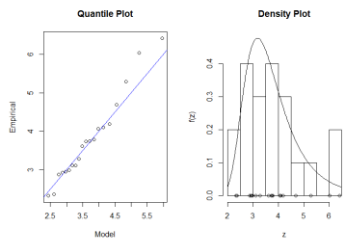


Figure 3. GEV Distribution.

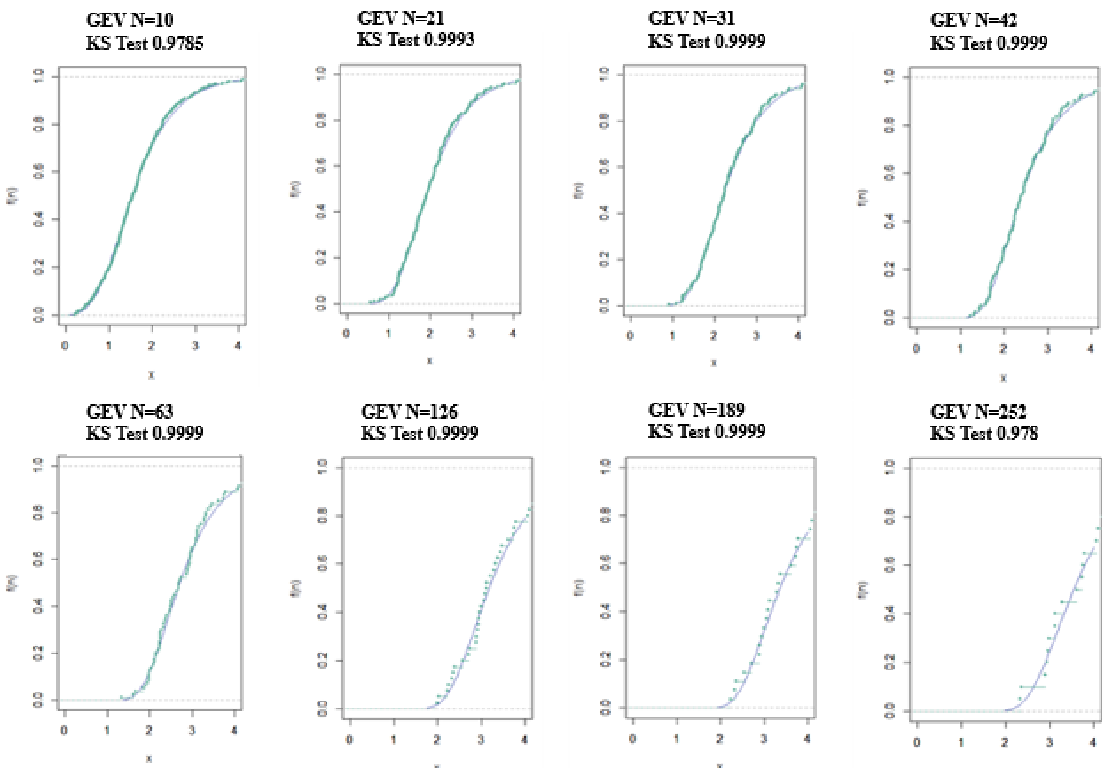


Figure 4. Maximum Likelihood Estimations (GEVD)

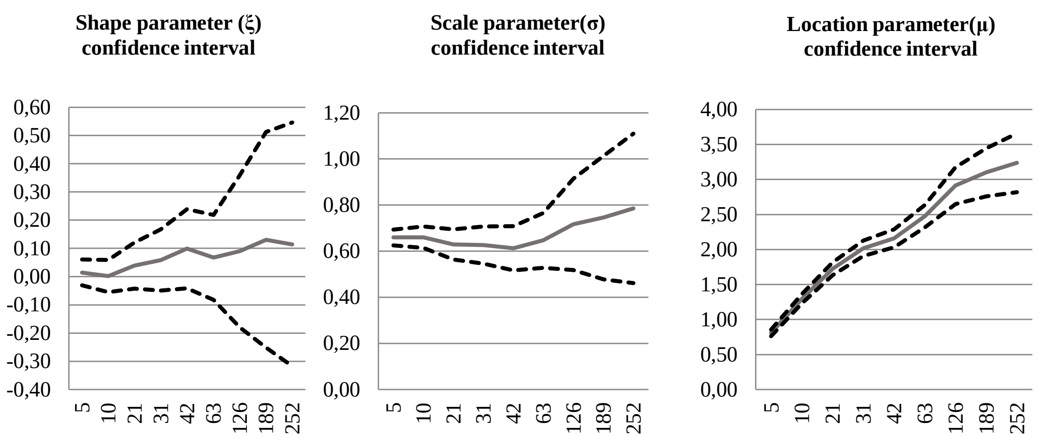


Figure 5. Tail Index (GEVD)

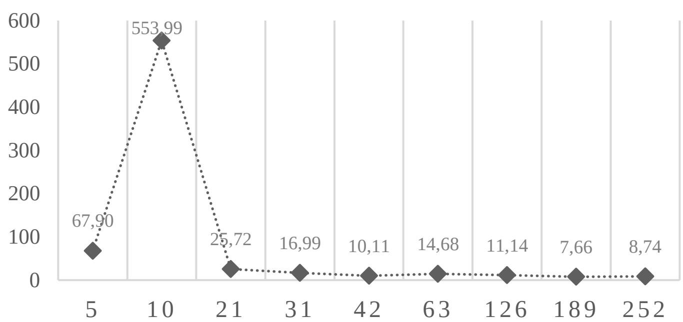


Figure 6. S&P500 Value at Risk estimates

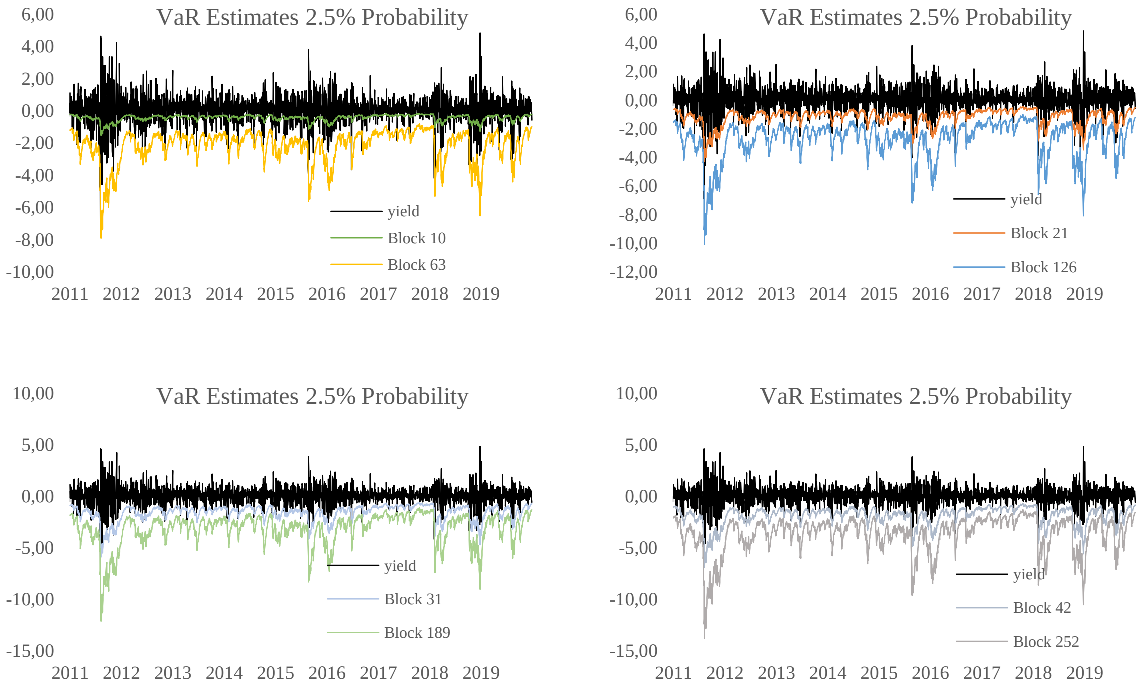


Figure 7. S&P500 Expected Shortfall estimates

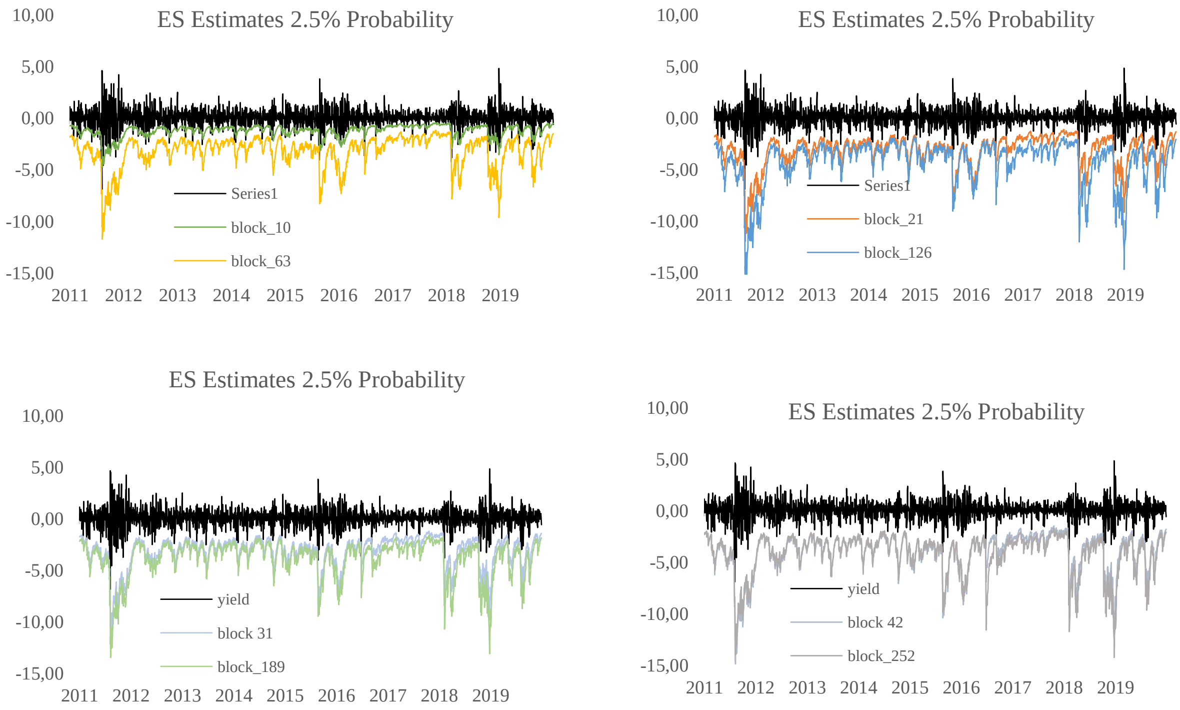


Table 1. Descriptive Statistics

	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque Bera
S&P 500	0.0158	0.0543	10.957	-9.469	1.1899	-0.230 (0.035)	8.652* (0.069)	15695 (0.00)

Note: This Table presents the descriptive statistics of the daily returns of the S&P500. The sample period is from January 3rd, 2000 to December 31st, 2019. The index return is calculated as R_t=100(ln(P_t)-ln(P_{t-1})) where P_t is the index level for period t. Standard errors of the skewness and excess kurtosis are calculated as $\sqrt{6/n}$ and $\sqrt{24/n}$ respectively. The JB statistic is distributed as the Chi-square with two degrees of freedom. (*) denotes significance at the 5% level.

Table 2. Maximum Likelihood Estimations (GEVD)

Block size	Number of Maximum	$\hat{\xi}$	$\hat{\sigma}$	$\hat{\mu}$	KS p-value	AD p-value	CvM p-value
5	1006	0.0147 (0.0227)	0.6594* (0.0171)	0.8082* (0.0233)	0.5393	0.01186	0.0122
10	503	0.0018 (0.0286)	0.6603* (0.0234)	1.2956* (0.0327)	0.9785	0.1692	0.1427
21	239	0.0388 (0.0409)	0.6298* (0.0326)	1.7241* (0.0450)	0.9993	0.7419	0.4948
31	162	0.0588 (0.0543)	0.6260* (0.0403)	2.0159* (0.0549)	0.9999	0.9506	0.92912
42	120	0.0989 (0.0699)	0.6128* (0.0473)	2.1587* (0.0634)	0.9999	0.9231	0.9369
63	80	0.0681 (0.0752)	0.6472* (0.0593)	2.4760* (0.0808)	0.9999	0.7096	0.6818
126	40	0.0897 (0.1346)	0.7164* (0.0999)	2.9132* (0.1313)	0.9999	0.8488	0.8939
189	27	0.1304 (0.1908)	0.7458* (0.1337)	3.1036* (0.1715)	0.9999	0.9914	0.9777
252	20	0.1144 (0.2158)	0.7856* (0.1621)	3.2388* (0.2099)	0.978	0.7168	0.6757

Note: We shade in light grey the cases where the p-value is under 5% and thus the null hypothesis is rejected. We mark with an asterisk the statistically significant parameters at the significance level of 5%. The standard error of the estimated parameters appears in parentheses.

Table 3. Descriptive Statistics of market risk estimations

Panel A. VaR										
	Confidence level 97.5%					Confidence level 99%				
Block size	Mean	Median	Std. Dev.	Max.	Min.	Mean	Median	Std. Dev.	Max.	Min.
10	-0.43	-0.37	0.19	-0.16	-1.52	0.04	0.03	0.03	0.24	0.01
21	-1.16	-1.00	0.54	-0.43	-4.29	-0.70	-0.60	0.33	-0.26	-2.57
31	-1.58	-1.35	0.74	-0.58	-5.88	-1.12	-0.97	0.53	-0.41	-4.21
42	-1.86	-1.59	0.87	-0.69	-6.90	-1.43	-1.23	0.67	-0.53	-5.33
63	-2.14	-1.84	1.01	-0.80	-7.90	-1.77	-1.52	0.83	-0.66	-6.52
126	-2.71	-2.33	1.28	-1.00	-10.09	-2.32	-1.99	1.10	-0.86	-8.68
189	-3.15	-2.72	1.51	-1.11	-12.12	-2.73	-2.35	1.30	-0.99	-10.36
252	-3.63	-3.13	1.72	-1.29	-13.75	-3.15	-2.70	1.49	-1.14	-11.83
Panel B. ES										
	Confidence level 97.5%					Confidence level 99%				
Block size	Mean	Median	Std. Dev.	Max.	Min.	Mean	Median	Std. Dev.	Max.	Min.
10	-1.22	-1.04	0.59	-0.42	-4.74	-0.71	-0.62	0.36	-0.22	-2.95
21	-3.12	-2.69	1.50	-1.10	-11.96	-3.31	-2.86	1.59	-1.16	-12.74
31	-3.16	-2.69	1.49	-1.20	-11.65	-3.16	-2.71	1.50	-1.15	-11.73
42	-4.00	-3.42	1.89	-1.48	-14.86	-3.88	-3.31	1.84	-1.45	-14.45
63	-3.17	-2.71	1.50	-1.16	-11.83	-3.16	-2.70	1.50	-1.15	-11.72
126	-4.24	-3.48	2.13	-1.79	-16.58	-4.17	-3.43	2.09	-1.77	-16.17
189	-3.86	-3.27	1.80	-1.59	-13.47	-3.80	-3.31	1.73	-1.40	-13.45
252	-4.16	-3.52	1.95	-1.74	-14.24	-4.18	-3.64	1.89	-1.62	-13.98

Note: This Table presents the descriptive statistics for VaR estimates (Panel A) and ES estimates (Panel B) at 99% and 97.5% confidence levels for every block size.

Table 4. ANOVA test

	F value	Pr(<F)	
VaR 99%	3111	0.00	***
VaR 97.5%	2123	0.00	***
ES 99%	1068	0.00	***
ES 97.5%	793.6	0.00	***

Note: Significance codes 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1' '. A significant p-value indicates that some of the estimated means are different from the other.

Table 5. Tukey HSD test VaR and ES estimates.

Panel A. Tukey Test. Differences between mean VaR estimates.

	VaR 99%							VaR 97.5%						
	10	21	31	42	63	126	189	10	21	31	42	63	126	189
21	0.000							0.000						
31	0.000	0.000						0.000	0.000					
42	0.000	0.000	0.000					0.000	0.000	0.000				
63	0.000	0.000	0.000	0.000				0.000	0.000	0.000	0.000			
126	0.000	0.000	0.000	0.000	0.000			0.000	0.000	0.000	0.000	0.000		
189	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	
252	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Panel A. Tukey Test. Differences between mean ES estimates.

	ES 99%							ES 97.5%						
	10	21	31	42	63	126	189	10	21	31	42	63	126	189
21	0.001							0.000						
31	0.001	0.033						0.000	0.995					
42	0.001	0.001	0.001					0.000	0.000	0.000				
63	0.001	0.035	1.000	0.001				0.000	0.977	0.999	0.000			
126	0.001	0.001	0.001	0.001	0.001			0.000	0.000	0.000	0.000	0.000		
189	0.001	0.001	0.001	0.677	0.001	0.001		0.000	0.000	0.000	0.089	0.000	0.000	
252	0.001	0.001	0.001	0.001	0.001	1.000	0.001	0.000	0.000	0.000	0.029	0.000	0.697	0.000

Note: This table presents the p-value of the statistics. We highlight in light grey the cases in which the null hypothesis is rejected and thus the pairs of block sizes provide a similar mean of the estimates.

Table 6. Backtesting VaR and ES S&P500

Panel A: 97.5% confidence level

	Exceptions		VaR						ES			
BLOCK SIZE	**Nº**	**%**	**LRuc**	**LRind**	**LRcc**	**BTC**	**DQ**	**MF**	**RC**	**NZ**	**BD**
10	531	23.46%	0.0000	0.0000	0.0000	0.0000	**0.1811**	0.0000	**0.3285**	0.0000	0.0000
21	171	7.56%	0.0000	**0.6221**	0.0000	0.0000	0.0454	0.0000	**0.1243**	0.0000	0.0011
31	75	3.31%	**0.1191**	**0.5456**	**0.2473**	0.0065	0.0151	0.0020	**0.0737**	0.0000	0.0004
42	54	2.39%	**0.8179**	**0.3820**	**0.6645**	**0.6355**	0.0024	0.0013	**0.0591**	0.0000	0.0000
63	34	1.50%	0.0306	**0.2865**	**0.0548**	**0.9988**	0.0080	**0.0705**	**0.0305**	0.0140	0.0004
126	14	0.62%	0.0000	**0.2393**	0.0000	**1.0000**	0.0016	**0.1912**	0.0132	0.0040	0.0000
189	8	0.35%	0.0000	**0.1244**	0.0000	**1.0000**	0.0000	**0.5858**	0.0074	**0.3388**	0.0000
252	4	0.18%	0.0000	**0.9374**	0.0000	**1.0000**	**0.9895**	**0.9999**	0.0039	**0.6817**	0.0000

Panel B: 99% confidence level

BLOCK SIZE	**Nº**	**%**	**LRuc**	**LRind**	**LRcc**	**BTC**	**DQ**	**MF**	**RC**	**NZ**	**BD**
10	1098	48.52%	0.0000	0.0000	0.0000	0.0000	**0.1306**	0.0000	**0.6618**	0.0000	**0.3101**
21	344	15.20%	0.0000	0.0000	0.0000	0.0000	**0.0975**	0,0000	**0.2264**	0.0000	**0.3813**
31	184	8.13%	0.0000	0.0000	0.0000	0.0000	**0.2545**	0.0000	**0.1279**	0.0000	**0.3040**
42	100	4.42%	0.0000	**0.6217**	0,0000	0,0000	**0.0190**	0.0000	**0.0960**	0.0000	0.0081
63	64	2.83%	0.0000	**0.5821**	0.0000	0.0000	0.0162	0.0039	**0.0541**	0.0000	**0.2757**
126	26	1.15%	**0.6467**	**0.4990**	**0.7164**	**0.2382**	**0.1192**	0.0042	0.0239	0.0000	0.0032
189	15	0.66%	**0.2577**	**0.2596**	**0.2792**	**0.9465**	0.0062	**0.1243**	0.0147	0.0358	0.0153
252	7	0.31%	0.0108	**0.8907**	0.0386	**0.9995**	**0.7923**	**0.3009**	0.0090	**0.2424**	0.0040

Note: The table shows the p-value for the following statistics: (i) the unconditional coverage test (LRuc); (ii) statistics for serial independence (LRind); (iii) the Conditional Coverage test (LRcc);); (iv) the backtesting criterion (BTC) and (v) the Dynamic Quantile test (DQ). For ES estimates the table shows the p-value for the following tests: McNeil and Frey (2000) (MF), Righi and Ceretta (2013) (RC), Nolde and Ziegel (2007) (NZ), and Bayer and Dimitriadis (2022) (DB). We highlight in bold the p-value in the cases where we cannot reject the null hypothesis at 5%.

Table 7. Backtesting VaR 99% and 97.5% all assets.

				Panel A: VaR 97.5%				
Indexes	**10**	**21**	**31**	**42**	**63**	**126**	**189**	**252**
CAC40	25.43%	5.17%	1.74%	0.83%	0.30%	0.04%	0.04%	0.04%
DAX-30	27.02%	4.06%	1.56%	0.73%	0.22%	0.00%	0.00%	0.00%
FTSE100	26.88%	4.19%	1.60%	0.43%	0.09%	0.00%	0.00%	0.00%
HSI	27.05%	5.87%	2.55%	1.03%	0.27%	0.09%	0.00%	0.00%
IBEX35	35.24%	6.97%	2.37%	0.90%	0.13%	0.04%	0.04%	0.04%
Merval	31.40%	6.48%	1.60%	0.18%	0.05%	0.00%	0.00%	0.00%
Nikkei	23.72%	5.63%	2.37%	0.64%	0.23%	0.05%	0.00%	0.00%
Exchange Rates								
£ /€	23,12%	7,64%	4,22%	2,43%	1,24%	0,43%	0,09%	0,00%
¥/€	28,50%	7,25%	2,30%	0,51%	0,21%	0,04%	0,00%	0,00%
$/€	26,81%	8,27%	4,30%	2,08%	0,87%	0,34%	0,34%	0,34%
Commodities								
Brent	30.48%	7.06%	2.94%	1.27%	0.45%	0.09%	0.14%	0.05%
Cooper	33.18%	9.05%	3.70%	1.56%	0.67%	0.00%	0.00%	0.00%
Gold	38.05%	9.55%	4.68%	2.59%	1.34%	0.31%	0.31%	0.45%
Silver	40.64%	9.44%	4.70%	2.44%	0.96%	0.30%	0.17%	0.09%
				Panel B: VaR 99%				
Indexes	**10**	**21**	**31**	**42**	**63**	**126**	**189**	**252**
CAC40	47.57%	14.17%	6.00%	2.74%	0.91%	0.22%	0.04%	0.04%
DAX-30	50.02%	11.59%	6.01%	3.29%	0.86%	0.04%	0.00%	0.00%
FTSE100	50.00%	12.10%	5.75%	2.46%	0.43%	0.04%	0.00%	0.00%
HSI	52.13%	15.54%	7.61%	3.72%	1.21%	0.13%	0.04%	0.00%
IBEX35	62.52%	20.22%	9.98%	3.87%	0.90%	0.09%	0.04%	0.04%
Merval	60.79%	19.99%	8.44%	2.10%	0.37%	0.05%	0.00%	0.00%
Nikkei	46.44%	15.66%	6.84%	3.47%	0.90%	0.14%	0.05%	0.00%
Exchange Rates								
£ /€	39.42%	14.51%	8.53%	4.78%	3.24%	1.41%	0.68%	0.34%
¥/€	65.78%	19.50%	8.11%	3.46%	0.90%	0.09%	0.04%	0.04%
$/€	52.39%	15.27%	9.60%	5.46%	2.47%	0.43%	0.26%	0.21%
Commodities								
Brent	59.11%	16.28%	7.91%	4.34%	1.36%	0.23%	0.14%	0.09%
Cooper	65.21%	26.90%	14.27%	6.20%	2.45%	0.13%	0.04%	0.00%
Gold	70.87%	24.49%	11.91%	7.14%	2.85%	0.89%	0.49%	0.54%
Silver	79.33%	27.33%	13.27%	7.40%	3.35%	0.91%	0.74%	0.35%

Note: We shade in light grey the cases where we cannot reject the null hypothesis for at least 4 out of 5 tests at 5%. The number indicates the percentage of exceptions.

Table 8. Backtesting ES 99% and 97.5% all assets.

	ES 99%								ES 97,5%							
Block size	10	21	31	42	63	126	189	252	10	21	31	42	63	126	189	252
Indexes																
CAC40	1	1	1	1	0	2	1	1	1	1	0	0	1	1	1	1
DAX30	1	1	0	0	0	1	1	1	1	1	0	0	2	0	0	0
FTSE100	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	0
HSI	0	0	0	0	0	2	1	1	0	0	0	0	1	1	0	0
Nikkei	0	0	0	0	0	2	1	1	1	0	0	0	1	1	0	1
Merval	0	0	0	0	1	1	1	1	0	0	0	1	1	0	0	0
IBEX35	2	2	1	0	1	1	1	1	1	1	0	0	2	1	1	1
Exchange rates																
GBP	3	1	1	1	1	3	2	0	1	1	3	4	3	2	0	0
JPY	1	3	1	2	2	3	1	0	1	1	1	3	3	1	0	0
USD	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1
Commodities																
Brent	1	1	0	0	0	1	2	1	1	0	0	0	0	1	2	1
Cobre	1	1	3	0	0	2	1	1	0	1	0	0	0	0	0	0
Gold	1	1	3	2	0	3	2	3	1	1	0	3	2	2	2	2
Silver	0	0	0	2	0	3	1	2	1	0	0	0	1	1	2	1

Note: The table shows the number of tests where we cannot reject the null hypothesis at 5%. The tests are McNeil and Frey (2000), Righi and Ceretta (2013), Nolde and Ziegel (2007), and Bayer and Dimitriadis (2022).

Table 9. Percentage of exceptions VaR estimates BMM

VaR estimates at 2.5% probability								
Assets/Block size	**10**	**21**	**31**	**42**	**63**	**126**	**189**	**252**
Brent	30,48%	7.06%	**2.94%**	1.27%	0,45%	0,09%	0,14%	0,05%
Cooper	33,18%	9,05%	3,70%	1,56%	0,67%	0,00%	0,00%	0,00%
Gold	38,05%	9,55%	4,68%	**2,59%**	1,34%	0,31%	0,31%	0,45%
Silver	40,64%	9,44%	4,70%	**2,44%**	0,96%	0,30%	0,17%	0,09%
S&P500	23,46%	7,56%	3,31%	**2,39%**	1,50%	0,62%	0,35%	0,18%
CAC40	25,43%	5,17%	1,74%	0,83%	0,30%	0,04%	0,04%	0,04%
DAX30	27,02%	4,06%	1,56%	0,73%	0,22%	0,00%	0,00%	0,00%
FTSE100	26,88%	4,19%	1,60%	0,43%	0,09%	0,00%	0,00%	0,00%
HSI	27,05%	5,87%	**2,55%**	1,03%	0,27%	0,09%	0,00%	0,00%
IBEX35	35,24%	6,97%	**2,37%**	0,90%	0,13%	0,04%	0,04%	0,04%
Merval	31,40%	6,48%	1,60%	0,18%	0,05%	0,00%	0,00%	0,00%
NIKKEI	23,72%	5,63%	**2,34%**	0,63%	0,23%	0,05%	0,00%	0,00%
GBP	23,12%	7,64%	4,22%	**2,43%**	1,24%	0,43%	0,09%	0,00%
JPY	28,50%	7,25%	**2,30%**	0,51%	0,21%	0,04%	0,00%	0,00%
USD	26,81%	8,27%	4,30%	**2,08%**	0,87%	0,34%	0,34%	0,34%
VaR estimates at 1% probability								
Assets/Threshold	**10**	**21**	**31**	**42**	**63**	**126**	**189**	**252**
Brent	59,11%	16,28%	7,91%	4,34%	**1,36%**	0,23%	0,14%	0,09%
Cooper	65,21%	26,90%	14,27%	6,20%	2,45%	0,13%	0,04%	0,00%
Gold	70,87%	24,49%	11,91%	7,14%	2,85%	**0,89%**	0,49%	0,54%
Silver	79,20%	27,28%	13,27%	7,40%	3,31%	**0,91%**	0,74%	0,35%
S&P500	48,52%	15,20%	8,13%	4,42%	2,83%	**1,15%**	0,66%	0,31%
CAC40	47,57%	14,17%	6,00%	2,74%	**0,91%**	0,22%	0,04%	0,04%
DAX30	50,02%	11,59%	6,01%	3,29%	**0,86%**	0,04%	0,00%	0,00%
FTSE100	50,00%	12,10%	5,75%	2,46%	0,43%	0,04%	0,00%	0,00%
HSI	52,13%	15,54%	7,61%	3,72%	**1,21%**	0,13%	0,04%	0,00%
IBEX35	62,52%	20,22%	9,98%	3,87%	**0,90%**	0,09%	0,04%	0,04%
Merval	60,79%	19,99%	8,44%	2,10%	0,37%	0,05%	0,00%	0,00%
NIKKEI	46,44%	15,66%	6,84%	3,47%	**0,90%**	0,14%	0,05%	0,00%
GBP	39,42%	14,51%	8,53%	4,78%	3,24%	**1,41%**	0,68%	0,34%
JPY	65,78%	19,50%	8,11%	3,46%	**0,90%**	0,09%	0,04%	0,04%
USD	52,39%	15,27%	9,60%	5,46%	2,47%	0,43%	0,26%	0,21%

Note: The Table presents the percentage of exceptions for a 2.5% level of probability. An exception occurs when $r_{t+1} < VaR_{\alpha}$. Cases in which the percentage of exceptions is close to the theoretical percentage in BMM are marked in bold.

Table 10. Percentage of exceptions VaR estimates POT

VaR estimates at 2.5% probability								
Assets/Block size	**87**	**88**	**89**	**90**	**91**	**92**	**93**	**94**
Brent	2,53%	2,49%	2,49%	2,49%	2,58%	2,58%	2,58%	2,58%
Cooper	2,45%	2,45%	2,50%	2,45%	2,45%	2,50%	2,50%	2,50%
Gold	2,85%	2,85%	2,85%	2,90%	2,90%	2,90%	2,90%	2,90%
Silver	2,65%	2,65%	2,61%	2,61%	2,57%	2,61%	2,65%	2,61%
S&P500	2,61%	2,61%	2,65%	2,61%	2,61%	2,61%	2,61%	2,70%
CAC40	1,17%	1,17%	1,17%	1,26%	1,30%	1,39%	1,35%	1,39%
DAX30	2,55%	2,59%	2,68%	2,68%	2,72%	2,77%	2,77%	2,81%
FTSE100	2,64%	2,68%	2,72%	2,72%	2,72%	2,81%	2,81%	2,81%
HSI	2,10%	2,24%	2,19%	2,24%	2,24%	2,24%	2,24%	2,24%
IBEX35	2,02%	2,07%	2,07%	2,07%	2,07%	2,02%	2,11%	2,11%
Merval	1,19%	1,19%	1,19%	1,19%	1,19%	1,19%	1,19%	1,19%
NIKKEI	2,12%	2,12%	2,16%	2,16%	2,21%	2,21%	2,21%	2,21%
GBP	2,69%	2,73%	2,73%	2,82%	3,03%	3,07%	3,07%	3,07%
JPY	3,33%	3,28%	3,33%	3,41%	3,41%	3,54%	3,67%	3,71%
USD	3,03%	3,03%	3,07%	3,07%	3,11%	3,24%	3,24%	3,28%
VaR estimates at 1% probability								
Assets/Threshold	**87**	**88**	**89**	**90**	**91**	**92**	**93**	**94**
Brent	1,40%	1,45%	1,45%	1,45%	1,45%	1,49%	1,58%	1,58%
Cooper	0,67%	0,67%	0,67%	0,67%	0,67%	0,71%	0,71%	0,71%
Gold	1,07%	1,12%	1,12%	1,12%	1,12%	1,12%	1,12%	1,12%
Silver	0,74%	0,74%	0,74%	0,74%	0,74%	0,74%	0,74%	0,74%
S&P500	1,15%	1,15%	1,15%	1,15%	1,15%	1,15%	1,15%	1,19%
CAC40	0,39%	0,39%	0,39%	0,39%	0,43%	0,43%	0,43%	0,43%
DAX30	1,04%	1,08%	1,08%	1,12%	1,12%	1,12%	1,12%	1,21%
FTSE100	1,30%	1,34%	1,43%	1,38%	1,43%	1,47%	1,47%	1,47%
HSI	0,45%	0,45%	0,45%	0,49%	0,45%	0,49%	0,49%	0,49%
IBEX35	0,65%	0,65%	0,65%	0,65%	0,73%	0,73%	0,73%	0,73%
Merval	0,18%	0,18%	0,18%	0,18%	0,18%	0,18%	0,18%	0,18%
NIKKEI	0,90%	0,90%	0,90%	0,90%	0,99%	1,04%	1,04%	0,99%
GBP	1,49%	1,49%	1,53%	1,53%	1,53%	1,58%	1,58%	1,62%
JPY	1,31%	1,31%	1,31%	1,31%	1,40%	1,40%	1,44%	1,40%
USD	1,08%	1,08%	1,08%	1,13%	1,13%	1,17%	1,22%	1,31%

Note: The Table presents the percentage of exceptions for a 1% level of probability. An exception occurs when $r_{t+1} < VaR_{\alpha}$.

Chapter V. Conclusions.

This dissertation aimed to address three areas within the Extreme Value Theory (EVT) framework: (i) evaluating the performance of the volatility model estimation under distribution with fat tails and skewness in improving VaR estimation results in the framework of the Conditional EVT, (ii) investigating the sensitivity of GPD quantiles and market risk measures (Value at Risk and Expected Shortfall) to threshold selection under the POT method, and (iii) analyzing the sensitivity of GEVD parameters and market risk quantification to the choice of different block sizes with BMM.

The present Chapter aims to conclude with the main findings and contributions of each section.

The study conducted in Chapter II of this Thesis revealed some evidence that the use of a heavy-tailed and skewed distribution yields better results in terms of VaR estimation accuracy, firm's loss function, and capital requirement compared to a symmetric distribution.

Nevertheless, the findings align with recent research suggesting that VaR estimation can be improved by assuming fat tail and skewed distributions, or, in other words, assuming a normal distribution for VaR estimation may underestimate risk, as financial return distribution is skewed and exhibits excess kurtosis.

Chapter III examines the relevance of threshold choice in measuring market risk using the conditional EVT and the Generalized Pareto distribution. Results show that parameter estimates are sensitive to the selected threshold, but GPD quantiles do not change much, especially for high quantiles (95^{th}, 96^{th}, 97^{th}, 98^{th} and 99^{th}), which are relevant in risk estimation. VaR and ES estimations are practically equivalent for a large set of thresholds, and there is not one optimal threshold, but rather a set of optimal thresholds that provide similar market risk measures.

What stands out the most is that the findings indicate that in estimating market risk, researchers and practitioners do not need to prioritize the selection of a particular threshold, as a wide range of options produces comparable risk estimations. The study has been extended to a set of 14 assets from alternative markets: 7 stock indexes (CAC40, DAX30, FTSE100, HangSeng, IBEX35, Merval and Nikkei), four commodities (Copper, Gold, Crude Oil Brent and Silver) and three rates exchange (£/€, $/€ and ¥/€), and it corroborates the results obtained for the S&P500 stock index.

However, while the quantification of risk is not primarily dependent on threshold selection, it is worth mentioning that for certain thresholds, small differences are found for the highest percentiles. This could be a valuable resource for financial institutions that may be interested in choosing the threshold that minimizes the capital requirements for market risk.

As an additional analysis, we calculate the minimum capital requirements for market risk based on the ES (99%) estimations. The results reveal that there is a set of thresholds that provides the same results, finding only minor differences for the highest percentiles.

Although this result opens up an opportunity for further investigation, this Chapter represents a novelty and thus, potentially the main contribution of this Thesis in the use of the POT method in the area of market risk measures.

Finally, Chapter IV focuses on the use of EVT and specifically the Block Maxima Method for estimating market risk. The study finds that the BMM does not provide accurate market risk estimates as is highly dependent on the block size selected.

We detect that the VaR estimations are highly sensitive to the block size selected for fitting GEV distribution. Both the smallest and the largest block sizes lead to inaccurate estimations of market risk. Only intermediate block sizes, among the selected in this study, seem to provide reasonable VaR estimations, although we must be cautious as the results are not robust, in the sense that no one block size performs well for the whole set of assets considered. In the case of Expected Shortfall, we found a strong risk overestimation and any block size provides appropriate risk estimations.

Therefore, the block size may be a critical aspect that should not be chosen arbitrarily. Consequently, and in line with the literature reviewed, BMM may not be the most reliable method among the EVT approaches for estimating market risk. The use of this method could require further research on optimal block size selection techniques. Also, we consider it could be worthwhile to explore deeper into this topic considering the stationarity of data under the BM method.

Chapter V.Conclusions.

We end Chapter IV with a comparison of VaR estimations between the BM and POT methods and conclude, as expected, that the POT method provides more consistent results across different thresholds, while the BM method is highly dependent on the selected block size to achieve a level of exceptions close to the theoretical one.

Based on the findings of our analysis, we can conclude that the POT method represents a robust and effective approach for estimating extreme events and measuring financial risk. This method allows for the accurate estimation of tail probabilities and can provide a valuable understanding of extreme events that other methods as BM may miss.

APPENDIX

Appendix A: Algorithm of VaR estimation using Block Maxima Method

We divide the period into m sub-periods with n observations on each period (i.e the block size).

Let α^* be a small upper tail probability showing a potential loss, and q_α^* be the $(1-\alpha^*)^{th}$ quantile of the sub-period maxima under the limiting GEV distribution, then

$$(1-\alpha^*) = \begin{cases} exp\left\{-\left[1+\frac{\xi_n(q_\alpha^*-\mu_n)}{\sigma_n}\right]^{-\frac{1}{\xi_n}}\right\}, for\ \xi \neq 0 \\ exp[-\exp\,(-\frac{q_\alpha^*-\mu_n}{\sigma_n}), for\ \xi = 0 \end{cases} \quad \text{(A1)}$$

Where $1+\frac{\xi_n(q_\alpha^*-\mu_n)}{\sigma_n} > 0\ if\ \xi \neq 0$. By logarithmic transformation, we have:

$$ln(1-\alpha^*) = \begin{cases} -\left[1+\frac{\xi_n(q_\alpha^*-\mu_n)}{\sigma_n}\right]^{-\frac{1}{\xi_n}}, for\ \xi \neq 0 \\ -exp(-\frac{q_\alpha^*-\mu_n}{\sigma_n}), for\ \xi = 0 \end{cases} \quad \text{(A2)}$$

And the quantile is

$$q_\alpha^* = \begin{cases} \mu_n + \frac{\sigma_n}{\xi_n}\left[1-\{-ln(1-\alpha^*)\}^{-\xi_n}\right], for\ \xi \neq 0 \\ \mu_n + \sigma_n\{-ln(1-\alpha^*)\} \qquad , for\ \xi = 0 \end{cases} \quad \text{(A3)}$$

The quantile q_n^* for a given probability α^* is the VaR for the sub-period maximum. Knowing that most asset returns have either weak serial correlations or no correlations at all, the relation between sub-period maxima and the observed return series r_i is as follows

$$1-\alpha^* = [1-P(r_i \leq q_\alpha^*)]^n \quad \text{(A4)}$$

Let $\alpha = P(r_i \leq q_n^*)$, thus $1-\alpha^* = (1-\alpha)^n$

From (A3) and (A4), we have

$$q_\alpha^* = \begin{cases} \mu_n + \frac{\sigma_n}{\xi_n}\left[1-\{-nln(1-\alpha)\}^{-\xi_n}\right], for\ \xi \neq 0 \\ \mu_n + \sigma_n\{-nln(1-\alpha)\} \qquad , for\ \xi = 0 \end{cases} \quad \text{(A5)}$$

Using maximum likelihood estimation, we can estimate the parameters μ, σ and ξ in (A5) under *i.i.d.* By the definition of VaR and (A5), we have

$$q_\alpha = \begin{cases} \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \{-nln(1-\alpha)\}^{-\hat{\xi}}\right], for\ \hat{\xi} \neq 0 \\ \hat{\mu} + \hat{\sigma}\{-nln(1-\alpha)\} \qquad , for\ \hat{\xi} = 0 \end{cases} \tag{A6}$$

From (A5), we know that VaR is related to the period length *n*.

The estimation of the dynamic VaR will be calculated as:

$$VaR_t(\alpha) = \mu_t + \sigma_t\, q_\alpha \tag{A7}$$

where μ_t and σ_t represent the conditional mean and the forecasted volatility of the returns respectively and q_α is the quantile function of the GEVD.

Appendix B: Algorithm of Expected Shortfall using Block Maxima Method

Definition 1. For a random variable x with continuous distribution function, Expected Shortfall or conditional Value at Risk equals the conditional expectation of x given that $x < VaR_\alpha$ (x), that is,

$$ES_\alpha(\alpha) = \mathbb{E}[X|\ X < VaR_\alpha] \tag{A8}$$

To obtain a dynamic ES, following the Equation (A7), then

$$ES_\alpha(\alpha) = \mu_t + \sigma_t \mathbb{E}[z_t|\ z_t < q_\alpha\ ] \tag{A9}$$

Definition 2. According to Nadarajah & Kotz (2008), the moments of a truncated distribution with upper bound B can be obtained via the integral

$$\mathbb{E}[x|\ x < B\ ] = \frac{1}{F(B)}\int_{-\infty}^{B} xf(x)dx \tag{A10}$$

where *F* is the cumulative distribution function (CDF), f is the density function of x (in our case *x* is the z_t, i.e.the standardized maximum returns) and *B* would be the quantile function q_α. Thus,

$$\mathbb{E}[\boldsymbol{z_t}|\ \boldsymbol{z_t} < \ \boldsymbol{q_n}] = -\frac{\boldsymbol{1}}{\boldsymbol{\alpha}}\int_{-\infty}^{-q_\alpha} \boldsymbol{z_t f(z_t)dz} \tag{A11}$$

Where $f(z_t)$ is the density function of standardized maximum returns.

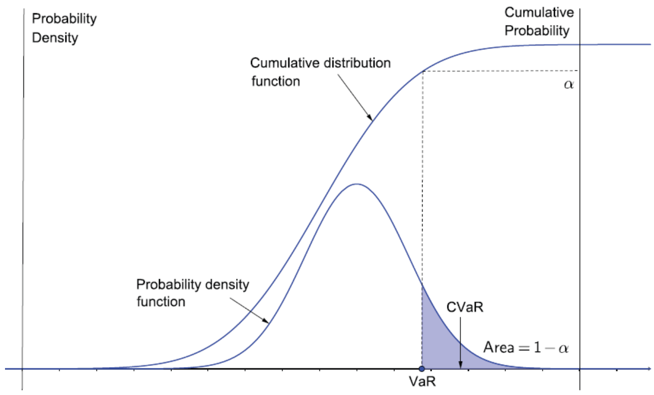


Illustration of the concepts of VaR and CvaR (right tail) from *C. Filippi et al. / Intl. Trans. in Op. Res. 27 (2020) 1277–1319.*

Definition 3. CDF and pdf of standardized returns.

Let y be the standardized returns, $y = \frac{r-\mu}{\sigma}$ we have the cumulative distribution function, $F_{n,t}(y) = 1 - [1 - F(z_t)]^{\frac{1}{n}}$ and density function

$$f_{n,t}(y) = \frac{1}{n}[1 - F(z_t)]^{\frac{1}{n}-1} f_{n,t}(z_t) \tag{A12}$$

Where $F_{n,t}(y)$ is the distribution function of standardized maximum returns $r_{n,t}$, $z_t = \frac{r_{n,t}-\mu_n}{\sigma_n}$ presented by Jenkinson (1955) for the *i.i.d* case,

$$F(z_t) = \begin{cases} 1 - exp\left[-(1+\xi z_t)^{1/\xi}\right], for\ \xi \neq 0 \\ 1 - exp[-exp(z_t)]\ , \qquad ,for\ \xi = 0 \end{cases} \tag{A13}$$

From (A13) we can obtain the derivative to get the probability density function of the standardized returns,

$$f_{n,t}(z_t) = \begin{cases} (1+\xi z_t)^{\frac{1}{\xi}-1} exp\left[-(1+\xi z_t)^{1/\xi}\right], for\ \xi \neq 0 \\ exp[z_t - \exp(z_t)]\ , \qquad ,for\ \xi = 0 \end{cases} \tag{A14}$$

Substituting expressions (A13) and (A14) into (A12) we have

$$f(z_t) = \frac{1}{n} exp\left[-(1+\xi z_t)^{\frac{1}{\xi}} \left(\frac{1}{n} - 1\right)\right] (1+\xi z_t)^{\frac{1}{\xi}-1} exp\left[-(1+\xi z_t)^{1/\xi}\right] \tag{A15}$$

Substituting (A15) in (A11) we have

$$\mathbb{E}[z_t |\ z_t < q_n] = -\frac{1}{n\alpha} \int_{-\infty}^{-q_n} z_t (1+\xi z_t)^{\frac{1}{\xi}-1} exp\left[-(\frac{1}{n}(1+\xi z_t)^{1/\xi}\right] dz \tag{A16}$$

By Corollary of Definition 1 and substituting (A16) in (A9) we have

$$ES_{\alpha}(\alpha) = -\frac{\sigma_t}{n\alpha}\int_{-\infty}^{-q_n} z_t(1+\xi z_t)^{\frac{1}{\xi}-1} exp\left[-\frac{1}{n}(1+\xi z_t)^{1/\xi}\right] dz - \mu_t \tag{A17}$$

Appendix C: Algorithm of VaR estimation using Peaks over Threshold

Definition 4. CDF of excesses.

Given a variable X of randomly distributed observations X_1, ..., X_n we define an exceedance as Xi > u where u is the threshold.

The distribution function F_u also called the conditional excess distribution function is defined as

$$F_u(y) = P(X - u \leq y|\ X > u); \tag{A18}$$

Where $y = x - u$ are the excesses.

The values taken by the random variable X are mostly between 0 and u, hence estimating F within this range usually presents no challenges. However, estimating the $F_u(y)$ can be problematic, as there are typically very few observations in this area. In such cases, the EVT can be highly useful, as it offers a robust solution for the conditional excess distribution function, as described in the following theorem.

$$For\ x \geq u$$

$$F(x) = P\{X \leq x\} = P\{u \leq X \leq x\} + P\{X \leq u\} =$$
$$= (1 - P\{X \leq u\}F_u(y) + P\{X \leq u\} \tag{A19}$$

Theorem 1. (Pickands (1975), Balkema and de Haan (1974)) For a large class of underlying distribution functions F the conditional excess distribution function Fu(y), for u large, is well approximated by:

$$F_u(y) \approx G_{\sigma\xi}, \qquad u \to \infty$$

$$\text{Where, } G_{\sigma\xi}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma}y\right)^{-\frac{1}{\xi}}, for\ \xi \neq 0 \\ 1 - e^{-y/\sigma}, \qquad , for\ \xi = 0 \end{cases} \tag{A20}$$

$G_{\sigma\xi}(y)$ is the so-called generalized Pareto distribution (GPD), ξ is the shape parameter and σ is the scale.

If x is defined as $x = u + y$, the GPD can also be expressed as a function of x

$$G_{\sigma\xi}(x) = 1 - (1 + \frac{\xi(x-u)}{\sigma})^{-\frac{1}{\xi}} \quad \text{(A21)}$$

From (A19) we have $F(x) = [1 - F(u)]G_{\sigma\xi} + F(u)$ (A22)

And then, $F_u(y) = \frac{F(x)-F(u)}{1-F(u)}$ (A23)

Assuming a GPD function for the tail distribution, we can derive mathematical formulas for the Value at Risk (VaR) and Expected Shortfall by expressing them as functions of the GPD parameters.

Replacing in (A22) $G_{\sigma\xi}$ by the GPD and $F(u)$ by the estimate $(n - Nu)/n$, where n is the total number of observations and Nu is the number of observations above the threshold u, we get:

$$\hat{F}(x) = 1 - \frac{Nu}{n} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-\frac{1}{\xi}} \quad \text{(A24)}$$

Inverting (A24) for a given probability α gives,

$$\widehat{VaR}(\alpha) = u + \frac{\hat{\xi}}{\hat{\sigma}}\left(\left(\frac{n}{Nu}\alpha\right)^{-\hat{\xi}} - 1\right) \quad \text{(A25)}$$

Appendix D: Algorithm of ES estimation using Peaks over Threshold

From the definition of Expected Shortfall, $ES\alpha = E(X \mid X > VaR\alpha)$ (A26)

We get, $\widehat{ES}\alpha = \widehat{VaR}(\alpha) + E(X - \widehat{VaR}(\alpha) \mid X > \widehat{VaR}(\alpha))$ (A27)

where the second term on the right is the expected value of the exceedances over $VaR(\alpha)$.

It is known that the *mean excess function* for the GPD with parameter $\xi < 1$ is:

$$e(z) = E(X - z \mid X > z) = \frac{\sigma + \xi z}{1-\xi}, \sigma + \xi z > 0 \quad \text{(A28)}$$

Similarly, given the definition (A26) and using expression (A27), for $z = VaR(\alpha) - u$ and X representing the excesses y over u we get

$$\widehat{ES}\alpha = \widehat{VaR}(\alpha) + \frac{\hat{\sigma} + \hat{\xi}(\widehat{VaR}(\alpha) - u)}{1-\hat{\xi}} = \frac{\widehat{VaR}(\alpha)}{1-\hat{\xi}} + \frac{\hat{\sigma} + \hat{\xi} u}{1-\hat{\xi}} \quad \text{(A29)}$$

BIBLIOGRAPHY

[1] Abad P., Benito S. and López C. (2015). Role of the loss function in the VaR comparison. *Journal of Risk Model Validation*, Vol. 9, nº 1, pp. 1-19. DOI: 10.21314/JRMV.2015.132

[2] Abad P., Benito S. and López-Martín, C. (2014). A comprehensive review of value at risk methodologies. *The Spanish Review of Financial Economic*, 12, 15-32. DOI: 10.1016/j.srfe.2013.06.001

[3] Abad, P. and Benito, S. (2013). A detailed comparison of value at risk estimates. *Mathematics and Computers in Simulation 94*, 258-276. DOI: 10.1016/j.matcom.2012.05.011

[4] Acerbi C. and Taasche D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7), 1487-1503. DOI: 10.1016/S0378-4266(02)00283-2

[5] Alberga, D., Shalita H. and Yosefb R (2008). Estimating stock market volatility using asymmetric GARCH models. Applied Financial Economics, 2008, 18, 1201–1208. DOI: 10.1080/09603100701604225

[6] Allen, D. E., Singh, A. K., & Powell, R. J. (2011). Extreme market risk-an extreme value theory approach. https://ro.ecu.edu.au/ecuworks2011/720/

[7] Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2013). Financial risk measurement for financial risk management. In Handbook of the Economics of Finance (Vol. 2, pp. 1127-1220). *Elsevier*. DOI: 10.1016/B978-0-44-459406-8.00017-2

[8] Artzner P., Delbaen F., Eber J.M. and Heath D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203-28. DOI: 10.1111/1467-9965.00068

[9] Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1997). A characterization of measures of risk. Cornell University Operations Research and Industrial Engineering. https://hdl.handle.net/1813/9070

[10] Ausín M. and Galeano P. (2007). Bayesian Estimation of the Gaussian Mixture GARCH Model. *Computational Statistics & Data Analysis*. Vol. 51, nº 5, pp. 2636-2652. DOI: 10.1016/j.csda.2006.01.006

[11] Bader, B., Yan, J., & Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics*, 12(1), 310-329. DOI: 10.1214/17-AOAS1092

[12] Bali T. and Theodossiou P. (2008). Risk Measurement Performance of Alternative Distribution Functions. *The Journal of Risk and Insurance*. Vol. 75, nº 2, pp. 411-437. DOI: 10.1111/j.1539-6975.2008.00266.x

[13] Balkema A., and De Haan L. (1974). Residual life time at great age. *The Annals of Probability, 2*(5), 792-804. DOI: 10.1214/aop/1176996548

[14] Bank for International Settlements (2014). BIS Annual Economic Report. 84th Annual Report, 2013/14 https://www.bis.org/publ/arpdf/ar2014e.pdf

[15] Barendse, S., Kole, E., & van Dijk, D. J. (2019). Backtesting value-at-risk and expected shortfall in the presence of estimation error. DOI: 10.2139/ssrn.3439309

[16] Basel Committee on Banking Supervision (1996). Supervisory framework for the use of "Backtesting" in conjunction with the internal models approach to market risk capital requirements. https://www.bis.org/publ/bcbs22.pdf

[17] Basel Committee on Banking Supervision (2011a). Interpretive issues with respect to the revisions to the market risk framework. http://www.bis.org/publ/bcbs193a.pdf

[18] Basel Committee on Banking Supervision (2011b). Messages from the academic literature on risk measurement for the trading book. http://www.bis.org/publ/bcbs_wp19.pdf

[19] Basel Committee on Banking Supervision (2012). Consultative Document: Fundamental review of the trading book. http://www.bis.org/publ/bcbs219.pdf

[20] Basel Committee on Banking Supervision (2013). Fundamental review of the trading book: A revised market risk framework. Basel, Switzerland: Bank for International Settlements (BIS). Available at http://www.bis.org/bcbs/publ/d305.htm

[21] Basel Committee on Banking Supervision (2014). Fundamental review of the trading book: outstanding issues. Technical report, Basel Committee on Banking Supervision.

[22] Basel Committee on Banking Supervision (2017). High-level summary of Basel III reforms. Basel, Switzerland: Bank for International Settlements (BIS). Available at: https://www.bis.org/bcbs/publ/d424_hlsummary.pdf

[23] Bayer, S., & Dimitriadis, T. (2022). Regression-based expected shortfall backtesting. *Journal of Financial Econometrics*, *20*(3), 437-471. DOI: 10.1093/jjfinec/nbaa013

[24] Bayer, S., and Dimitriadis T. (2020) 'esback': Expected Shortfall Backtesting. R package version 0.3.0. Available at https://cran.r-project.org/web/packages/esback/esback.pdf

[25] Beirlant J, Goegebeur Y, Segers J. and Teugels, J. (2004). Statistics of extremes: Theory and applications, Wiley, London. ISBN: 978-0-471-97647-9.

[26] Beirlant, J., Kijko, A., Reynkens, T., Einmahl, J.H.J.(2018) Estimating the maximum possible earthquake magnitude using extreme value methodology: the Groningen case. *Natural Hazards* 169, 1–23. DOI: 10.1007/s11069-017-3162-2

[27] Bekiros and Georgoutsos (2005). Estimation of value-at-risk by extreme value and conventional methods: A comparative evaluation of their predictive performance. *Journal of International Financial Markets Institutions and Money* 15(3):209-228. DOI: 10.1016/j.intfin.2004.05.002

[28] Benito Muela, S., López-Martín, C. and Navarro Cervantes, M. Á. (2017). The role of the skewed distributions in the framework of extreme value theory (EVT). *International Business Research*, 10(11), 88-102. DOI: 10.5539/ibr.v10n11p88

[29] Benito Muela, S., López-Martín, C. and Navarro Cervantes, M. Á. (2023). Assessing the importance of the choice threshold in quantifying market risk under the POT approach (EVT). *Risk Management* 25, 6. DOI: 10.1057/s41283-022-00106-w

[30] Benito S. (2013). Evaluating an EGARCH Model with Fat Tails, Skewness and Leverage in Forecasting VaR. Journal of Contemporary Management. Article ID: 1929-0128-2015-03-67-14.

[31] Bera, A.K and Higgins, M.L (1993). Arch Models: properties, estimation and Testing. Journal of Economic Surveys, Vol.7, No.4, pp.305-62. DOI: 10.1111/j.1467-6419.1993.tb00170.x

[32] Black F. (1976). Studies in stock price volatility changes. Proceedings of the *1976 Business Meeting of the Business and Economics Statistics Section*, American Association: 177-181.

[33] Bollerslev T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, 21, pp. 307-327. DOI: 10.1016/0304-4076(86)90063-1

[34] Bollerslev T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *Review of Economics and Statistics*, 69, pp. 542-547. DOI: 10.2307/1925546

[35] Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. *Handbook of econometrics*, *4*, 2959-3038. DOI: 10.1016/S1573-4412(05)80018-2

[36] Bommier, E. (2014). Peaks-over-threshold modelling of environmental data. Available at: http://uu.diva-portal.org/smash/get/diva2:760802/FULLTEXT01.pdf

[37] Boshnakov, G.N. (2019). Compute Expected Shortfall and Value at Risk for Continuous Distributions. ‘cvar’ R Package. Available at https://CRAN.R-project.org/package=cvar

[38] Brie, L., Genest, B., & Arsac, M. (2018). Back-Testing of Expected Shortfall: Main Challenges and Methodologies. DOI: 10.2139/ssrn.3174711

[39] Brooks, C., Clare, A. D., Dalle Molle, J.W. and Persand, G. (2005). A comparison of extreme Value theory approaches for determining value at risk. *Journal of Empirical Finance, 12*(2), 339-352. DOI: 10.1016/j.jempfin.2004.01.004

[40] Bystörm Hans N. E. (2004). Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. *International Review of Financial Analysis*, 133-152. DOI: 10.1016/j.irfa.2004.02.003

[41] Caballero-Megido, C., Hillier, J., Wyncoll, D., Bosher, L., & Gouldby, B. (2018). Comparison of methods for threshold selection for extreme sea levels. *Journal of Flood Risk Management*, 11(2), 127-140. DOI: 10.1111/jfr3.12296

[42] Caires, S. (2009). A comparative simulation study of the annual maxima and the peaks-over-threshold methods. *Deltares report* 1200264-002 for Rijkswaterstaat, Waterdienst.http://resolver.tudelft.nl/uuid:143b0f1e-f61e-44ab-8da3 9241970d915b

[43] Campbell, S. D. (2005). A review of backtesting and backtesting procedures. https://www.federalreserve.gov/econres/feds/a-review-of-backtesting-and-backtesting-procedures.htm

[44] Carreau, J., Naveau, P., Neppel, L. (2017). Partitioning into hazard subregions for regional peaks-over threshold modeling of heavy precipitation. *Water Resources Research*, American Geophysical Union, 53 (5), pp.4407 - 4426. ff10.1002/2017WR020758ff. ffhal-01874218

[45] Carvalhal, A., & Mendes, B. V. (2003). Value-at-risk and extreme returns in Asian stock markets. *International Journal of Business*, 8(1). DOI: 10.2139/ssrn.420266

[46] Cebrián, A. C., and Abaurrea, J. (2006). Drought analysis based on a marked cluster Poisson model, *Journal of Hydrometeorology*, 7, 713–723. https://www.jstor.org/stable/24910906

[47] Cebrián, A. C., Denuit, M. and Lambert, P. (2003). Generalized Pareto fit to the society of actuaries large claims database, *North America Actuarial Journal*, 7(3), 18–36. DOI: 10.1080/10920277.2003.10596098

[48] Chan, Stephen & Nadarajah, Saralees & Afuecheta, Emmanuel. (2015). An R Package for Value at Risk and Expected Shortfall. Communications in Statistics - Simulation and Computation. 45. 00-00. DOI: 10.1080/03610918.2014.944658.

[49] Chavez-Demoulin, V. and Embrechts, P. (2004). Smooth extremal models in finance and insurance. *Journal of Risk and Insurance*, 71(2), pp.183-199. DOI: 10.1111/j.0022-4367.2004.00085.x

[50] Chen C, Gerlach R, Lin E, Lee W. (2012). Bayesian forecasting for financial risk management, pre and post the global financial crisis. *Journal of Forecasting* **31**(8): 661-687. DOI: 10.1002/for.1237

[51] Chen, James Ming (2018). "On Exactitude in Financial Regulation: Value-at-Risk, Expected Shortfall, and Expectiles" *Risks* 6, no. 2: 61. DOI: 10.3390/risks6020061

[52] Cheng L., AghaKouchak, A., Gilleland, E. and Katz, R.W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic change*, 127(2), 353. DOI: 10.1007/s10584-014-1254-5

[53] Cheng, L. and AghaKouchak, A. (2014). Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. Scientific reports, 4, 7093. DOI: 10.1038/srep07093

[54] Choulakian, V., and Stephens, M.A. (2001). Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics*, 43(4), 478–484. DOI: 10.1198/00401700152672573

[55] Christoffersen, P. (1998). Evaluating interval forecasting. International Economic Review, 39, 841-862. DOI: 10.2307/2527341

[56] Christoffersen, P. (2003). Elements of financial risk management. *Academic Press.* ISBN: 9780128102350.

[57] Coles S. (2001). An introduction to statistical modeling of extreme values (Vol.208). *British Library Cataloguing in Publication Data. Springer Series in Statistics*, pp:78-84. ISBN: 978-1-4471-3675-0.

[58] Cont, R. (2005). Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models. Social Science Research Network. DOI: 10.1007/978-3-540-34625-8_10.

[59] Creal, D., Koopman, S. and Lucas, A. (2010), A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations, No 10-032/2, Tinbergen Institute Discussion Papers, Tinbergen Institute. DOI: 10.1198/jbes.2011.10070

[60] Creal, D., Koopman, S., and Lucas, A. (2013), Generalized Autoregressive Score Models with Applications. Journal of Applied Econometrics, 28, issue 5, p. 777-795. DOI: 10.1002/jae.1279

[61] Cruz, M. G. (2002). Modeling, measuring and hedging operational risk (Vol. 4). Wiley. ISBN: 978-0-471-51560-9

[62] Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, 18(3-4), 257-271. DOI: 10.1016/0022-1694(73)90051-6

[63] Danielsson, J. (2011). Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab. John Wiley & Sons. ISBN: 978-0-470-66943-3

[64] Danielsson, J., James, K. R., Valenzuela, M., & Zer, I. (2016). Model risk of risk models. *Journal of Financial Stability*, *23*, 79-91. DOI: 10.1016/j.jfs.2016.02.002

[65] Dannielsson J., de Haan L., Peng L. and de Vries C.G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis* 76, 226-248. DOI: 10.1006/jmva.2000.1903

[66] Darling, D. A. (1955). The Cramer-Smirnov test in the parametric case. The Annals of Mathematical Statistics, 1-20. Available at: https://www.jstor.org/stable/2236752

[67] Davis, M. H. (2016). Verification of internal risk measure estimates. *Statistics & risk modeling*, *33*(3-4), 67-93. DOI: 10.1515/strm-2015-0007

[68] Davison, A. and Smith, R. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society*, 52, 3, 393-442. DOI: 10.1111/j.2517-6161.1990.tb01796.x

[69] De Zea Bermudez, P., Amaral Turkman, M.A. & Turkman, K.F. (2001). A predictive approach to tail probability estimation, *Extremes*, 4, 295–314. DOI: 10.1023/A:1016546027962

[70] Diebold, F. X., Schuermann, T., & Stroughair, J. D. (1998). Pitfalls and opportunities in the use of extreme value theory in risk management. In decision technologies for computational finance (pp. 3-12). *Springer*, Boston, MA. DOI: 10.1007/978-1-4615-5625-1_1

[71] Dkengne, P. S., Girard, S., & Ahiad, S. (2020). An automatic procedure to select a block size in the continuous generalized extreme value model estimation. https://hal.inria.fr/hal-02952279

[72] Domonkos, P., and Piotrowicz, K. (1998). Winter temperature characteristics in central Europe, *International Journal of Climatology*, 18, 1405–1417. DOI: 10.1002/(SICI)1097-0088(19981115)18:13<1405::AID-JOC323>3.0.CO;2-D

[73] Dowd, K. (2007). *Measuring market risk*. John Wiley & Sons.

[74] Drees, H., de Haan, L. and Resnick, S. (2000). How to make a Hill plot. *Annals of Statistics*, 28(1), 254–274. DOI: 10.1214/aos/1016120372

[75] DuMouchel, M. (1983). Estimating the stable index α in order to measure tall thickness: A critique. *The Annals of Statistics*, 11(4), 1019-1031. DOI: 10.1214/aos/1176346318

[76] Dupuis, D. J. (1999). Exceedances over high thresholds: A guide to threshold selection, *Extremes* 1(3): 251–261. DOI: 10.1023/A:1009914915709

[77] Elliott, G., & Timmermann, A. (Eds.). (2013). *Handbook of economic forecasting*. Elsevier.

[78] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). Modelling extremal events. *British actuarial journal*, *5*(2), 465-465. DOI:10.1017/S1357321700000532

[79] Embrechts, P., Resnick, S. I., & Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. North American Actuarial Journal, 3(2), 30-41. DOI: 10.1080/10920277.1999.10595797

[80] Engeland, K., Hisdal, H., & Frigessi, A. (2004). Practical extreme value modelling of hydrological floods and droughts: a case study. Extremes, 7(1), 5-30. DOI: 10.1007/s10687-004-4727-5

[81] Engle R, Ng V. (1993). Measuring and Testing the Impact of News on Volatility. *Journal of Finance* **48**: 1749-1778. DOI: 10.1111/j.1540-6261.1993.tb05127.x

[82] Engle R. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica*, 50, pp. 987-1008. DOI: 10.2307/1912773

[83] Engle R. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, 22(4), 367-381. DOI: 10.1198/073500104000000370

[84] Engmann, S., & Cousineau, D. (2011). Comparing distributions: the two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test. *Journal of applied quantitative methods*, *6*(3). Available at: http://www.jaqm.ro/issues/volume-6,issue-3/pdfs/1_engmann_cousineau.pdf

[85] Ergün, A. T. and Jun, J. (2010). Time-varying higher-order conditional moments and forecasting intraday VaR and expected shortfall. *The Quarterly Review of Economics and Finance*, 50(3), 264-272. DOI: 10.1016/j.qref.2010.03.003

[86] Feller, W. An introduction to probability theory and its applications. 1, 2nd.

[87] Fernandez, Carmen and Steel, Mark F.J., On Bayesian Modelling of Fat Tails and Skewness. Available at SSRN: http://ssrn.com/abstract=821 or http://dx.doi.org/10.2139/ssrn.821

[88] Fernández, V. (2003). Extreme value theory and value at risk. Revista de Análisis Económico, 18(1). https://www.rae-ear.org/index.php/rae/article/view/24

[89] Ferreira A. de Haan L and Peng L. (2003). On optimising the estimation of high quantiles of a probability distribution. *Statistics: A Journal of Theoretical and Applied Statistics.* 37(5), 401-434. DOI: 10.1080/0233188021000055345

[90] Filippi, C., Guastaroba, G., & Speranza, M. G. (2020). Conditional value-at-risk beyond finance: a survey. *International Transactions in Operational Research*, *27*(3), 1277-1319. DOI: 10.1111/itor.12726

[91] Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, *44*(4), 1680-1707. DOI: 10.1214/16-AOS1439

[92] Fissler, T., Ziegel, J. F., & Gneiting, T. (2015). Expected Shortfall is jointly elicitable with Value at Risk-Implications for backtesting. DOI: 10.48550/arXiv.1507.00244

[93] Flugentiusson, H. (2012). Push it to the limit. Testing the usefulness of extreme value theory in electricity markets. http://lup.lub.lu.se/student-papers/record/3166413

[94] Fontnouvelle, P., Rosengren, E. and Jordan, J. (2007). Implications of alternative operational risk modeling techniques. NBER Chapters in: *The Risks of Financial Institutions*, pages 475-505, National Bureau of Economic Research, Inc. ISBN: 0-226-09285-2

[95] Gaines, S. D., & Denny, M. W. (1993). The largest, smallest, highest, lowest, longest, and shortest: extremes in ecology. *Ecology*, *74*(6), 1677-1692. DOI: 10.2307/1939926

[96] Gencay, R., & Selçuk, F. (2004). Extreme value theory and Value-at-Risk: Relative performance in emerging markets. International Journal of Forecasting, 20(2), 287-303. DOI: 0.1016/j.ijforecast.2003.09.005

[97] Gerstengarbe, F. W. and Werner, P.C. (1989). A method for statistical definition of extreme value regions and their application to meteorological time series. *Zeitschrift Fuer Meteorologie* 39 (4): 224–6. ISSN 0084-5361

[98] Gerstengarbe, F. W. and Werner, P.C. (1991). Some critical remarks on the use of extreme-value statistics in climatology, *Theoretical and Applied Climatology*, 44(1), 1–8. DOI: 10.1007/BF00865546

[99] Ghalanos, A. & Kley, T. (2020) Univariate GARCH Models. 'rugarch' R Package. Available at https://CRAN.R-project.org/package=rugarch

[100] Ghorbel, A., & Trabelsi, A. (2008). Predictive performance of conditional extreme value theory in value-at-risk estimation. *International Journal of Monetary Economics and Finance*, *1*(2), 121-148. DOI:10.1504/IJMEF.2008.019218

[101] Gilleland, E., Katz, R. (2016) 2.0: An Extreme Value Analysis Package in R Journal of Statistical Software, Volume 72, Issue 8.'extRemes' R Package. Available at https://cran.r-project.org/web/packages/extRemes/extRemes.pdf

[102] Gilli, M., Këllezi, E., & Hysi, H. (2006). A data-driven optimization heuristic for downside risk minimization. Swiss Finance Institute Research Paper, (06-2). DOI: 10.2139/ssrn.910233

[103] Gneiting, Tilmann. "Making and evaluating point forecasts." *Journal of the American Statistical Association* 106.494 (2011): 746-762. DOI: 10.1198/jasa.2011.r10138

[104] Gumbel, E. J. (1958). Statistical theory of floods and droughts. *Journal of the Institution of Water Engineers and Scientists*, (12), 157-184.

[105] Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *Journal of. Multivariate Analysis*., 32, 177–203. DOI: 10.1016/0047-259X(90)90080-2

[106] Harvey A. and Sucarrat G. (2013). EGARCH models with fat tails, skewness and leverage. Computational Statistics and Data Analysis, pp. 1-19 DOI: 10.1016/j.csda.2013.09.022

[107] Harvey, Andrew and Chakravarty, Tirthankar, (2008). Beta-t-(E)GARCH, Cambridge Working Papers in Economics, Faculty of Economics, University of Cambridge. Available at: http://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe0840.pdf

[108] Heinrich,C. (2014) The mode functional is not elicitable, *Biometrika*, Volume 101, Issue 1, March 2014, Pages 245–251. DOI: 10.1093/biomet/ast048

[109] Ho A.K. and Wan A.T. (2002). Testing for covariance stationarity of stock returns in the presence of structural breaks: an intervention analysis. *Applied Economics Letters* 9(7), 441-447. DOI: 10.1080/13504850110090210

[110] Holzmann, H., & Eulert, M. (2014). The role of the information set for forecasting—with applications to risk management. *The Annals of Applied Statistics*, *8*(1), 595-621. DOI: 10.1214/13-AOAS709

[111] Iriondo A. (2017). Análisis de sensibilidad de las medidas de riesgo financiero VaR y CVaR ante cambios en el rendimiento umbral en el marco del método POT (Peak over Threshold). Master students' final work, Faculty of Economic and Business Administration (UNED).

[112] Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, *81*(348), 158-171. DOI: 10.1002/qj.49708134804

[113] Jobst, A.A. (2007). Operational risk – The sting is still in the tail but the poison depends on the dose. *Journal of Operational Risk*, 2(2), 3-59. DOI: 10.5089/9781451868036.001

[114] Jorion P. (2001). Value at risk: the new benchmark for managing financial risk. McGraw-Hill. ISBN: 0070700427

[115] Karpouzos, D. K., Kavalieratou, S. and Babajimopoulos, C. (2010). Trend analysis of precipitation data in Pieria region (Greece), *European Water*, 30, 31–40.

[116] Katz, R., Parlange, M. and Naveau, P. (2002). Statistics of extremes in hydrology, *Advances in Water Resources*, 25(8-12), 1287–1304. DOI: 10.1016/S0309-1708(02)00056-8

[117] Kellner, R., & Rösch, D. (2016). Quantifying market risk with Value-at-Risk or Expected Shortfall?–Consequences for capital requirements and model risk. *Journal of Economic Dynamics and Control*, *68*, 45-63. DOI: 10.1016/j.jedc.2016.05.002

[118] Kendall, M. G. (1938). A new measure of rank correlation, *Biometrika*, 30(1/2), 81–93. DOI: 10.1093/biomet/30.1-2.81

[119] Kharin, V. V., Zwiers, F., Zhang, X. and Wehner, M. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change* 119 345–357. DOI: 10.1007/s10584-013-0705-8

[120] Khokhlov, V. (2018). Conditional Value-at-Risk for Uncommon Distributions. DOI: 10.2139/ssrn.3200629

[121] Kjellson, B. (2013). Forecasting Expected Shortfall: An Extreme Value Approach. https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=3735573&fileOId=3735814

[122] Kohl, Matthias & Ruckdeschel, Peter & Horbenko, Nataliya. (2018). RobExtremes: Optimally robust estimation for extreme value distributions. 'RobExtreme' R Package. Available at https://CRAN.R-project.org/package=RobExtremes

[123] Kupiec P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 2, 73-84. DOI: 10.3905/jod.1995.407942

[124] Langousis, A., A. Mamalakis, M. Puliga, and R. Deidda (2016), Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database, *Water Resources Research*, 52, 2659–2681, DOI:10.1002/2015WR018502

[125] Lasch, P., Lindner, M., Ebert, B., Flechsig, M., Gerstengarbe, F.W., Suckow, F., and Werner, P.C. (1999). Regional impact analysis of climate change on natural and managed forests in the Federal State of Brandenburg, Germany, *Environmental Modeling and Assessment.* 4(4), 273–286. DOI: 10.1023/A:1019024619886

[126] Leder, N., Smirčić, A., & Vilibić, I. (1998). Extreme values of surface wave heights in the Northern Adriatic. *Geofizika*, *15*(1), 1-13. https://hrcak.srce.hr/17693

[127] Lee, Gary G. J. and Engle, Robert F. (1993). A Permanent and Transitory Component Model of Stock Return Volatility. Available at SSRN: http://ssrn.com/abstract=5848

[128] Li F., van Gelder P.H.A.J.M., Ranasinghe R., Callaghan D.P. and Jongejan R.B. (2014). Probabilistic modelling of extreme storms alongthe Dutch coast. *Coastal Engineering*, 86, 1–13. DOI:10.1111/jfr3.12296.

[129] Lindholm, D. (2015). On Value-at-Risk and the more extreme: A study on quantitative market risk measurements. http://www.urn:nbn:se:uu:diva-256173

[130] Longin, F. (1998). Value at Risk: Une nouvelle approche fondée sur les valeurs extrêmes. Annales d'économie et de statistique, 23-51. DOI: 10.2307/20076150

[131] Longin, F. M. (1996). The asymptotic distribution of extreme stock market returns. *Journal of business*, 383-408. https://www.jstor.org/stable/2353373

[132] Longin, F. M. (2000). From value at risk to stress testing: The extreme value approach. *Journal of Banking & Finance*, *24*(7), 1097-1130. DOI: 10.1016/S0378-4266(99)00077-1

[133] Lopez J.A. (1998). Testing your Risk Tests. *Financial Survey*, May-June (1998) pp. 18-20.

[134] Lopez J.A. (1999). Methods for Evaluating Value-at-Risk Estimates, *Federal Reserve Bank of San Francisco Economic Review*, 2 pp. 3-17. https://www.proquest.com/openview/64d4fc2891b74f26f0447a77ab299a9c/1?pq-origsite=gscholar&cbl=41365

[135] Loretan M. and Phillips P.C.B. (1994). Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *Journal of Empirical Finance* 1(2), 211-248. DOI: 10.1016/0927-5398(94)90004-3

[136] Louzis, D. P., Xanthopoulos-Sissinis, S., and Refenes, A. P. (2012). Stock index value-at-risk forecasting: A realized volatility extreme value theory approach. *Economics Bulletin, 32*(1), 981-991. RePEc:ebl:ecbull:eb-11-00870

[137] MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., & Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, 55(6), 2137-2157. DOI: 10.1016/j.csda.2011.01.005

[138] Madsen, H., Pearson, C.P and Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling, *Water Resources Research*, 33(4), 759–769. DOI: 10.1029/96WR03849

[139] Mandelbrot, Benoît. (1963), The Variation of Certain Speculative Prices, The Journal of Business, 36, issue. Available at: https://www.jstor.org/stable/2350970?origin=JSTOR-pdf

[140] Marinelli, C., d'Addona, S., & 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. DOI: 10.1142/S0219024907004548

[141] McAleer M., Jimenez J.A. and Pérez A. (2013). International Evidence on GFC-robust Forecast for Risk Management under the Basel Accord. *Econometric Institute*

Report EI 2011-04, Erasmus University Rotterdam, Econometric Institute. DOI: 10.1002/for.1269

[142] McAleer M., Jiménez-Martin J. and Pérez-Amaral T. (2010a). A Decision Rule to Minimize Daily Capital Charges in Forecasting Value-at-Risk, *Journal of Forecasting*, 29, pp. 617-634. DOI: 10.1002/for.1167

[143] McNeil A. (1998). Calculating Quantile Risk Measures for Financial Time Series Using Extreme Value Theory. Departmet of Mathematics, ETS. Swiss Federal Technical University E-Collection. http://e-collection.ethbib.etchz.ch/

[144] 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*, nº 7, pp. 271-300. DOI: 10.1016/S0927-5398(00)00012-8

[145] McNeil, A. J., Frey, R. and Embrechts, P. (2005). Quantitative risk management: concepts, techniques, and tools. *Princeton Series in Finance.* ISBN: 9780691166278

[146] Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3-46). NBER. ISBN 0-87014-202-X

[147] Miron, D. and Tudor C. (2010). Asymmetric Conditional Volatility Models: Empirical Estimation and Comparison of Forecasting Accuracy, Journal for Economic Forecasting. RePEc:rjr:romjef:v::y:2010:i:3:p:

[148] Mögel, B., and Auer, B. R. (2018). How accurate are modern value-at-risk estimators derived from extreme value theory? *Review of Quantitative Finance and Accounting*, *50*(4), 979-1030. DOI: 10.1007/s11156-017-0652-y

[149] Morgan J.P. (1996). Riskmetrics Technical Document, 4th Ed. New York.

[150] Moscadelli, M. (2004). The modelling of operational risk: Experience with the analysis of the data collected by the Basel Committee. DOI: 10.2139/ssrn.557214

[151] Nadarajah, S., & Kotz, S. (2008). Estimation methods for the multivariate t distribution. *Acta Applicandae Mathematicae*, *102*(1), 99-118. DOI: 10.1007/s10440-008-9212-8

[152] Nadarajah, S., Chan, S. (2017) Computes 26 Financial Risk Measures for Any Continuous Distribution. 'Risk' R Package. Available at https://search.r-project.org/CRAN/refmans/Risk/html/Risk-package.html

[153] Nadarajah, S., Zhang, B., & Chan, S. (2014). Estimation methods for expected shortfall. *Quantitative Finance*, *14*(2), 271-291. DOI: 10.1080/14697688.2013.816767

[154] Naveau, P.,Huser, R., Ribereau, P. and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research* 52 2753-2769. DOI: 10.1002/2015WR018552

[155] Neftci S. (2000). Value at risk calculations, extreme events and tail estimation. *The Journal of Derivatives*. 7(3), 23-27. DOI: 10.3905/jod.2000.319126

[156] Nelson D. (1991). Conditional Heteroskedasticity in Asset Returns: a New Approach. *Econometrica*, Vol. 59, nº 2, pp. 347-370. DOI: 10.2307/2938260

[157] Nolde, N., & Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The annals of applied statistics*, *11*(4), 1833-1874. DOI: 10.1214/17-AOAS1041

[158] Northrop, P.J., Coleman, C.L. (2014). Improved threshold diagnostic plots for extreme value analyses. *Extremes* 17, 289–303. DOI: 10.1007/s10687-014-0183-z

[159] Novales, A., and Garcia-Jorcano, L. (2019). Backtesting extreme value theory models of expected shortfall. *Quantitative Finance*, *19*(5), 799-825. DOI: 10.1080/14697688.2018.1535182

[160] Omran M.F. and McKenzie E. (2010). Heteroscedasticity in stock returns data revisited: Volume versus GARCH effects. *Applied Financial Economics* 10(5):553-60. DOI: 10.1080/096031000416433

[161] Ou, S., & Yi, D. (2009, July). Robustness analysis and algorithm of expected shortfall based on extreme-value block minimum model. In *2009 International Conference on Business Intelligence and Financial Engineering* (pp. 288-292). IEEE. DOI: 10.1109/BIFE.2009.73

[162] Özari, Ç., Eren, Ö., & Saygin, H. (2019). A new methodology for the block maxima approach in selecting the optimal block size. *Tehnički vjesnik*, 26(5), 1292-1296. DOI: 10.17559/TV-20180529125449

[163] Pagan, A. (1996). The econometrics of financial markets. Journal of empirical finance, 3(1), 15-102. DOI: 10.1016/0927-5398(95)00020-8

[164] Papalexiou, S. M., Koutsoyiannis, D., & Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, 17(2), 851-862. DOI: 10.5194/hess-17-851-2013

[165] Patton, A. J., Ziegel, J. F., & Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of econometrics*, *211*(2), 388-413. DOI: 10.1016/j.jeconom.2018.10.008

[166] Patton, E. (2001). What good is a volatility model? *Quantitative Finance*, 1(2), 237-245. DOI: 10.1016/B978-075066942-9.50004-2

[167] Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119–131. http://www.jstor.org.ezproxy.uned.es/stable/2958083

[168] Polanski A, Stoja E. (2010). Incorporating higher moments into Value-at-Risk forecasting. *Journal of Forecasting* Vol.29, nº 6, pp. 523-535. DOI: 10.1002/for.1155

[169] Pratiwi, N., Iswahyudi, C., & Safitri, R. I. (2019, May). Generalized extreme value distribution for value at risk analysis on gold price. In *Journal of Physics: Conference Series* (Vol. 1217, No. 1, p. 012090). IOP Publishing. DOI: 10.1088/1742-6596/1217/1/012090

[170] Qian Yiping, Lin Xiang, Chen Zhiya. Application of BMM model in operation risk measurement of commercial banks [J]. *Statistics & Decision*, 2010, (7): 68-71.

[171] Queensley C. Chukwudum, Peter Mwita & Joseph K. Mung'atu (2019). Optimal threshold determination based on the mean excess plot, *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2019.1624772

[172] Reiss, R.D. and Thomas, M. (2007). Statistical analysis of extreme values: With applications to insurance, finance, hydrology, and other fields, *Birkhauser*, Boston.

[173] Resnick, S. I. (1997). Heavy tail modeling and teletraffic data, *The Annals of Statistics* 25(5): 1805–1869. https://www.jstor.org/stable/2959003

[174] Richard H. Gerlach, Cathy W.S Chen and Nancy Y.C Chan (2011). Bayesian Time-Varying Quantile Forecasting for Value-at-Risk in Financial Markets. Journal of business and Economics Statistics, 29:4, 481-492, DOI: 10.1198/jbes.2010.08203

[175] Righi, M., & Ceretta, P. S. (2013). Individual and flexible expected shortfall backtesting. *Journal of Risk Model Validation*, *7*(3), 3-20. DOI: 10.2139/ssrn.2155659

[176] Robinzonov, N. (2013). Risk Management Using R. Available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=71ddb2a6463baa81a46e71dc54b2ba49034127b7

[177] Rodríguez, Gabriel (2017). Extreme value theory: An application to the Peruvian stock market returns, *Revista de Métodos Cuantitativos para la Economía y la Empresa*, ISSN 1886-516X, Universidad Pablo de Olavide, Sevilla, Vol. 23, pp. 48-74. http://hdl.handle.net/10419/195412

[178] Rubio, F. J.; Steel, M. F. J. Bayesian modelling of skewness and kurtosis with Two-Piece Scale and shape distributions. Electron. J. Statist. 9 (2015), no. 2, 1884--1912. DOI:10.1214/15-EJS1060. http://projecteuclid.org/euclid.ejs/1440680330

[179] Ruggiero P., Komar P. D. and Allan J. C. (2010). Increasing wave heights and extreme value projections: The wave climate of the US Pacific Northwest. *Coastal Engineering*, 57(5), pp. 539-552. DOI: 10.1016/j.coastaleng.2009.12.005

[180] Rydén, J. (2011). Statistical analysis of temperature extremes in long-time series from Uppsala. *Theoretical and applied climatology*, 105(1), 193-197. DOI: 10.1007/s00704-010-0389-1

[181] Saeb, A. (2018) Goodness of Fit Test for Continuous Distribution Functions. R Package 'gnFit'. Available at https://CRAN.R-project.org/package=gnFit

[182] Santinelli, L., Morio, J., Dufour, G., & Jacquemart, D. (2014). On the sustainability of the extreme value theory for WCET estimation. In 14th International Workshop on Worst-Case Execution Time Analysis. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. DOI: 10.4230/OASIcs.WCET.2014.21

[183] Sarma M, Thomas S, Shah A. (2003). Selection of Value-at-Risk Models. *Journal of Forecasting* **22**(4) 337-358. DOI: 10.1002/for.868

[184] Scarrot, A and McDonald, A (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT – Statistical Journal*. Volume 10, Number 1, March 2012, 33–60. DOI: 10.57805/revstat.v10i1.110

[185] Schneider, L.F., Krajina, A. & Krivobokova, T. (2021). Threshold selection in univariate extreme value analysis. *Extremes* 24, 881–913 (2021). DOI: 10.1007/s10687-021-00405-7

[186] Singh, A. K., Allen, D. E., & Robert, P. J. (2013). Extreme market risk and extreme value theory. *Mathematics and computers in simulation*, 94, 310-328. DOI: 10.1016/j.matcom.2012.05.010

[187] Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1), 67-90. DOI: 10.1093/biomet/72.1.67

[188] Solari, S. and Losada, M.A. (2012). A unified statistical model for hydrological variables including the selection of threshold for the peak over threshold method. *Water Resources Research* 48. DOI: 10.1029/2011WR011475

[189] Stephens, M. A. (1970). Use of the Kolmogorov–Smirnov, Cramer–Von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(1), 115-122. DOI: 10.1111/j.2517-6161.1970.tb00821.x

[190] Szubzda, F., & Chlebus, M. (2019). Comparison of block maxima and peaks over threshold Value-at-Risk models for market risk in various economic conditions. *Central European economic journal*, 6(53), 70-85. DOI: 10.2478/ceej-2019-0005

[191] Tanaka, S., and Takara, K. (2002). A study on threshold selection in POT analysis of extreme floods, IAHS AISH Pub., 271, 299–304. Available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ce6e56b14e37512c73ed65ce3beb5bc5ddb75e4c

[192] Thompson, P., Cai, Y., Reeve, D. and Stander, J. (2009). Automated threshold selection methods for extreme wave analysis. *Coastal Engineering* 56 (10):1013–21. DOI :10.1016/ j.coastaleng.2009.06.003.

[193] Tsay, R. S. (2005). *Analysis of financial time series*. John Wiley & Sons.

[194] Tsay, R.S. (2010). Extreme Values, Quantiles, and Value at Risk. DOI: 10.1002/9780470644560.ch7

[195] Van Zyl J.M. (2011). Application of the Kolmogorov–Smirnov test to estimate the threshold when estimating the extreme value index. *Communications in Statistics-Simulation and Computation*, 40(2), 199-207. DOI: 10.1080/03610918.2010.533227

[196] Wadsworh, J. L. and Tawn, J. A. (2012). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. J. Royal *Statistical. Society. Series B. Statistical. Methodology*. 74 543–567. DOI: 10.1111/j.1467-9868.2011.01017.x

[197] Wang, J., You, S., Wu, Y., Zhang, Y., & Bin, S. (2016). A method of selecting the block size of BMM for estimating extreme loads in engineering vehicles. *Mathematical Problems in Engineering*. DOI: 10.1155/2016/6372197

[198] Wang, Q. J. (1991). The POT model described by the generalized Pareto distribution with Poisson arrival rate, *Journal of Hydrology*, 129(1-4), 263-280. DOI: 10.1016/0022-1694(91)90054-L

[199] Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, *16*(2), 419-441. DOI: 10.1111/j.1467-9965.2006.00277.x

[200] Werner, P. C., and Gerstengarbe, F.W. (1997). Proposal for the development of climate scenarios, *Climate Research*, 8, 171–182. DOI: 10.3354/CR008171

[201] Wimmerstedt, L. (2015). Backtesting Expected Shortfall: the design and implementation of different backtests. urn:nbn:se:kth:diva-172444

[202] Wyncoll, D. and Gouldby, B. (2013). Application of a multivariate extreme value approach to system flood risk analysis. *Journal of Flood Risk Management*, 1, 3-10. DOI: 10.1111/jfr3.12069

[203] Zhang M.H. and Cheng Q.S. (2005). An Approach to VaR for Capital Markets with Gaussian Mixture. *Applied Mathematics and Computations*. Vol. 168, pp 1079-1085. DOI: 10.1016/j.amc.2004.10.004

[204] Žiković, S., and Aktan, B. (2009). Global financial crisis and VaR performance in emerging markets: A case of EU candidate states-Turkey and Croatia. *Zbornik radova Ekonomskog fakulteta u Rijeci: časopis za ekonomsku teoriju i praksu*, 27(1), 149-170. https://hrcak.srce.hr/38423

[205] Zikovic, Sasa and Filer, Randall K., (2012). Ranking of VaR and ES Models: Performance in Developed and Emerging Markets. CESifo Working Paper Series No. 3980. Available at SSRN: http://ssrn.com/abstract=2171673

[206] Zivot, E., & Wang, J. (2006). Modeling financial time series with S-PLUS (Vol. 2). New York: Springer. DOI: 10.1007/978-0-387-32348-0