ACRN Journal of Finance and Risk Perspectives 11 (2022) 79-104

Contents lists available at SCOPUS

ACRN Journal of Finance and Risk Perspectives

journal homepage: http://www.acrn-journals.eu/

A comparison of market risk measures from a twofold perspective: accurate and loss function

Sonia Benito Muela*, Carmen López-Martin, Raquel Arguedas-Sanz

National Distance Education University (UNED)

ARTICLE INFO

Article history:
Received 13 May 2020
Revised 01 July 2021 and 06 October 2022
Accepted 07 November 2022
Published 04 January 2023

Keywords:
Expected shortfall
Value at Risk
APARCH model
Backtesting
Skewed distributions

JEL: C13, C58, G11, G15, G17

ABSTRACT

Under the new regulation based on Basel solvency framework, known as Basel III and Basel IV, financial institutions must calculate the market risk capital requirements based on the Expected Shortfall (ES) measure, replacing the Value at Risk (VaR) measure. In the financial literature, there are many papers dedicated to compare VaR approaches but there are few studies focusing in comparing ES approaches. To cover this gap, we have carried out a comprenhensive comparative of VaR and ES models applied to IBEX-35 stock index. The comparison has been carried out from a twofold perspective: accurate risk measure and loss functions. The results indicate that the method based on the conditional Extreme Value Theory (EVT) is the best in estimating market risk, outperforming Parametric method and Filter Historical Simulation.

Introduction

Over the last decades the financial services industry has undergone significant transformation due to internal and external factors, including business model transformation, adoption of advances technologies, changing regulatory frameworks, etc.

Since the global financial crisis of 2008, there has also been an ever-growing need for financial institutions to accurately assess their exposure to financial risks (Summinga and Narsoo, 2019), especially under stressed conditions or combined with other interrelated triggering events (Danielsson et al., 2014; Deloitte, 2017). The market risk

* Corresponding author.
E-Mail address: soniabm@cee.uned.es
ORCID: 0000-0001-9409-0868

https://doi.org/10.35944/jofrp.2022.11.1.005
ISSN 2305-7394

This work was supported by the Spanish Ministerio de Ciencia e Innovación (PID2020-113367RB-I00) and Young Talent Research Project UNED 2022 (076-044355 ENER-UK).

regulatory pre-crisis models were build up on assumptions (stochastic processes governing market prices, volatility models, etc) that did not adecuately capture tail-risk events, hence underestimating posible losses in stressed conditions (assuming market liquidity and large diversification effects across asset classes, etc). The crisis revelead substantial weaknesses in the banking system and the prudential framework, leading to risk-taking unsupported by adequate capital and liquidity buffers (Committee on the Global Financial System, CGFS, 2018).

Regulators have responded to the crisis by reforming the global prudential framework and enhancing supervisión in order to increase banks' resilience and financial stability worlwide. The Basel Committee on Banking Supervision[1] (BCBS), in its October 2013 consultative paper, and subsequent versions published thereafter, for revised market risk framework, suggested new ways of dealing with market risk in banks' trading and banking books.

Under the new regulation based on Basel solvency framework (BCBS, 2012, 2016a, 2017a, 2017b, 2019), known as Basel III, also coined as Basel III-revised or Basel IV[2] (see Feridun and Ozün, 2020), financial institutions must calculate the market risk capital requirements based on the Expected Shortfall (ES) measure, replacing the Value at Risk (VaR) measure based on internal models legitimized by the supervisory authorities since 1998 (BCBS, 1996; Hubbert, 2012; Acerbi and Szekey, 2014; Chen, 2014; Szylar, 2014; Chang et al, 2019; Rossignolo, 2019). Table 1 shows a summary of the evolution of the Basel capital requirements. The BCBS estimated that the new rules will result in an approximate median capital increase of 22% and a weighted average capital increase of 40% (BCBS, 2016b:7), compared with the previous framework.

[1] The Basel Committee on Banking Supervision (BCBS) was established in 1974 by the central bank governors of the G10 countries as a forum for regular cooperation on banking supervision. It is currently composed of central banks and high-level representatives of the supervisory authorities of the following member countries: Argentina, Australia, Belgium, Brazil, Canada, China, France, Germany, Hong Kong SAR, India, Indonesia, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, Russia, Saudi Arabia, Singapore, South Africa, Spain, Sweden, Switzerland, Turkey, Great Britain and the United States. The BCBS is not a legislative body. Its role is to develop supervisory recommendations, which only take effect when are adopted by national authorities in each jurisdiction

[2] A critical analysis of the new Basel minimum capital requirements for market risk can be found in Orgeldinger (2017). In the context of credit risk, Binder and Lehner (2020) study the impact of the Basel IV regulations on risk weight densidity, one of the main topics of the revision of the Basel III framework in order to reduce the excessive volatility of risk weights for credit risk which is also associated to the introduction of parameter floors and output floors (BCBS, 2017a).

Table 1. Evolution of Basel standards. Capital requirements and approximations

BIS Agreement	Objective	Issue date	Implementation date
	Pre-financial crisis		
Basel I (BIS I)	Creating a general framework focus on *Credit risk* (CAR: 8% RWA)	1988	1992
Ammended to Basel I	Incorporation *of market risk*. Internal VaR models	1996	1998
Basel II (BIS II)	Provide a more risk-sensitive framework, which is structured in 3 pillars: • Pillar I: minimum capital requirements for *credit, market and operational risks* - Standardized approach & internal risk models (VaR) • Pillar II: Supervisory review process • Pillar III: Market discipline	1999	2007
	Post-financial crisis		
Basel 2.5	Adjust capital requirements for market risk: • Stress VaR • Capital burden for incremental risk	2009	2011
Basel III	Strengthen Basel II to adopt post-financial crisis reforms: • Reviews of internal and standardized model approximations. • Strengthening the solvency ratio • Incorporation of illiquidity risk with two new ratios: LCR and NSFR • Límits between banking and trading book. • CET 1: 2-4% • Capital buffer: 2.5-7% • Countercyclical buffer: 0-2.5% • Leverage ratio: 3%	2009/ 2010	2022 initially but deferred by one year to 1 January 2023 (BCBS, 2020)
	Basel III-revised (also called Basel IV)		
Basel III-revised (coined as Basel IV)	Market risk: Replacement of VaR by ES as a regulatory risk measure	2016 (*)	
	Elements of the package that includes the following items: • Credit risk: a revised standardised approach and revisions to the internal ratings-based (IRB) approach • Credit valuation adjustment (CVA) framework revised; • Operational risk: a revised standardised approach • Revisions to the measurement of the leverage ratio • An aggregate output floor, which will ensure that banks' RWA generated by internal models are no lower than 72.5% of RWA as calculated by the Basel III Framework's standardised approaches.	2017	

Notes: CAR: Minimum regulatory capital requirement; RWA: Risk-weighted assets; ES: Expected Shortfall; LCR Liquidity Coverage Coefficient; NSFR: Stable Net Funding Ratio; CET 1: Common Equity Tier 1 own resources
(*) Previous versions in 2012/2013
Source: Own elaboration

VaR has traditionally been the main risk measure to compute the market risk of financial assets. Portfolio VaR represents the maximum amount an investor may lose that is likely to occur at a given confidence level over a specified period (holding period). It provides a measure of the loss frequency (Acerbi and Tasche, 2002) but it does not provide an estimate of the loss severity.

The main limitations of VaR as a measure of financial risk are (i) VaR does not report the magnitude of loss when it is greater than VaR, which means that it fails to capture tail risks and (ii) VaR is not a coherent risk measure by not meeting the axiom of subaditivity, as demonstrated by Artzner et al (1999). This non-subaditivity of VaR implies that it may be the case that, contrary to the classical idea of diversification, the total VaR of a portfolio could be greater than the sum of the VaR of each of its components. Both problems are solved with an alternative risk measure, the Expected Shortfall (ES), which is consistent and does report losses beyond the VaR, by measuring the average of the losses when they exceed the VaR[3].

Expected Shortfall (ES) measure is defined as the expected return on an asset/portfolio conditional on the return being below a given quantile of its distribution (its VaR). Unlike VaR measure, ES is a coherent measure of risk (Artzner et al., 1997, 1999). This feature along with the observation shows that under the stress conditions as the observed during the global financial crisis, VaR forecasts (99 percentil) were exceeded multiple times, explaining the change in the regulation fostered by the Basel Committee on Banking Supervision. In the new solvency framework ES (97.5th percentile, one-tailed confidence level) is calibrated on a stressed period to reduce procyclicality (Zhang, 2016).

The methodologies developed to calculate a portfolio VaR are (i) the variance-covariance approach also called Parametric method, (ii) the Non-parametric approach (e.g. Historical Simulation), and (iii) the Semiparametric approach (e.g. Extreme Value Theory, Filter Historical Simulation and CaViaR method). All these models can be extended to calculate ES measure.

In the financial literature, there are many papers dedicated to compare VaR approaches (see Abad et al., 2014 for an exhaustive review) but there are few studies focusing in comparing ES approaches. Regarding to this last measure, most of the studies focus on proposing different tests to evaluate the validity of the ES estimations[4].

Some comparison studies of ES approaches can be found in McNeil and Frey (2000), Yamai and Yoshiba (2002a, 2002b), Harmantzis et al. (2006), Rigui and Ceretta (2015), Zikovic and Filer (2012), Liu and Kuntjoro (2015) and Sobreira and Louro (2020). McNeil and Frey (2000) and Yamai and Yoshiba (2002a, 2002b) compare the approach based on the conditional Extreme Value Theory (EVT) with the Gaussian model and conclude that the former provides more accurate ES estimacions. Harmantzis et al. (2006) compare the performance of three ES models: Gaussian model, Historical Simulation (HS) and the approach based on the EVT. They conclude that Gaussian model understimates ES measure meanwhile the HS model and the conditional EVT provide more accurate estimations. Rigui and Ceretta (2015) compare a large set of models: (i) Parametric method below several distribution (ii) EVT and (iii) CaViaR method. The models are estimated using unconditional and conditional volatility models. They find that there is a predominance of conditional models over the unconditional models.

[3] Yamai and Yoshiba (2002a, b) of the Bank of Japan were the first regulators to take into account the great problem of VaR, its non-subaditivity and after their comparative analysis with VaR, laid the foundations of the ES as a possible future regulatory measure. Some authors such as Daníelsson et al (2001) and Basak and Shapiro (2001), showed that VaR had potentially destabilizing effects on the economy. Danielsson and Zhou (2017) considered that the biggest drawback of VaR is not its non-subsidiarity but that the main disadvantage of VaR compared to ES lies in its greater ease of being manipulated without incurring regulatory breaches.

[4] These tests are based on exception frequency test (Du and Escanciano, 2017; Colletaz et al., 2013; Moldenhauer and Pitera, 2017), exception frequency and independence tests (Du and Escanciano, 2017; Costanzino and Curran, 2015, 2018; Kratz et al., 2018; Emmer et al., 2015, Clift et al., 2016; Patton et al., 2019), exception magnitude tests (McNeil and Frey, 2000; Wong, 2008, 2010) and exception frequency and magnitude test (Acerbi and Szekely, 2014; Linconvil and Chiann, 2018). A comprehensive review of most of these test can be found in García-Jorcano (2017) and Novales and Garcia-Jorcano (2019).

Zikovic and Filer (2012) compare several ES models: (i) parametric model, some non-parametric models as (ii) HS and (iii) Kernel Historical approach, and two semiparametric approaches (iv) FHS method and (v) the unconditional and conditional EVT approach. The results show that for a large number of models there is no statistically significant difference. The top performers are the approach based on the conditional EVT with a GARCH model for modeling volatility and the models based on volatility updating. Liu and Kuntjoro (2015) evaluate the performance of three family models: (i) parametric approach below Gaussian distribution and Student-t distribution (ii) Historical Simulation (HS) which is a non parametric approach and (iii) the EVT (semi-parametric approach). From these three families, nine models of estimation are developed as they consider conditional and unconditional volatility measure. The study is carried out in three periods. Among all nine models, the unconditional EVT model was the only suitable model for all the three evaluation periods. Unexpectedly, the result also indicated that conditional models did not improve the accuracy of ES estimates compared to corresponding unconditional models. Sobreira and Louro (2019) compare the performance of different methodologies: HS, Riskmetrics, Parametric approach below different distributions and the EVT approach. The results show that the combination of asymmetric GARCH with EVT performs the best in forecasting ES measure, specially for more conservative coverage levels. Overall, all of these studies indicate that the approach based on the conditional EVT is the best in estimating ES measure.

In this paper we compare the performance of three relevant approaches for forecasting VaR and ES. The methods included in the comparison are: (i) parametric approach and two semiparametric approach: (ii) the approach based on the conditional EVT and (iii) Filter Historical Simulation (FHS). In the case of parametric method, five distribution have been considered: Normal, Student-t (symmetric) distribution (STD), the skewness student-t distribution (SSTD), generalized error distribution (GED) and the skewness generalized error distribution (SGED) of Theodossiou (2001). To estimate portfolio market risk a forecast volatility is required, for which we use an APARCH model. This model has been estimated below different distribution: Gaussian, Student-t, GED, skew Student-t and skew GED. Thus for each method we have five risk measures.

Our objective is to carry out a systematic analysis that simultaneously consider different approaches and different distribution hypotheses for modeling volatility. Unlike the aforementioned papers the comparison of these three methodologies is made both in terms of their ability to provide accurate risk measure and in terms of loss functions. To undertake this study, we focus on the Spanish stock market, which has not been previously analysed in the literature on ES estimation models throughout the current decade and, particularly in the recent period 2014-2017 that follows the financial crisis.

The rest of the paper is organized as follows. Section II describes the methodology, which includes the risk measures applied, the volatility model used, the backtesting of VaR and ES forecasts and loss functions. Section III and IV present the data and empirical results, respectively, and Section V ends with the conclusions.

Methodology

Risk Measures

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:

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

where $\tilde{\sigma}_t^2 = E(z_t^2|\Omega_{t-1})$ and z_t has the conditional distribution function G(z), $\mathrm{G(z)} = \mathrm{P}(\mathrm{z_t} < \mathrm{z}|\Omega_{t-1})$. It can be assumed that $\sigma_t = \sigma$ for all *t* or that σ_t has a probability density $\Pr\,(\sigma_t|\Omega_{t-1})$. In this paper, we consider the later. The VaR with a given probability $\alpha \in (0, 1)$, denoted by VaR(α), is defined as the α quantile of the probability distribution of financial returns:

$$F\big(VaR(\alpha)\big) = \Pr\big(X_t \leq VaR(\alpha)\big) = \alpha \tag{2}$$

There are two methods to estimate this quantile: (1) inverting the distribution function of financial returns $F(x)$ and (2) inverting the distribution function of innovations $G(z)$. With regard to the later, it is also necessary to estimate σ_t^2

$$VaR_t(\alpha) = F^{-1}\,(\alpha) = \mu_t + \alpha_t G^{-1}(\alpha) \tag{3}$$

Therefore, a VaR model involves the specification of $F(x)$ or σ_t^2 and $G(z)$. Parametric method, the approach based on the conditional Extreme Value Theory approach and the Filter Historical Simulation approach focus on estimating[5] $G(z)$. The ES measure with a given probability $\alpha \in (0,1)$, denoted by ES(α), 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:

$$ES_t(\alpha) = E_{t-1}[Z_t|Z_t \leq VaR_t(\alpha)] \tag{4}$$

or

$$ES_t(\alpha) = \mu_t + \alpha_t E_{t-1}[Z_t|Z_t \leq G^{-1}(\alpha)] \tag{5}$$

Table 2 summarizes the market risk capital requirements (based on VaR metrics first and ES currently) proposed by the Basel Committee regulations over time.

[5] We assume that $\mu_t = \mu$.

Table 2. Market risk capital requirements under Basel Accords: internal models approach

1996

The daily capital requirement for general market risk (CRM) would be the maximum between the previous day's VaR and k times the average daily VaR over the past 60 days, where k was a multiplier in the range between 3 and 4. This multiplier, set by the banking supervisors, was conditional on the results of the validation exercises of the models (*backtesting*, see table below):

$$CRM_t = \max\left\{VaR_{t-1}, k \cdot \sum_{i=1}^{60} \frac{1}{60} \cdot VaR_{t-i}\right\}$$

$$3 \leq k \leq 4$$

VaR should be obtained at the 99% confidence level using a 10-day holding period.
The historical observation period for calculating the VaR had to be at least one year.
The supervisor did not prescribe a particular type of model for the estimation of VaR, being able to use the method of variances / covariances, historical simulation or Monte Carlo, among others.

2009

An additional capital requirement for market risk was set (coined as *"Stress VaR"*, VaR^s) based on the results of *stress testing* scenarios, in addition to that indicated in the previous formula based on VaR. *Stress VaR* (calculated at least weekly using a 10-day horizon and with a confidence of 99%) aimed to replicate the VaR that would have been generated in the bank's current portfolio if market risk factors had experienced a period of 12 months of continuous financial stress.
The capital requirement for market risk is calculated according to:

$$CRM_t = \max\left\{VaR_{t-1}, k \cdot \sum_{i=1}^{60} \frac{1}{60} \cdot VaR_{t-i}\right\} + max\left\{VaR^s_{t-1}, k \sum_{i=1}^{60} \frac{1}{60} VaR^s_{t-i}\right\}$$

2012 to date

Replacement of the VaR by the ES metric. The ES is calculated, on a daily basis**,** at a confidence level of 97.5. The ES for a liquidity horizon has to be calculated from an ES with a 10-day liquidity

The ES has to be calibrated for a period of stress. Banks can specify a small number of risk factors that has to explain a minimum of 75% of the variation in the ES model. The ES for capital risks is calculated as follows: $ES = ES_{R,S} \cdot \left(ES_{F,C}/ES_{R,C}\right)$

The ES for regulatory capital purposes is equal to the ES based on the periods of stress observations using a small group of risk factors multiplied by the ratio of the ES measure based on the most recent 12-month observation period with an entire set of risk factors.

The aggregate capital requirement for market modellable risk factors (*IMCC*) is equal to the maximum of the most recent observation and a weighted average of the previous 60 días ($IMCC_{avg}$) scaled by a multiplier (*m*) that is fixed at a minimum of 1.5 (up to 2 depending on the *backtesting* results (see table):

$$CRM_t = \max\left\{IMCC_{t-1}, m \cdot IMCC_{avg}\right\}$$

Backtesting zones in the Basel Capital Accords.

Zone	Number of exceptions	*Backtesting*-dependent multiplier 1996/2009 (*k*)	*Backtesting*-dependent multiplier 2012 (*m*)
Green	0 a 4	3	1.5
Yellow/ Amber	5	3.4	1.7
	6	3.5	1,76
	7	3.65	1.83
	8	3.75	1,88
	9	3.85	1.92
Red	10 or more	4	2

In the following lines we describe the general characteristics of the methods that we compare in this paper.

(i) Parametric approach.

This approach assumes that financial returns follow a known distribution function. Below this method $G^{-1}(\alpha)$ is the percentile α of the distribution assumed. In this study we have considered five types of distribution: (i) Normal, (ii) Student's t-distribution, (ii) skew Student-t, (iv) GED and (v) skew GED. With this method $E[Z_t|Z_t \leq G^{-1}(\alpha)]$ can be calculate as follow

$$E_{t-1}[Z_t|Z_t \leq G^{-1}(\alpha)] = \frac{1}{\alpha}\int_{-\infty}^{G^{-1}(\alpha)} zg(z)dz = \frac{1}{\alpha}\int_{0}^{\alpha} G^{-1}(s)ds \qquad (6)$$

where $g()$ is the density function of the innovations.

The main advantage of this method is that it is very easy to implement, especially when a normal distribution is assumed for innovations, although in this case the results are usually somewhat poor. These results improve when fat tail and asymmetric distributions are assumed, in which case the degree of difficulty for risk estimation increases.

(ii) Filter Historical Simulation (FHS)

Filter Historical Simulation (FHS) was proposed by Barone-Adesi et al. (1999). According to this method $G^{-1}(\alpha)$ can be calculated as follow:

$$G^{-1}(\alpha) = Quantile\left\{(\hat{z}_t)_{t=1,}^{n}, \alpha\right\} \qquad (7)$$

where $(\hat{z}_t)_{t=1}^{n}$ is a innovation sample simulated by bootstrapping. Thus, the right term in Equation 7 is the quantile α of the innovations simulated sample. Replacing (7) in (3) we obtain the VaR measure. In this case we calculate $E[Z_t|Z_t \leq G^{-1}(\alpha)]$ by approximation (see Emmer et al., 2015):

$$E[Z_t|Z_t \leq G^{-1}(\alpha)] = \frac{1}{4}\{q_{\alpha}(\hat{z}_t) + q_{0.75\alpha+0.25}(\hat{z}_t) + q_{0.5\alpha+0.5}(\hat{z}_t) + q_{0.25\alpha+0.75}(\hat{z}_t)\} \qquad (8)$$

being q_{α} the α quantile of the innovations simulated sample $(\hat{z}_t)_{t=1}^{n}$. The ES measure will be obtained by replacing the above expression in [5].

This approach does not make strong assumptions about the distribution of the returns portfolio, so that it can accommodate wide tails, skewness and any other non-normal features. The disadvantages of this approach is that the results slightly dependent on the data set. Besides, unlike other non-parametric approaches like Historical Simulation, FHS take volatility background into account.

(iii) The approach based on the conditional Extreme Value Theory.

Within the EVT context, there are two approaches to study the extreme events. One of them is the direct modeling of the distribution of minimum or maximum realizations (McNeil, 1998). The other one is modeling the exceedances of a particular threshold (Peaks Over Threhold method (POT)). This last approach is generally considered to be the most useful for practical applications due to the more efficient use of the data for the extreme values. In the Appendix of the paper we provide a detailed description of this methodology. In this paper we use 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:

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

where μ_t and $\tilde{\sigma}_t$ represent the conditional mean and the conditional standard deviation of the returns[6] and $G^{-1}(1-\alpha)$ is the quantile $(1-\alpha)$ of the GPD. The ES of a portfolio at $(1-\alpha)$% probability will be calculated as:

$$ES_t(\alpha) = \mu_t + \tilde{\sigma}_t E_{t-1}[Z_t | Z_t \geq G^{-1}(1-\alpha)] \tag{10}$$

and

$$ES_t = \mu_t + \tilde{\sigma}_t \left[\frac{G^{-1}(1-\alpha)}{1-\xi} + \frac{\beta - \xi u}{1-\xi} \right] \tag{11}$$

where ξ and β are the shape parameter and ths escala parameter of the Generalized Pareto distribution (GPD).

The same as FHS, this method captures some characteristics of the financial returns as curtosis and skew and it take into account changes in volatility (conditional ETV). The disadventages of this method is that it depends on the extreme return distribution assumption. Besides, its results depend on the extreme data set.

APARCH model

The APARCH model (Asymmetric Power ARCH model) was proposed by Ding et al (1993). This model can well express volatility clustering, fat tails, excess kurtosis, the leverage effect, and the Taylor effect. The latter effect is named after Taylor (1986) who observed that the sample autocorrelation of absolute returns was usually larger than that of squared returns. The APARCH equation is,

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i (|\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta \tag{12}$$

where ω, α_i, γ_i, β_j and δ are additional parameters to be estimated. The parameter γ_i reflects the leverage effect $(-1 < \gamma_i < 1)$. A positive (resp. negative) value of γ_i means that past positive (resp. negative) shocks have a deeper impact on current conditional volatility than past negative (resp. positive) shocks. The parameter δ plays the role of a Box-Cox transformation of σ_t $(\delta > 0)$.

The APARCH equation is supposed to satisfy the following conditions, *i*) $\omega > 0$ (since the variance is positive), $\alpha_i \geq 0$, $i = 1, 2, \ldots, q$, $\beta_j \geq 0$, $j = 1, 2, \ldots, p$. When $\alpha_i = 0$, $i = 1, 2, \ldots, q$, $\beta_j = 0$, $j = 1, 2, \ldots, p$, then $\sigma^2 = \omega$, *ii*) $0 \leq \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j \leq 1$. The APARCH is a general model because it has great flexibility, having as special cases, among others, GARCH and GJR-GARCH models.

[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.

Backtesting

Backtesting VaR

Many authors are concerned about the adequacy of the VaR and ES measures, especially when they compare several methods. Papers, which compare the VaR methodologies commonly use two alternative approaches: the basis of the statistical accuracy tests and/or loss functions (Sarma et al., 2003; Angelidis and Degiannakis, 2007; Abad et al., 2015).

a.- Accuracy tests

We check the accuracy of different estimates of VaR using five standard tests which are the most usual procedures: the unconditional coverage (LR_{uc}) test (Kupiec, 1995), the conditional coverage (LR_{cc}) test and the independence test (LR_{ind}) of Christoffersen (1998), Backtesting Criterion Statistic (BTC) and *Dynamic Quantile* (DQ) test of Engle and Manganelli (2004). To implement all these tests, the exception indicator (I_t) must be defined. If r_t represents the returns and $VaR(\alpha)$ is the VaR obtained with a given probability $\alpha \in (0,1)$, we have an exception when $r_t < VaR_t(\alpha)$and then I_t is equal to one (zero otherwise).

Kupiec (1995) shows that assuming the probability of an exception is constant, then the number of exceptions ($x = \sum I_t$) follows a binomial distribution $B(N, \alpha)$, where N is the number of observations. An accurate $VaR(\alpha)$ measure should produce an unconditional coverage ($\hat{\alpha} = \sum I_t/N$) equal to α percent. Thus, the null hypothesis of this test is $\hat{\alpha} = \alpha$ for the unconditional coverage test has as a null hypothesis $\hat{\alpha} = \alpha$, with a likelihood ratio statistic:

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

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

Christoffersen (1998) developed a *conditional coverage test*. This jointly examines whether the percentage of exceptions is statistically equal to the one expected and the serial independence of I_{t+1}. He proposed an independence test, which aimed to reject VaR models with clustered violations. The LR_{cc} test examines jointly whether the model generates a correct proportion of failures (LR_{uc}) and whether the exceptions are statistically independent from one another (LR_{ind}). The independence property of exception is an essential property because the measures of risk must reply automatically to any new information; a model that does not consider this factor would provoke clustering of exceptions. The likelihood ratio statistic of the conditional coverage test is $LR_{cc} = LR_{uc} + LR_{ind}$, which is asymptotically distributed $\chi^2(2)$. The LR_{ind} statistic is the likelihood ratio statistic for the hypothesis of serial independence against first-order Markov dependence. The LR_{ind} statistic is $LR_{ind} = 2[logL_A - logL_o]$ 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, $\pi_{01} = \frac{N_{01}}{N_{01}+N_{00}}$ and $\pi_{11} = \frac{N_{11}}{N_{11}+N_{10}}$. The likelihood function under the null hypothesis ($\pi_{01} = \pi_{11} = \pi = \frac{N_{11}+N_{01}}{N}$) is $L_0 = (1-\pi)^{N_{00}+N_{10}}(\pi)^{N_{01}+N_{11}}$.

A similar test to Kupiec test for the significance of the departure of $\hat{\alpha}$ from α is the BTC. In this case, what is compared to check if the VaR estimates are accurate is to compare if the number of observed exceptions is equal to the number of expected exceptions H_o: $N\alpha \leq N\hat{\alpha}$. The statistic (Z) known as Backtesting Criterion, is defined by the expression:

$$Z = \frac{N\hat{\alpha} - N\alpha}{\sqrt{N\alpha(1-\alpha)}} \qquad (14)$$

which is asymptotically distributed as a normal with mean zero and variance one, $Z \sim N(0,1)$.

Finally, the DQ test examines whether the exception indicator is uncorrelated with any variable that belongs to the information set Ω_{t-1} available when the VaR was calculated[7]. This is a Wald test of the hypothesis that all slopes are zero in a regression of the exception indicator variable on a constant, five lags and the VaR estimate.

$$I_t = \beta_0 + \sum_{i=1}^{5} \beta_i I_{t-1} + \mu VaR + \varepsilon_t \qquad (15)$$

We consider that a model is accurate when it passes all tests.

b.- Loss functions

The backtesting procedures based on certain statistical tests present a drawback; they only show whether the VaR estimates are accurate, so this toolbox does not allow us to rank the models. Backtesting based on the loss function pays attention to the magnitude of the failure when an exception occurs. Lopez (1998, 1999), who is a pioneer in this area, proposes to examine the distance between the observed returns and the forecasted VaR(α). This difference represents the loss that has not been covered. The loss functions enable the financial manager to rank the models. The model that minimises the total loss will be preferred to the other models. Lopez (1999) proposed a general form of the loss function:

$$L_t = \begin{cases} f(r_t, VaR) & if \quad r_t < VaR \\ g(r_t, VaR) & if \quad r_t \geq VaR \end{cases} \qquad (16)$$

where $f(r_t, VaR)$ and $g(r_t, VaR)$ are functions such that $f(r_t, VaR) \geq g(r_t, VaR)$, thereby penalising to a greater extent those cases where the real returns fall below the VaR estimations. Lopez proposed different functional forms for $f(r_t, VaR)$ and $g(r_t, VaR)$. In this study, we apply two loss functions: Lopez's magnitude loss function and Lopez's lineal loss function.

Lopez's Magnitude loss function (LF1) has the following quadratic specification where large failures are penalised more than small failures:

$$\text{LF1} = \begin{cases} 1 + (VaR_t - r_t)^2 & if \quad r_t < VaR_t \\ 0 & otherwise \end{cases} \qquad (17)$$

Lopez's lineal loss function (LF2) is calculated as follows:

$$\text{LF2} = \begin{cases} (VaR_t - r_t) & if \quad r_t < VaR_t \\ 0 & if \quad r_t \geq VaR_t \end{cases} \qquad (18)$$

Caporin (2008) notes that there is an open issue with the functions aforementioned. For this author, what is important is not the losses uncovered but their relative size. For this purpose, the author proposes three different loss function. In this study we have applied the following specification:

[7] A more detailed description of these tests can be found in Abad et al. (2014).

$$\text{LF3}=\begin{cases}\left|1-\left|\frac{r_t}{VaR_t}\right|\right| & if \quad r_t<VaR_t\\ 0 & if \quad r_t\geq VaR_t\end{cases} \tag{19}$$

Backtesting ES

a.- Accuracy test

In order to check the accuracy of conditional ES, we use McNeil and Frey (2000) test which is the most successful in the literature. 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 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. 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 population mean μ_y and variance σ_y, by applying the central limit theorem, the statistics for testing the null hypothesis are given by

$$t=\frac{\bar{y}}{\frac{S_y}{\sqrt{N}}}\sim t_{N-1} \tag{20}$$

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

b.- Loss functions

In the same way to Lopez's loss function for VaR, Nieto and Ruiz (2008) introduced the loss function calculated with respect of the ES. These authors obtained the loss function with respect to ES, when an exception occurs. The loss function for ES (L) has the following expression:

$$LF4=\begin{cases}(r_t-ES_t)^2 & if \quad r_t<VaR_t\\ 0 & if \quad r_t\geq VaR_t\end{cases} \tag{21}$$

Assuming ES is the new risk measure, we generalize Lopez's lineal loss function (LF2) and Caporini loss function (LF4) calculated with respect ES. Thus, we introduce the following expressions to quantify loss function for ES:

$$\text{LF5}=\begin{cases}1+(VaR_t-r_t)^2 & if \quad r_t<VaR_t\\ 0 & otherwise\end{cases} \tag{22}$$

$$\text{LF6}=\begin{cases}(ES_t-r_t) & if \quad r_t<VaR_t\\ 0 & if \quad r_t\geq VaR_t\end{cases} \tag{23}$$

$$\text{LF7}=\begin{cases}\left|1-\left|\frac{r_t}{ES_t}\right|\right| & if \quad r_t<VaR_t\\ 0 & if \quad r_t\geq VaR_t\end{cases} \tag{24}$$

The model which minimizes the LF will be the best.

Data

The data consist of the IBEX35[8] 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 3, 2000, through December 29, 2017. The full data period is divided into a learning sample (January 3rd, 2000 to December 31th, 2013) and a forecast sample (January 3rd, 2014 to December 29th, 2017). Thus, we work with 4568 observations and generate 1022 out-of-sample VaR and ES forecasts with a confidence level of α=97.5%[9].

Figure 1 presents the evolution of the daily index and returns of the IBEX35. The index shows a sawtooth profile alternating periods with upward slope and 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 IBEX35 was particularly high from 2008 to 2009, coinciding with the period known as the Global Financial Crisis. In the last years of the sample, we observe a period that is more stable.

The basic descriptive statistics are provided in Table 3. The unconditional mean of daily return is negative and very close to zero (-0.0032%). The skewness statistic is negative, implying that the distribution of daily returns is skewed to the left. The kurtosis coefficient, with a value close to 9, shows that the yields do not follow a normal distribution as this 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).

[8] The IBEX35 is the main reference stock market index of the Spanish stock market (https://www.bolsasymercados.es/esp/Home). It is an index weighted by stock market capitalization that is made up of the 35 most liquid companies listed on the Spanish stock market.

[9] Under the new regulation based on the Basel III solvency framework (BCBS, 2019), ES (α=0.975) replaces VaR (α=0.99). The comparison conducted in this paper is not intended to compare both measures in the new framework but to evaluate these risk measures in terms of accuracy and model risk. For that reason the confidence level chosen (α=0.975) is the same for ES and VaR.

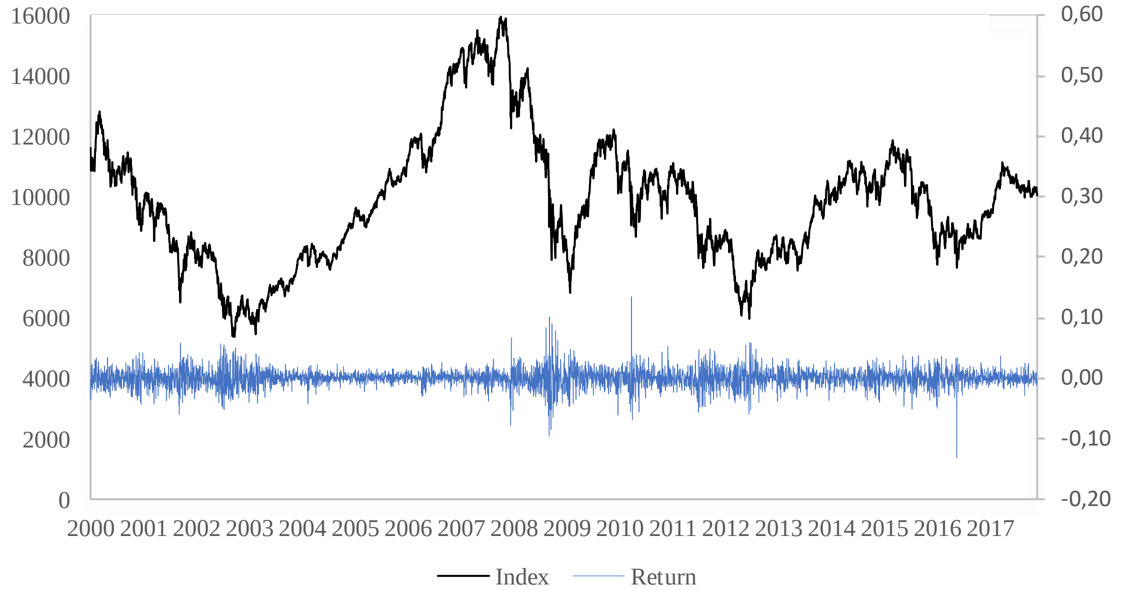

Figure 1. IBEX35

Table 3. Descriptive statistics of the daily returns

	Mean (%)	Median	Maximum	Minimum	Std.	Skewness	Kurtosis	Jarque-
IBEX35	-0.0032*	0.0621	0.1348	-0.1319	0.0149	-0.0796*	8.819*	6451

Note: (*) denotes significance at the 5% level.

Empirical results

Market risk estimations (VaR and ES)

In this section we analyse the market risk estimations that we have obtained for the IBEX35 returns from the models analysed. Figure 2 reports these estimations for each of the distributions (Normal, Student-t, GED, Skew Student-t and Skew GED) we have considered. Panel (a) reports VaR estimations join to the returns and Panel(b) reports ES estimations also join to the returns.

Un inspection visual of these figures suggest that the market risk of the IBEX35 is time-varying moving between -2% and -8% mostly. Apparently, no large differences are observed in the estimates obtained by each of the models used for their estimation. However, these differences exist and are revealed when some descriptive statistics of the estimates are analyzed. Table 4 reports the mean, median, standard deviation and minimum of the VaR estimations. Table 5 reports these statistics for the ES estimations.

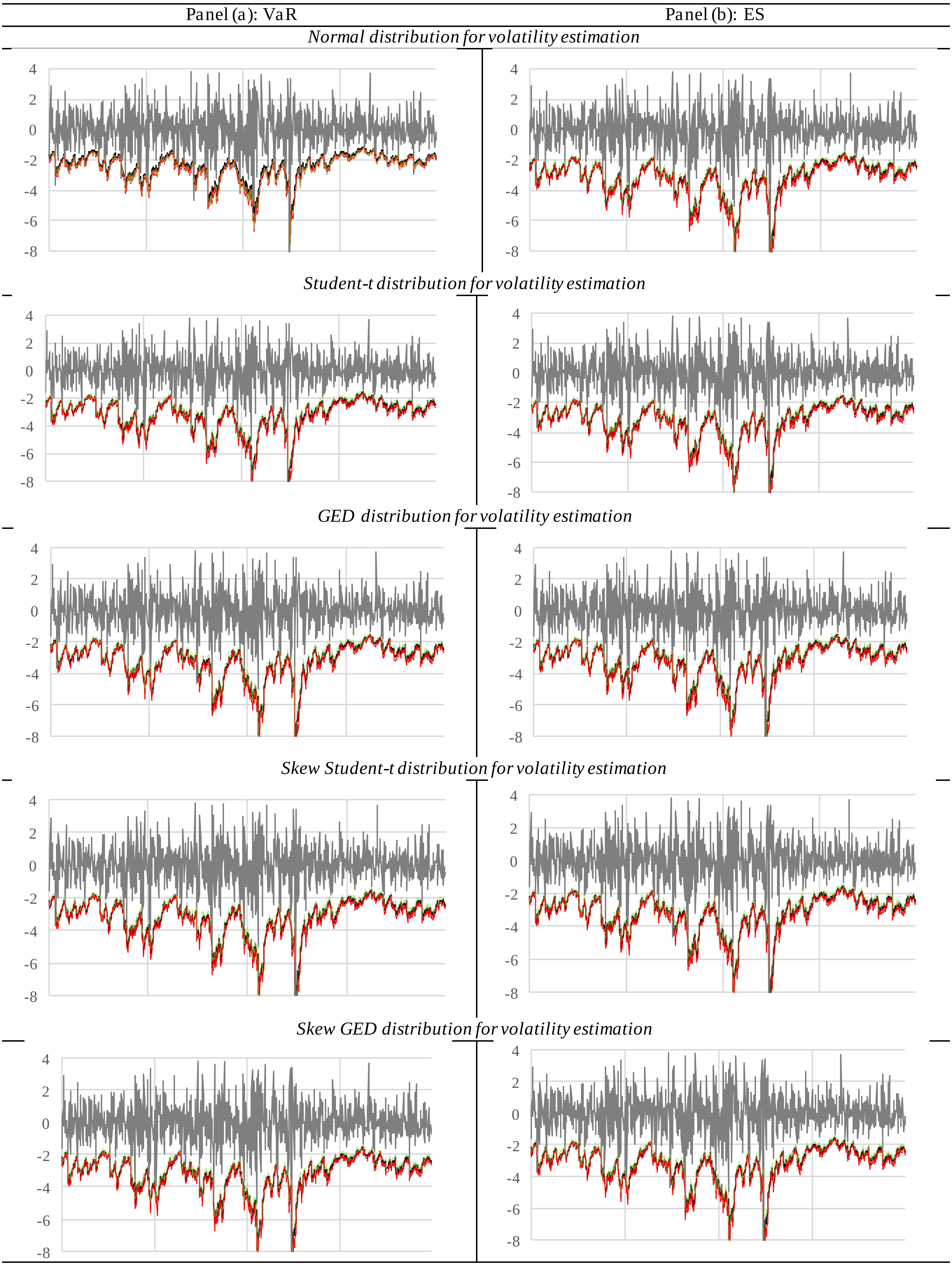

Figure 2. VaR and ES estimations below different models

Note: The table shows the IBEX-35 daily retuns (in grey) and VaR (Panel a) and ES (Panel b) forecasts from parametric model (in black), EVT model (in red) and FHS model (in green).

Table 4. Descriptive statistics of the VaR estimations

	PARAMETRIC				EVT				FHS			
	mean	median	std	min	mean	median	std	min	mean	median	std	min
Normal	-2.39	-2.19	0.85	-6.89	-2.60	-2.39	0.94	-7.49	-2.53	-2.32	0.91	-7.60
Student-t	-2.41	-2.21	0.88	-6.88	-2.60	-2.37	0.96	-7.45	-2.49	-2.30	0.89	-6.66
GED	-2.45	-2.25	0.89	-7.11	-2.60	-2.38	0.96	-7.54	-2.51	-2.30	0.91	-7.44
SSD	-2.55	-2.33	0.93	-7.32	-2.60	-2.37	0.95	-7.45	-2.52	-2.31	0.92	-7.46
Skew GED	-2.57	-2.37	0.90	-6.98	-2.60	-2.38	0.95	-7.54	-2.53	-2.30	0.92	-7.59

Table 5. Descriptive statistics of the ES estimations

	PARAMETRIC				EVT				FHS			
	mean	median	std	min	mean	median	std	min	mean	median	std	min
Normal	-2.85	-2.62	1.02	-8.22	-3.35	-3.09	1.20	-9.97	-2.99	-2.75	1.07	-8.67
Student-t	-3.09	-2.84	1.12	-8.90	-3.36	-3.09	1.22	-9.90	-2.94	-2.70	1.05	-7.86
GED	-3.05	-2.80	1.10	-8.94	-3.36	-3.09	1.22	-10.02	-2.97	-2.72	1.07	-8.63
SSD	-3.29	-3.02	1.19	-9.55	-3.36	-3.09	1.21	-9.90	-2.99	-2.73	1.08	-8.55
Skew GED	-3.18	-2.93	1.11	-8.69	-3.36	-3.09	1.21	-10.02	-2.99	-2.73	1.08	-8.73

The analysis of descriptive statistics reveals some interesting issues. First of all, we observed that the market risk estimations obtained with Parametric model are hight sensitive to the distribution assume for the returns. Normal distribution provides the lowest market risk estimacions while the skewed distributions provides the highest. For instance, in average, the VaR estimation below a Normal distribution is -2.39 becoming -2.57 when we assume a Skew GED distribution. These differences are still higher when we calculate the ES. In average, the ES estimation from a Normal distribution is -2,85, while the forescast losses below a Skew GED becomes -3.18%.

Unlike the Parametric method, in average terms the market risk estimations obtained from the method based on the EVT and FHS method do not depend much on the assumed distribution, although occasionally some differences can be observed.

Second, the variability of the estimated losses is also sensitive to the distribution assumed for the yields. Again, the normal provides the most stable risk estimates, followed by the Student-t and/or GED and finally the skewed distributions. Third, in all the cases analyzed, the mean is lower than the median, indicating that there are large losses in the left tail.

To last, we also observed large differences between the VaR estimates obtained under the parametric method and those obtained with EVT and FHS. Between these last two methods, however, the differences on average do not exceed 5 basis points. It should also be noted that in the case of the parametric method, the differences are reduced when we consider distributions with fat and asymmetric tails. In the case of the normal, the differences average almost 20 basic points.

In the case of the ES, the method based on the EVT provides and FHS provide market risk estimations very different. The former forecast losses around 3.3% while FHS forecast losses around 3,0%.

In the followings section we analyse the accurate of the market risk estimations and we use loss functions to discern which of the methods yields better estimates.

Backtesting results

In this section, the accuracy of risk measures is analysed. The risk measures, VaR and ES, were obtained from three different models: (i) parametric method and two semiparametric approaches: (ii) the aproach based on the conditional EVT and (iii) Filter Historical Simulation. In the case of parametric method, five distributions have been considered: Normal, Student-t (symmetric) (STD) distribution, the skewness Student-t distribution (SSTD), generalized error distribution (GED) and the skewness generalized error distribution (SGED) of Theodossiou (2001). All these methods require a forecast volatility to estimate portfolio market risk to which we use an APARCH model. This model has been estimated below different distribution: Normal, STD, GED, SSTD and SGED.

Thus for each method we have five risk measures each one coming from a different distribution used to model return volatility. These measures are obtained one day ahead at the 97.5% confidence level.

To evaluate the accuracy of VaR estimates, five tests have been applied: LRuc, BTC, LRind, LRcc and DQ. Table 6 reports the p-value of these tests, joint to the number and percentage of exceptions.

Table 6. Accuracy tests: VaR (α=97,5%)

	Normal	STD	GED	SSTD	SGED
			Panel (a) Parametric apporach		
Nº exceptions	**30**	**30**	**30**	**28**	**27**
% exceptions	**2.94**	**2.94**	**2.94**	**2.74**	**2.64**
LR_{uc}	0.57	0.57	0.57	0.75	0.85
BTC	0.27	0.27	0.27	0.35	0.38
LR_{ind}	0.37	0.37	0.37	0.41	0.42
LR_{cc}	0.57	0.57	0.57	0.67	0.71
DQ	0.07	0.07	0.07	0.27	0.23
			Panel (b): Conditional EVT		
Nº exceptions	**26**	**27**	**28**	**26**	**27**
% exceptions	**2.54**	**2.64**	**2.74**	**2.54**	**2.64**
LR_{uc}	0.95	0.85	0.75	0.95	0.85
BTC	0.40	0.38	0.35	0.40	0.38
LR_{ind}	0.44	0.42	0.41	0.44	0.42
LR_{cc}	0.74	0.71	0.67	0.74	0.71
DQ	0.49	0.16	0.28	0.13	0.24
			Panel (c) Filter Historical Simulation (FHS)		
Nº exceptions	**28**	**30**	**28**	**30**	**28**
% exceptions	**2.74**	**2.94**	**2.74**	**2.94**	**2.74**
LR_{uc}	0.75	0.57	0.75	0.57	0.75
BTC	0.35	0.27	0.35	0.27	0.35
LR_{ind}	0.41	0.37	0.41	0.37	0.41
LR_{cc}	0.67	0.57	0.67	0.57	0.67
DQ	0.27	0.09	0.16	0.07	0.26

Note: Table shows the p-value of the statistics: (i) the unconditional coverage test (LRuc); (ii) the back-testing criterion (BTC); (iii) statistics for serial independence (LRind); (iv) the Conditional Coverage test (LRcc) and (v) the Dynamic Quantile test (DQ).

The three methods underestimate risk as they provide a percentage of exceptions higher than it is expected (2,5%) but it is the approach based on the conditional extreme value theory which provide a percentage closer to the expected one.

According to accuracy tests, VaR estimates are accurate for all approaches regardeless the distribution assumed for modeling volatility. The null hypothesis has not been rejected by any test.

In a second stage, and after checking the accuracy of the VaR measurements, we have calculated the loss functions (Table 7) . The model that provides the lowest loss function value is the best. Figure 3 shows the loss functions for each model.

Thus, according to Lopez's loss function, the best model for estimating VaR measure is the based on the conditional extreme value theory (EVT). This result is obtained regardeless of the distribution assumed for modeling return volatility. In particular, the lowest value for loss functions is obtained for EVT under SSTD. Parametric model is the worst model for three distributions (Normal, STD and GED) and FHS has the worst behaviour for skewness distributions (SSTD and SGED).

Table 7. Loss functions for VaR

VaR		**Parametric**	**EVT**	**FHS**
Lopez's magnitude loss function (LF1)	Normal	160.11	**143.6**	152.05
	STD	158.03	**144.27**	150.75
	GED	156.37	**145.46**	152.01
	SSTD	148.36	**143.29**	154.19
	SGED	146.47	**144.4**	148.71
Lopez's lineal loss function (LF2)	Normal	29.61	**23.93**	26.09
	STD	29.00	**24.05**	26.09
	GED	28.08	**24.02**	26.54
	SSTD	25.44	**24.01**	26.21
	SGED	24.63	**23.94**	26.29
Caporini's loss function (LF3)	Normal	12.74	**9.35**	10.63
	STD	12.42	**9.47**	10.51
	GED	11.81	**9.43**	10.94
	SSTD	10.23	**9.43**	10.78
	SGED	9.68	**9.37**	10.50

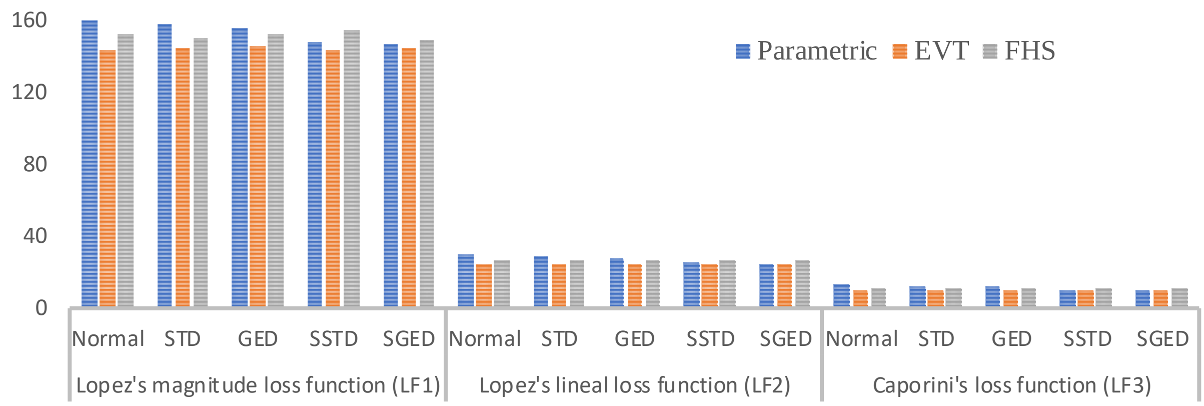

Figure 3. Loss functions for VaR

To test whether the ES estimations are correct, we use the test proposed by McNeil and Frey (2000). The results of this test are displayed in Table 8. According to this tests all approaches provide correct estimations of the ES measure as in no case we find evidence against the null hypothesis that says that average of the discrepancy measure is equal to zero.

Table 8. Backtesting for Expected Shortfall estimations

	ES (α=97,5%)				
	Normal	STD	GED	SSTD	SGED
Parametric	0,95	0,86	0,88	0,74	0,82
EVT	0,73	0,73	0,69	0,75	0,71
FHS	0,91	0,9	0,93	0,9	0,92

Note: Table shows the p-value of the McNeil and Frey test (2000).

Then, we have obtained the four loss functions presented in Subsection 2.3. Table 9 shows the results for each model.

Table 9. Loss functions for ES

ES		Parametric	EVT	FHS
Nieto and Ruiz loss function (LF4)	Normal	107,35	**91,77**	103,65
	STD	98,43	**90,93**	103,07
	GED	100,33	**91,39**	103,60
	SSTD	92,08	**90,99**	104,55
	SGED	95,76	**91,35**	101,34
López's magnitude loss function (LF5)	Normal	126,35	**98,77**	118,65
	STD	111,43	**97,93**	118,07
	GED	114,33	**98,39**	119,60
	SSTD	100,08	**97,99**	117,55
	SGED	104,76	**98,35**	116,34
Lopez's lineal loss function (LF6)	Normal	18,54	**13,20**	16,47
	STD	15,33	**13,17**	16,59
	GED	15,88	**13,18**	17,14
	SSTD	13,52	**13,17**	16,75
	SGED	14,22	**13,18**	16,71
Caporini's loss function (LF7)	Normal	6,49	**3,86**	5,55
	STD	4,91	**3,84**	5,60
	GED	5,18	**3,85**	5,76
	SSTD	4,00	**3,84**	5,74
	SGED	4,36	**3,85**	5,51

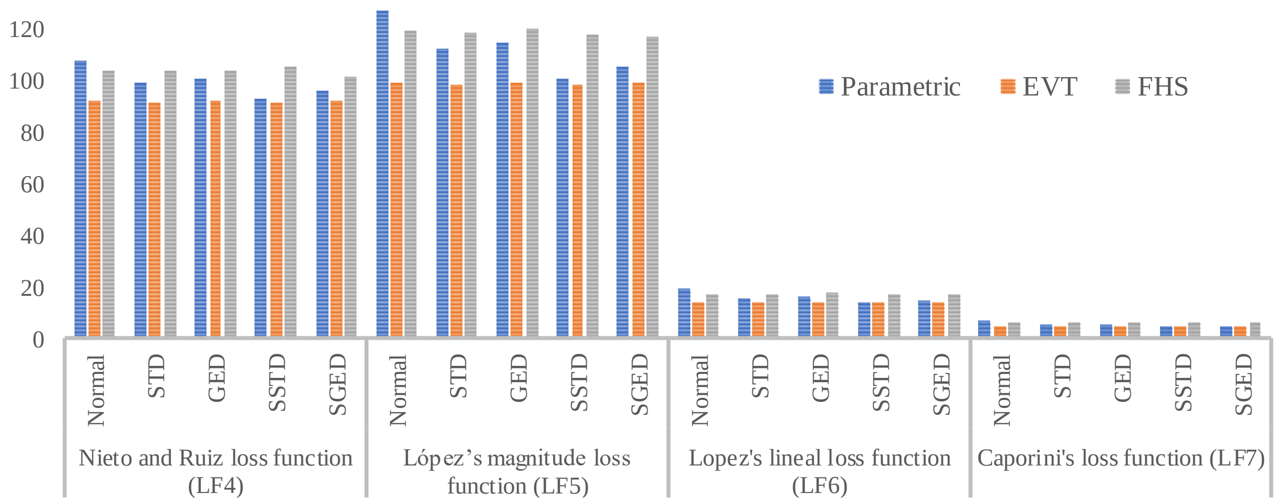

Figure 4. Loss functions for ES

Regarding the loss functions we find that the approach based in the conditional EVT model provide the lowest losses. This result is irrespective of the distribution assumed for modeling volatility. In particular, the lowest value for loss functions is obtained for EVT under STD. The worst model is FHS since it provides the highest values of the loss function for all distribution functions except for the normal distribution.

To sum up, in the analize period we find that according to the accurate test all the models (a total of fithteen) used to estimate market risk, given by VaR and ES measures, provide accurate risk estimations. But not all of them capture equally the tail risk. According to the loss function, the model that performed better in capturing tail risk is the approach based on the Extreme Value Theory, both in VaR estimation and ES estimation. This results is in line with the VaR literature which indicate that this model is the best performed in VaR estimate (Ergun and Jun, 2010; Ozun et al., 2010 and Tolikas et al., 2007). The novelty of our results is linked to the ES models due to the fact of the lack or very limited comparative papers in this field.

Conclusions

Under the new regulation based on Basel solvency framework (BCBS, 2012, 2016a, 2017a, 2019), known as Basel IV, financial institutions must calculate the market risk capital requirements based on the Expected Shortfall (ES) measure, replacing the Value at Risk (VaR) measure based on internal models, legitimized by the supervisory authorities since 1998 (BCBS, 1996; Hubbert, 2012; Szylar, 2014; Acerbi and Szekey, 2014). In the financial literature, there are many papers dedicated to compare VaR approaches but there are few studies focusing in comparing ES approaches.

Our study aims to cover this gap by carrying out a comprenhensive comparative of VaR and ES models. The methods included in the comparison are: (i) parametric approach; (ii) the approach based on the conditional EVT and (iii) Filter Historical Simulation. In the case of parametric method, five distribution have been considered: Normal, Student-t (symmetric) distribution (STD), the skewness student-t distribution (SSTD), generalized error distribution (GED) and the skewness generalized error distribution (SGED) of Theodossiou (2001). To estimate portfolio market risk a forecast volatility is required to which we use an APARCH model. This model has been estimated below different distribution: Gaussian, Student-t, GED, skew Student-t and skew GED. Thus, for each method we have five risk measure. Our objective is to carry out a systematic analysis that simultaneously considers different approaches and different distribution hypotheses for modeling volatility.

For this study we focus on the Spanish stock market, which has not been previously analysed in the literature on ES estimation models. The analysis period goes from January 3rd, 2014 to December 29th, 2017. The results show that all the models provide accurate risk estimations. But not all of them capture equally the tail risk. According to the loss function, the model that performed better in capturing tail risk is the approach based on the EVT, both in VaR estimation and ES estimation. This results is in line with the VaR literature which indicates that this model is the best performed in VaR estimate. The novelty of our results is in connection to the ES models as the comparative papers in this field are scarce. We also detect that the conditional EVT entails less risk model that Parametric method and FHS, as the former is less sensitive to the distribution assumed for modeling volatility. Finally, we find that in line with the literature, the VaR measure entails less model risk than the ES measure. This fact imply that there are more possibilities of regulatory arbitrage with this new measure than with VaR measure.

References

Abad, P., Muela, S. B., & Martín, C. L. (2015). The role of the loss function in value-at-risk comparisons. *The Journal of Risk Model Validation*, 9(1), 1. https://doi.org/10.21314/JRMV.2015.132

Abad, P., Benito, S., & López, C. (2014). A comprehensive review of Value at Risk methodologies. *The Spanish Review of Financial Economics*, 12(1), 15–32. https://doi.org/10.1016/j.srfe.2013.06.001

Acerbi, C., & Szekey, B. (2014). Backtesting Expected Shortfall - Introducing three model-independent, non-parametric backtest methodologies for Expected Shortfall. Working paper MSCI Inc. Available at https://www.msci.com/documents/10199/22aa9922-f874-4060-b77a-0f0e267a489b

Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, 26, 1487–1503. https://doi.org/10.1016/S0378-4266(02)00283-2

Angelidis, T., and Degiannakis, S. (2007). Backtesting VaR models: a two-stage procedure. *The Journal of Risk Model Validation* 1(2), 27–48. https://doi.org/10.21314/JRMV.2007.007

Artzner, P., Delbaen, F., Eber, J.M., & Heath, D. (1997). Thinking coherently. *Risk*, 10(11), 68–71.

Artzner, P., Delbaen, F., Eber, J.M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203–228. https://doi.org/10.1111/1467-9965.00068

Balkema, A., & de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2, 792–804.

Barone-Adesi, G., Giannopoulos, K., & Vosper, L. (1999). VaR without Correlations for Nonlinear Portfolios. *Journal of Futures Markets*, 19, 583-602.

Basak, S. y Saphiro, A. (2001). Value-at-Risk based management: Optimal policies and asset prices. *The Review of Financial Studies*, 14 (2), 371-405. https://doi.org/10.1093/rfs/14.2.371

Basel Committee on Banking Supervision (BCBS) (2020). Implementation of Basle standards. Basle, Switzerland: Bank for International Settlements (BIS). Available at: https://www.bis.org/bcbs/publ/d510.htm

Basel Committee on Banking Supervision (BCBS) (2019). Minimum capital requirements for market risk. Basel, Switzerland: Bank for International Settlements (BIS). Available at: https://www.bis.org/bcbs/publ/d457.pdf

Basel Committee on Banking Supervision (BCBS) (2017a). Basel III: Finalising post-crisis reforms. Basel, Switzerland: Bank for International Settlements (BIS). Available at: https://www.bis.org/bcbs/publ/d424.htm

Basel Committee on Banking Supervision (BCBS) (2017b). 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

Basel Committee on Banking Supervision (BCBS) (2016a). Minimum capital requirements for market. Basel, Switzerland: Bank for International Settlements (BIS). Available at: http://www.bis.org/bcbs/publ/d352.htm

Basel Committee on Banking Supervision (BCBS) (2016b). Explanatory note on the revised minimum capital requirements for market risk. Basel, Switzerland: Bank for International Settlements (BIS).

Basel Committee on Banking Supervision (BCBS) (2012). Fundamental review of the trading book: A revised market risk framework. Basel, Switzerland: Bank for International Settlements (BIS). Available at: https://www.bis.org/bcbs/publ/bcbs212.pdf

Basel Committee on Banking Supervision (BCBS) (1996). Amendment to the capital accord to incorporate market risks, Switzerland: Bank for International Settlements (BIS). Available at: https://www.bis.org/publ/bcbs24.pdf.

Bekiros, S., & Georgoutsos, D. (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 & Money*. 15 (3), 209-228. https://doi.org/10.1016/j.intfin.2004.05.002

Binder, Ch. & Lehner, O.(2020). The Problem of Heterogeneity within Risk Weights: Does Basel IV contain the Solution? . *ACRN Journal of Finance and Risk Perspectives*, 8, 183-205. https://doi.org/10.35944/jofrp.2019.8.1.012

Black, F. (1976). Studies in stock price volatility changes. In: Proceedings of the 1976 Business Meeting of the Business and Economics Statistics Section, American Association, 177–181.

Brooks, C., Clare, A., Dalle Molle, J., & Persand, G. (2005). A comparison of extreme value theory approaches for determining value at risk. *Journal of Empirical Finance*, 12, 339–352. https://doi.org/10.1016/j.jempfin.2004.01.004

Caporin, M. (2008). Evaluating value-at-risk measures in the presence of long memory conditional volatility. *The Journal of Risk*, 10(3), 79-110. Retrieved from https://www.proquest.com/scholarly-journals/evaluating-value-at-risk-measures-presence-long/docview/197276647/se-2

Chang, Ch. L.; Jiménez, J.A.; Maasoumi, E.; McAleer, M. & Perez-Amaral, T. (2019). Choosing expected shortfall over VaR in Basel III using stochastic dominance *International Review of Economics and Finance*, 60, 95-113. https://doi.org/10.1016/j.iref.2018.12.016

Chen, J.M. (2014). Measuring market risk under the basel accords: *VaR*, stressed *VaR*, and expected shortfall. *AESTIMATIO, the IEB International Journal of Finance*, 8, 184-201.

Christoffersen, P. (1998). Evaluating interval forecasting. *International Economic Review*, 39, 841-862. https://doi.org/10.2307/2527341

Clift, S.S., Costanzino, N., & Curran, M. (2016). Empirical performance of backtesting methods for expected shortfall. Available at SSRN: https://ssrn.com/abstract=2618345 or http://dx.doi.org/10.2139/ssrn.2618345

Colletaz, G., Hurlin, C., & Pérignon, C. (2013). The risk map: A new tool for validating risk models. *Journal of Banking & Finance*, 37(10),3843–3854. https://doi.org/10.1016/j.jbankfin.2013.06.006

Committee on the Global Financial System (CGFS) (2018). Structural changes in banking after the crisis. CGFS Papers nº 60, Bank for International Settlements, January.

Cont, R. (2006). Model uncertainty and its impacto on the pricing of derivative instruments. *Mathematical Finance*, 16, 519-547. https://doi.org/10.1111/j.1467-9965.2006.00281.x

Costanzino, N., & Curran, M. (2015). Backtesting general spectral risk measures with application to expected shortfall. *Journal of Risk Model Validation*, 9(1), 21–31. http://dx.doi.org/10.2139/ssrn.2514403

Costanzino, N., & Curran, M. (2018). A simple traffic light approach to backtesting expected shortfall. *Risks*, 6(2). https://doi.org/10.3390/risks6010002

Danielsson, J. & Zhou, C. (2017). Why risk is so hard to measure? Systemic Risk Centre Discussion Paper 36 (February), London School of Economics. http://eprints.lse.ac.uk/62002/1/dp-36.pdf

Danielsson, J., James, K., Valenzuela, M., & Zer, I. (2014). Model risk and the implications for risk management, macroprudential policy and financial regulations. VoxEU.org, 8. Avalilable at https://voxeu.org/article/model-risk-risk-measures-when-models-may-be-wrong

Danielsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O. & Shin, H. S. (2001). An Academic response to Basel II. Financial Markets Group (FGM) Special Paper N.º 130, London School of Economics.

Deloitte (2017). Model Risk Management. Driving the value in modelling. Risk Advisory, april. Available at https://www2.deloitte.com/content/dam/Deloitte/fr/Documents/risk/deloitte_model-risk-management_plaquette.pdf

Ding, Z., Granger, C. & Engle, R.F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1, 83-106. https://doi.org/10.1016/0927-5398(93)90006-D

Du, Z., & Escanciano, J. C. (2017). Backtesting expected shortfall: Accounting for tail risk. *Management Science*, 63(4), 901–1269. https://doi.org/10.1287/mnsc.2015.2342

Engle, R., & Manganelli, S. (2004). CaViaR: Conditional autoregressive Value at Risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4), 367-381. https://doi.org/10.1198/073500104000000370

Embrechts, P., Resnick, S., & Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 26, 30–41. https://doi.org/10.1080/10920277.1999.10595797

Emmer, S., Kratz, M., & Tasche, D. (2015). What is the Best Risk Measure in Practice? A comparison of Standard Measures. *Journal of Risk*, 18(2), 31-60. https://ssrn.com/abstract=2799732

Ergun, A., & Jun, J. (2010). Time-varying higher-order conditional moments and fore-casting intraday VaR and expected shortfall. *The Quarterly Review of Economics and Finance*, 50, 264–272. https://doi.org/10.1016/j.qref.2010.03.003

Feridun, M. & Ozün, A. (2020). Basel IV implementation: a review of the case of the European Union. *Journal of Capital Markets Studies 4(*1), 7-24. https://doi.org/10.1108/JCMS-04-2020-0006

Francq, C., & Zakoïan, J. M. (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics*, 184(1), 158-173. https://doi.org/10.1016/j.jeconom.2014.06.019

García-Jordano, L. (2017). Sample sizes, skewness and leverage effects in Value at Risk and Expected Shortfall estimation. PhD Doctoral Thesis. Available at: http://eprints.ucm.es/46253/1/T39548.pdf

Harmantzis, F.C., Miao, L., & Chien, Y. (2006). Emprirical study of value-at-risk and expected shortfall models with heavy tails. *The Journal of Risk Finance*, 7(2), 117-135. https://doi.org/10.1108/15265940610648571

Heres, A.R. (2017). Riesgo de modelo en la estimación del VaR y CVaR: Aplicación a carteras de renta variable y carteras de deuda. Master Thesis 017/013. Master en Banca y Finanzas Cuantitativas. Universidad Complutense de Madrid. Available at: https://www.uv.es/bfc/TFM2017/13%20Ana%20Regina%20Heres%20%20car%20.pdf

Hubbert, S. (2012). Essential mathematics for market risk management. (Vol. 642). John Wiley & Sons.

Jorion, P. (2001). Value at Risk: The New Benchmark for Managing Financial Risk. McGraw-Hill.

Kratz, M., Lok, Y.H., & McNeil, A.J. (2018). Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. *Journal of Banking & Finance*, 88, 393–407. https://doi.org/10.1016/j.jbankfin.2018.01.002

Kupiec, P. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *Journal of Derivatives*, 2, 73-84. Available at: https://ssrn.com/abstract=7065

Lincoviln, J.E. & Chiann, Ch. (2018). Testes para avaliação das previsões do value-at-risk e expected shortfall. *RBFin - Brazilian Review of Finance*, 17(4), 56-76.

Liu, H., & Kuntjoro, S. (2015). Measuring Risk with Expected Shortfall. Comparison of Expected Shortfall and Value at Risk. Master Thesis. Master´s Programme in Finance. Lund University. Available at: http://lup.lub.lu.se/student-papers/record/5470705

Lopez, J.A. (1998). Testing your risk tests. *Financial Survey*, 18–20.

Lopez, J.A. (1999). Methods for evaluating value-at-risk estimates. Federal Reserve Bank of San Francisco Economic Review 2, 3–17.

Louzis, D.P., Xanthopoulos-Sisinis, S., & Refenes, A.P. (2012). Stock index Value-at-Risk forecasting: A realized volatility extreme value therory approach. *Economics Bulletin*, 32(1), 981-991.

McNeil, A. (1998). Calculating Quantile Risk Measures for Financial Time Series Using Extreme Value Theory. Department of Mathematics, ETS. Swiss Federal Techni-cal University E-Collection.

McNeil, A.J. & Frey, R. (2000). Estimation Of Tail-Related Risk Measures For Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance*, 7, (3-4), 271–300. https://doi.org/10.1016/S0927-5398(00)00012-8

Mögel, B., & Auer, B.R. (2018). How accurate are modern Value-at-Risk estimators derived from extreme value theory? *Review of Quantitative Finance and Accounting*, Springer, 50(4), 979-1030. https://doi.org/10.1007/s11156-017-0652-y

Moldenhauer, F., & Pitera, M. (2017). Backtesting expected shortfall: is it really that hard?. Cornell University . Available at https://arxiv.org/abs/1709.01337

Nieto M.R., & Ruiz, E. (2008). Measuring financial risk: comparison of alternative procedures to estimate VaR and ES. Working Paper 08-73, Statistics and Econometrics Series, 26. Available at: https://e-archivo.uc3m.es/bitstream/handle/10016/3384/ws087326.pdf?sequence=1

Novales, A., & García-Jorcano (2019). Backtesting Extreme Value Theory Models of Expected Shortfall. *Quantitative Finance*, 19 (5), 799-825. https://doi.org/10.1080/14697688.2018.1535182

Orgeldinger, J. (2017). Critical Analysis of the New Basel Minimum Capital Requirements for Market Risk. *Italian Journal of Science & Engineering 1*(1). https://doi.org/10.28991/esj-2017-01111

Ozun, A., Cifter, A., & Yilmazer, S. (2010). Filtered extreme-value theory for value-at-riskestimation: evidence from Turkey. *Journal of Risk Finance*, 11, 164–179. https://doi.org/10.1108/15265941011025189

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. https://doi.org/10.1016/j.jeconom.2018.10.008

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119–131. https://doi.org/10.1214/aos/1176343003

Righi, M.B., & Ceretta, P.S. (2015). A comparison of Expected Shortfall estimation models. *Journal of Economics and Business*, 78, 14-47. https://doi.org/10.1016/j.jeconbus.2014.11.002

Rossignolo, A.F. (2019). Basel IV A gloomy future for Expected Shortfall risk models. Evidence from the Mexican Stock Market. *Revista Mexicana de Economía y Finanzas Nueva Época*, 14, 559-582. https://doi.org/10.21919/remef.v14i0.423

Sarma, M., Thomas, S., & Shah, A. (2003). Selection of Value-at-Risk models. *Journal of Forecasting*, 22(4), 337-358. https://doi.org/10.1002/for.868

Sobreira, N., & Louro, R. (2020). Evaluation of volatility models for forecasting Value-at-Risk and Expected Shortfall in the Portuguese stock market. *Finance Research Letters*. 32, 101098. https://doi.org/10.1016/j.frl.2019.01.010

Summinga-Sonagadu, R., & Narsoo, J. (2019). Risk Model Validation: An Intraday VaR and ES Approach Using the Multiplicative Component GARCH. *Risks*, 7(10). https://doi.org/10.3390/risks7010010

Szylar, C. (2014). Handbook on market risk. John Wiley & Sons, Hoboken.

Taylor, S. (1986). Modeling Financial Time Series. Chichester, UK: John Wiley and Sons.

Theodossiou, P. (2001). Skewness and Kurtosis in Financial Data and the Pricing of Options. Working Paper. Rutgers University.

Tolikas, K., Koulakiotis, A., & Brown, R. (2007). Extreme risk and value-at-risk in the German stock market. *European Journal of Finance*, 13, 373–395. https://doi.org/10.1080/13518470600763737

Wong, W. K. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking & Finance*, 32, 1404–1415. https://doi.org/10.1016/j.jbankfin.2007.11.012

Wong, W. K. (2010). Backtesting value-at-risk based on tail losses. *Journal of Empirical Finance*, 17(3), 526-538. https://doi.org/10.1016/j.jempfin.2009.11.004

Yamai, Y., & Yoshiba, T. (2002a). On the Validity of Value-at-Risk: Comparative Analyses with Expected Shortfall. *Monetary and Economic Studies*, 20(1), 57-86. Available at: https://ideas.repec.org/s/ime/imemes.html

Yamai, Y., & Yoshiba, T. (2002b). Comparative analyses of expected shortfall and Value-at-Risk: their estimation error, decomposition, and optimization. *Monetary and Economic Studies*, 20(1), 87-122. Available at https://ideas.repec.org/a/ime/imemes/v20y2002i1p87-121.html

Zhang, H. (2016). Market Risk Modeling Framework under Basel. *Commercial Banking Risk Management*, 35–52. https://doi.org/10.1057/978-1-137-59442-6_2

Žikovic, S. & Filer, R.K. (2012). Ranking of VaR and ES Models: Performance in Developed and Emerging Markets. *Czech Journal of Economics and Finance*, 63(4), 327-359. Available at: https://ideas.repec.org/s/ces/ceswps.html

Appendix: Extreme Value Theory

Extreme value theory (EVT) is a powerful and yet fairly robust framework in which to study the tail behaviour of a distribution. Even though EVT has previously found large applicability in climatology and hydrology, there is also a number of extreme value studies in the finance literature in recent years (Novales and Garcia-Jorcano, 2019; Mögel and Auer, 2018; Louzis et al., 2012; Brooks et al., 2005; Bekiros and Georgoutsos, 2005; Embrechts et al., 1999).

Within the EVT context, there are two approaches to study the extreme events. One of them is the direct modeling of the distribution of minimum or maximum realizations. The other one is modeling the exceedances of a particular threshold. This last method is called Peaks Over Threshold (POT). In the next lines we describe this approach.

In general, we are not only interested in the maximum of observations, but also in the behaviour 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 $y = X_t - u$.

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 $F_u(y) = P\big((X-u) \leq y | X > u\big)$ 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:

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

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

$$G_{\beta.\xi}(\mathrm{y}) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta} y\right)^{-\frac{1}{\xi}} & if\ \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\beta}\right) & if\ \ \xi = 0 \end{cases} \tag{2}$$

where ξ and $\beta > 0$ are the shape parameter and the scale parameter, respectively. 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 $F_u(\mathrm{y})$ 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

$$F(x) = 1 - \frac{N_u}{n}\left(1 + \frac{\xi}{\beta}(x-u)\right)^{-\frac{1}{\xi}} \tag{3}$$

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

$$q_\alpha = u + \frac{\beta}{\xi}\left(\left(\frac{n}{N_u}(1-\alpha)\right)^{-\xi} - 1\right) \tag{4}$$

The distributional choice is motivated by a theorem (Balkema & 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, appropriately normalized maximum sample converge 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 behaviour.