eman ta zabal zazu

Universidad del País Vasco | Euskal Herriko Unibertsitatea

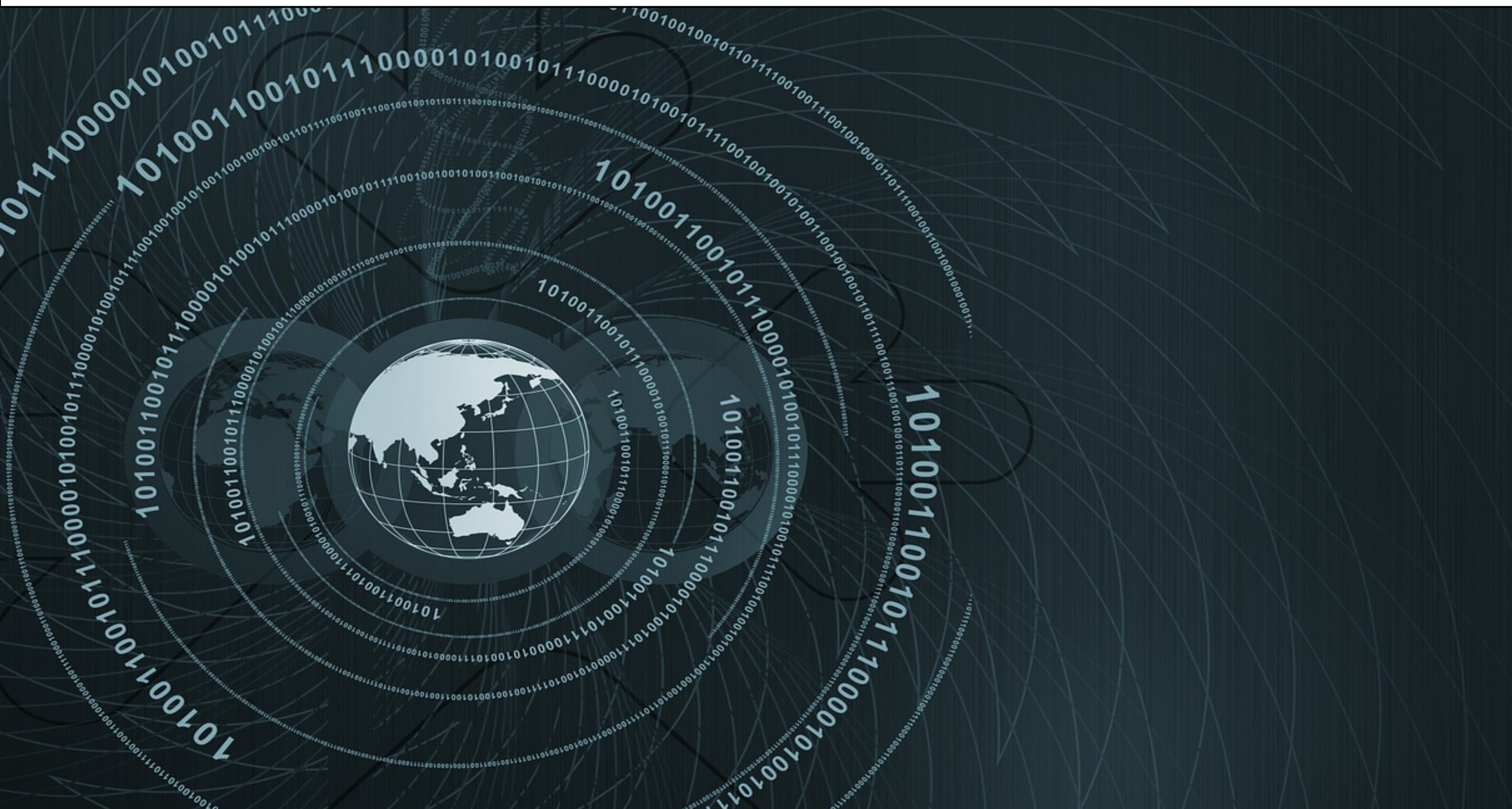
Doctoral Thesis

FOUR ESSAYS ON FINANCIAL RISK QUANTIFICATION

ANDRÉS MORA VALENCIA

Thesis Supervisors:
ANTONIO DÍAZ
GONZALO GARCÍA-DONATO

October, 2019



Doctoral Thesis

"FOUR ESSAYS ON FINANCIAL RISK QUANTIFICATION"

ANDRÉS MORA VALENCIA

Thesis Supervisors:

ANTONIO DÍAZ

GONZALO GARCÍA-DONATO

October, 2019

ACKNOWLEDGEMENTS

First of all, Thank God without Him I am nothing! Then, I would like to thank my thesis advisors, Professors Drs. Antonio Díaz and Gonzalo García-Donato for their invaluable guidance, patience, confidence and support.

I would want also to thank the professors of the Master and Ph.D Programme in Quantitative Finance and Economy for all the acquired knowledge.

I am also very thankful to Professors Drs. Esther B. Del Brío and Javier Perote for their support and helping me finishing one chapter of this thesis.

I am also grateful for conversations and suggestions of Professor Dr. Alfonso Novales, collaboration of Professors Dr. Miguel Angel Martínez, Susan Orbe and Angel Pardo.

This thesis is dedicated in the memoriam of my father Guillermo, to my mother Hilda and my brother Diego Fernando.

Contents

LIST OF TABLES

LIST OF FIGURES

Introduction of the Thesis

Financial risk is related to the potential loss that a financial institution or investor may incur due to adverse variations in financial variables (factors). The main financial risks are credit, market and operational risk. Credit risk may be defined as the potential loss that a borrower fails to meet its obligations (BCBS, 1999), whereas market risk is the potential loss resulting from adverse movements of market prices (BCBS, 2016), and operational risk is defined as the potential loss arising from inadequate or failed internal processes, people and systems or from external events (BCBS, 2011).

The adequate assessment of these risks is the base for regulatory and economic capital, i.e. the amount of money provisioned by financial institutions to buffer the potential losses. One of the main concerns of regulators is the procurement of financial stability, which can be threatened by financial crises. That is why the necessity to examine the statistical and mathematical properties of risk measures. Value-at-Risk (VaR) has been the standard risk measure in the financial industry during twenty years since its inception (in the so-called Basel I in 1996) and can be defined as the maximum loss (in "normal" market conditions) given a confidence level and time horizon. However, it is very well-known that VaR does not satisfy a desirable property for risk measures to be coherent, the so-called subadditivity or diversification property. Moreover, VaR can be seen as the minimum loss given the worst losses. For these reasons, this risk measure has been criticized for its inability to capture tail risk. A possible solution is provided with Expected Shortfall (ES) measure, which is quantified as the average losses given that losses have exceeded VaR, and ES is proven to be a coherent risk measure.

The aim of regulators is that financial institutions can raise their capital provisions after the events occurred in the global financial crisis (subprime and sovereign debt crises). To this end, the Financial Review of Trading Book (FRTB) was initially released in 2012 bringing challenges to the financial industry. With the new regulation, the Basel Committee proposed to switch from VaR to ES for market risk purposes and scrap Advanced Measurement Approach (AMA) for operational risk quantification. Besides the standard proposal of Basel Committee to estimate each type of risk, the regulator allows financial institutions the use of their own internal models. The acceptance of the latter models depends on validation procedures such as backtesting and stress testing processes. For VaR there are “relatively” simple methods to perform backtesting given that this risk measure satisfies the elicitability property, a desirable property to perform forecasting. ES is not elicitable by itself (Gneiting, 2011), but is jointly elicitable with VaR (Fissler and Ziegel, 2016) what makes the validation procedure of ES models not that straightforward as in the VaR case. Currently the regulator advises calculating ES as risk measure and perform VaR backtesting at 97.5% and 99% confidence levels. A barometer of financial system is the performance of the so-called global systemically important banks (G-Sibs). It is considered that a failure on one of these banks may trigger a financial crisis. In 2018, the US G-Sibs exhibited 11 VaR breaches in aggregate, a high amount of violations when it is expected 2.5 (6.25) exceptions for 99%-VaR (97.5%-VaR).[1]

According to the final FRTB framework published on January 14, 2019 the Basel Committee estimates a median increase of 16%, and a weighted average of 22% in market risk capital requirements compared with the previous Basel 2.5 framework based on an impact study carried out by the Committee with data of 2017. A similar study resulted in an increasing around 40% with the original FRTB version. The initial implementation date of the new rules was January 2019, but it was postponed until January 2022. Despite of the challenges above-mentioned, banks see more attractive to develop their own internal models rather than adopting the standardized approach. Though the latter approach is less complex it is more risk-sensitive than internal model approach (IMA), and it is expected an increase in average of 30% in capital requirements with respect to Basel 2.5 with the standard approach. A study in 2016 found market risk capital would increase by 1.5 times by utilizing IMA models and 2.4 times under the standardized approach.[2] However, a recent (and undisclosed) study of International Swaps and Derivatives Association (ISDA) suggests an increase of three times higher.[3]

Regarding operational risk, this type of risk is gaining relevance each time. By end of 2018, 32.4% of the risk-weighted assets (RWA) for the eight US G-Sibs corresponded to operational risk capital. Since Basel II the most predominant models are the Loss Distribution Approach (LDA) classified as one of the Advanced Measurement Approach

[1] “At US G-Sibs, 11 VAR breaches in 2018” report by Alessandro Aimone. Available at https://www.risk.net/risk-quantum/6412711/at-us-g-sibs-11-var-breaches-in-2018.
[2] “Can bankers stop the trading book killer?” report by Samuel Wilkes. Available at https://www.risk.net/our-take/6645301/can-bankers-stop-the-trading-book-killer
[3] “Revealed: FRTB impact three times higher than expected” report by Samuel Wilkes. Available at https://www.risk.net/regulation/6947536/revealed-frtb-impact-three-times-higher-than-expected

(AMA). The other two suggested approaches by the Basel Committee are the Basic Indicator Approach (BIA) and the Standardized Approach (TSA), which basically depends on the gross income of the financial institutions. Under LDA models, it is common to employ Extreme Value Theory (EVT), for modelling the tails, and copulas for risk aggregation to quantify VaR at 99.9% of confidence level of the whole entity. One of the main critiques of LDA models is that this is not a forward-looking risk measure. Moreover, by employing different distributions to fit the severities may lead different capital buffers with the same data. The new proposal of the Basel Committee is the Standardized Measurement Approach (SMA) that will replace the three current methodologies, and as in the case for the new rules for market risk the objective of the Committee is that financial institutions increase the capital buffer. The objective seems to be possibly achieved according to a recent publication of the European Banking Authority (EBA). The study found that total annual losses overshot operational risk capital in ten occasions with the current regulation. With the new proposal it would have been only three cases. The analysis was performed in 146 banks for three years (2015 – 2017).[4]

Given the above framework, more study is needed to provide regulators and practitioners the adequate tools to assess the risk measures and how to validate it. This Thesis aims to analyze the performance of VaR and recent proposals of validating ES with different distributional models for market risk. Furthermore, this project also reviews the advantages and drawbacks of the new guidelines for operational risk and proposes the Median Shortfall as a risk measure.

The first chapter "Risk quantification in turmoil markets" analyzes VaR backtesting by employing five distributions: the Gaussian distribution, the Student's t distribution, the generalized Pareto distribution (GPD), the α-stable distribution and the g-and-h distribution. The latter two distributions do not have a closed-form expression for its probability density function (pdf), then parameter estimation is challenging. We examine two estimation methods for stable distribution based on the characteristic function. One method is subject to Maximum Likelihood Estimation (MLE) and the other method employs regression (RegK). For g-and-h distribution, the popular method to estimate its

[4] "Policy Advice on the Basel III Reforms: Operational Risk". Available at https://eba.europa.eu/documents/10180/2886865/Policy+Advice+on+Basel+III+reforms+-+Operational+Risk.pdf

parameters is based on quantiles. In particular, we propose a robust regression to estimate the *h* parameter. The results show that GPD and α-stable distributions perform well for risk measurement purpose. A version of this chapter has been published in "Risk Management Journal" co-authored with Antonio Díaz and Gonzalo García-Donato.

The second chapter "Quantifying Risk in Traditional Energy and Sustainable Investments" aims to examine the ability of the recent proposal of Basel Committee of ES risk measure to appropriately quantify market risk in the stock returns. This proposal poses a challenge to academics and practitioners, due to the procedure to validate ES measures is still an open question. The data in the application section consists of portfolio of sustainable energy assets and another portfolio of traditional energy asset, more specifically the Dow Jones Sustainability Index the Bloomberg World Oil & Gas Index, respectively. To this end, we perform a novel backtesting procedure based on multinomial tests (Kratz et al., 2018) for different value-at-risk (VaR) levels rather than performing a binomial test for each VaR level, since it is proven that ES can be approximated as a weighted sum of different VaR levels (Emmer et al., 2015). The results show that fat-tailed distributions, e.g. stable and generalized Pareto distributions, outperform for both sustainable and traditional portfolios. A version of this chapter has been published in "Sustainability Journal" co-authored with Antonio Díaz and Gonzalo García-Donato.

The third chapter "Risk quantification for Commodity ETFs: Backtesting Value-at-Risk and Expected Shortfall" studies the risk assessment of alternative methods for a wide variety of Commodity ETFs divided in two parts. We implement well-known as well as and recently proposed backtesting techniques for both VaR and ES under extreme value theory (EVT), parametric, and semi-nonparametric techniques to univariate ETFs in the first part. The application of Gram-Charlier distribution to ES is introduced in this chapter and for this purpose we derive a straightforward closed form of ES. For the ES validation, we employ the method proposed by Fissler and Ziegel (2016) as an alternative of the proposal of Kratz et al. (2018) employed in the previous chapter. We show that, for the confidence levels recommended by Basel Accords, EVT and Gram-Charlier expansions have the best coverage test and skewed-t and Gram-Charlier the best relative performance. Hence, we recommend the application of the above-mentioned distributions to mitigate regulation concerns about global financial stability and commodities risk assessment. The second part expands the previous part from the univariate (Gram-Charlier distribution) to the multivariate case (semi-nonparametric - SNP distribution) and studies the risk assessment of ETF portfolios. We applied the SNP model with dynamic conditional correlations (DCC) and EGARCH models and implement the Fissler and Ziegel (2016) methodology to validate ES to commodity ETF portfolios formed by bivariate combinations of two metals (Gold and Silver ETFs) and energy (Oil ETF). Results support that multivariate SNP-DCC model outperforms multivariate normal distribution and provides accurate risk measures for commodity ETFs. A version of the univariate part of this chapter has been published in "International Review of Financial Analysis" and the multivariate part in "European Journal of Finance" co-authored with Esther B. Del Brío and Javier Perote.

Finally, the fourth chapter "A note on SMA vs. LDA-AMA: The dawning of a new regulation" summarizes the main research on Loss Measurement Approach (LDA), which suggests the estimation of 99.9%-VaR for operational risk. Furthermore, advantages and drawbacks of the Basel Committee proposal (SMA model) are detailed in this chapter. Moreover, it is suggested the estimation of 99.9%-Median Shortfall instead of the Standard Measurement Approach. A version of this chapter has been published in "Journal of Operational Risk".

Then, the idea of this thesis is to provide some analysis of validation results for the new proposal (ES risk measure). Since it is not clear how to validate the ES amounts for market risk proposes, the results of this document is important for risk managers, the Basel Committee, main financial regulators (the Federal Reserve, the UK's Prudential Regulation Authority, the European Banking Authority and Japan's Financial Services Agency), and local regulators. Finally, for operational risk purposes, the last chapter reviews the literature about the most common methodology to assess 99.9%-OpVaR which is the LDA technique. Moreover, it is proposed the Median Shortfall as risk measure when Loss Distribution Approach is employed to quantify provision for operational losses.

References

BCBS (1999). Principles for the Management of Credit Risk. Available at https://www.bis.org/publ/bcbs54.pdf.

BCBS (2016). Minimum capital requirements for market risk. Available at https://www.bis.org/bcbs/publ/d352.pdf.

BCBS (2011). Principles for the Sound Management of Operational Risk. Available at https://www.bis.org/publ/bcbs195.pdf.

Emmer, S., Kratz, M., Tasche, D., 2015, What is the best risk measure in practice?, Journal of Risk, 18, 31–60.

Fissler, T., Ziegel, J.F. 2016. Higher order elicitability and Osband's principle. Annals of Stastics, 44 (4), 1680-1707.

Gneiting, T. 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106, 746-762.

Kratz, M., Lok, Y., McNeil, A.J., 2018, Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. Journal of Banking & Finance 88(C), 393-407.

BCBS (1999). Principles for the Management of Credit Risk. Available at https://www.bis.org/publ/bcbs54.pdf.

BCBS (2016). Minimum capital requirements for market risk. Available at https://www.bis.org/bcbs/publ/d352.pdf.

BCBS (2011). Principles for the Sound Management of Operational Risk. Available at https://www.bis.org/publ/bcbs195.pdf.

Chapter 1. Risk quantification in turmoil markets[5]

Abstract

The aim of this paper is to examine the performance of the Value-at-Risk (VaR) measure under different distributional models in the highly demanding context of the recent financial crisis. This task is one of the main challenges of the financial industry. In addition to the normal and Student's t distributions, we analyze three distributions especially appropriate for capturing tail risk: the generalized Pareto distribution (GPD), the α-stable distribution and the g-and-h distribution. We also address the problem of efficiently estimating the parameters of these distributions. Our backtesting analysis shows that GPD and α-stable distributions perform well for this risk measurement purpose.

Keywords: Backesting VaR, g-and-h, alpha-stable, EVT-POT

1.1 Introduction of Chapter 1

Due to the recent financial crises (subprime and sovereign debt) and the recent plunge in oil prices, financial institutions are searching for better models to quantify investment risk. Value-at-Risk (VaR) is the standard risk measure in the financial industry, but regulatory entities have recently expressed concern about its inability to capture tail risk. These entities have proposed employing the Expected Shortfall (ES) model.[6] However, it is not clear how to evaluate the goodness of the ES risk model (backtesting), and financial institutions must continue to perform backtesting based on VaR models. In fact, the Fundamental Review of Trading Book (BIS 2016) states that the backtesting requirements remain based on the 1-day static VaR measure.[7]

---

[5] A version of this chapter has been published in Risk Management Journal, co-authored with Antonio Díaz and Gonzalo García-Donato.

[6] VaR is defined as the α quantile of a relevant profit and loss (P&L) distribution to assess the risk exposure of single investments and portfolios. The expected shortfall is the conditional expectation of loss given that the loss is beyond the VaR level.

[7] One of the key enhancements of the revised market risk framework (see BIS 2016) consists of a shift from VaR to an ES measure of risk under stress. The document comments that the use of ES will help ensure a more prudent capture of "tail risk" and capital adequacy during periods of significant financial market stress. ES must be computed daily for each trading desk using a 97.5th percentile, one-tailed confidence level. However, backtesting requirements are based on comparing each desk's 1-day static VaR measure

The normal distribution was initially suggested to fit financial asset returns, but the observed return distribution exhibits heavier tails than the normal distribution. For this reason, Student's t distribution has been employed as a benchmark model, according to the recent literature. However, these distributions are symmetric around the mean. To capture tail risk, the literature proposes several distributions. Among the studies that apply heavy tail distributions in measuring market risk, McNeil and Frey (2000) find that the VaR from a generalized Pareto distribution (GPD) performs well on several probability levels. Kiesel *et al*. (2003) find that results obtained by peak over threshold (POT) are similar to the empirical quantiles. Angelidis *et al*. (2004) compare the performance of normal, Student's t and generalized error distributions (GED) and show that leptokurtic distributions behave well. In a similar study, Bekiros and Georgoutsos (2005) show that GPD performs better than other distributions (normal, Student's-t and GED) to calculate VaR at 99.5%. Other examples of good performance for risk quantification obtained by extreme value theory (EVT) distributions are Byström (2004), Lauridsen (2000), Ourir and Snoussi (2012), and De Jesús *et al*. (2013), among others.

Combinations or variations of EVT distributions also perform well for risk quantification purposes. For instance, Chavez-Demoulin and McGill (2012) and Herrera and Schipp (2014) employ Hawkes processes with GPD. The authors find that this type of process provides an appropriate estimation of risk measures, although the classical POT model performs better than the Hawkes process in low-volatility periods. Santos *et al*. (2013) propose the Duration-based Peak over Threshold (DPOT) and Peaks over Random Threshold (PORT) methods. The results show that DPOT and PORT perform better than classical POT, according to the unconditional coverage test. Wei *et al*. (2013) compare a multifractal volatility model with several GARCH-type models. The backtesting results show that the proposed method performs better than GARCH models. Herrera and Shipp (2014) utilize self-exciting marked point processes and find that these models provide accurate risk measures.

Several studies show that the α-stable distribution performs better than other distributions for a number of purposes. Khindanova *et al*. (2001) conclude that the VaR-α-stable model performs better than the VaR-normal model. Marinelli *et al*. (2007) compare the VaR-α-stable and VaR-GPD models and find that the α-stable distribution

(calibrated to the most recent 12 months' data, equally weighted) at both the 97.5^{th} percentile and the 99^{th} percentile, using at least one year of current observations of the desk's one-day P&L.

is better to compute VaR at 95%; however, the models present similar results when the VaR at 99% is calculated. Güner *et al*. (2010) employ hedge fund returns and compare the out-of-sample results between the α-stable and Student's t distribution. The results show the superiority of the α-stable distribution. Rachev *et al*. (2010) (RRS hereafter) compare the α-stable distribution with the normal distribution, Student's t distribution and GPD, and the results again show that the α-stable distribution performs best. However, this result could be driven by the threshold choice.

Another alternative distribution considered in the literature is the skewed t. Kuester *et al*. (2006) compare the performance of the skewed-t distribution with the normal distribution, Student's t distribution and GPD. The results show that the GPD and skewed t distribution perform well. Stavroyiannis *et al*. (2012) analyze the skewed t and Pearson's type-IV distributions. They find that Pearson's type IV distribution is better than the skewed-t distribution when calculating VaR at high confidence levels. These results are in line with those of Grigoletto and Lisi (2009, 2011). Based on Zhang and Cheng (2005), Haas (2009) studies both Gaussian and Student's t mixtures. The author finds that mixtures of Student's t distributions perform better than Gaussian mixtures. On the other hand, Ausín et al. (2014) show how to perform Bayesian predictions of VaR. The authors develop a Bayesian semiparametric approach employing mixtures of Gaussian distributions with a prior Dirichlet process, and they find that their proposal allows for superior flexibility when capturing the stylized facts of financial returns.

Another flexible model with good backtesting performance is Gram-Charlier (GC) distribution, which is an expansion of normal distribution in terms of Hermite polynomials (see for instance Polanski and Stoja, 2010; Del Brio et al., 2014).

In this paper, we consider three distributions to appropriately capture tail risk, i.e., the α-stable, g-and-h and generalized Pareto distributions. Unfortunately, the parameter estimation for these distributions is challenging, particularly when the density is not analytically expressible. This is the case of α-stable and g-and-h distributions. The classical maximum likelihood method cannot be applied directly, and the numerical extraction of the density is necessary. In addition, the estimation process may be computationally expensive. We use these distributions to measure VaR and examine friendlier and more efficient estimation methods. Koutruvelis (1980) proposed a linear regression on the sample characteristic function to estimate the α-stable parameters, whereas the quantile method developed by Hoaglin (1985) is used to estimate the g-and-h parameters. In addition, GPD parameter estimates depend on the threshold selection and then its quantile quantification. There are some techniques for choosing the threshold, but there is no method for choosing the "optimal" threshold. This drawback is also examined in this paper. We illustrate the relevance of the threshold choice using the RRS (2010) paper as a benchmark. These authors fix the threshold at 1.02% based on a result obtained by Goldberg and Weinstein (2008). This value is far removed from the range of values proposed by the vast majority of studies in the literature.

In the process of estimating the three distributions proposed to capture tail risk, we use a two-step procedure (McNeil and Frey, 2000) using an ARMA(1,1)-GARCH(1,1) process. Most of the abovementioned empirical studies that apply the POT method employ this procedure. It is based on Diebold *et al*. (1998), which consists of filtering the financial asset returns by an AR(1)-GARCH(1,1) process in the first step. In the second step, the GPD is fitted to the standardized residuals. Jalal and Rockinger (2008) show that this technique presents very good forecasting properties when calculating VaR. In this line, Berkowitz and O'Brien (2002) find that a VaR-ARMA(1,1)-GARCH(1,1)-normal model is satisfactory to forecast risk measures for main US commercial banks. Pérignon *et al.* (2008) examine Canadian commercial bank data and obtain a similar result.

Our sample includes five indexes that proxy the portfolio investment of US stocks, European stocks expressed in US dollars, US equity options, foreign exchange rates, and commodities during the November 2006–December 2015 period. This period is chosen because it includes the financial market turmoil in 2007 and 2008, which led to the most severe financial crisis since the Great Depression. The crisis spread rapidly from financial markets to the real economy. Many large financial institutions in the United States and Europe were bailed out by national governments, and others failed. Part of this period was disastrous for the financial sector, and the risk management practices of financial institutions have been subjected to considerable criticism.

This paper aims at providing useful insight regarding the use of distributions that appropriately capture tail risk in measuring VaR. We empirically test whether the ARMA-GARCH-VaR model and the α-stable, g-and-h and generalized Pareto distributions succeeds in the backtesting analysis of different financial assets. We start by replicating the RRS (2010) analysis to compare the results for alternative ARMA-GARCH specifications. We show that the threshold choice is a key issue. In addition, we observe that the ARMA(1,1)-GARCH(1,1) is a sound model to capture the volatility process. The results of ARMA-GARCH-α-stable are consistent with RRS (2010) and suggest an adequate VaR prediction under the POT method.

We proceed to formally investigate which distributions improve the performance of the VaR measure using the backtesting technique. One novelty of this research is the comparison of the α-stable distribution with other parametric distributions. Few studies are devoted to analyzing this distribution from a market risk measurement perspective. Moreover, to our knowledge, this is the first work to include the g-and-h distribution in a backtesting framework. In addition, we propose a robust regression to estimate the *h* parameter in the g-and-h distribution case.

Our analysis has certain similarities with that provided by Del Brio et al. (2014). The authors examine the performance of heavy-tailed innovations applied in developed and emerging market indices. They show that GC and GPD exhibit good performance in the backtesting results. Although we also base the analysis on normal, Student's t (benchmark cases in the literature) and GPD, we consider other heavy-tailed distributions such as g-and-h and alpha-stable distributions applied to different financial assets. Additionally, we explore other estimation methods different from the classical MLE for distributions which the pdf is not expressible. Moreover, our study includes violation ratio, independence test and 95% confidence interval of the exceptions in the backtesting results.

The rest of the paper is organized as follows: Section 2 presents the models and the VaR methodology, Section 3 addresses the estimation methods for the distributions presented in the previous section, Section 4 analyzes the data and the results on VaR backtesting, and Section 5 concludes.

1.2 The ARMA-GARCH-VaR model

The most relevant stylized facts of daily stock returns (Granger and Ding 1995, Rydén *et al*. 1998, Cont 2001, McNeil *et al*. 2005, among others) are (1) an absence of a linear autocorrelation, (2) an unconditional (conditional, in some cases) distribution of returns exhibiting heavy tails, (3) profit and loss asymmetry, (4) volatility clustering, and (5) a slow decay of autocorrelation in absolute returns. Since the GARCH model is a flexible tool useful to capture the aforementioned properties (Teräsvirta and Zhao 2011), stock returns are filtered by a GARCH process, and the VaR measure is employed to quantify risk. Current risk measures are based on statistical measures that describe the conditional loss distribution, and the one most widely used in the financial industry is the VaR measure. VaR can be defined as the maximum loss that can be expected of a portfolio over a time horizon (1 day or 10 days for market risk and regulatory purposes) given a specific level, $100(1-\alpha)$ % probability (confidence level). Assume that R represents the returns random variable and has a cumulative distribution function (cdf) F. Then, VaR can be defined as

$$VaR_{1-\alpha}(R) = \inf\{r \in \mathbb{R}: F(r) \geq \alpha\}. \tag{1}$$

In other words, $VaR_{1-\alpha}(R) = F^{-1}(\alpha)$ is the α quantile of returns cdf. As it is commonly used, it is assumed that returns follow $R_{t+1} = \mu_{t+1} + \sigma_{t+1}Z_{t+1}$ (see, for instance, McNeil *et al*. 2005). Given probability level α, the VaR can be expressed as

$$VaR_{1-\alpha} = \mu_{t+1} + \sigma_{t+1}q_{\alpha}(Z_{t+1}), \tag{2}$$

where Z_{t+1} is the innovations process with G cdf centered at zero and has unit variance, and $q_{\alpha}(Z_{t+1})$ is the α quantile of Z_{t+1}. Let R_t be the log returns, and assume that R_t is a covariance stationary process. Then, the ARMA(1,1)-GARCH(1,1) process is given by

$$\mu_{t+1} = \theta_0 + \theta_1\mu_{t+1} + \theta_2\varepsilon_{t+1} + \varepsilon_{t+1},$$

$$Z_{t+1} = \varepsilon_{t+1}\sigma_{t+1}, \qquad \varepsilon_{t+1} \sim G(0,1),$$

$$\sigma_{t+1}^2 = \beta_0 + \beta_1\varepsilon_{t+1}^2 + \beta_2\sigma_t^2. \tag{3}$$

Conditional mean μ_{t+1} is modeled by an ARMA(1,1) process, and a GARCH(1,1) process is employed for conditional volatility σ_{t+1}. The ARMA(1,1)-GARCH(1,1) has been shown to be an appropriate model due to its forecasting abilities, among other GARCH-type models (Lunde and Hansen, 2005). The parameters are usually estimated by the ML method.[8]

[8] For a review of ARMA-GARCH models, see, for instance, Li et al. (2002). Examples and more details of ARMA-GARCH models applied in risk management can be found in McNeil et al. (2005) and Jondeau et al. (2007).

In this paper, the model presented in expression (3) is combined with expression (2) to quantify VaR as follows:

$$VaR_{1-\alpha} = \mu_{t+1} + \sigma_{t+1} q_\alpha(Z_{t+1}), \tag{4}$$

where $q_\alpha(Z_{t+1})$ is the α-quantile of Z_{t+1}.

1.1. Distributions

For the innovations distribution (*G* in equation 3), the following distributions are analyzed: (i) normal, (ii) Student's t, (iii) α–stable, (iv) generalized Pareto, and (v) g-and-h.

(i) Normal distribution

The density of a standard normal distribution is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{5}$$

(ii) Student's t distribution

$$t(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{n-2}\right)^{-\frac{\nu+1}{2}}, \tag{6}$$

where Γ represents the gamma function, and ν is the degree of the freedom parameter. In financial applications, ν is found to vary between 6 and 10 (Danielsson 2011, p. 41). When $\nu \to \infty$, Student's t distribution is approximated by a normal distribution.

(iii) Stable distribution

The stable distribution is usually described by its characteristic function because its probability density function (pdf) is not expressible. There are three cases for known closed-form expressions for their densities: the normal, Cauchy and Lévy distributions.

The characteristic function is given by

$$E\left[e^{i\theta X}\right] = \begin{cases} exp\left(-\sigma^\alpha |\theta|^\alpha \left(1 - i\beta(sign\theta)\tan\frac{\pi\alpha}{2}\right) + i\mu\theta\right) & \text{if } \alpha \neq 1, \\ exp\left(-\sigma|\theta| \left(1 + i\beta\frac{2}{\pi}(sign\theta)\ln|\theta|\right) + i\mu\theta\right) & \text{if } \alpha = 1, \end{cases} \tag{7}$$

where the sign function is defined as

$$sign\theta = \begin{cases} 1 & \text{if } \theta > 0, \\ 0 & \text{if } \theta = 0, \\ -1 & \text{if } \theta < 0. \end{cases} \tag{8}$$

The four parameters are index of stability (characteristic exponent) $\alpha \in (0,2]$, skewness parameter $\beta \in [-1,1]$, scale parameter $\sigma > 0$, and location parameter $\mu \in \mathbb{R}$.

In our applications, the standard α-stable is employed, i.e., $\mu = 0$ and $\sigma = 1$. When $\alpha = 2$ and $\beta = 0$, the α-stable distribution is the normal distribution, and its tail has a finite variance. The smaller the value of the index stability, the heavier the distribution tail. If $1 < \alpha < 2$, the mean is finite, but the distribution exhibits infinite variance; however, when $\alpha < 1$ the mean is not finite.

Since the density is not expressible in closed form, the parameter estimation is not straightforward. Usually, three methods are employed: quantile method, generalized method of moments (GMM), and maximum likelihood (ML). The quantile method was initially proposed by Fama and Roll (1971). McCulloch (1986) developed a technique based on five quantiles of a sample to estimate α and β without asymptotic bias, and it is restricted to $\alpha > 0.6$. The method of moments is applied to the empirical characteristic function, e.g., Carrasco and Florens (2000) introduce GMM for a continuum of moments conditions. On the other hand, numerical methods are utilized to extract the pdf when ML is applied. One method to approximate the pdf is the fast Fourier transform (FFT) approach (Mittnik *et al*. 1999), and the other is the direct integration method (Nolan 2001).

(iv) The generalized Pareto distribution (GPD)

The cdf of the GPD is given by

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi\, x/\beta)^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - \exp(-x/\beta), & \text{if } \xi = 0 \end{cases}, \tag{9}$$

where ξ is the shape parameter, and β is the scale parameter.

If $\xi \geq 0$, then $\beta > 0$ and $x \geq 0$. However, if $\xi < 0$, then $0 \leq x \leq -\beta/\xi$. When $\xi > 0$, the GPD is the Pareto distribution; when $\xi = 0$, it is the exponential distribution; and when $\xi > 0$, the distribution is the Pareto type II distribution. Heavy-tailed empirical distributions usually follow a GPD with positive shape parameter $\xi > 0$.

A useful result in EVT is the Pickand, Balkema and de Haan Theorem (see, for example, Embrechts *et al*. 1997). Given financial loss data, exceedances distribution F_u (losses above threshold u) can be approximated by a GPD when the threshold tends toward the endpoint of the distribution. That is, given distribution function F (which belongs to the maximum domain of attraction of the generalized extreme value, GEV distribution), exceedances distribution F_u converges to a GPD when threshold u is progressively increased.

According to McNeil *et al*. (2005), the VaR or α-quantile is obtained from

$$q_\alpha(Z) = u + \frac{\beta}{\xi}\left[\left(\frac{1-\alpha}{N_u/n}\right)^{-\xi} - 1\right], \tag{10}$$

where u is the chosen threshold, β and ξ are the scale and shape parameters, respectively, N_u is the threshold exceedances, and n is the sample size. Therefore, N_u/n is an empirical estimator for the excess distribution.

EVT is successfully applied to financial risk measures. There are two types of methods: block maxima and peaks over threshold (POT). The POT method is very useful to analyze extreme losses that exceed high threshold u. Because of its efficient use of data, POT is the technique most widely employed. This approach was introduced by Smith (1989), Davison and Smith (1990), and Leadbetter (1991). The selection of the threshold is a drawback of the POT method because there is a tradeoff between the bias and variance of the GPD parameter estimates. There are several alternatives to solve this drawback (see, for instance, Beirlant *et al*. 2004); however, there is no method for choosing the "optimal" threshold. Graphical techniques, such as the visualization of the mean excess plot, may lead to an imprecise selection of the threshold when the distribution is far from the GPD (Ghosh and Resnick 2010). Among others, Chavez-Demoulin (1999), McNeil and Frey (2000), and Araújo-Santos and Fraga-Alves (2013) suggest choosing the threshold at the 10th percentile of the standardized residuals of the log returns.

(v) g-and-h distribution

The g-and-h distribution was introduced by Tukey in 1977 and has been successfully applied in finance (see Dutta and Perry, 2006 and the references therein). Let

$Z \sim N(0,1)$ be a standard normal random variable. Random variable X is g-and-h distributed with g and h parameters ($g \neq 0$ and $h \in \mathbb{R}$) if

$$X = \frac{\exp(gZ)-1}{g} \exp(hZ^2/2), \tag{11}$$

and it is $X \sim$g-and-h denoted. The g parameter controls the amount and asymmetry direction, while the h parameter accounts for kurtosis. The larger the value of g, the more skewed the distribution, while the distribution exhibits more elongation at higher values of h.

Several distributions can be obtained by varying the values of the parameters. A special case of the g-and-h distribution is the normal distribution when $g = h = 0$. Martinez (1981), Martinez and Iglewicz (1984), Hoaglin (1985), Dutta and Babbel (2002), and others have studied the properties of the g-and-h distribution. The pth percentile of g (g_p) is given by $g_p = -\left(\frac{1}{Z_p}\right) \log\left(\frac{X_{1-p} - X_{0.5}}{X_{0.5} - X_p}\right)$, and it is common to estimate g as the median of different values of g_p. Another property is that given a certain value g, the h value is given by $\log \frac{g(X_p - X_{1-p})}{\exp(gZ_p) - \exp(-gZ_p)} = h(Z^2/2)$. Then, an estimate for h can be obtained as the slope of the regression $\log \frac{g(X_p - X_{1-p})}{\exp(gZ_p) - \exp(-gZ_p)}$ vs. $(Z^2/2)$.

In this paper, a robust regression is proposed to estimate the h parameter. An important result of Degen *et al*. (2007) is concerned with EVT and the g-and-h distribution. The authors find that high quantile estimation by the POT method is generally inaccurate if the data follow a g-and-h distribution. For this reason, the g-and-h distribution is also included in the backtesting. The quantile estimation is based on Degen *et al*. (2007, Section 2.1). Since $k(x) = \frac{\exp(gx)-1}{g} \exp(hx^2/2)$ is strictly increasing (for $h > 0$), the cdf of a g-and-h random variable X can be written as

$$F(x) = \Phi\left(k^{-1}(x)\right), \tag{12}$$

where Φ denotes the standard normal cdf. Then, the quantile is given by

$$q_\alpha = k\left(\Phi^{-1}(\alpha)\right). \tag{13}$$

1.3 Estimation methods

The ML method is generally employed to estimate the parameters when the distribution presents a closed-form expression for the pdf. Various numerical methods exist to maximize the (log) likelihood function; however, it can exhibit several local maxima or be relatively flat. Moreover, several distributions do not have an analytical expression for the pdf, e.g., the α-stable and g-and-h distributions. In some cases, a numerical extraction of the likelihood function is possible, but it may be computationally expensive. The financial returns are filtered by an ARMA(1,1)-GARCH(1,1) process (Berkowitz and O'Brien, 2002; Güner et al., 2010), and its parameters are estimated by the ML method for the normal and Student's t distributions. A two-step procedure is applied for the general Pareto, α-stable and g-and-h distributions. In the first step, the returns are filtered using QML, and then, the distributions are fitted to the standardized residuals from the first step. This procedure has been shown to provide accurate results (see, e.g., Ergen, 2015).

1.4 Estimation method for the normal and Student's t distributions

The likelihood function is given by

$$L(\Theta; R) = \prod_{t=1}^{n} \frac{1}{\sigma_t} g\left(\frac{r_t - \mu_t}{\sigma_t}\right). \tag{14}$$

where σ_t follows a GARCH(1,1) process and μ_t an ARMA(1,1) process, Θ represents the parameter set to be estimated, and g is the distribution of the disturbances (normal or Student's t, in this case).

To obtain the estimations, the first-order conditions for the (log) likelihood functions are solved using numerical algorithms (usually a Newton-Raphson modification such as BFGS or BHHH methods).

1.5 Estimation method for the GPD

To estimate the GPD parameters through maximum likelihood estimation (MLE), the loglikelihood function is given by

$$\log L(Z; \xi, \beta) = -N_u \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{N_u} \log\left(1 + \xi \frac{Z_j}{\beta}\right), \tag{15}$$

where Z is the standardized residuals from (3), N_u is the number of exceedances over specific threshold u, and β and ξ are the scale and shape parameters, respectively.

As mentioned, literature usually uses a threshold u, which corresponds to the 10% of the left tail of the standardized residual distribution.

1.6 Estimation method for the α-stable distribution

To estimate the α-stable parameters, two methods are employed. One is based on MLE (henceforth α-stable-ML), where the pdf is estimated from the characteristic function using direct integration, following Nolan (2001). Another method, proposed by Koutrouvelis (1980), is also employed. The latter is based on regression (henceforth α-stable-RegK) over the sample characteristic function, $\hat{\phi}_t = \frac{1}{n}\sum_{j=1}^{n} \exp\{itx_j\}$. Because of the law of large numbers, the sample characteristic function is a consistent estimator for the characteristic function (Čižek et al., 2011, p. 31). The α-stable characteristic function obtains

$$\log(-\log|\ \phi(t)\ |^2) = \log(2\sigma^\alpha) + \alpha \log(t). \tag{16}$$

As expression (16) depends only on α and σ, a regression is proposed to estimate these parameters. The regression equation is $y_k = m + \alpha\omega_k + \varepsilon_k$, where $y = \log(-\log|\ \phi(t)\ |^2)$, $m = \log(2\sigma^\alpha)$, $\omega = \log(t)$, and ε_k is the error term. Koutrouvelis (1980) proposes $t_k = \frac{\pi k}{25}$, where $k = 1, \ldots, K$, and K ranges between 9 and 134 for different α values and sample sizes. Once α and σ have been estimated, μ and β can be estimated. The real and imaginary part of the characteristic function are given by (for $\alpha \neq 0$)

$$\mathrm{Re}\phi(t) = \exp(-|\ \sigma t\ |^\alpha)\cos\left[\mu t + |\ \sigma t\ |^\alpha \beta \mathrm{sgn}(t) \tan\frac{\pi\alpha}{2}\right], \tag{17}$$

$$\mathrm{Im}\phi(t) = \exp(-|\ \sigma t\ |^\alpha)\sin\left[\mu t + |\ \sigma t\ |^\alpha \beta \mathrm{sgn}(t) \tan\frac{\pi\alpha}{2}\right], \tag{18}$$

Then, (17) and (18) can obtain

$$\arctan\frac{\mathrm{Im}\phi(t)}{\mathrm{Re}\phi(t)} = \mu t + \beta\sigma^\alpha \tan\frac{\pi\alpha}{2} sgn(t)|\ \sigma t\ |^\alpha. \tag{19}$$

Another regression is run to obtain μ and β estimates. The regressions are repeated, with $\widehat{\alpha}, \widehat{\beta}$, $\hat{\sigma}$ and $\hat{\mu}$ as the initial conditions, until a convergence criterion is satisfied. The 25% of the truncated mean is employed as the initial estimate for μ, and α is obtained from Fama and Roll (1971), as suggested by Koutrouvelis (1980).

1.7 Estimation method for the g-and-h distribution

To estimate the g parameter, the methodology proposed by Hoaglin (1985) is utilized in our applications. This methodology is based on quantiles and compared with other estimation methods (such as ML and method of moments). The method based on quantiles is probably more accurate, fitting the tails of the distribution, and it is adequate for the g-and-h distribution, which is a transformation of the standard normal distribution. Other methods have been proposed to estimate the parameters, numerical estimations of the likelihood function (Rayner and MacGillivray, 2002) and Bayesian computations (Haynes and Mengersen, 2005).

Our paper estimates the h parameter as the slope of a robust regression $\log \frac{g(X_p - X_{1-p})}{\exp(gZ_p) - \exp(-gZ_p)}$ vs $Z_p^2/2$ and not the classical linear regression.

1.4 Backtesting in turmoil markets

In this section, we assess the VaR performance using backtesting applied to three sample data and the abovementioned estimation methods for the five distributions: normal, Student's t, GPD, α-stable and g-and-h.

As input, we use returns of a US stock exchange index (the Dow Jones Industrial Average, DJIA), an European stock index (the DJ EuroStoxx 50 Index) expressed in US dollars, US equity market volatility (the Chicago Board Options Exchange Volatility Index, CBOE VIX), which can be interpreted as a proxy of an equity option portfolio,[9] the US dollar spot exchange rate versus the Swiss Franc (CHF/USD), and a commodity

[9] The VIX or S&P 500 option implies volatility is usually used as an expected future volatility measure or an investor "fear gauge" index. The volatility is the main determinant of the price of an equity option. The higher the volatility of the underlying asset (the equity market), the higher the price of the option.

index investment (the Bloomberg Commodity Index).[10] We consider alternative investments of a typical financial institution. These five indexes replicate diversified portfolios on financial assets. Most financial institutions invest in portfolios of US stocks, European stocks, US equity options, foreign currencies, and commodities.

The sample period includes one of the worst global financial crises since the 1930s. This was not only a financial crisis but an economic crisis. The bailouts and defaults of some financial institutions focused policy makers' attention on the risk management practices of financial institutions. New Basel III rules (sometimes called Basel IV rules) emerged from this period of stress. Thus, our sample period seems to be especially appropriate to compute tail risk.

1.5 Backtesting

Backtesting is the most common technique utilized by the financial industry to validate internal VaR models. The idea is to calculate the number of times the actual losses have exceeded the estimated VaRs. These exceedances are called violations or exceptions. It is expected that the number of exceptions is approximately 1% of cases when a 99% VaR is calculated. If the percentage of exceptions is higher (lower) than 1%, then the VaR model underestimates (overestimates) risk. The one-day-ahead VaR is calculated by implementing a rolling window, usually of 250 or 500 observations. The backtesting is based on the assumptions that the number of exceptions is generated by an i.i.d. Bernoulli process. The random variable (the number of exceptions, X) is 1 if the actual loss is greater than the predicted VaR; otherwise, it is 0. Then, the indicator function is

$$I_{1-\alpha,t+1} = \begin{cases} 1, & if\ r_{t+1} < VaR_{1-\alpha,t+1} \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

If backtesting period T is very large, X follows a normal distribution, and a 95% confidence interval for the expected number of exceptions is given by

$$E(X) = T\alpha,$$

[10] The most popular commodity investment strategy is to invest in a basket of commodities in a given commodity index. In fact, a number of market participants and policy-makers contend that commodity index investment was a major driver of the 2007-2008 spike in commodity futures prices.

$$var(X) = T\alpha(1 - \alpha). \tag{21}$$

The test of unconditional coverage (Kupiec, 1995) enables testing whether the realized deviation rate from the VaR is in line with the confidence interval. Additionally, Christoffersen (1998) develops a conditional coverage test that represents an incorporated test of the hypothesis of unconditional coverage and independence. The author tests for bunching, i.e., a greater probability of deviation from VaR happening after a previous deviation from VaR (hit sequences).

1.6 Capturing the volatility process

To examine the performance of different methodologies in the backtesting and illustrate how relevant the threshold choice in EVT is, we perform a comparison of different alternatives using the RRS (2010) paper as a benchmark. These authors compare three fat-tailed methodologies, i.e., stable Paretian, Student's t with ν = 5 (degrees of freedom) and EVT, alongside the normal distribution model, in a backtesting exercise. The models are filtered for auto-regression and volatility clustering based on ARMA-GARCH. In this section, we replicate the RRS (2010) paper using the same distributions and almost the same dataset, but we consider five alternative methods: the ARMA-GARCH-normal; the ARMA-GARCH-t with fixed three degrees of freedom (ν = 3) and a changing number of degrees of freedom estimated in each step of the rolling window; the ARMA-GARCH-POT; the ARMA-GARCH-α-stable estimated by two different methods; and the ARMA-GARCH-g-and-h.

In addition to the additional models we fit, we illustrate the relevance of the parameter choice in the cases of Student's t and POT. These authors employ a strange value for the threshold in the latter model (u = 1.02%). Goldberg and Weinstein (2008) suggest this threshold. These authors use a particular method to select this value from a specific set of data. However, there is no a commonly accepted threshold selection method. Most of the literature proposes a threshold value equivalent to 5% or 10% of the left tail distribution. This range of values with an appropriate data size yields accurate parameter estimates of the GPD. Thus, we prefer to use the standard value proposed by the literature, i.e., u = 10%.

We also try to find evidence about the method that allows the best adjustment to the market conditions during the sample period, providing the most realistic VaR forecasts with exceedances within the confidence interval. Additionally, this preliminary comparison allows us to fix the value of the threshold (u) in the ARMA-GARCH-POT that we use in the next section.

The dataset comprises 1,671 Dow Jones Industrial Average (DJIA) Index prices from August 2003 to December 2009. The log returns are calculated as $r_t = 100\log(P_t/P_{t-1})$, where P_t is the price at time t, for a total of 1,670 observations. Table 1 presents the main descriptive statistics. As expected, the statistics show that the DJIA log returns exhibit heavier tails (excess kurtosis is 12.37) than the normal distribution.

Table 1: Descriptive statistics of DJIA returns

Statistics	**DJIA**
Mean	0.0080
Median	0.0298
Standard deviation	1.2476
Variance	1.5564
Skewness	0.0351
Excess kurtosis	12.3739
Minimum	−8.2005
Maximum	10.5083

The table presents descriptive statistics for the daily logarithmic returns of the DJIA index during the August 2003–December 2009 period.

The parameters of the α-stable distribution are estimated using both ML (α-stable-ML) and the Koutrouvelis method (α-stable-RegK). In the ML case, the pdf is obtained by direct integration of the characteristic function. In the estimation of the POT method, the two-step procedure proposed by McNeil and Frey (2000) is applied. The threshold (u) is fixed to the 10% percentile of the standardized residuals. The threshold selection and its consequences in estimating the GPD parameters are drawbacks of this method. Though there exist visualization methods to choose the threshold (Hill plot and mean excess plot), these methods have problems, and there is no "optimal" method to choose the threshold. This remains an open question in EVT. For that reason, we suggest choosing $u = 10\%$

(as in other EVT studies) on the left-tail distribution; this allows an adequate VaR prediction under the POT method, according to our results.

Table 2 compares the results of the backtesting conducted by RRS (2010) (left column) with the results for the five proposed models in our paper (rest of the columns). Although RRS (2010) change the length of the rolling window from 500 to 3,000 days in the POT model, we use a 500-day rolling window in all models. Therefore, the backtesting period is 1,170 days, and the expected number of exceptions is 11.70 when calculating 99% VaR. The standard error is 3.40, and the 95% confidence interval is (5.03, 18.37).

In general, we obtain a number of exceptions for the ARMA-GARCH-normal distribution, similarly to RRS (2010). Also consistent with RRS, the estimated exceptions for the ARMA-GARCH-α-stable distribution is the closest to the expected number of violations. However, there are some differences from other models. Although the number of exceptions for the ARMA-GARCH-t distribution are seemly similar, RRS (2010) employ $\nu = 5$, while our study uses $\nu = 3$. Nevertheless, when the degrees of freedom parameter is estimated in each step of the rolling window, we obtain 15 exceptions, which is within the 95% confidence interval [5, 18].

For the POT method, RRS (2010) employ an unusual threshold, $u = 1.02\%$, and a different moving window of 3,000 observations. As an alternative, we choose a threshold $u = 10\%$ and maintain a consistent size, i.e., 500 observations, for the moving window for all analyzed models. While RRS (2010) reveal a poor performance of EVT, our study obtains 14 exceptions, which lays in the 95% confidence interval [5, 18]. This simple model comparison shows that the results of RRS (2010) may have been caused by a poor threshold and rolling days. Therefore, the 10^{th} percentile of the data is employed in the following application.

The right side of Table 2 presents the violation ratio (VR) and the statistics and p-values for the unconditional coverage (UC) and independence (ind) tests. The VR is calculated as the ratio between the violations obtained by each VaR model and the expected violations. The best model is the one in which the VR is closest to 1, according to the expected number of exceptions. If the VR is higher (lower) than 1, the model underestimates (overestimates) risk. The VR is formally tested by the UC test, and the VaR model is correct according to the exceptions ratio under the null hypothesis. Though a VaR model may be adequate in the sense of the coverage test, the exceptions may occur in successive days. This could cause the bankruptcy of a financial institution with a greater likelihood than if the exceptions happened independently. For this reason, an independence hypothesis test is also performed. The null hypothesis is the independent occurrence of the exceptions.[11]

[11] More details about these hypothesis tests can be found in Christoffersen (1998 and 2003) and Danielsson (2011).

Table 2: Backtesting results for DJIA log returns: # Exceptions, Violation ratio and Unconditional Coverage and Independence Tests

	RRS (2010)	Our analysis					
Model	Exceptions	Exceptions	*VR*	LR_{uc}	p-value	LR_{ind}	p-value
ARMA-GARCH-Normal	20	27	2.3076	14.7603	0.0001	1.2768	0.2585
ARMA-GARCH-t:							
· fixed ν	3 {ν = 5}	4 {ν = 3}	0.3419	6.8647	0.0088	0.0274	0.8684
· changing ν	-	15	1.2820	0.8632	0.3528	0.3899	0.5323
ARMA-GARCH-POT	1 {*u* =1.02% and 3,000 days rolling window}	14 {*u* =10%}	1.1966	0.4296	0.5121	0.3394	0.5602
ARMA-GARCH- α-stable:							
· ML	11	12	1.0256	0.0077	0.9300	0.2489	0.6178
· RegK	-	10	0.8547	0.2624	0.6085	0.1726	0.6778
ARMA-GARCH-g-and-h	-	8	0.6837	1.3294	0.2489	0.1102	0.7398
Backtesting period	1,130 days	1,170 days					
95% confidence interval for the number of exceptions	[4, 17]	[5, 18]					

The left column of the table replicates results shown by RRS (2010) in their Figure 6. The footnote of the figure says: "*Note: the backtesting period is 1,130 days and the 95% confidence interval for the number of exceedances is [4, 17]*". The rest of the columns show our own results: the number of exceptions, the violation ratio (*VR*), and the likelihood ratios for both the unconditional coverage test (LR_{uc}) and the independence part of the conditional coverage hypothesis (LR_{ind}). Both tests are proposed by Christoffersen (1998). RRS (2010) employ ν=5 to estimate ARMA-GARCH-t, while our study uses ν=3. Alternatively, we also evaluate ARMA-GARCH-t by estimating the degrees of freedom parameter in each step of the rolling window. α-stable-ML indicates that the parameters have been estimated using ML (the pdf is obtained by direct integration of the characteristic function), and α-stable-RegK uses the Koutrouvelis method to estimate the parameters. Dataset for our estimates: daily logarithmic returns of the DJIA index during the backtesting period July 2005 to December 2009 (1,170 days) obtained from 500-day rolling windows.

Table 3: Descriptive Statistics for CHF/USD, EuroStoxx50, VIX and Commodity Index returns

Statistics	CHF/USD	EuroStoxx50	VIX	BCOM
Mean	0.0063	-0.0111	-0.0340	-0.0323
Median	-0.0131	0.0000	0.3044	0.0000
Standard deviation	0.6449	1.5278	7.2746	1.1176
Variance	0.4159	2.3344	52.9201	1.2490
Skewness	-0.0991	0.0257	-0.7431	-0.3296
Kurtosis	4.9506	8.2881	6.8337	6.1266
Minimum	-3.4831	-8.2078	-49.6007	-6.4023
Maximum	2.4405	10.4376	35.0588	5.6474

The table presents descriptive statistics for the daily logarithmic returns of the US dollar spot exchange rate versus the Swiss Franc (CHF/USD), the DJ Euro Stoxx 50 Index (EuroStoxx50), the CBOE Volatility Index (VIX) and the Bloomberg Commodity Index (BCOM). Period from November 2006 to December 2015.

As expected, VaR models that do not satisfy the confidence interval for the number of exceptions are the same models where the null hypothesis is rejected for the UC test (fourth column of Table 3). These models are ARMA-GARCH-normal and ARMA-GARCH-t (ν = 3). The independent occurrence of the exceptions hypothesis cannot be rejected for all models. This means that ARMA(1,1)-GARCH(1,1) is a sound model to capture the volatility process.

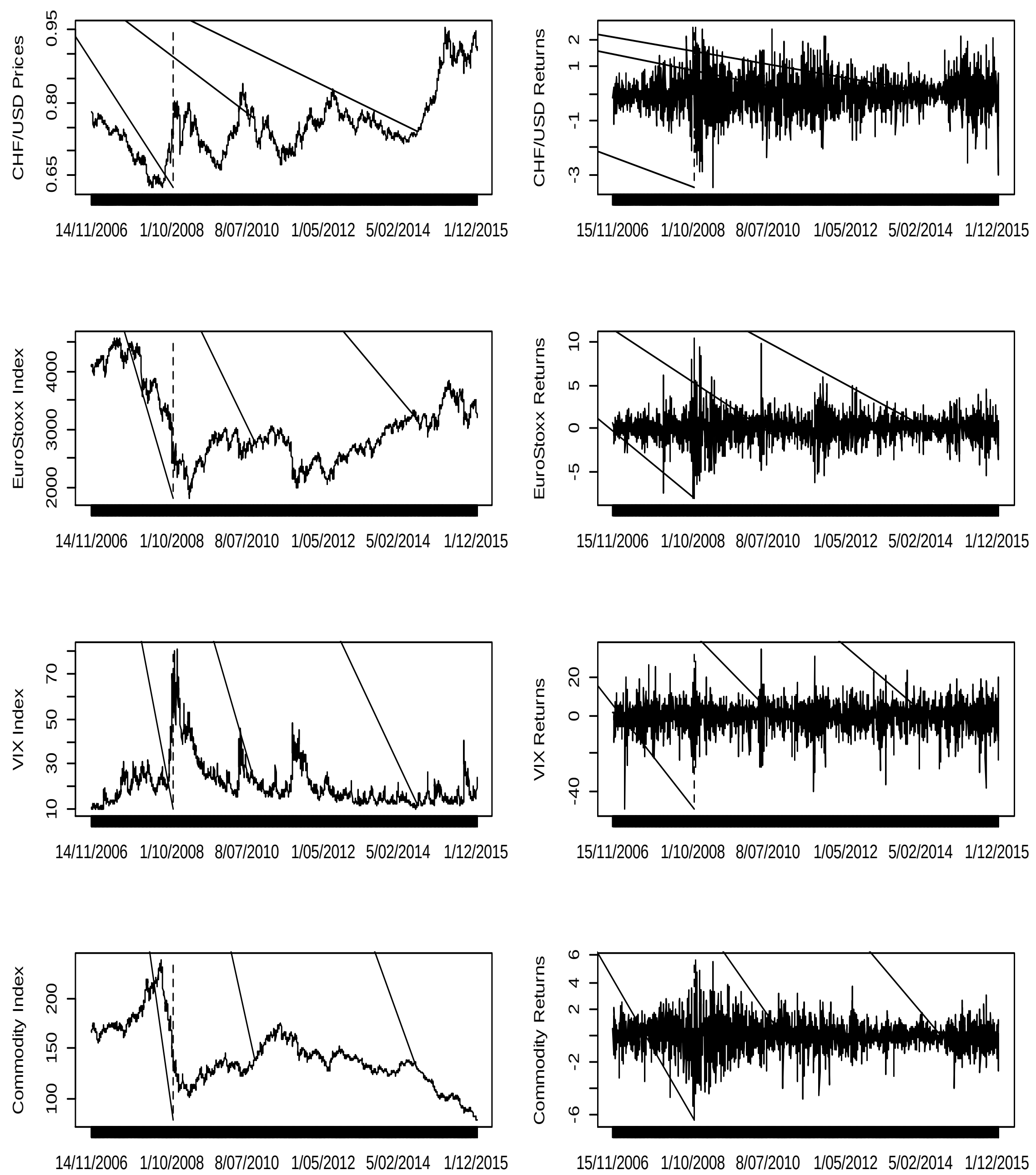
1.7 Backtesting for different portfolios

In this section, we analyze the performance of the considered methods to compute the VaR on four diversified portfolios during a large time period that includes the worst financial crisis since the Great Depression. The period from November 2006 to December 2015 contained a global financial crisis and economic crisis and a European debt crisis. Therefore, these characteristics make this period an ideal context for backtesting VaR. The selection of the period is deliberate; the period allows us to start the backtesting immediately after the bankruptcy of Lehman Brothers (September 15, 2008). The initial 500-day window to compute the VaR includes the beginning of the turmoil in August 2007 and the failure of this financial services firm.

Financial institutions allocate financial assets into classes. Four of the most usual classes are European stocks, US equity options, foreign currencies, and commodities. We consider three indexes and a foreign exchange rate as proxies of the typical well-diversified portfolios that financial institutions hold. The data correspond to the US dollar spot exchange rate versus the Swiss Franc (CHF/USD), the DJ Euro Stoxx 50 Index (henceforth EuroStoxx50), the Chicago Board Options Exchange Volatility Index (VIX) and the Bloomberg Commodity Index (BCOM).

Our period consists of 2,370 working days. We compute 2,369 logarithmic returns as $r_t = 100\log(P_t/P_{t-1})$ from the daily prices of each time series. Figure 1 shows the prices in the upper panel and the returns in the lower panel. The vertical line represents the beginning of the backtesting period. The figures depict the extreme volatility that characterizes the period. Table 3 shows descriptive statistics for the analyzed time series. The returns exhibit heavier tails than normal. EuroStoxx50 returns show the largest kurtosis, and VIX returns are the most dispersed (standard deviation is 7.27).

Figure 1: Prices and returns of the four portfolios

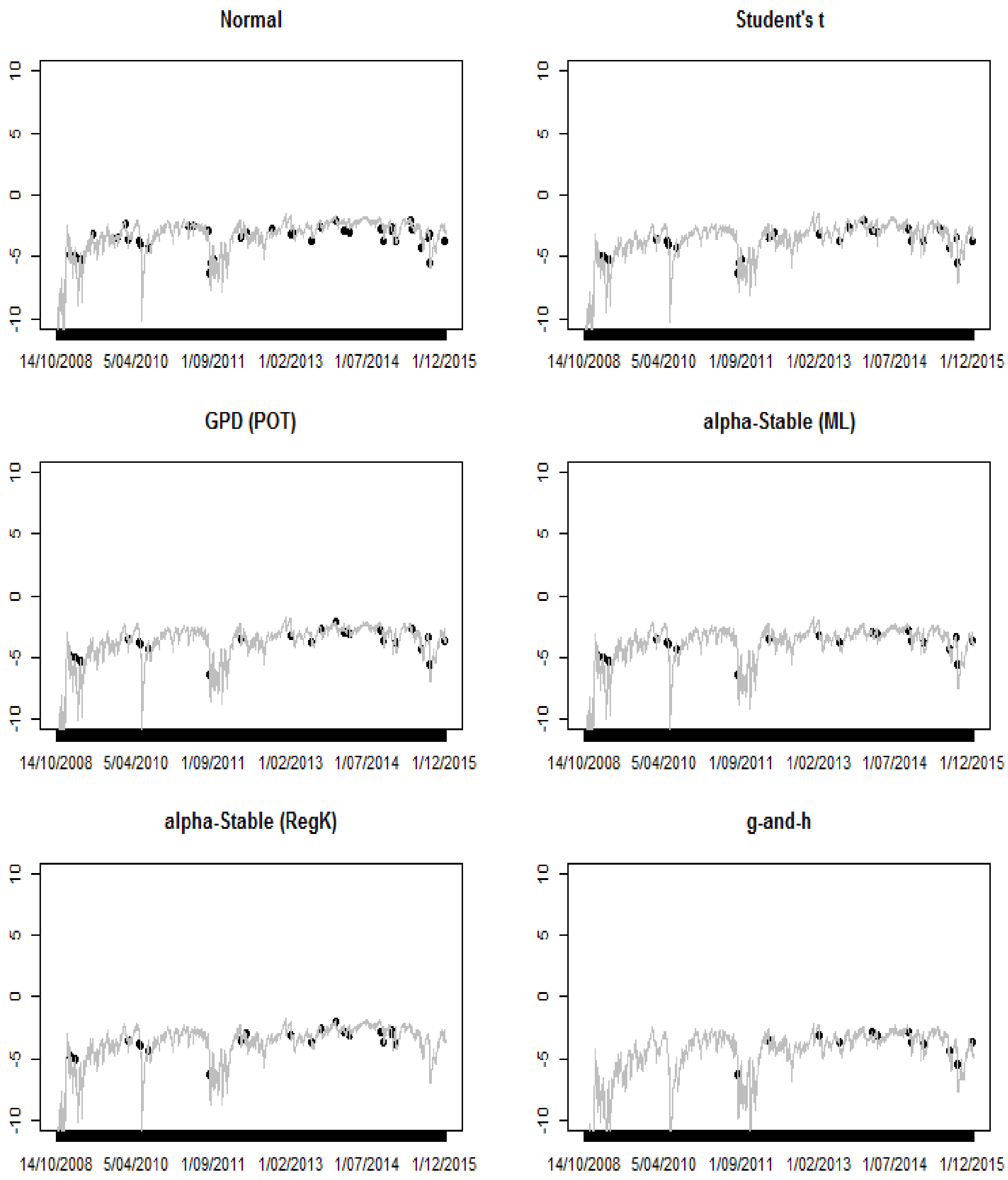


Daily prices and logarithmic returns of the US dollar spot exchange rate versus the Swiss Franc (CHF/USD), the DJ Euro Stoxx 50 Index (EuroStoxx) expressed in US dollars, the CBOE Volatility Index (VIX) and the Bloomberg Commodity Index (BCOM). Period from November 2006 to December 2015.

We use the same distributions described in the previous section to compute the VaR and evaluate the goodness of each VaR model forecast. We also learn from the results of the previous section to choose the *v* and *u* parameters. In the case of ARMA-GARCH-t, degrees of freedom *v* is estimated in each step of the rolling window. In the case of ARMA-GARCH-POT, the value of threshold *u* is fixed at the commonly used 10%. The window size for the rolling window used during the backtesting period is 500 days. Then, the backtesting analysis includes 1,869 days, from October 2008 to December 2015. The expected number of exceptions is approximately 19 when a 99% VaR is calculated, and the 95% confidence interval for the exceptions is between 10 and 27.

Figure 2 plots the temporal evolution of the VaR of the EuroStoxx50 returns and highlights where the exceptions are for each estimation model. Tables 4 and 5 show the numerical results of the backtesting analysis. ARMA-GARCH-normal performs well only for CHF/USD and commodity returns. The normal distribution produces good backtesting results when the asset returns exhibit not as heavy tails as in the case of CHF/USD and commodity returns. Their kurtosis values are 4.95 and 6.12, respectively, less than the other two asset returns where the normal model under-predicts risk. On the other hand, ARMA-GARCH-t, ARMA-GARCH-g-and-h and ARMA-GARCH-α-stable(RegK) do not work well for VIX returns. The g-and-h distribution tends to over-predict risk for the analyzed asset returns because the violation ratio is less than one. Both ARMA-GARCH-POT and ARMA-GARCH-α-stable(ML) perform well for the analyzed time series.

Figure 2: Backtesting for EuroStoxx-50 returns: VaR estimates and exceptions



Dataset: daily VaR estimates and exceptions during the backtesting period October 2008 to December 2015 (1,870 days) obtained from daily logarithmic returns of the DJ EuroStoxx 50 index expressed in US dollars using 500-day rolling windows.

Table 4: Backtesting Results for four portfolios and five estimation methods: number of exceptions

95% confidence interval for number of exceptions = [10, 27].

Expected number of exceptions = 19.

Model	CHF/USD	EuroStoxx50	VIX	BCOM
ARMA-GARCH-normal	25	40	57	26
ARMA-GARCH-t (changing *v*)	25	26	36	17
ARMA-GARCH-POT (u = 10%)	23	23	23	17
ARMA-GARCH-α-stable-ML	20	20	13	19
ARMA-GARCH-α-stable-RegK	20	13	8	16
ARMA-GARCH-g-h	14	12	10	14

The number of degrees of freedom *v* in ARMA-GARCH-t is estimated in each step of the rolling window. The threshold in ARMA-GARCH-POT is fixed in u = 10%. α-stable-ML indicates that the parameters have been estimated using ML (the pdf is obtained by direct integration of the characteristic function), and α-stable-RegK uses the Koutrouvelis method to estimate the parameters. Dataset: daily logarithmic returns of the CHF/USD, the DJ Euro Stoxx 50 Index (EuroStoxx50), the CBOE Volatility Index (VIX) and the Bloomberg Commodity Index (BCOM). The backtesting period ranges from October 2008 to December 2015 (1,869 days) obtained from 500-day rolling windows.

Table 5: Unconditional Coverage and independence tests for the different portfolios

Model	VR	LR_{uc}	p-value	LR_{ind}	p-value
			Panel a: CHF/USD		
Normal	1.34	1.9459	0.1630	0.8953	0.3440
Student’s t	1.34	1.9459	0.1630	0.8953	0.3440
POT (u = 10%)	1.23	0.9353	0.3335	1.1356	0.2866
α-stable-ML	1.07	0.0906	0.7633	0.4329	0.5106
α-stable-RegK	1.07	0.0906	0.7633	0.4329	0.5106
g-and-h	0.75	1.3018	0.2538	0.2114	0.6456

		Panel b: EuroStoxx 50			
Normal	2.14	18.4976	0.0000	1.1733	0.2787
Student's t	1.39	2.5745	0.1086	0.7340	0.3916
POT (*u* = 10%)	1.23	0.9353	0.3334	0.5735	0.4489
α -stable-ML	1.07	0.0906	0.7633	0.4329	0.5106
α -stable-RegK	0.69	1.9585	0.1617	0.1822	0.6695
g-and-h	0.64	2.7702	0.0960	0.1552	0.6936
		Panel c: VIX			
Normal	3.05	51.2959	0.0000	0.8082	0.3686
Student's t	1.93	12.7403	0.0003	1.7213	0.1895
POT (*u* = 10%)	1.23	0.9353	0.3335	4.6534	0.0310
α -stable-ML	0.69	1.9584	0.1616	0.1822	0.6694
α -stable-RegK	0.43	7.8649	0.0050	0.6882	0.7931
g-and-h	0.53	4.9127	0.0267	0.1076	0.7428
		Panel d: Commodity			
Normal	1.39	2.5745	0.1086	0.7340	0.3916
Student's t	0.91	0.1592	0.6899	0.3123	0.5763
POT (*u* = 10%)	0.91	0.1592	0.6899	0.3123	0.5763
α -stable-ML	1.01	0.0051	0.9427	0.3904	0.5320
α -stable-RegK	0.85	0.4111	0.5214	0.2765	0.5990
g-and-h	0.75	1.3018	0.2538	0.2114	0.6456

The number of degrees of freedom *v* in ARMA-GARCH-t is estimated in each step of the rolling window. The threshold in ARMA-GARCH-POT is fixed at *u* = 10%. α-stable-ML indicates that the parameters have been estimated using ML (the pdf is obtained by direct integration of the characteristic function), and α-stable-RegK uses the Koutrouvelis method to estimate the parameters. *VR* is the violation ratio. LR_{uc} is the likelihood ratio test for unconditional coverage. LR_{ind} is the likelihood ratio test of the independence part of the conditional coverage hypothesis. Both tests are proposed by Christoffersen (1998). Dataset: daily logarithmic returns of the CHF/USD, the DJ Euro Stoxx 50 Index (EuroStoxx50) expressed in US dollars, the CBOE Volatility Index (VIX) and the Bloomberg Commodity

Index (BCOM). The backtesting period ranges from October 2008 to December 2015 (1,869 days) obtained from 500-day rolling windows.

The UC test shown in Table 5 confirms the results obtained in Table 4. The null hypothesis of the independence test cannot be rejected in any model. The GPD and α-stable distributions perform well for the analyzed cases in this paper. However, special care must be taken when estimating its parameters. POT-GPD accuracy depends on the threshold selection. As observed in the previous application, a careless selection of the threshold yields poor backtesting results. The ML estimation is not directly applied to the α-stable distribution because its pdf is not analytically expressible; however, Nolan (2001) developed fast and accurate techniques to estimate the parameters of the α-stable distribution, as confirmed with the good results of the backtesting procedure.

1.5 Conclusions of Chapter 1

The best model to capture tail events and to quantify accurate risk measures is still an open question in quantitative risk management. The literature shows that several distributions capture the skewness and heavy tails of financial asset returns. In our study, we examine the performance of Student's t, GPD, α-stable and g-and-h distributions to address one of the main financial risk management challenges. The current VaR models have been subjected to a great deal of criticism since the bailouts and failure of several financial institutions during the last financial crisis. We examine the forecasting ability of these distributions through a backtesting analysis of an extremely demanding sample period. Additionally, the estimation of the α-stable distribution is a challenge. We check two alternative estimation methods: the ML method (which, according the literature, provides more accurate results than other methods but is computationally expensive) and methods developed by Nolan (2001) that provide both fast and accurate results.

Consistent with the financial literature, we conclude that the GPD, α-stable and g-and-h distributions perform well for heavy-tailed data in our sample period. Therefore, we also recommend the use of the g-and-h distribution that incorporates skewness and kurtosis in the VaR calculation. Although there is no closed expression for its pdf, the parameter estimation of it is simpler than the α-stable case. However, the risk measures obtained by the g-and-h model are very conservative for not data that are not so heavy tailed. For this case, the GPD and the α-stable distribution also work well. The POT method employs exceedances over a certain threshold, while the α-stable employs all data. The threshold selection is a drawback in EVT because there is a tradeoff between bias and variance in estimating the GPD parameters. Nevertheless, satisfactory VaR results are obtained when the 10^{th} percentile is chosen as a threshold. Future research will be conducted to test coherent risk measures with the models analyzed in this paper.

References

Angelidis, T., Benos, A., Degiannakis, S. 2004. The use of GARCH models in VaR estimation. Statistical Methodology 1(1-2), 105-128.

Araújo Santos, P. and Fraga Alves, M.I. 2013. Forecasting Value-at-Risk with a duration-based POT method. Mathematics and Computers in Simulation 94 (C), 295-309.

Araújo Santos, P., Fraga-Alves I., Hammoudeh, S. 2013. High quantiles estimation with Quasi-PORT and DPOT: An application to value-at-risk for financial variables. North American Journal of Economics and Finance 26, 487-496.

Ausín, M.C., Galeano, P., Ghosh, P. 2014. A semiparametric Bayesian approach to the analysis of financial time series with applications to value at risk estimation. European Journal of Operational Research, 232 (2), 350-358.

Bank for International Settlements, 2016. Minimum capital requirements for market risk. Basel Committee on Banking Supervision. January.

Beirlant, J. Goegebeur, Y., Segers, J., Teugels, J. 2004. Statistics of Extremes: Theory and Applications. John Wiley Sons.

Bekiros, S. D., Georgoutsos D. A. 2005. Estimation of Value-at-Risk by exteme value and conventional methods: a comparative evaluation of their predictive performance. Int. Fin. Markets, Inst. And Money 15 209-228.

Berkowitz, J., O'Brien, J. 2002. How accurate are Value-at-Risk models at commercial banks? Journal of Finance 57, 1093-1111.

Byström, H. N. E. 2004. Managing extreme risks in tranquil and volatile markets using conditional extreme value theory, International Review of Financial Analysis. 13(2), 133-152.

Carrasco, M., Florens J. P. 2002. Efficient GMM Estimation Using the Empirical Characteristic Function. Working Paper, Institut d'Economie Industrielle, Toulouse.

Čížek, P., Härdle, W. K., Weron, R. 2011. Statistical Tools for Finance and Insurance. Springer, Berlin.

Chavez-Demoulin, V. 1999. Two problems in environmental statistics: Capture-recapture analysis and smooth extremal models. Ph.D. thesis. Department of Mathematics, Swiss Federal Institute of Technology, Lausanne.

Chavez-Demoulin, McGill, J.A. 2012. High-frequency financial data modeling using Hawkes processes. Journal of Banking & Finance 36 (12) 3415-3426.

Christoffersen, P.. 1998. Evaluating Interval Forecasts. International Economic Review 39, 841-862.

Christoffersen, P. 2003. Elements of Financial Risk Management, San Diego, CA: Academic Press.

Cont, R. 2001. Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance, 1, 223-236.

Danielsson, J. 2011. Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab. Wiley-Blackwell.

Davison, A.C. Smith, R.L. 1990. Models for exceedances over high thresholds, J. R. Stat. Soc. Ser. B Stat. Methodol. 52 393–442.

De Jesús, R., Ortiz, E., Cabello, A. 2013. Long run peso/dollar exchange rates and extreme value behavior: Value at Risk modeling. The North American Journal of Economics and Finance 24, 139-152.

Degen, M., Embrechts, P., Lambrigger, D. 2007. The quantitative modeling of operational risk: between g-and-h and EVT. ASTIN Bulletin 37(2) 265-291.

Del Brio, E.B., Mora-Valencia, A., Perote, J. VaR performance during the subprime and sovereign debt crises: An application to emerging markets. Emerging Markets Review 20, 23-41.

Diebold, F. X., Schuermann, T., Stroughair, J. D. 1998. Pitfalls and opportunities in the use of extreme value theory in risk management. Working Paper, 98-10, Wharton School, University of Pennsylvania.

Dutta, K., Babbel D. F. 2002. ón Measuring Skewness and Kurtosis in Short Rate Distributions: The Case of the US Dollar London Inter Bank óffer Rates. Wharton Financial Institutions Center Working Paper.

Dutta, K., Perry, J. 2006. A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital. Federal Reserve Bank of Boston, Working Paper No. 06-13.

Embrechts, P., Klüppelberg, C., Mikosch, T. 1997. Modelling Extremal Events for Insurance and Finance. Berlin: Springer-Verlag.

Ergen, I. 2015. Two-step methods in VaR prediction and the importance of fat tails. Quantitative Finance, 15 (6), 1013-1030.

Fama, E. F., Roll, R. 1971. Parameter estimates for symmetric stable distributions. Journal of the American Statistical Association 66, 331-338.

Ghosh, S., Resnick, S. 2010. A discussion on mean excess plots. Stochastic Processes and their Applications 120, 1492-1517.

Goldberg L.G.M., Weinstein, J. 2008. Beyond value at risk: forecasting portfolio loss at multiple horizons. Journal of Investment Management 6(2), 73–98.

Granger, C. W. J., Ding, Z. 1995. Some properties of absolute returns. An alternative measure of risk. Annales d'économie et de statistique, 40, 67-91.

Grigoletto, M., Lisi, F. 2009. Looking for skewness in financial time series. The Econometrics Journal 12, 310-323.

Grigoletto, M., Lisi, F. 2011. Practical implications of higher moments in risk management. Statistical Methods and Applications 20, 487-506.

Güner, B., Rachev, S., Edelman, D., Fabozzi, F. J. 2010. Bayesian Inference for Hedge Funds with Stable Distribution of Returns. En Rethinking Risk Measurement and Reporting: Volume II (Ed. K. Böcker), 95-136.

Haas, M. 2009. Value-at-Risk via mixture distributions reconsidered. Applied Mathematics and Computation, 215 (6) 2103-2119.

Haynes, M., Mengersen, K. 2005. Bayesian estimation of g-and-k distributions using MCMC. Comput. Stat 20 (1): 7-30.

Herrera, R., Schipp, B. (2014). Statistics of extreme events in risk management: The impact of the subprime and global financial crisis on the German stock market. The North American Journal of Economics and Finance, 29, 218–238.

Hoaglin, D. C. 1985. Summarizing Shape Numerically: The g-and-h Distributions. Chapter 11 en Exploring Data Tables Trends, and Shapes. Eds. Hoaglin, Mosteller, and Tukey. New York, NY: John Wiley.

Jalal A., Rockinger M. 2008. Predicting tail-related risk measures: The consequences of using GARCH filters for non GARCH data. Journal of Empirical Finance, 15(5), 868-877.

Jondeau, E., Poon, S.-H., Rockinger M. 2007. Financial Modeling Under Non-Gaussian Distributions, Springer Verlag.

Khindanova, I., Rachev, S., Schwartz, E. 2001. Stable modeling of value at risk. Mathematical and Computer Modelling 34(9-11), 1223-1259.

Kiesel, R., Perraudin, W., Taylor, A. 2003. An Extreme Analysis of VaRs for Emerging Market Benchmark Bonds. En: Credit Risk: Measurement, Evaluation and Management, eds: G.Bol, et al., Physica-Verlag.

Koutrouvelis, I.A. (1980). Regression-type Estimation of the Parameters of Stable Laws. J Am Stat Ass 75 918 – 928.

Kuester, K., Mittnik, S., Paolella, M. S. 2006. Value-at-risk prediction: a comparison of alternative strategies. Journal of Financial Econometrics 4, 53-89.

Kupiec, P. 1995. Techniques for Verifying the Accuracy of Risk Management Models. Journal of Derivatives 3, 73–84.

Lauridsen, S. 2000. Estimation of Value at Risk by Extreme Value Methods. Extremes 3 (2), 107-144.

Leadbetter, M.R. 1991. On a basis for ‘peaks over thresholds’ modeling, Statist. Probab. Lett. 4 357–362.

Li, W. K., Ling, S., McAleer, M. 2002. Recent Theoretical Results for Time Series Models with GARCH Errors. Journal of Economic Surveys 16 (3) 245-269.

Lunde, A, Hansen, P. R. 2005. A forecast comparison of volatility models: Does anything beat a garch(1,1)?. Journal of Applied Econometrics 20 (7) 873–889.

Marinelli, C., D’Addona, S., Rachev, S. 2007. A comparison of some univariate models for Value-at-Risk and expected shortfall. International Journal of Theoretical and Applied Finance, 10 (6).

Martínez, J. 1981. Some Applications of Robust Scale Estimators, Doctoral Thesis, Temple University, Department of Statistics. Philadelphia.

Martínez, J. & Iglewicz, B. 1984. Some properties of the Tukey g and h family of distributions, Communications in Statistics: Theory and Methods 13 (3) 353-369.

McCulloch, J. H. 1986. Simple consistent estimators of stable distribution parameters, Communications in Statistics - Simulations 15, 1109-1136.

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, 271-300.

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

Mittnik, S., Rachev, S. T., Doganoglu, T., Chenyao, D. 1999. Maximum likelihood estimation of stable Paretian models, Mathematical and Computer Modelling 29: 275-293.

Nolan, J. P. 2001. Maximum likelihood estimation and diagnostics for stable distributions, in E. Barndorff-Nielsen, T. Mikosch, S. Resnick (eds.), Lévy Processes, Brikhäuser, Boston.

Ourir, A., Snoussi, W. 2012. Markets liquidity risk under extremal dependence: Analysis with VaRs methods. Economic Modelling 29 (5) 1830-1836.

Perignon, C., Deng, Z. Y., Wang, Z. J. 2008. Do banks overstate their Value-at-Risk? Journal of Banking & Finance 32, 783-794.

Polanski, A., Stoja, E. 2010. Incorporating higher moments into value-at-risk forecasting. Journal of Forecasting 29, 523–535.

Rachev, S., Racheva-Iotova, B., Stoyanov, S. 2010. Capturing fat tails. Risk Magazine, May 2010, 72-77.

Rayner, G. D., MacGillivray, H. L. 2002. Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. Statistics and Computing 12, 57-75.

Rydén, T., Teräsvirta, T., Asbrink, S. 1998. Stylized facts of daily return series and the hidden markov model. Journal of Applied Econometrics 13, 217-244.

Stavroyiannis, S. Makris, I. Nikolaidis, V. Zarangas, L. 2012. Econometric modeling and value-at-risk using the Pearson type-IV distribution. International Review of Financial Analysis 22, 10-17.

Smith, R.L. 1989. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone, Statist. Sci. 4 367–393

Teräsvirta, T., Zhao, Z. 2011. Stylized facts of return series, robust estimates and three popular models of volatility. Applied Financial Economics 21, 67-94.

Wei, Y., Cheng, W., Lin Y. 2013. Measuring daily Value-at-Risk of SSEC index: A new approach based on multifractal analysis and extreme value theory. Physica A: Statistical Mechanics and its Applications 392 (9), 2163-2174.

Zhang, M.-H., Cheng, Q.-S. 2005. An Approach to VaR for capital markets with Gaussian mixture. Applied Mathematics and Computation 168, 1079-1085.

Chapter 2. Quantifying risk in traditional energy investments[12]

Abstract

As investors are increasingly concerned about the risks associated to the environmental and financial impact of burning of fossil fuels (oil, natural gas and coal), we examine the ability of the recently adopted expected shortfall (ES) risk measure to properly quantify market risk for a sustainable fossil-free stock index and a highly consumable fuel dependent index. We obtain evidence that a newly proposed backtesting procedure for the ES based on multinomial tests is an adequate and simple method to validate these risk measure when applied to a highly volatile stock index. Backtesting results of the ES show that flexible heavy-tailed distribution α–stable performs well for modelling the loss distribution. These results are even better off when the variances of fossil fuel price returns are included as external regressors in the GARCH model variance equation. In this case, the ES computed from the four considered loss distributions perform properly.

Keywords: Oil and gas industry; Expected shortfall; Backtesting; Sustainability index

2.1 Introduction of Chapter 2

The debate on the role of fossil fuels in climate change affects all facets of society. The financial industry is also being affected, both by a progressive increase in awareness of the potential impact of climate change on investments, and by the risk of fossil fuels becoming "stranded", i.e. unburned or in the ground, as regulation increases. The traditional energy industry is currently exposed to downside risks from write-offs or revaluations of these unsustainable assets. However, companies in the traditional energy industry have been used for diversification purposes and have demonstrated their potential to provide high realized returns along with high volatility as commodity prices rise or fall. While there is a move towards divestment in fossil fuels, replacing investment in the traditional energy sector with other sustainable investments, individual and institutional investors seek to balance risk and expected return.[13]

---

[12] A version of this chapter has been published in Sustainability Journal, co-authored with Antonio Díaz and Gonzalo García-Donato.

[13] For instance, the Rockefeller Family Fund publicly announced its decision to divest from fossil fuels. In addition, a report by Moody's (Adams et al., 2016) notes that 175 oil and mining companies were on below

The existing literature proposes several risk measurement tools to provide financial institutions, risk managers and market participants with appropriate technical approaches to measure the risk of the financial markets. It is therefore important for these market players to adequately quantify the potential economic loss of their investments. In the case of the traditional energy industry, the literature focuses risk quantification based on Value-at-Risk (VaR), but there is scarce work regarding the new trend-setting topic of the Expected Shortfall (ES) backtest. In this paper, we examine the use and validation of the ES or Conditional Value-at-Risk (CVaR), as the risk measure recently recommended by banking regulators, in two broadly diversified investments, one in the traditional energy industry and one excluding fossil fuel companies, during the last decade. In addition, we implement a new ES backtesting procedure based on multinomial tests.

The main focus of our research in this paper is to examine the ability of ES risk measures to correctly quantify the risk the investments in the oil, gas and coal industry investments. ES is defined as the expected loss conditional on the loss being greater than the VaR level. In January 2016, financial regulators propose the use of the ES instead of VaR to prudently capture tail risk and capital adequacy. This change is challenging for portfolio and risk managers because it is not clear which validation method the regulator and the industry should use to test the proposed risk measure, i.e., it is not clear how to evaluate the goodness of the ES risk measure. Currently, there is a vivid debate in academia and the financial industry about how to validate internal models in regulatory capital under ES calculation. In this paper, we apply the new method proposed by Kratz et al. (2018) to validate ES. As this risk measure can be approximated as a weighted sum of different levels of VaRs; this method consists of utilizing a multinomial test instead of several independent binomial tests.

Our paper makes four contributions to the literature on risk measurement in the context of the traditional energy markets. The high volatility of the stock returns of the companies of energy industry provides a suitable and demanding dataset to examine the performance of the proposed ES backtesting technique. First, we employ several GARCH models to adequately model the risk of a broadly diversified portfolio of traditional energy industry stocks over a long period of time that includes periods of calm, turmoil and severe financial and economic crises. Second, the behavior of this portfolio is examined in relation to the behavior of a sustainable equity portfolio to provide guidance to investors at a time when a divestment

investment grade watch in early 2016, mainly because of the shift from carbon-intensive fossil fuel to renewable energy investment, i.e. transition risk, which affects oil prices.

movement is observed in the fossil fuel industry. Third, we applied a new ES backtesting procedure based on multinomial tests for different VaR levels instead of performing a binomial test for each VaR level as in the scarce previous literature on financial markets. To the best of our knowledge, this is the first attempt to apply multinomial tests on traditional energy and sustainable stock indexes. Fourth, we analyze the inclusion of exogeneous variables to improve the performance of the forecast volatility model as corroborated by the backtesting analysis.

We proxy the traditional energy industry through the S&P 500 Oil, Gas and Consumable Fuels Index, called Traditional Index (TI), and the divestment movement in fossil fuels thorough the FTSE Developed ex Fossil Fuel Index, called Sustainable Index (SI). A simple descriptive analysis reveals that the global portfolio excluding fossil fuel industry assets performs financially better than the portfolio of assets related to the traditional oil and gas industry over the last decade. However, TI outperformances during the crisis and subsequent period of uncertainty, which is a good feature for portfolio and risk managers seeking to diversify overall risk of their portfolios.

We consider four statistical models, normal, Student's t, α–stable, and generalized Pareto, to examine the variability of negative log-returns of two broadly diversified stock indexes. The one-day-ahead VaR and ES are calculated by applying a rolling window of 250 observations. Thus, the length of the backtesting period for both indexes is 2709 days with an expected number of exceptions of 27.09 for a 99%-VaR. We compute the well-known binomial tests for VaR at 99% and the two new multinomial tests, i.e. the Pearson and Nass statistics, proposed by Kratz et al. (2018), for 97.5%-ES backtesting.

The results of the multinomial and binomial tests show that flexible heavy-tailed distribution α–stable performs well for data employed in our paper. This is important for regulation purposes and for practitioners. For market risk, the main difficulty is the ES backtesting method as noted above, but this work sheds light on solving this problem. Moreover, we find that including variance of unsustainable asset returns as external regressors in the GARCH model help improve the backtesting results. In this case, the ES computed from the four considered loss distributions perform properly.

The rest of the paper is organized as follows: we present a survey of the relevant literature in Section 2 of the paper. Section 3 presents the models and the backtesting methodology, Sections 4 and 5 analyze the data and the results on ES backtesting, and Section 6 concludes the paper.

2.2 Literature review

There is an academic literature on the modelling of the risk of highly volatile prices of both energy commodities and energy stocks and derivatives. Energy commodity markets are naturally vulnerable to significant price changes. It is therefore important to model these price fluctuations and implement an effective tool for managing energy price risk. VaR has become a popular risk measure in the financial industry. The internal model approach under the Basel II framework proposes VaR as a risk measure to gauge the amount of assets needed to cover possible losses, i.e., the minimum regulatory capital requirements. A variety of works have been published on risk quantification applied to different financial assets (e.g., stocks, bonds, commodities, and derivatives), and several backtesting methods have been proposed to validate VaR models (see for instance Christoffersen, 1998; for different VaR forecasting tests).[14]

VaR answers the question of how much we can lose with a given probability over a given time horizon.[15] The popularity of this instrument is essentially due to its conceptual simplicity. VaR reduces the risk associated with any portfolio to a single number, the loss associated with a given probability. In addition, VaR helps portfolio managers determine the most appropriate risk management policy for each situation. Thus, VaR is the primary tool used to forecast extreme declines in returns and is often used for designing optimal risk management strategies.

Previous literature examines the use of VaR to measure risks in energy markets. They use different return distributions in order to estimate VaR from oil and carbon prices (e.g., Cabedo and Moya, 2003, Ewing and Malik, 2010, Feng, Wei, Wang, 2012). Some studies consider GARCH specifications to model heavy-tailed and asymmetric return distributions for VaR estimation from energy commodity prices (e.g., Costello et al., 2008, Fan et al., 2008, Hung et al., 2008, Marimoutou, Raggad, and Trabelsi, 2009, Youssef, Belkacem, and Mokni, 2015, Bunn et al., 2016, Lux et al., 2016, Wang, Liu, Ma, and Wu, 2016, Lyu et al., 2017, Ewing, Malik and Anjum, 2018) and from their financial derivative prices and even from carbon dioxide emission allowance prices (e.g., Nomikos and Pouliasis, 2011, Segnon et al.,

[14] The idea is to calculate the number of times the actual losses have exceeded the estimated VaRs. It is expected that the number of exceptions is approximately 1% of cases when a 99% VaR is calculated. If the percentage of exceptions is higher (lower) than 1%, then the VaR model underestimates (overestimates) risk.

[15] VaR is defined as the α quantile of a relevant profit and loss (P&L) distribution to assess the risk exposure of single investments and portfolios. It estimates how much a portfolio might lose, given normal market conditions, over a target horizon such that there is a low, prespecified probability that the actual loss will be larger.

2017). Alternative methodologies to capture downside risks for crude oil prices are also used (e.g., Huang et al., 2008, Chiu et al., 2010, He et al., 2011, Herrera, 2013). For multivariate analysis cases, recent papers propose copula approaches to model dependence between different crude oil markets or between crude oil and other energy markets (e.g., Aloui et al., 2014, Jäschke, 2014, Zolotko et al., 2014, Ghorbel et al., 2014, Lu et al., 2014, González-Pedraz, 2014).

Nevertheless, financial regulatory entities have recently expressed concern about the inability of VaR to capture tail risk. It is not a "coherent" measure of risk because it does not satisfy the property of "subadditivity" (Artzner et al., 1999). In addition, VaR does not allow the magnitude of losses suffered above the threshold to be known. In January 2016, the Basel Committee on Banking Supervision changes from requiring banks to calculate market risk capital on the basis of VaR to using ES on the behavior of market variables during a 250-day period of stressed market conditions.[16] ES, C-VaR or expected tail loss is the expected loss conditional on the loss being worse than the VaR loss. As with VaR, ES attempts to provide a unique number that summarizes the total risk in a portfolio. The use of ES poses a challenge to portfolio and risk managers because it is not clear which validation method the regulator and the industry will employ to test the proposed risk measure, i.e., it is not clear how to evaluate the goodness of the ES risk measure. A backtesting method for ES is not as straightforward as in the case of VaR, due to ES not satisfying the elicitability property (e.g., Weber, 2006; Gneiting, 2011). An appropriate scoring function that this risk measure can minimize does not exist. In fact, the Basel Committee proposes to use ES to calculate capital requirements, but instead proposes to carry out the backtest using a VaR measure.[17]

Literature on the use of ES in the energy industry is scarce. Some studies use ES constraints in the optimization programs to choose investment projects (e.g., Bruno and Sagastizábal, 2011, Tekiner-Mogulkoc et al., 2015, Hemmati et al., 2016, Lu et al., 2016; Roustai, et al., 2018). Other papers include the ES as risk objective function in the estimation of hedging strategies to reduce price volatility risk into energy markets (e.g., Alizadeh et al., 2008, Conlon and Cotter, 2013, Chai and Zhou, 2018). The authors suggest that ES should be an appropriate metric accounting for some properties of the energy assets. Finally, Youssef et

[16] The "Fundamental Review of the Trading Book" (Basel Committee on Banking Supervision, January 2016) changes the measure to use for determining market risk capital. Instead of VaR with a 99% confidence level, expected shortfall (ES) with a 97.5% confidence level is proposed.

[17] The "Fundamental Review of the Trading Book" states that the backtesting requirements continue to be based on the 1-day static VaR measure considering 250 days of (rolling) window size.

al. (2015) apply both VaR and ES to model the price risk of four energy commodities. To backtest ES, they use a circular bootstrap method from the one-sided test proposed by McNeil and Frey (2000). They conclude that the forecasted ES measure captures actual shortfalls in a satisfactory manner.

There are several important differences between these prior papers and our own research in this paper. First, we examine the correct measurement of the risk borne by the equity portfolios of companies in the traditional energy sector and of companies in all sectors, excluding those related to the fossil fuel sector. Second, we analyze a new ES backtesting method in a highly volatile financial asset. Third, we study the inclusion of exogeneous variables to improve the performance of the forecast volatility model.

2.3 The Model and Methodology

VaR and ES approaches model the left tail of the return distribution or, similarly, the right tail of the loss distribution. The losses or negative log-returns over the next day are defined here as $L_{t+1} = -100\log(P_{t+1}/P_t)$, where P_t represents the corresponding index prices. As it is commonly employed in the literature (see, e.g., McNeil et al., 2005), we suppose that conditional on the location-scale parameters μ_{t+1} and σ_{t+1}, negative log-returns follow $L_{t+1} = \mu_{t+1} + \varepsilon_{t+1}$, and the innovations are $\varepsilon_{t+1} = \sigma_{t+1} Z_{t+1}$. The random variables Z_{t+1} are assumed to be independently distributed with a common cumulative distribution function (CDF) G that, for certain cases, depends on unknown parameters. We discuss several possibilities for G in the next section. The parameter μ_{t+1} is modeled by an ARMA(1,1) process, and a GARCH(1,1) process is employed for σ_{t+1}, that is,

$$\mu_{t+1} = \theta_0 + \theta_1 \mu_t + \theta_2 \varepsilon_t + \varepsilon_{t+1},$$

$$\sigma_{t+1}^2 = \beta_0 + \beta_1 \varepsilon_t^2 + \beta_2 \sigma_t^2 + \gamma_1 var_t^{oil} + \gamma_2 var_t^{gas} + \gamma_3 var_t^{coal}, \tag{1}$$

where θ_1 and θ_2 are the parameters associated of AR(1) and MA(1) respectively. Apart from the variables of the standard GARCH(1,1) model, the variances of oil, gas and coal price returns are considered as external regressors. Thus, our empirical results consider two methods of backtesting. One method excludes the external regressors (i.e. $\gamma_1 = \gamma_2 = \gamma_3 = 0$) from the

GARCH model, and the other method takes into account these variables in the variance equation of the GARCH model. Given a probability level α, the VaR can be expressed as

$$VaR_\alpha = \mu_{t+1} + \sigma_{t+1} q_\alpha, \tag{2}$$

where q_α is the α quantile of G. The ARMA-GARCH model is implemented by using rugarch package in R.[18]

3.1. Distributions and estimation strategy

In the ARMA(1,1)-GARCH(1,1) setting above, the location and variability of negative log-returns are modeled through the parameters μ_{t+1} and σ_{t+1}. The distribution G should be free of any such parameters (to avoid identifiability issues) and must account for other important features, such as asymmetry and/or kurtosis. In particular, the statistical models we consider are: (i) normal (used for comparative purposes), (ii) Student's t, (iii) α–stable, and (iv) generalized Pareto.

(vi) Normal distribution

The CDF of a standard normal distribution is given by

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \tag{3}$$

(vii) Student's t distribution

The CDF of a Student's t distribution is given by

$$H(x) = \int_{-\infty}^{x} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt, \tag{4}$$

where Γ represents the gamma function, and $\nu > 0$ is the degrees of freedom parameter that controls the kurtosis (small values of ν correspond to heavier tails). The Cauchy distribution is a particular case when $\nu = 1$.

(viii) α–stable distribution

The α–stable distribution is commonly described by its characteristic function, since the probability density function (PDF) is not available in closed-form.

[18] Ghalanos, A. (2018). rugarch: Univariate GARCH models. R package version 1.4-0.

$$E\left[e^{itX}\right] = \begin{cases} exp\left(-\left(1 - i\beta(sign(t))\tan\frac{\pi\alpha}{2}\right)\right) & \text{if} \quad \alpha \neq 1, \\ exp\left(|t|\left(1 + i\beta\frac{2}{\pi}(sign(t))\ln|t|\right)\right) & \text{if} \quad \alpha = 1, \end{cases} \tag{5}$$

where the $sign(t)$ function is defined as 1 if $t > 0$; 0 if $t = 0$ and -1 otherwise.

The parameters in this distribution are the index of stability (characteristic exponent) $\alpha \in (0,2]$ and a skewness parameter $\beta \in [-1,1]$. There are three cases with known closed-form expressions for their densities: the normal (when $\alpha = 2$ and $\beta = 0$), Cauchy ($\alpha = 1$ and $\beta = 0$), and Lévy distributions ($\alpha = 1/2$ and $\beta = 0$). The smaller the value of α, the heavier the distribution tail. The stable package for R developed by Nolan is employed to fit Stable distribution.

(ix) The generalized Pareto distribution (GPD)

The CDF of the GPD is given by

$$F_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi\, x/\beta)^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - \exp(-x/\beta), & \text{if } \xi = 0' \end{cases} \tag{6}$$

where ξ is the shape parameter and β is the scale parameter. When $\xi > 0$, the GPD is the Pareto distribution; when $\xi = 0$, it is the exponential distribution; and when $\xi < 0$, the distribution is the Pareto type II distribution. Heavy-tailed empirical distributions usually follow a GPD with a positive shape parameter $\xi > 0$.

When G is either a normal or a Student's t distribution, the parameters for the ARMA(1,1)-GARCH(1,1) and for the innovations are estimated jointly by employing the Maximum Likelihood (ML) estimation. A two-step approach is used to estimate the parameters for the cases where G is either a α-stable or a generalized Pareto distribution. First, the Quasi-ML (QML) method is used to estimate the parameters in the ARMA(1,1)-GARCH(1,1), thus allowing estimations of the underlying innovations to be produced, say, $\hat{\varepsilon}_{t+1}$. Specific methods are then performed in a second step to estimate the parameters in G:

- For the α-stable distribution, the ML approach is employed by using the direct integration method in Nolan (2001).
- For the generalized Pareto distribution, the peaks over threshold (POT) method is employed to estimate the parameters. According to McNeil et al. (2005), the VaR or α-quantile is obtained from

$$q_\alpha(Z) = u + \frac{\beta}{\xi}\left[\left(\frac{1-\alpha}{T_u/T}\right)^{-\xi} - 1\right], \tag{7}$$

where u is the chosen threshold, β and ξ are the scale and shape parameters, respectively, T_u is the threshold exceedances, and T is the sample size. Therefore, T_u/T is an empirical estimator for the excess distribution. In this paper, the threshold is chosen as the 10th percentile of the standardized residuals of the negative log-returns as is typical in the literature (Chavez-Demoulin, 1999, Nomikos and Pouliasis, 2011, McNeil et al., 2000, Araújo-Santos et al., 2013, Díaz et al., 2017). The evir package in R is employed to implement the EVT-GPD model.

3.2 Backtesting ES

As mentioned above, the method to be used to validate the results of the application of the ES remains an open question. Fissler et al. (2016) show that ES and VaR are jointly elicitable, and the authors propose a scoring function that is more complicated than the well-known scoring function for VaR. Comparative tests can then be performed following the Diebold-Mariano test (e.g., Del Brio et al., 2017). Based on the Monte Carlo simulations, Acerbi et al. (2014) propose other tests for ES; following the argument that VaR and ES are jointly elicitable, in this paper, we employ the simple approach proposed by Kratz et al. (2018) to validate ES calculations in an implicit manner. ES can be approximated by a weighted sum of VaR levels (Emmer et al., 2015), and then, a multinomial test can be performed rather than the binomial test for each VaR level. This paper extends applications of Kratz et al. (2018) to traditional energy and sustainable indexes. Moreover, our work considers an ARMA-GARCH model with external regressors to filter the negative log-returns of the analyzed assets, whereas Kratz et al. (2018) employ the ARCH and GARCH models.

Following the Kratz et al. (2018) notation, ES can be calculated as in Acerbi et al. (2002)

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 q_u(L)\,du. \tag{8}$$

A simple approximation can be obtained from different quantiles (Emmer et al., 2015)

$$\mathrm{ES}_\alpha(L) \approx \frac{1}{4}\left[q_\alpha(L) + q_{0.75\alpha+0.25}(L) + q_{0.5\alpha+0.5}(L) + q_{0.25\alpha+0.75}(L)\right],$$

where $q_\alpha(L) = \text{VaR}_\alpha(L)$. Kratz et al. (2018) then propose backtesting for ES by simultaneously backtesting multiple VaR estimates. Backtesting is based on multinomial tests of VaR exceptions. It is worthwhile to mention that the approximation can be generalized as

$$\text{ES}_\alpha(L) \approx \frac{1}{N}\sum_{i=1}^{N} q_{\alpha+\frac{i-1}{N}(1-\alpha)}(L), \tag{9}$$

where N is the number of quantiles to be used in the approximation. Although a higher N results in a better estimation of ES, simulations performed by Kratz et al. (2018) show that four quantiles provide reasonable size and power for the backtest. It is also noteworthy that the previous notation implies that risk measures are calculated over the loss distribution, i.e., the right tail of the distribution.

The number of exceptions (violations) are estimated given a certain model (distribution) and for each confidence level. As is typical in the literature, the exception indicator at each time t is defined as a function that takes value 1 if a loss has exceeded the VaR level. That is,

$$I_{t,i} = I_{\{L_t > \text{VaR}_{\alpha_i,t}\}}, \tag{10}$$

where α_i is as follows:

$$\alpha_i = \alpha + \frac{i-1}{N}(1-\alpha), \tag{11}$$

for $i = 1, \dots, N$, with $\alpha_0 = 0$ and $\alpha_{N+1} = 1$.

Then, the number of exceptions X_t at each time t is given by

$$X_t = \sum_{j=1}^{N} I_{t,i}. \tag{12}$$

As the number of exceptions follows a multinomial distribution, the unconditional coverage property can be written as

$$X_t \sim \text{MN}\big(1, (\alpha_1 - \alpha_0, \dots, \alpha_{N+1} - \alpha_N)\big), \tag{13}$$

Our interest is a measure that counts the outcomes $\{0,1,\dots,N\}$ with probabilities $\alpha_1 - \alpha_0, \dots, \alpha_{N+1} - \alpha_N$ that sum to one. The cell counts O_j are then given by

$$O_j = \sum_{t=1}^{n} I_{\{X_t = j\}}, \tag{14}$$

where n is the backtesting period, and $j = 0,1,\dots,N$. Then, the random vector should follow the multinomial distribution

$$(O_0, \dots, O_N) \sim \mathrm{MN}\big(n, (\alpha_1 - \alpha_0, \dots, \alpha_{N+1} - \alpha_N)\big). \tag{15}$$

The null and alternative hypotheses are given by

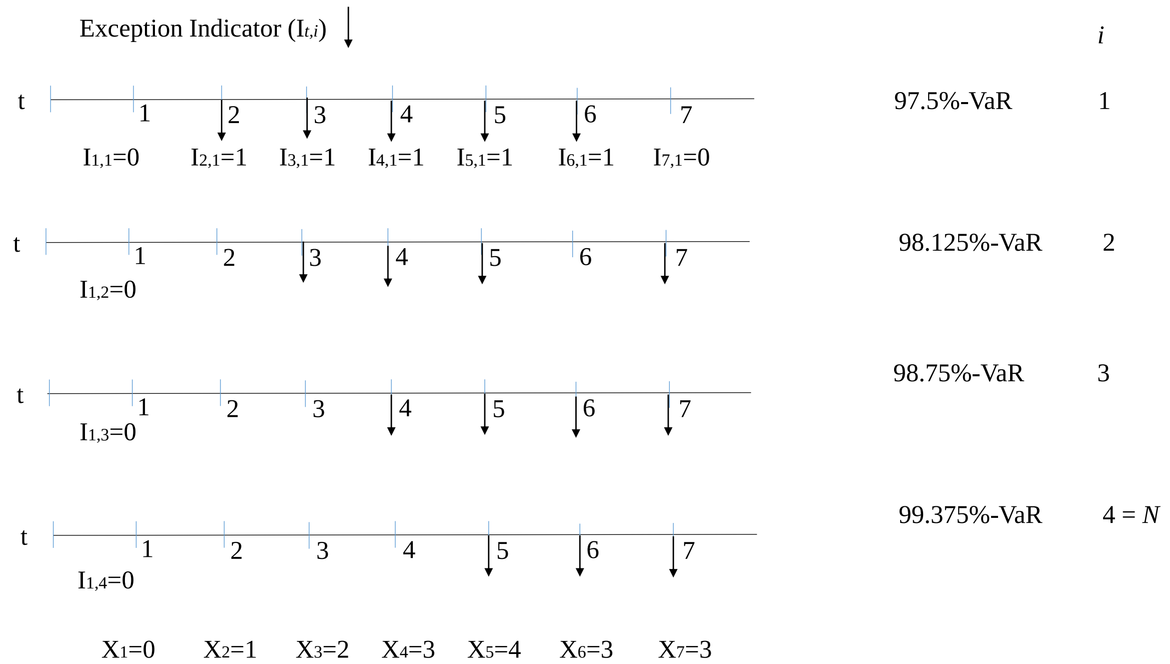
H0: $\theta_i = \alpha_i$ for $i = 1, \dots, N$

H1: $\theta_i \neq \alpha_i$ for at least one $i \in \{1, \dots, N\}$,

where $0 = \theta_0 < \theta_1 < \cdots < \theta_{N+1} = 1$ is an arbitrary sequence of parameters from a specific model, and $(O_0, \dots, O_N) \sim \mathrm{MN}\big(n, (\theta_1 - \theta_0, \dots, \theta_{N+1} - \theta_N)\big)$ (Kratz et al., 2018).

Figure 3 illustrates an example for our case (97.5%-ES and $N = 4$).

Figure 3. Graphical representation for multiple VaR backtesting



For this case, the number of exceptions X_t for each t is calculated as

$$X_{t=1} = I_{1,1} + I_{1,2} + I_{1,3} + I_{1,4} = 0,$$

$$X_{t=2} = I_{2,1} + I_{2,2} + I_{2,3} + I_{2,4} = 1,$$

$$\vdots$$

$$X_{t=7} = I_{7,1} + I_{7,2} + I_{7,3} + I_{7,4} = 3. \tag{16}$$

The cell counts O_j are given by

$$O_0 = \sum_{t=1}^{n} I_{\{X_t=0\}} = 1, \text{ since } X_1 = 0,$$

$$O_1 = \sum_{t=1}^{n} I_{\{X_t=1\}} = 1, \text{ since } X_2 = 1,$$

$$O_2 = \sum_{t=1}^{n} I_{\{X_t=2\}} = 1, \text{ since } X_3 = 2,$$

$$O_3 = \sum_{t=1}^{n} I_{\{X_t=3\}} = 3, \text{ since } X_4, X_6, X_7 = 3,$$

$$O_4 = \sum_{t=1}^{n} I_{\{X_t=4\}} = 1, \text{ since } X_5 = 4. \tag{17}$$

There are several multinomial tests; the most common is the Pearson chi-squared test, for which the test statistic S_N follows a χ_N^2 distribution under the null hypothesis:

$$S_N = \sum_{j=0}^{N} \frac{\left(O_{j+1} - n[\alpha_{j+1} - \alpha_j]\right)^2}{n[\alpha_{j+1} - \alpha_j]} \sim \chi_N^2. \tag{18}$$

The null hypothesis is rejected at a prespecified type I error κ when $S_N > \chi_N^2(1-\kappa)$.

Another test is the Nass test, which is an improvement over the previous test when cell probabilities are small (Kratz et al., 2018). The test statistic is χ_υ^2 distributed under the null hypothesis:

$$\frac{2\mathrm{E}(S_N)}{\mathrm{var}(S_N)} S_N \sim \chi_\upsilon^2, \tag{19}$$

where $\mathrm{E}(S_N) = N$, $\mathrm{var}(S_N) = 2N - \frac{N^2+4N+1}{n} + \frac{1}{n}\sum_{j=0}^{N} \frac{1}{[\alpha_{j+1} - \alpha_j]}$ and $\upsilon = \frac{2\left(\mathrm{E}(S_N)\right)^2}{\mathrm{var}(S_N)}$.

The null hypothesis is rejected at a prespecified type I error κ when $\frac{2\mathrm{E}(S_N)}{\mathrm{var}(S_N)} S_N >$
$\chi_\upsilon^2(1-\kappa)$.

2.4 Data

The idea of the empirical section is to compare the validation of the risk model of investments that only include, or alternately exclude, shares of companies related to the fossil fuel industry. Therefore, our data comprise two sets of daily prices detailed as follows. We prepare one of the datasets to consider the companies that are not exposed to unsustainable assets. We refer to these as the sustainable index (SI). It should capture the stock return behavior of sustainable companies. This first set of data corresponds to FTSE Developed ex Fossil Fuel Total Return Index. This index is a part of the Sustainability and Environmental, Social and Governance (ESG) indexes of FTSE Russell (other indexes with similar characteristics can be found at its website). This index is designed to represent the performance of FTSE All-World Index constituents after the exclusion of companies that have some exposure of revenues and/or reserves to fossil fuels. The second set of data is obtained from the S&P 500 Oil, Gas and Consumable Fuels Index.[19] This index includes companies in the energy sector engaged in the exploration, production, refining, marketing, storage and transportation of oil, gas, coal and consumable fuels. It is used as proxy of the whole oil and gas industry and is referred to as the traditional oil and gas index (TI).

Both indexes are capitalization-weighted and enable us to study the risk of investing in broadly diversified portfolios that include or exclude the traditional energy industry. The price data comprise information from July 31, 2006 to November 16, 2018 for a total of 3,210 price observations.[20] Moreover, we are interested in the effect of variability of main stranded asset price returns in the variance behavior of the indexes. To this end, prices of oil, gas and coal have been collected for the same period of SI and TI indexes.[21] The abovementioned data are obtained from Bloomberg platform.

Table 6. Descriptive statistics for daily stock returns of the Portfolio SI (sustainable industry) and Portfolio TI (traditional oil and gas industry)

Statistics	SI Index	TI Index	Oil returns	Gas returns	Coal returns
Mean	0.024	0.007	-0.008	-0.019	0.016
Median	0.070	0.000	0.000	0.000	0.000
Standard deviation	1.002 (15.85)	1.638 (25.91)	2.324 (36.75)	3.034 (47.97)	1.500 (23.72)
Skewness	-0.489	-0.261	0.134	0.617	0.697
Excess Kurtosis	9.003	13.747	4.968	5.873	44.624
Minimum	-6.785	-16.294	-13.065	-18.054	-20.729
Maximum	8.648	17.208	16.410	26.771	23.841

Annual volatility in parentheses

[19] "Standard and Poor's 500 Oil, Gas and Consumable Fuels Index is a capitalization-weighted index. The index was developed with a base level of 10 for the 1941-43 base period. The parent index is SPXL3. This is a GICS Level 3 Industries. Standard and Poor's 500 (Industry) Index is a capitalization-weighted index. The index is designed to measure performance of the broad domestic economy through changes in the aggregate market value of 500 stocks representing all major industries. The index was developed with a base level of 10 for the 1941-43 base periods." Source: Bloomberg LP.

[20] The selection of the indexes and period is restricted to availability of data from Bloomberg platform.

[21] The Generic 1st 'CL' Futures (CL1), Generic 1st 'NG' Futures (NG1) and Richards Bay Coal Futures (XO1) are obtained for oil, gas and coal prices, respectively.

The descriptive statistics show that analyzed stock index returns exhibit very well-known stylized facts for financial asset daily returns. The mean and median returns of the TI are close to zero, but the SI shows a mean of 0.024% and a median of 0.071%. In terms of daily volatility, the TI shows a standard deviation which is approximately two-thirds greater than the SI. The index returns distributions display fat tails, since excess kurtosis is higher than 0. Moreover, the distributions are negative skewed, which implies more negative extreme values. Figure 4 and 5 depict the returns for both indexes. The index returns exhibit similar characteristics and are remarkably affected by subprime crisis, as exhibited by the high volatility in approximately 2007 and 2008, and the financial problems faced by most companies in oil and gas industry. For the stranded asset returns, gas presents the higher volatility, whereas coal exhibit fatter tails among the three assets, and all the analyzed asset returns display positive skewed distributions.

Figure 4. Sustainable Index (SI) returns

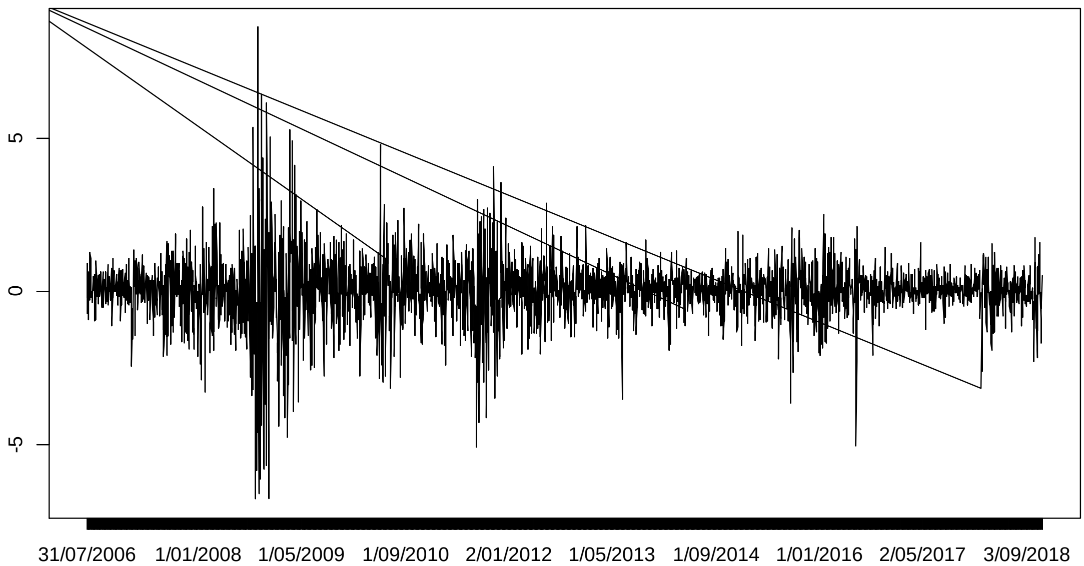
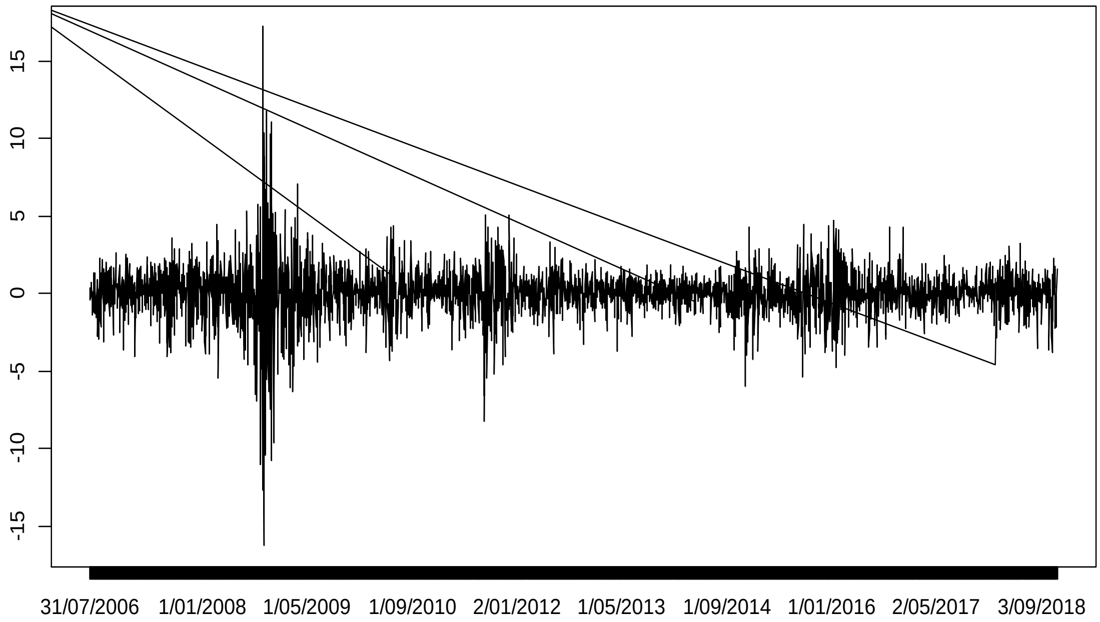
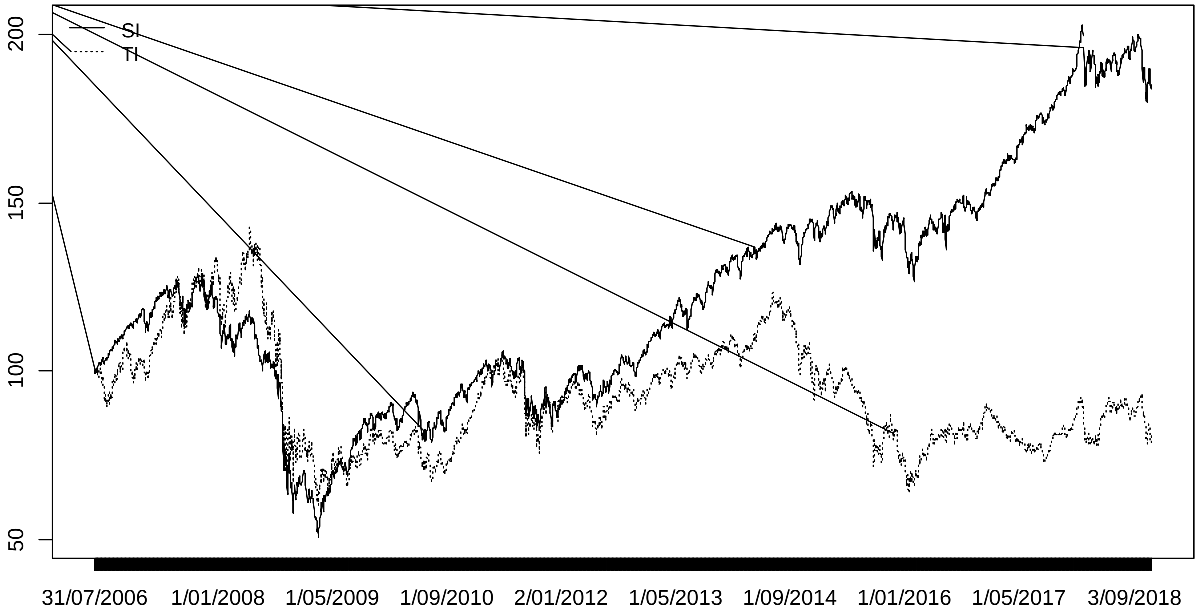


Figure 5. Traditional oil and gas industry Index (TI) returns



To show the temporal evolution of accumulated returns during the sample period, Figure 6 depicts the value of an initial investment of 100 on each of the indexes on July 31, 2006. In the financial crisis period, TI investment value is higher than SI investment. However, after such period, particularly since 2012, SI investment has clearly outperformed TI investment. According to the compound annual growth rate (CAGR), the performance of SI investment is better off than TI investment during the analyzed period. Considering a year of 252 days (12.7 years for 3,210 days) and final values (on November 16, 2018) for SI and TI of 184.95 and 80.89 respectively, the CAGR for SI and TI investments are 4.95% and −1.65% respectively.

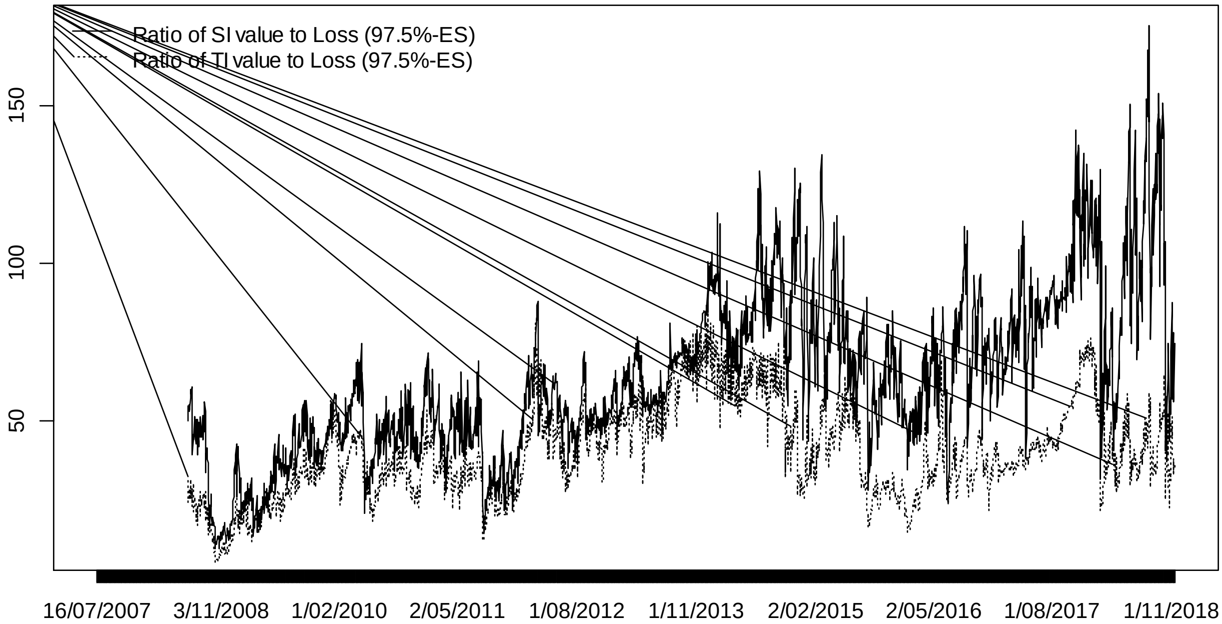
Figure 6. Value of an initial investment of 100 on each index



The relationship between risk, as measured by standard deviation, and the rate of return of the traditional oil and gas industry index shows a worse performance than that observed for the sustainable asset index. Alternatively, we also analyze the risk-return combination considering ES as the measure of risk. The relation value of SI investment to potential loss outperforms the relation of TI investment to loss during the analyzed period, when loss is estimated based on 97.5%-ES.[22] Figure 7 shows the evidence abovementioned.

[22] Values of 97.5%-ES estimated for Stable model are employed when external regressors are considered in the GARCH model. Thus, potential loss is calculated as $Loss_t = ES_t^{97.5}(Index\ Value_t)$.

Figure 7. Ratio of value index to potential loss



Beyond other considerations relating to climate change awareness or the risk of further regulation of the fossil fuel industry, i.e. from a strictly financial point of view, the global sustainable portfolio excluding fossil fuel industry assets performs better than the portfolio of assets related to the traditional oil and gas industry over the last decade. Anyway, this relatively poor performance on the traditional oil and gas industry assets is not necessarily a bad result. An interesting feature for portfolio managers is the outperformance of this index during the crisis and subsequent period of uncertainty. During these periods, the correlation of these assets with the rest of the market tends to decrease, being very low or even negative. Therefore, they are assets to be included in a portfolio to diversify the overall risk.

2.5 Empirical Results

This section presents the results of computing VaR and ES from both datasets, SI and TI stock indexes, and provides the ES backtesting analysis in comparison with traditional backtesting methods for VaR. The backtesting is classified in two cases. One method considers the ARMA-GARCH model with different innovations presented in Section 3.1. to filter the returns of SI and TI indexes, whereas the second method includes the variance of stranded asset (oil, gas and coal) returns as independent variables in the variance equation of GARCH model to filter the returns of the indexes. In-sample estimation results for the whole period are presented in Table 7. The results show that the parameters related to oil, gas and coal variances are statistically not significant. In other words, variance of stranded assets does not seem to have explanatory power in the variance of SI and TI returns. In fact, the estimation of ARMA-GARCH parameters when the variances of unsustainable assets are not included in the variance equation does not vary significantly comparing the estimation when the external regressors are taken into account. In what follows, we analyze whether the inclusion of these regressors help improve the backtesting results.

Table 7. In-Sample results of ARMA-GARCH model fit to the analyzed indexes

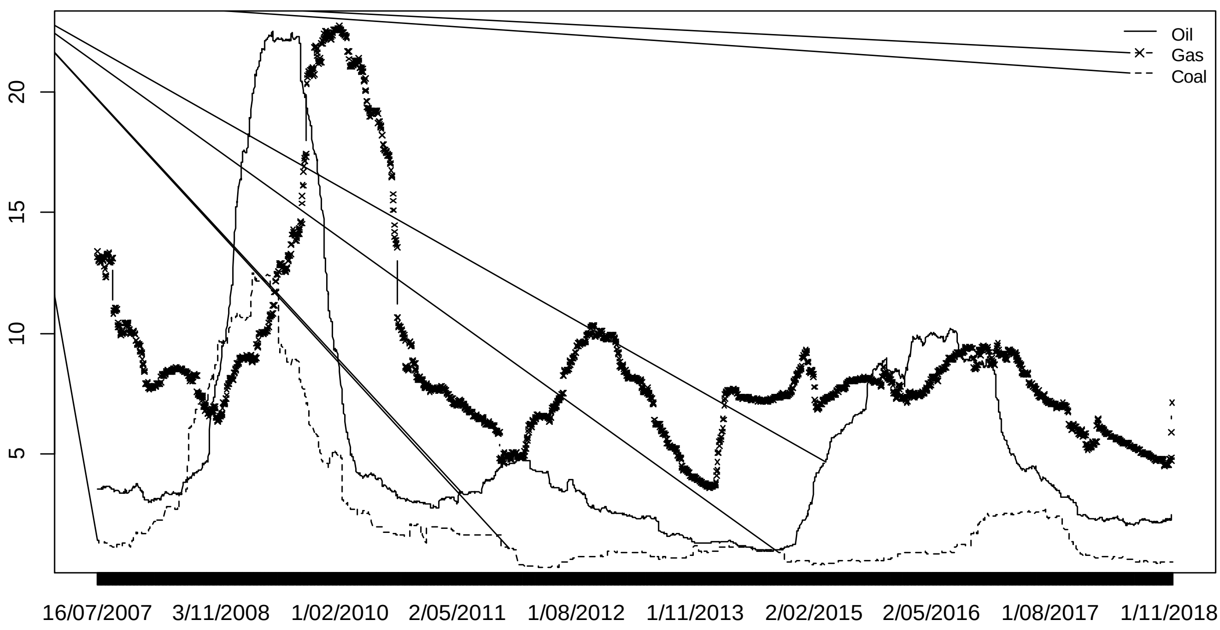
Panel A: ARMA-GARCH Estimation		
ARMA-GARCH with Gaussian innovations	TI Index	SI Index
θ_0	0.036 (0.067)	0.058 (0.000)
θ_1	-0.082 (0.887)	0.023 (0.857)
θ_2	0.028 (0.961)	0.101 (0.435)
β_0	0.023 (0.000)	0.009 (0.000)
β_1	0.084 (0.000)	0.105 (0.000)
β_2	0.907 (0.000)	0.888 (0.000)
ARMA-GARCH with Student's t innovations	TI Index	SI Index
θ_0	0.050 (0.007)	0.071 (0.000)
θ_1	-0.040 (0.938)	-0.067 (0.656)
θ_2	-0.013 (0.980)	0.188 (0.206)
β_0	0.019 (0.002)	0.006 (0.004)
β_1	0.080 (0.000)	0.101 (0.000)
β_2	0.913 (0.000)	0.898 (0.000)
ν	6.923 (0.000)	5.414 (0.000)
Panel B: ARMA-GARCH Estimation Considering External Regressors in the variance equation		
ARMA-GARCH with Gaussian innovations	TI Index	SI Index
θ_0	0.035 (0.667)	0.059 (0.000)
θ_1	-0.081 (0.887)	0.024 (0.855)
θ_2	0.027 (0.961)	0.102 (0.429)
β_0	0.023 (0.003)	0.005 (0.144)
β_1	0.083 (0.000)	0.109 (0.000)
β_2	0.907 (0.000)	0.881 (0.000)
γ_1	0.000 (0.999)	0.000 (0.425)
γ_2	0.000 (0.999)	0.000 (0.277)
γ_3	0.000 (0.999)	0.000 (0.999)
ARMA-GARCH with Student's t innovations	TI Index	SI Index
θ_0	0.050 (0.007)	0.071 (0.000)
θ_1	-0.040 (0.938)	-0.068 (0.652)
θ_2	-0.013 (0.980)	0.189 (0.205)
β_0	0.019 (0.003)	0.006 (0.093)
β_1	0.080 (0.000)	0.103 (0.000)
β_2	0.913 (0.000)	0.898 (0.000)

γ_1	0.000 (0.999)	0.000 (0.999)
γ_2	0.000 (0.999)	0.000 (0.999)
γ_3	0.000 (0.999)	0.000 (0.999)
ν	6.923 (0.000)	5.311 (0.000)

P-values in parentheses

The first 250 returns of oil, gas and coal assets are employed to obtain the first set of values of their respective variances, and a rolling window of 250 observations is implemented to estimate the rest of the variances. That is, the initial range of data from August 1, 2006 to July 15, 2007 is employed in order to calculate the variances that act as external regressors. Figure 8 shows the estimated variance of the stranded asset returns. Variability of oil and coal returns were mainly affected by the subprime crisis, and high volatility in the gas returns is observed posterior that date. The correlation between logreturns (estimated variance of logreturns) between oil and gas is 0.18 (0.40), for oil and coal is 0.16 (0.78), and for gas and coal is 0.02 (0.32) for the analyzed period.[23]

Figure 8. Variance of Oil, Gas and Coal price returns



The one-day-ahead VaR and ES are calculated by also implementing a rolling window of 250 observations, then the initial window size is ranged from July 16, 2007 to June 29, 2008. The backtesting period for both analyzed indexes (SI and TI) ranged from June 30, 2008 to

[23] The period of July 16, 2007 and November 16, 2018 is considered to estimate the correlations among asset logreturns.

November 16, 2018, and its length period is 2709 days. Thus, the expected number of exceptions is 27.09 (approximated to 27 in Table 3 and 4) for both indexes when calculating 99%-VaR.

Table 8 presents the results of testing VaR and ES for both the SI and the TI when the variance of unsustainable assets is not considered in the GARCH model. We compute the well-known binomial tests for VaR at 99% and the two new multinomial tests, i.e. the Pearson and Nass statistics, proposed by Kratz et al. (2018), for 97.5%-ES backtesting. In the case of the sustainable index, the binomial test for VaR rejects the Student's t and GPD models, since both models overpredict risk for the index returns. In most applications of market risk quantification, results of EVT techniques based on GPD model are favorable. However, in this case, the binomial test for 99%-VaR rejects the good performance of this model. A plausible reason is that the amount of observations (in the tail of the empirical distribution) employed to fit the GPD, which is 25 in each step of the rolling window. It is very well-known that parameter estimation depends on the threshold selection, which is still an open question in EVT, and this drawback is discussed for instance in Díaz et al. (2017). Backtesting ES of the SI, the results of Pearson and Nass tests do not reject the good performance of normal, Stable and GPD models, but Student's t model does not perform satisfactorily, which is consistent with backtesting of VaR results for the same index (Table 8, Panel a).

In the case of the traditional oil and gas industry index returns, only the Student's t model does not perform well according to the binomial test for 99%-VaR (Table 3, Panel b), whereas all the models perform well for 97.5%-ES backtesting, according to Pearson and Nass statistics.

Table 8. Comparison of 99%-VaR and 97.5%-ES (implicit) backtesting for the Sustainable Index (SI) and the Traditional Oil and Gas Industry (TI)

Model	99% VaR		97.5% ES						
	EE	Violations	O_0	O_1	O_2	O_3	O_4	Pearson	Nass
Panel a) Sustainable Index (SI):									
Normal	27	33 (0.270)	2644	13	17	10	25	7.60	7.39
Student's-t	27	11 (0.000)	2658	9	21	10	11	9.71	9.45
Stable	27	32 (0.357)	2631	21	19	16	12	2.84	2.76
GPD-POT	27	15 (0.011)	2654	17	20	10	8	8.17	7.94
Panel b) Traditional Oil & Gas Industry (TI):									
Normal	27	32 (0.357)	2648	9	17	14	21	5.22	5.07
Student's-t	27	14 (0.005)	2657	14	8	16	14	5.87	5.71
Stable	27	32 (0.357)	2638	9	21	17	24	7.65	7.44
GPD-POT	27	20 (0.151)	2660	13	13	9	14	6.18	6.01

EE stands for Expected Exceptions. The critical value for the Pearson test is 9.49, and that for the Nass test is 9.31. The P-value for the binomial test in parenthesis. O_j ($j = 0,1,2,3,4$) counts the times the number of exceptions (X_t) are equal to j for each time t and for all VaR levels. Backtesting periods include 2709 days for the SI and the TI portfolios.

Table 9 replicates analysis of Table 8, but the new analysis considers the external regressors in the equation of variance (Equation [1]). Although the binomial test for 99%-VaR still rejects the good performance of the Student's t model, all other models now exhibit a reasonable performance for VaR and ES tests. This is an important result, since there is evidence that employing external regressors (variance of stranded asset returns) help improve risk model validations for the analyzed data in our paper.

Table 9. Comparison of 99%-VaR and 97.5%-ES (implicit) backtesting for the Sustainable Index (SI) and the Traditional Oil and Gas Industry (TI). Considering external regressors in the variance equation of GARCH model.

Model	99% VaR		97.5% ES						
	EE	Violations	O_0	O_1	O_2	O_3	O_4	Pearson	Nass
Panel a) Sustainable Index (SI):									
Normal	27	33 (0.270)	2641	15	16	13	24	4.13	4.02
Student's-t	27	17 (0.036)	2655	12	21	10	11	7.40	7.20
Stable	27	32 (0.357)	2633	19	17	17	23	2.45	2.39
GPD-POT	27	20 (0.151)	2649	15	21	11	13	4.21	4.10
Panel b) Traditional Oil & Gas Industry (TI):									
Normal	27	35 (0.144)	2643	11	18	14	23	4.83	4.70
Student's-t	27	23 (0.417)	2650	18	10	18	13	3.91	3.81
Stable	27	34 (0.199)	2630	15	21	18	25	5.16	5.02
GPD-POT	27	25 (0.683)	2655	14	10	11	19	5.75	5.59

EE stands for Expected Exceptions. The critical value for the Pearson test is 9.49, and that for the Nass test is 9.31. The P-value for the binomial test in parenthesis. O_j ($j = 0,1,2,3,4$) counts the times the number of exceptions (X_t) are equal to j for each time t and for all VaR levels. Backtesting periods include 2709 days for the SI and the TI portfolios.

We also conduct the simple backtest of ES, commonly used in the literature, as a robustness check of the new multinomial test previously applied to validate ES. Table 10 shows the results of independent individual binomial backtests of VaR for four confidence levels equal to and higher than that used in the 97.5% ES estimate. This analysis is performed for both indexes without considering external regressors in variance equation. This methodology based on independent testing for different confidence levels provides results similar to those obtained in the multinomial test for three loss distributions. However, it indicates that the GPD model overpredicts risk when VaR is calculated at 98.75% and 99.375% (97.5% and 98.125%) confidence levels for SI (TI) index return as can be seen in Table 10, Panel a (Panel b). Anyway, the results for the Pearson and Nass tests for 97.5%-ES displayed in Table 8, which rejects the Student's t model and shows that the normal and Stable models perform well for both indexes, are also confirmed by the individual binomial tests.

Table 10. Exceptions obtained for each VaR level for the Sustainable Index (SI) and the Traditional Oil and Gas Index (TI)

Model	97.5%VaR	98.125%VaR	98.75%VaR	99.375%VaR
Panel a) Sustainable Index (SI):				
	[52;84] EE = 68	[37;65] EE = 51	[23;45] EE = 34	[9;25] EE = 17
Normal	65	52	35	25
Student's-t	51	42	21	11
Stable	78	57	38	22
GPD-POT	55	38	18	8
Panel b) Traditional Oil & Gas Industry (TI):				
	[52;84] EE = 68	[37;65] EE = 51	[23;45] EE = 34	[9;25] EE = 17
Normal	61	52	35	21
Student's-t	52	38	30	14
Stable	71	62	41	24
GPD-POT	49	36	23	14

EE stands for Expected Exceptions. 95% confidence interval in brackets for each expected number of exceptions. Backtesting periods include 2709 days for the SI and the TI portfolios.

Table 11 presents same results as Table 10 considering variance of unsustainable asset returns as independent variables in the variance equation of the GARCH model. Only Student's t model is rejected when 98.75%-VaR is calculated for SI index returns. Once again, the risk model performance is better off when the variances of stranded asset returns are included as regressors in the variance equation to assess VaR at different confidence levels.

Table 11. Exceptions obtained for each VaR level for the Sustainable Index (SI) and the Traditional Oil and Gas Index (TI). Considering external regressors in the variance equation of GARCH model.

Model	97.5%VaR	98.125%VaR	98.75%VaR	99.375%VaR
Panel a) Sustainable Index (SI):				
	[52;84] EE = 68	[37;65] EE = 51	[23;45] EE = 34	[9;25] EE = 17
Normal	68	53	37	24
Student's-t	54	42	21	11
Stable	76	57	40	23
GPD-POT	60	45	24	13

Panel b) Traditional Oil & Gas Industry (TI):

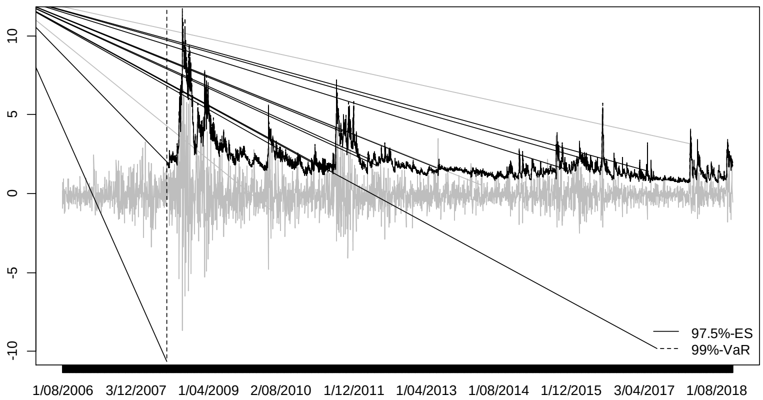
	[52;84] EE = 68	[37;65] EE = 51	[23;45] EE = 34	[9;25] EE = 17
Normal	66	55	37	23
Student's-t	59	41	31	13
Stable	79	64	43	25
GPD-POT	54	40	30	19

EE stands for Expected Exceptions. 95% confidence interval in brackets for each expected number of exceptions. Backtesting periods include 2709 days for the SI and the TI portfolios.

Finally, Figure 9 shows the comparison of 99%-VaR and 97.5%-ES (with external regressors in the variance equation) for each analyzed model applied to SI returns.[24] As expected, 99%-VaR is similar to 97.5%-ES for the Gaussian case; however, it is noted that 97.5%-ES is higher than 99%-VaR for stable and GPD cases. This result corroborates one of the arguments used by the Basel Committee to defend the use of ES to calculate the market risk of a financial institution.

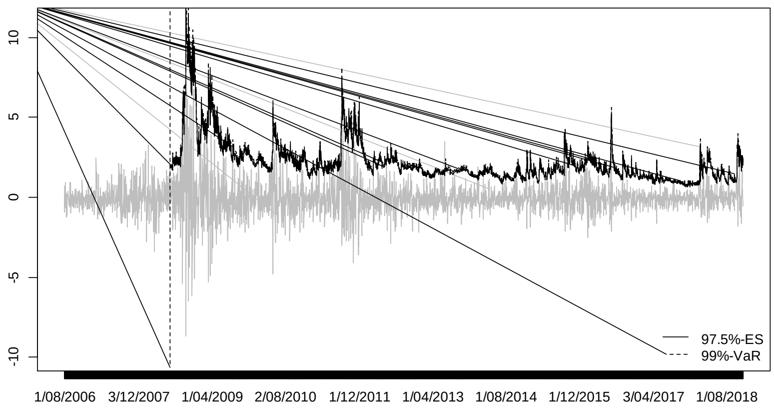
Figure 9. Comparison of 99%-VaR and 97.5%-ES for each model for SI negative log-returns

Panel A: 99%-VaR and 97.5%-ES for Gaussian model

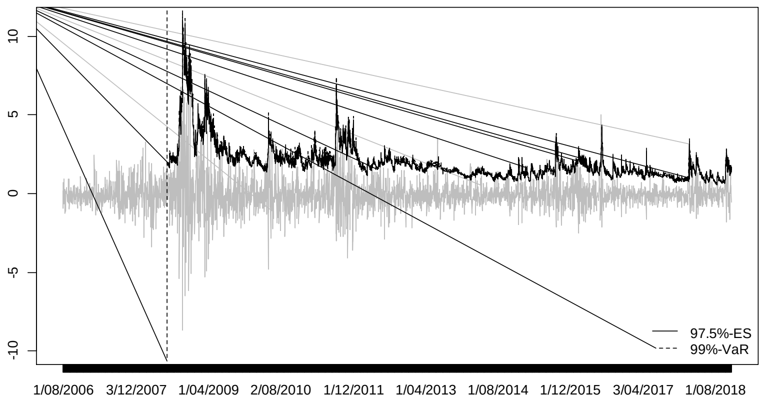


Panel B: 99%-VaR and 97.5%-ES for Student's t model

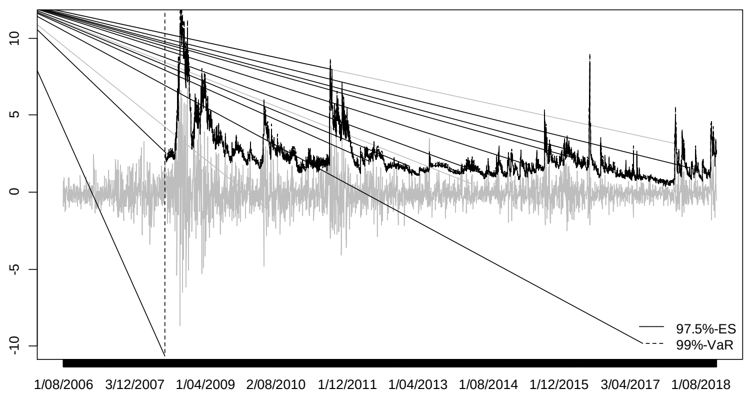
[24] Similar results are obtained for Portfolio TI returns and are available upon request, and for both indexes when external regressors are not included in the variance equation in GARCH model.



Panel C: 99%-VaR and 97.5%-ES for stable models



Panel D: 99%-VaR and 97.5%-ES for POT model



2.6 Conclusions of Chapter 2

Energy assets are a class of assets widely used by portfolio and risk managers for diversification purposes, although these assets have higher volatilities than other types of stocks and are affected by a trend towards divestment caused by greater global awareness of the environmental and financial impact of climate change. For this reason, we analyze the appropriate quantification of risk for portfolios formed by companies in the traditional energy sector and compare them with portfolios that exclude these assets. The validation of ES as a risk measure recently proposed by financial regulators is a topical issue and its implementation on highly volatile assets is an experiment of maximum interest. Thus, in this paper we shed some light on how to solve the problem of ES backtesting.

We consider two broadly diversified stock indexes, one of fossil fuel-related companies and another that includes sustainable companies from all sectors except these traditional energy companies. The sustainability stock index financially outperforms the traditional energy stock index, although the latter offers good results in times of turmoil and crisis in the markets. This is undoubtedly an interesting feature for portfolio and risk managers.

We examine several distributions to perform the ES backtesting procedure proposed by Kratz et al. (2018) when applied to both stock indexes. The validation of models shows that including the variances of the stranded asset (oil, gas and coal) returns as external regressors in the variance equation of GARCH model help improve the backtesting results of the analyzed models. In general, Stable model performs well in all cases. Most studies that employ EVT techniques based on GPD model show adequate results according to backtesting results in market risk quantification. However, in our application when window size is equal to 250 observations, the latter model does not perform well for 99%-VaR backtesting in sustainable index. Nevertheless, when external regressors are considered in the variance equation, the GPD model performs satisfactorily.

The backtesting procedure is based on multinomial tests for different VaR levels rather than performing a binomial test for each VaR level, since ES can be approximated in terms of multiple VaRs (Emmer et al., 2015). We obtain evidence that the multinomial test is an adequate and simple method to validate ES models as presented in this paper. This simple approach leads to an implicit manner for ES backtesting, and it is suggested for regulatory purposes. The current regulation proposes performing VaR backtesting at 97.5% and 99% confidence levels to validate ES calculations. The latter approach was an initial solution for the non-elicitability of ES, which made many in academia and the financial industry think that ES could not be backtested. Future research can be conducted to compare other ES tests such as those proposed by Acerbi et al. (2012 and 2014).

References

Alizadeh, A.H., Nomikos, N.K., Pouliasis, P.K., 2008. A Markov regime switching approach for hedging energy commodities. J. Bank. Financ. 32, 1970–1983.

Adams, C., Wilson, J., Vandevelde, M. (2016). Moody's puts 175 energy and mining companies on downgrade watch. Financial Times.

Acerbi, C., Székely, B., 2014, Back-testing expected shortfall. Risk Magazine, December.

Acerbi, C., Tasche, D., 2002, On the coherence of expected shortfall. Journal of Banking and Finance, 26, 1487–1503.

Aloui, R., Aïssa, M.S.B., Hammoudeh, S., Nguyen, D.K., 2014, Dependence and extreme dependence of crude oil and natural gas prices with applications to risk management. Energy Economics, 42, 332-342.

Araújo Santos, P., Fraga Alves, M.I., 2013, Forecasting Value-at-Risk with a duration-based POT method. Mathematics and Computers in Simulation, 94 (C), 295-309.

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

Bruno, S.V.B, Sagastizábal, C. (2010), Optimization of real asset portfolio using a coherent risk measure: application to oil and energy industries. Optim Eng, 12, pp. 257-275.

Bunn, D., Andresen, A., Chen, D., Westgaard, S., 2016, Analysis and Forecasting of Electricity Price Risks with Quantile Factor Models, The Energy Journal, 37 (1), 101-122.

Cabedo, J. D., & Moya, I. (2003). Estimating oil price 'value at risk' using the historical simulation approach. Energy Economics, 25, 239–253.

Chai, S., Zhou, P., 2018, The Minimum-CVaR strategy with semi-parametric estimation in carbon market hedging problems. Energy Economics, 76, 64-75.

Chavez-Demoulin, V., 1999, Two problems in environmental statistics: Capture-recapture analysis and smooth extremal models. Ph.D. thesis. Department of Mathematics, Swiss Federal Institute of Technology, Lausanne.

Chiu, Y-C., Chuang, I-Y., Lai, J-Y., 2010, The performance of composite forecast models of value-at-risk in the energy market. Energy Economics, 32 (2), 423-431.

Christoffersen, P., 1998, Evaluating interval forecasts. International Economic Review 39 (4), 841–862.

Conlon, T., Cotter, J., 2013. Downside risk and the energy hedger's horizon. Energy Econ. 36, 371–379.

Costello, A., Asem, E., Gardner, E., 2008, Comparison of historically simulated VaR: Evidence from oil prices, 30 (5), 2154-2166.

Del Brio, E.B., Mora-Valencia, A., Perote, J., 2017, Risk quantification for commodity ETFs: Backtesting value-at-risk and expected shortfall. International Review of Financial Analysis (forthcoming)

Díaz, A., García-Donato, G., Mora-Valencia, A., 2017, Risk quantification in turmoil markets. Risk Management, 19 (3), 202-224.

Emmer, S., Kratz, M., Tasche, D., 2015, What is the best risk measure in practice?, Journal of Risk, 18, 31–60.

Ewing, B. T., & Malik, F. (2010). Estimating volatility persistence in oil prices under structural breaks. Financial Review, 45, 1011–1023.

Ewing, B.T., Malik, F., Anjum, H. (2018). Forecasting value-at-risk in oil prices in the presence of volatility shifts. Review of Financial Economics forthcoming.

Fan, Y., Zhang, Y-J., Tsai, H-T., Wei, Y-M., 2008, Estimating 'Value at Risk' of crude oil price and its spillover effect using the GED-GARCH approach. Energy Economics, 30 (6), 3156-3171.

Feng, Z.H., Wei, Y.M., Wang, K., 2012. Estimating risk for the carbon market via extreme value theory: an empirical analysis of the EU ETS. Appl. Energy 99, 97–108.

Fissler, T., Ziegel, J.F., 2016, Higher order elicitability and Osband's principle. The Annals of Statistics, 44 (4), 1680–1707.

Ghorbel, A., Trabelsi, A., 2014, Energy portfolio risk management using time-varying extreme value copula methods. Economic Modelling, 38, 470-485.

Gneiting, T., 2011, Making and evaluating point forecasts. Journal of the American Statistical Association, 106, 746-762.

González-Pedraz, C., Moreno, M., Peña, J.I., 2014, Tail risk in energy portfolios. Energy Economics, 46, 422-434.

He, K., Li, K.K., Yen, J., 2011, Value-at-risk estimation of crude oil price using MCA based transient risk modeling approach. Energy Economics, 33 (5) 903-911.

Hemmati, R., Saboori, H., Saboori, S. (2016). Stochastic risk-averse coordinated scheduling of grid integrated energy storage units in transmission constrained wind-thermal systems within a conditional value-at-risk framework. Energy, 113, 762-775.

Herrera, R., 2013, Energy risk management through self-exciting marked point process. Energy Economics, 38, 64-76.

Huang, D., Yu, B., Fabozzi, F.J., Fukushima, M., 2008, CAViaR-based forecast for oil price risk. Energy Economics, 31 (4), 511-518.

Hung J.C., Lee M.C., Liu H.C. (2008), Estimation of value-at-risk for energy commodities via fat-tailed GARCH models. Energy Economics, 30 (3), 1173-1191.

Jäschke, S., 2014, Estimation of risk measures in energy portfolios using modern copula techniques. Computational Statistics & Data Analysis, 76, 359-376.

Kratz, M., Lok, Y., McNeil, A.J., 2018, Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. Journal of Banking & Finance 88(C), 393-407.

Lu, X.F., Lai, K.K., Liang, L., 2014, Portfolio value-at-risk estimation in energy futures markets with time-varying copula-GARCH model. Annals of Operations Research, 219, 333–357.

Lu, Z.G., Qi, J.T., Wen, B., Li, X.P., 2016. A dynamic model for generation expansion planning based on conditional value-at-risk theory under low-carbon economy. Electr. Power Syst. Res. 141, 363–371.

Lux, T., Segnon, M., Gupta, R., 2016, Forecasting crude oil price volatility and value-at-risk: Evidence from historical and recent data. Energy Economics, 56, 117-133.

Lyu, Y., Wang, P., Wei, Y., Ke, R., 2017, Forecasting the VaR of crude oil market: Do alternative distributions help? Energy Economics, forthcoming.

Marimoutou, V., Raggad, B., Trabelsi, A., 2009, Extreme Value Theory and Value at Risk: Application to oil market. Energy Economics, 31 (4) 519-530.

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

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, 271-300.

Nolan, J.P., 2001, Maximum likelihood estimation and diagnostics for stable distributions, in E. Barndorff-Nielsen, T. Mikosch, S. Resnick (eds.), Lévy Processes, Brikhäuser, Boston.

Nomikos, N., Pouliasis, P., 2011, Forecasting petroleum futures markets volatility: The role of regimes and market conditions. Energy Economics, 33 (2), 321-337.

Roustai, M., Rayati, M., Sheikhi, A., Ranjbar, A., 2018. A scenario-based optimization of Smart Energy Hub operation in a stochastic environment using conditional-value-at-risk. Sustain. Cities Soc. 39, 309–316

Segnon, M., Lux, T., Gupta, R., 2017, Modeling and forecasting the volatility of carbon dioxide emission allowance prices: A review and comparison of modern volatility models. Renewable and Sustainable Energy Reviews, 69, 692-704.

Tekiner-Mogulkoc, H., Coit, D.W., Felder, F.A., 2015. Mean-risk stochastic electricity generation expansion planning problems with demand uncertainties considering conditional value-at-risk and maximum regret as risk measures. Electr. Power Energy Syst. 73, 309–317.

Wang, Y., Liu, L., Ma, F., & Wu, C. (2016). What the investors need to know about forecasting oil futures return volatility. Energy Economics, 57, 128–139.

Weber, S., 2006, Distribution-invariant risk measures, information, and dynamic consistency. Mathematical Finance, 16 (2) 419-441.

Youssef, M., Belkacem, L., Mokni, K., 2015, Value-at-Risk estimation of energy commodities: A long-memory GARCH–EVT approach. Energy Economics, 51, 99-110.

Zolotko, M., Okhrin, O., 2014, Modelling the general dependence between commodity forward curves. Energy Economics, 43, 284-296.

Chapter 3. Risk quantification for Commodity ETFs: Backtesting Value-at-Risk and Expected Shortfall[25]

Abstract

This paper studies the risk assessment of alternative methods for a wide variety of Commodity ETFs and portfolios based on them. We implement well-known, as well as recently proposed backtesting techniques, for both value-at-risk (VaR) and expected shortfall (ES) under parametric and semi-nonparametric techniques. The application of the latter to ES is introduced in this paper and for this purpose we derive a straightforward closed form of ES. We show that, for the confidence levels recommended by Basel Accords, Student's t and Gram-Charlier present the best relative performance for individual Commodity ETFs. Moreover, we show that multivariate semi-nonparametric distribution performs better than multivariate normal distribution for portfolios of Commodity ETFs. Hence, we recommend the application of the above mentioned leptokurtic distributions to mitigate regulation concerns about global financial stability and financialization of commodity business throughout an adequate risk assessment.

Keywords*:* Value-at-risk; Expected shortfall; Backtesting; Gram-Charlier expansion; SNP-DCC; Financialization.

3.1 Introduction of Chapter 3

Commodity prices experienced a boom in the mid-2000s triggered by the need of metal supplies (such as iron, aluminum, copper, steel, and zinc) and energy from China. During this period, large commodity traders in 2010 such as Vitol, Glencore or Cargill had benefited from the positive returns of commodities, typically 6-8% a year.

[25] A version of this chapter has been published in International Review of Financial Analysis Journal (univariate case), and European Journal of Finance (multivariate case) co-authored with Esther B. Del Brio and Javier Perote

Recently, there has been a sharp decline in the price of these assets because of both the increase in supply and the slower economic growth in emerging markets. By end of August 2015 all main commodities (energy, industrial metals and agriculture) prices had plunged in a 10-20% range year-to-date. Slowdown of Chinese investment has mainly affected copper producers (e.g., Chile), since China demands 45% of the output. As a consequence, by the end of September 2015 Glencore's share price had plummeted by almost a third. In 2013, Deutsche Bank and JP Morgan also suffered huge losses in based metals businesses. In addition, developed markets based on mining and energy industries like Canada and Australia have also been affected by the decline in commodity prices.

However, the volatility in the commodity market is not only affected by demand and supply of consumers and producers. Derivatives and Commodity Exchange Traded Funds (ETFs) have been employed by financial market participants to invest in commodity-linked instruments and its speculation has been the reason of the rapid increase of oil prices in 2007 and part of 2008 (Hume et al., 2016). In fact, in the last decades, the presence of capital markets and financial industry with the so-called "financialization" of commodity business has distorted commodity prices and contributed to the bubble formation in these markets.

With the boom of commodity prices, several ETFs were created in order to track the price of main commodities such as gold, silver and oil, and then allow investors to buy commodities as trading stocks. ETFs were created in the early 90s and can be defined as baskets of securities that are traded on a stock exchange like individual stocks. Initially, ETFs replicated equity products, but other types of assets such as fixed income, credit, emerging markets, commodities, among others, are also employed as underlying assets. The main advantage of these financial instruments is the low-cost diversification and easier accessibility of certain asset classes, but also they feature high liquidity and tradability.

On the other hand, commodity assets present higher volatilities than equity assets even in "relatively" calm periods, because commodities cannot be easily stored. The high volatility of the underlying assets also produces a higher uncertainty on the Commodity ETFs market. In other words, there is the so-called decay or slippage (as it is known in the industry jargon), which is the rapidly decline in ETFs price. For instance, gold ETFs fell closer to 33% in 2014, and most of sales in 2013 came from SPDR Gold Shares ETF (one of the analyzed data in this paper), mainly caused by the announcement of the Federal Reserve about reducing its quantitative easing plan and the global sentiment.[26] Financialization has helped to convert commodities into tradeable securities (e.g. Commodity ETFs), and thus has contributed to the recent decline in commodity prices and its high volatility (Tang and Xiong, 2012; Singleton, 2013 among others).

From the regulatory point of view, the Financial Stability Board (FSB) and Bank for International Settlements (BIS) published separated reports in April 2011 and expressed their concerns about the potential systemic impact of the ETFs industry. Possible financial stability risks are related to the complexity and relative opacity of these instruments, which make difficult the risk assessment according to the regulators' reports. Furthermore, the UN Conference on Trade and Development in 2015 has criticized the financialization of commodities highlighting its impact on speculative bubbles inflation.

Since 1996, Value-at-risk (VaR) has been the standard market risk measure to assess regulatory capital requirements. This risk measure, defined as the maximum loss given certain probability level and time horizon (usually 99% and one or ten days, respectively), has been criticized because it does not fulfill the 'subadditivity' property. This property states that the total risk of a portfolio should be less than or equal to the sum of risks of the individual portfolio assets, i.e. a 'coherent' risk measure should be consistent with risk diversification. For this reason, other coherent risk measures have been proposed, the Expected Shortfall (ES) or conditional VaR (CVaR), defined as the conditional expectation of losses for losses that has exceeded a certain VaR threshold, being the most widely used. As a matter of fact, the VaR's inability to accurately capture tail risk during the global financial crisis made the Basel Committee on Banking Supervision (BCBS) reconsider the use of VaR and publish the review of trading book rules – the so-called Basel 2.5 – on May 2012 with the aim to replace VaR by ES.

However, ES also feature disadvantages compared to VaR. By definition, ES is always higher than VaR at a given confidence level, and can be huge depending on the assumed distribution to fit the losses. Though many banks have been used ES to assess regulatory capital, ES is very sensitive to extreme events and may result in unstable capital numbers at high confidence levels – for this reason, ES is commonly computed at 97.5%. In addition, ES is not as intuitive as VaR, and thus its figures are not easy to explain to the board members. Furthermore, ES does

[26] Though Commodity ETFs have been used to amplify bets – especially on oil, gold and natural gas price, our study analyzes non-leveraged Commodity ETFs. Leveraged Commodity ETFs (LETFs) will be the focus in a future research.

not satisfy the ‘elicitability’ property – i.e the capacity of determination for optimal forecasts (Emmer et al., 2015), which generated a vivid debate on whether ES can be backtestable. On the contrary, VaR satisfies such property and there are very well-known backtesting methods for this measure. Therefore, the discussion resulted in a tradeoff between coherence and elicitability, the former being preferred by regulators. However, recent studies (see Section 2) have shown that ES can be backtestable, although the proposed methods are less appealing than the straightforward backtesting alternatives for VaR.

This paper contributes to the literature by providing a comparison between both risk measures (VaR and ES) through a wide variety of methodologies including parametric distributions (Gaussian, Student’s t and Skewed-t), semi-nonparametric distributions – a couple of expansions of Gram-Charlier (GC) series – in univariate and multivariate frameworks. In all these cases we implement backtesting techniques to study model performance and risk assessment of Commodity ETFs, designed in response to the concerns about global financial stability. So far, risk management studies have been devoted to commodities directly or commodity derivatives. Hence, and to our knowledge, this is the first paper which analyzes risk management applied to Commodity ETFs. Furthermore, we derive a closed form to calculate ES for the Gram-Charlier distribution, which facilitates the implementation of ES on this type of models. All in all, the contribution of this paper is three-fold. First, we propose a jointly (i.e. more efficient) estimation of Gram-Charlier parameters and ARMA-EGARCH process for risk calculation purposes. Previous studies have implemented two-step procedures and estimated in a first step an AR(MA)-GARCH process by Quasi Maximum Likelihood (QML) and the remaining density parameters in a second step. Second, we extend the study in Del Brio et al. (2011), by applying SNP-DCC in a risk measure framework. Third, the results of this work have interesting implications for practitioners and regulators, since our application implements appropriate backtesting techniques to evaluate the performance of ES for individual assets and portfolios, which may help to mitigate main concerns of regulators about commodity financialization.

The rest of the paper is organized as follows: Section 2 reviews the literature on risk measures and their applications to commodity assets, Section 3 describes the risk models and VaR and ES methodologies, Section 4 presents the empirical performance of the models for forecasting VaR and ES with a sample of six different Commodity ETFs, Section 5 compares the ES backtesting performance for three bivariate portfolios, Section 6 discusses the main results of this paper in light of the new regulation and Section 7 concludes.

3.2 Literature Overview

This section presents an overview of recent advances on risk measures focusing on their properties and applications to commodity markets.

VaR has been the standard risk measure for financial industry since mid-1990s and a vast literature has been devoted to it – see, e.g. Jorion (2006). The use of ES was proposed by Artzner et al. (1997) and its properties studied in Acerbi and Tasche (2002), and Rockafellar and Uryasev (2002). As discussed in the previous section, VaR is not a coherent risk measure since it does not satisfy subadditivity, unlike ES.[27] Empirical studies have shown that under extreme fluctuations VaR may underestimate risk but ES may overcome this shortcoming; however, ES strongly depends on the estimation accuracy (Yamai and Yoshiba, 2005). As a consequence, capital calculated from ES might be less stable than that of the VaR when the losses exhibit fat tails. This evidence is consistent with Cont et al. (2010), who find that ES is less robust and more data-sensitive than VaR. Nevertheless, under a new notion of (qualitative) robustness, Embrechts et al. (2015) show that ES is more robust than VaR.

Recently, the 'elicitability' of risk the measures has been discussed because of its importance for studying model performance in forecasting applications. Formally, a risk measure is elicitable if it minimizes a suitable expected scoring function (Bellini and Bignozzi, 2015). Gneiting (2011) shows that ES does not satisfy this property and thus finding accurate techniques for backtesting ES could be challenging. In this line, Acerbi and Székely (2014) provide three methods to backtest ES and show that ES can be used for model testing, but still elicitability is important to make comparisons of different models. Particularly, if elicitability is not satisfied there is no consistent scoring function and therefore it is not possible to establish which model has the best performance. Very recently, Fissler et al. (2016) propose two methods based on conditional elicitability of ES, one of them is applied in this paper.

On the other hand, Bellini et al. (2014) and Ziegel (2014) finds that expectiles (introduced by Newey and Powell, 1987) are both coherent and elicitable risk measures. Expectiles are derived by employing asymmetric least squares and can be used to estimate VaR and ES (Taylor, 2008).

[27] See Artzner et al. (1999) for more details about the properties that a risk measure should satisfy to be coherent.

This measure has the advantage of assuming a free distribution, but its concept is less intuitive than VaR and ES.

As discussed above, every risk measure has its own advantages and drawbacks. Emmer et al. (2015) overview the desired properties of risk measures (coherence, elicitability, robustness, comonotonicity and additivity, among others) and their impact on capital allocation. These authors conclude that ES is a good risk measure in practice, although the discussion of the best risk measure is still an open question and depends on the characteristics to be evaluated. For instance, in a recent work Koch-Medina and Munari (2016) warn about the use of ES for capital adequacy and portfolio risk measurement since ES does not necessarily guarantee protection of liability holders.

Recent studies on risk management applications to commodity markets have undertaken different approaches and have been applied to different assets (Giot and Laurent, 2003; Hung et al., 2008; Chiu et al., 2010; Aloui and Mabrouk, 2010; Youssef et al. (2015), Andriosopoulos and Nomikos, 2015; Aloui and Jammazi, 2015), although as far as we know they have never been used for modeling Commodity ETFs. All these papers, after testing a wide variety of models, point to the adequacy of different (asymmetric and long memory) GARCH models with (skewed) heavy-tailed distributions for risk assessment. Particularly, Giot and Laurent (2003) and Aloui and Mabrouk (2010) show that the APARCH (or FIAPARCH) model with skewed-t innovations outperforms other alternatives in several commodity and energy markets. Similarly, Youssef et al. (2015) support a FIAPARCH with EVT for estimating both VaR and ES – a circular bootstrap being employed to backtest ES. On the other hand, Hung et al. (2008) find that GARCH models with heavy-tailed innovations provide accurate risk measures for energy commodities and Steen et al. (2015) and Andriosopoulos and Nomikos (2015) show that quantile regression techniques and Monte Carlo simulations, respectively, outperform many other alternatives for modeling risk for most commodities and, particularly those of energy. Finally, other authors find evidence in favor Hull-White (Chiu et al., 2010) and three wavelet-based **(**Aloui and Jammazi, 2015**)** models for Brent and WTI crude oil prices.

Previous mentioned studies are related to an application for individual assets. Regarding portfolio risk, Lu et al. (2014) combine copula (t-Copula, Gaussian copula and Symmetric Joe-Clayton copula) with GARCH-type models to estimate VaR of an equally weighted portfolio formed by crude oil futures and natural gas futures. The results show that t-Copula performs well and skewed-t has a better fit than normal and Student-t for individual assets.

Our study is related to all this literature, but we focus on the comparison between traditionally employed models in commodity markets with semi-nonparametric techniques (both univariate and multivariate), which have never been employed for such purposes despite being very flexible to accurately account for salient empirical regularities of financial data (particularly leptokurtosis and wavy-thicked tails).

3.3 Models and Methodology

It is well-known that financial returns are highly leptokurtic, negatively skewed and exhibit clustering and persistence in conditional volatility.[28] Commodity ETFs are not an exception on such patterns and thus non-Gaussian models seem also appropriate for providing accurate risk measures. This paper compares the relative performance of different parametric and semi-nonparametric methods for measuring both VaR and ES. In what follows we review these models, their implementation for computing VaR and ES (remarkably, the ES application for the GC distribution, which is introduced in this paper) and the tests for relative performance based on backtesting techniques.

Commodity returns present a predictable component in the conditional mean that has traditionally been modeled according to simple ARMA structures. Squared returns, however, exhibit particular dynamics (conditional heteroskedasticity, volatility clustering and long memory) that have been extensively studied since Engle (1982) and Bollerslev (1986) introduced ARCH and GARCH models. As we focus on VaR and ES performance due to the distributional hypotheses, we implement standard models in risk management applications for the first and second conditional moments. Particularly, our applications assume an ARMA(1,1) for modeling conditional mean – equation (2) – and an EGARCH(1,1) for modeling conditional variance – equation (3), the latter process being able to capture the 'leverage effect' or the asymmetric impact in volatility from negative and positive shocks (Nelson, 1991).

Therefore, we model asset returns (r_t) as in equation (1), where μ_t and σ_t^2 are the conditional mean and variance, respectively, and z_t are the standardized errors defined in equation (4), which we assume to be distributed according to a certain distribution *G*. The complete model is described below.

[28] See, e.g. Cont (2001) for a description of the main stylized empirical regularities of the data.

$$r_t = \mu_t + \sigma_t z_t \ , \quad (1)$$

$$\mu_t = \varphi + \phi\mu_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t \ , \quad (2)$$

$$\log \sigma_t^2 = \omega + \alpha(|z_{t-1}| - \mathbb{E}[|z_{t-1}|]) + \gamma z_{t-1} + \beta \log \sigma_{t-1}^2 \quad (3)$$

$$z_t = \varepsilon_t/\sigma_t, \ \ z_t \sim G(0,1), \quad (4)$$

where φ, ϕ and θ are the parameters of the ARMA(1,1), and ω, α, β and γ the parameters of the EGARCH(1,1) and ε_t a white noise. For the sake of comparison, different standardized (i.e. zero mean and unit variance) density specifications are considered for G.

3.1. Distributional hypotheses

Besides the Gaussian density, included as benchmark, we compare the most widely used parametric distributions to capture leptokurtosis (Student's t) and skewness (skewed-t). Furthermore, we also incorporate the semi-nonparametric approach, which has been shown to accurately feature financial returns distribution (Mauleón and Perote, 2000) but, as stated above, we are unaware of any previous application for modeling Commodity ETFs. The outstanding performance of this method is based on the asymptotic property of the GC type A series for approximating any frequency function under weak regularity conditions (Kendall and Stuart, 1977). All these probability density functions (pdf) are described below.

(i) Gaussian pdf:

$$\phi(z_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{{z_t}^2}{2}}. \quad (5)$$

(ii) Student's t pdf:

$$t(z_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{{z_t}^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \quad (6)$$

where Γ is the gamma function and ν is the degrees of freedom parameter.

(iii) Skewed-t pdf (Fernández and Steel, 1998):

$$g(z_t) = \begin{cases} -\frac{2}{\gamma+\frac{1}{\gamma}} t(\gamma z_t) & for \quad z_t < 0, \\ \frac{2}{\gamma+\frac{1}{\gamma}} t\left(\frac{z_t}{\gamma}\right) & for \quad z_t \geq 0, \end{cases} \quad (7)$$

where γ is the shape parameter, which incorporates the skewness, and $t(z_t)$ is the Student's t pdf in equation (6).

(iv) GC Type A pdf:

$$f(z_t, \boldsymbol{d}) = (1 + \sum_{s=2}^{S} d_s H_s(z_t))\phi(z_t), \tag{8}$$

where $\phi(z_t)$ denotes the normal pdf in equation (5), $\boldsymbol{d}' = (d_1, \dots, d_S) \in \mathbb{R}^S$ is a vector of parameters such that $f(z_t, \boldsymbol{d}) \geq 0$ and H_s is the Hermite polynomial (HP) of order *s*, which is defined in terms of the *s*th order derivative of $\phi(z_t)$ as

$$\frac{d^s \phi(z_t)}{d{z_t}^s} = (-1)^s H_s(z_t)\phi(z_t). \tag{9}$$

In particular, the first eight HP are: $H_1(z_t) = z_t$, $H_2(z) = {z_t}^2 - 1$, $H_3(z) = {z_t}^3 - 3z_t$, $H_4(z_t) = {z_t}^4 - 6{z_t}^2 + 3$, $H_5(z_t) = {z_t}^5 - 10{z_t}^3 + 15z_t$, $H_6(z_t) = {z_t}^6 - 15{z_t}^4 + 45{z_t}^2 - 15$, $H_7(z_t) = {z_t}^7 - 21{z_t}^5 + 105{z_t}^3 - 105z_t$, $H_8(z_t) = {z_t}^8 - 28{z_t}^6 + 210{z_t}^4 - 420{z_t}^2 + 105$.

These polynomials form an orthonormal basis, thus satisfying the orthogonality property,

$$\int H_s(z_t) H_j(z_t) \phi(z_t)\, dz_t = 0 \;\; \forall s \neq j\ , \tag{10}$$

which is the basis of the characterization of GC series as a pdf and its parameters in terms of its density moments. For instance, even non-central moments depend on the even d_s parameters (e.g. d_2 accounts for the variance, d_4 for the excess kurtosis and the rest of the even parameters capture higher-order moments) and skewness is captured by the odd parameters. A well-known shortcoming of this expansion is that it is not necessarily positive for all the values of the parameters, which calls for the implementation of positive transformations (Gallant and Nychka, 1987; León et al. 2009) or positivity constraints (Jondeau and Rockinger, 2001). However, we use the original GC Type A expansion in equation (8), which is simpler and more useful for risk measure applications.

Most of the empirical studies that implement GC density in finance – mainly for option valuation, e.g. Jarrow and Rudd (1982) – truncate the expansion in *n* = 4 and employ only two terms of the expansion, d_3 and d_4. These models, as well as the other parametric distributions, were jointly estimated with the ARMA(1,1)-EGARCH(1,1) models – equations (2) and (3) – by maximum likelihood (ML) techniques.

3.2. VaR measures

The estimated VaR with a confidence level α is computed as a linear transformation of the estimated α-quantile, $\hat{q}_\alpha(z_t)$, of the assumed standardized distributions in the above section. Therefore, the predicted VaR for the variable *r* at the time horizon *t*+1 and with confidence level α is given in equation (11).

$$\mathrm{VaR}_{t+1}^{\alpha} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{q}_\alpha(z_{t+1}), \tag{11}$$

where $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ are the predictions for the mean and standard deviation conditioned on the available information at time *t*, Ω_t, obtained from the ARMA-EGARCH model described in equations (2) and (3). This measure amounts for the maximum expected loss of the variable in *t*+1 obtained with a probability α.

The α-quantiles can be easily estimated by numerical integration and, for the GC density, may be obtained as those $\hat{q}_\alpha(z_t) = \varphi^{-1}(\alpha)$ satisfying

$$\varphi\big(\hat{q}_\alpha(z_t)\big) = \int_{-\infty}^{\hat{q}_\alpha(z_t)} \phi(z_t)dz_t - \phi\big(\hat{q}_\alpha(z_t)\big)\sum_{s=2}^{S} d_s H_{s-1}\big(\hat{q}_\alpha(z_t)\big) = \alpha, \tag{12}$$

as a direct application of equation (10).

3.3. ES measures

The ES for a given confidence level α is the expected return of the variable on the worst α per cent of the cases. In terms of the standardized distribution of z_t it can be calculated as

$$ES_\alpha = \frac{1}{1-\alpha}\int_{\alpha}^{1} VaR_u(z_t)du. \tag{13}$$

Thus, the expected shortfall for the variable *r* at the time horizon *t*+1 and with confidence level α is obtained through the following transformation

$$ES_{t+1}^{\alpha} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\widehat{ES}_\alpha(z_{t+1}), \tag{14}$$

where $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ are the predictions for the mean and standard deviation conditioned on Ω_t. The ES estimates are usually obtained by numerical integration, but in most cases the use of closed forms can simplify the estimation procedures. Particularly, closed forms of most of the cases studied in Section 3.1 are displayed below.[29]

(i) Gaussian ES:

$$ES_\alpha = \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha}, \tag{15}$$

where ϕ is the pdf of standard normal – equation (5) – and $\Phi^{-1}(\alpha)$ is its corresponding α-quantile.

(ii) Student's t ES:

$$ES_\alpha = \frac{t_\nu\left(g_\nu^{-1}(\alpha)\right)}{1-\alpha}\left(\frac{\nu+\left(g_\nu^{-1}(\alpha)\right)^2}{\nu-1}\right), \tag{16}$$

where t_ν is the pdf of standard Student's t – equation (6), $g_\nu^{-1}(\alpha)$ its corresponding α-quantile and ν the degrees of freedom parameter.

(iii) GC Type A ES:

$$ES_\alpha = \phi\left(\varphi^{-1}(\alpha)\right)\left[1 + \sum_{s=2}^{S} d_s\left[H_s\left(\varphi^{-1}(\alpha)\right) + sH_{s-2}\left(\varphi^{-1}(\alpha)\right)\right]\right]. \tag{17}$$

where ϕ is the pdf of standard normal – equation (5) – and $\varphi^{-1}(\alpha)$ is the α-quantile of the GC Type A distribution –equation (8). The proof is provided in the Appendix A.

3.4. Backtesting methods

3.4.1. Tests for VaR

a. Bernoulli coverage test

The VaR backtesting is usually based on the assumption that the number of exceptions is generated by an iid Bernoulli process. The violations or exceptions is 1 if the actual loss is greater than the predicted VaR, otherwise it is 0. The null hypothesis of Bernoulli Coverage test indicates that the VaR model is correct according to the exceptions ratio.

A coverage test may also be performed by the likelihood ratio (LR), which asymptotically follows a chi-square distribution with one degree of freedom,

[29] The ES under the Skewed-t in the empirical application of next section was obtained by numerical integration.

$$LR = -2\log\left\{\frac{(1-p)^{T_0}p^{T_1}}{(1-T_1/T)^{T_0}(T_1/T)^{T_1}}\right\} \sim \chi_1^2, \quad (18)$$

where p is the proportion of violations to be tested – 1% and 2.5% in next section's empirical application. T_1 and T_0 are the number of ones (violations) and zeros (no violations) in the backtesting period of size T. The p-value is then calculated as:

$$p - value = 1 - F(LR), \quad (19)$$

where F is the cumulative distribution function (cdf) of a chi-square random variable with one degree of freedom.

b. Relative Comparison for VaR

The relative performance of a risk measure is based on a scoring function. This function is denoted as $S(x, y)$, where x is the set of point forecasts given by a certain model and y are the verifying observations. The scoring function can be viewed as a prediction error function that has to be minimized. A scoring function consistent for VaR is (Gneiting, 2011)

$$S(\mathrm{VaR}_{t+1}^{\alpha}, r_{t+1}) = \left(I_{\{r_{t+1}<VaR_t^{\alpha}\}} - \alpha\right)\left(Ç(VaR_t^{\alpha}) - Ç(r_{t+1})\right), \quad (20)$$

where Ç is a strictly increasing function. This scoring function can be used to rank VaR forecasts and to implement Diebold and Mariano's (1995) test for relative VaR model performance. The Diebold-Mariano's test statistic (DM) is calculated as

$$DM = \frac{\bar{d}}{s.e.(d)}, \quad (21)$$

where $\bar{d}$ denotes the mean of d_t which is

$$d_t = S^{(i)}(\mathrm{VaR}_{t+1}^{\alpha}, r_{t+1}) - S^{(j)}(\mathrm{VaR}_{t+1}^{\alpha}, r_{t+1}), \quad (22)$$

where $S^{(i)}$ and $S^{(j)}$denote the scores obtained from VaR models (i) and (j) respectively, $s.e.(d)$ denotes the standard error of the statistics and requires the implementation of a heteroskedasticity and autocorrelation consistent (HAC) variance estimator (e.g., Newey-West estimator). The null hypothesis is $DM = 0$, and the alternative is $H_1^{(i)}: DM < 0$ and $H_1^{(j)}: DM > 0$. Since the limiting distribution of DM statistic is standard normal the null hypothesis is rejected for smaller values than -1.96 at 5% of significance level indicating the outperformance of model (i) and the opposite indicates that model (j) is better. The application in next section uses a squared error function for the scoring function.

3.4.2. Tests for ES

a. t-test

A first test for ES is based on the violation residuals which is calculated as

$$K_{t+1} = \left(\frac{L_{t+1} - ES_t^{\alpha}}{ES_t^{\alpha} - \mu_{t+1}}\right) I_{\{L_{t+1} > VaR_t^{\alpha}\}}, \tag{23}$$

where L_{t+1} is the actual loss, ES_t^{α} is the estimated ES given in equation (14), μ_{t+1} is the conditional mean of the model – equation (2). The indicator function $I_{\{L_{t+1} > VaR_t^{\alpha}\}}$ takes value 1 when the actual loss has exceeded the estimated VaR, and 0 otherwise. Then, the null hypothesis of zero mean violation residuals may be tested by a simple t-test on this variable (McNeil et al., 2015),

$$t - stat = \frac{\bar{K}}{s/\sqrt{T}}, \tag{24}$$

where $\bar{K}$ is the sample mean of the violation residuals of size T, and s denotes its standard deviation. The p-value is calculated as

$$p - value = 2\text{Prob}(t > |t - stat|). \tag{25}$$

b. Relative Comparison for ES

Fissler et al. (2016) show that VaR and ES are jointly elicitable and therefore a possible scoring function for both arguments is given by

$$S(\text{VaR}_{t+1}^{\alpha}, \text{ES}_{t+1}^{\alpha}, r_{t+1}) = \left(I_{\{r_{t+1} < VaR_t^{\alpha}\}} - \alpha\right)\left(G_1(VaR_t^{\alpha}) - G_2(r_{t+1})\right) + \frac{1}{\alpha} G_2(\text{ES}_{t+1}^{\alpha}) I_{\{r_{t+1} < VaR_t^{\alpha}\}} (\text{VaR}_{t+1}^{\alpha} - \text{ES}_{t+1}^{\alpha}) + G_2(\text{ES}_{t+1}^{\alpha})(\text{ES}_{t+1}^{\alpha} - \text{VaR}_{t+1}^{\alpha}) - \mathcal{g}_2(\text{ES}_{t+1}^{\alpha}), \tag{26}$$

where $\mathcal{g}_2' = G_2$, G_1 and G_2 are continuously differentiable functions[30], e.g. $G_1(x) = x$ and $G_2(x) = \exp(x)$, see Fissler and Ziegel (2016) for more details. In this paper, a homogeneous of degree zero scoring function is employed, since Patton and Sheppard (2009) show that this type of scoring functions presents good size and power properties in volatility applications. In particular, $G_1(x) = 0$ and $\mathcal{g}_2(x) = \log x$. For such scoring function, the relative performance of the models can be assessed by implementing DM test in equation (21).

[30] G_1 is weakly increasing function and G_2 is strictly increasing function.

3.4 Risk quantification for individual commodity ETFs

4.1. Data

The sample comprises nine years of daily prices from January 2007 to January 2016 for Gold, Silver, Oil, Agriculture, Energy and Broad Commodity ETFs. All data were obtained from Bloomberg; further details on the data are provided in Appendix B. Table 12 displays the descriptive statistics for continuously compounded returns of these series, defined as r_t=100log(P_t/P_{t-1}), where P_t represents ETFs prices at time *t*.

Table 12. Descriptive statistics of Commodity ETFs

	Gold	**Silver**	**Oil**	**Agriculture**	**Energy**	**Broad**
Mean	0.0244	0.0045	-0.0731	-0.0099	-0.0406	-0.0279
Median	0.0433	0.0882	-0.0171	0.0000	0.0000	0.0000
Standard deviation	1.2642 (19.99)	2.1821 (34.50)	2.1963 (34.73)	1.2680 (20.05)	1.7523 (27.71)	1.3127 (20.75)
Min	-9.1905	-19.8349	-11.2996	-8.9990	-9.02391	-6.9341
Max	10.6974	13.4733	9.1691	7.6675	8.8686	6.6484
Skewness	-0.2567	-0.9751	-0.1966	-0.2513	-0.1912	-0.3048
Excess Kurtosis	6.0464	7.6419	2.4764	5.4450	2.6507	2.6222

Annual volatility in parentheses

Descriptive statistics confirm that Commodity ETFs feature the same regularities as most financial data: Location statistics (daily mean and median) are around zero reflecting market efficiency; they exhibit negative skewness (i.e. more probability mass around positive returns, but also more extreme values at the left tail of negative returns); high volatility and kurtosis (Oil ETF featuring the highest volatility followed by Silver ETF, the latter having the highest excess kurtosis). Figures 10 and 11 display prices and returns, respectively, which give a clear idea about the instability and volatility clustering of the data.

Figure 10. Commodity ETFs prices

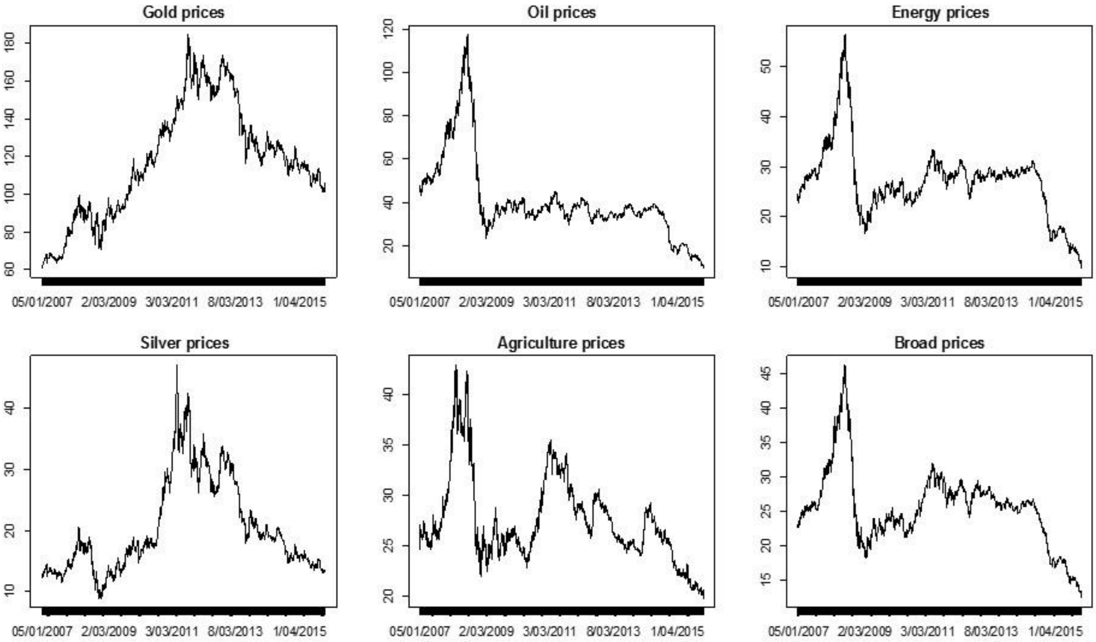
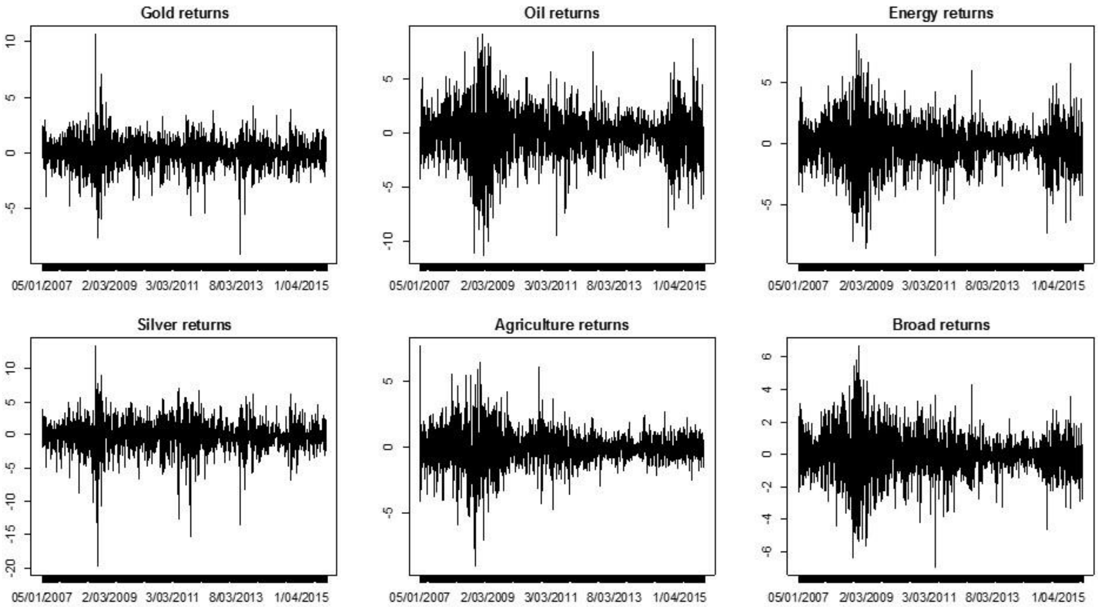


Figure 11. Commodity ETFs returns



4.2. Model Performance for Commodity ETFs

This section presents the parameter estimation of the ARMA(1,1)-EGARCH(1,1) model under the alternative distributional hypotheses in Section 3.1 and the model comparison according to the tests mentioned in Section 3.2. We examine backtesting for both VaR (at 99% and 97.5%) and ES (at 97.5%), and discuss the relative model assessment.

4.2.1. Model estimates

Table 13 presents the ML estimates of the parameters of the ARMA(1,1)–EGARCH(1,1) model under four distributional hypotheses: Gaussian (Panel A), Student's t (Panel B), Skewed-t (Panel C) and GC (Panel D). It is noteworthy that, unlike Del Brio and Perote (2012), where parameters are estimated in two steps, we estimate ARMA-EGARCH and GC innovation parameters jointly by ML. P-values for testing the significance of every parameter are given in parentheses. The estimates of the AR (ϕ) and MA (θ) structures are significant for most cases of Commodity ETF series. The values of the EGARCH(1,1) parameters are also statistically significant, γ is positive indicating that positive innovations generate more volatility than negative shocks, and β is close to one for all series, reflecting the persistence and clustering in volatility. The estimates for the shape (degrees of freedom), v, and skew parameter (ξ) of the Student's t distributions are significant and reveal the leptokurtic and (negatively) skewed nature of the data, which is confirmed by the estimates of the parameters of the GC densities, d_3 (skewness coefficient) and d_4 (excess kurtosis).

Table 13. Estimates of ARMA(1,1)-EGARCH(1,1) models

	Gold	Silver	Oil	Agriculture	Energy	Broad
			Panel A: Gaussian			
μ	0.023 (0.000)	-0.013 (0.766)	-0.038 (0.248)	-0.032 (0.057)	-0.029 (0.252)	-0.037 (0.068)
φ	0.952 (0.000)	0.044 (0.153)	-0.725 (0.000)	0.490 (0.000)	-0.858 (0.000)	-0.880 (0.000)
ϕ	-0.965 (0.000)	-0.014 (0.669)	0.685 (0.000)	-0.538 (0.000)	0.819 (0.000)	0.858 (0.000)
ω	0.014 (0.000)	0.051 (0.000)	0.013 (0.000)	0.002 (0.113)	0.008 (0.000)	0.004 (0.008)
α	-0.010 (0.236)	-0.016 (0.164)	-0.051 (0.000)	0.007 (0.344)	-0.046 (0.000)	-0.023 (0.002)
β	0.981 (0.000)	0.971 (0.000)	0.991 (0.000)	0.997 (0.000)	0.993 (0.000)	0.994 (0.000)
γ	0.145 (0.00)	0.203 (0.000)	0.101 (0.000)	0.114 (0.000)	0.098 (0.000)	0.103 (0.000)
			Panel B: Student's t			
μ	0.042 (0.022)	0.038 (0.242)	-0.019 (0.574)	-0.028 (0.024)	-0.004 (0.857)	-0.020 (0.301)
φ	0.449 (0.000)	-0.337 (0.000)	-0.715 (0.000)	0.402 (0.000)	-0.837 (0.000)	-0.855 (0.000)
ϕ	-0.489 (0.000)	0.322 (0.000)	0.672 (0.000)	-0.457 (0.000)	0.795 (0.000)	0.827 (0.000)
ω	0.003 (0.115)	0.012 (0.000)	0.008 (0.000)	-0.001 (0.671)	0.003 (0.091)	0.000 (0.870)
α	0.009 (0.373)	0.011 (0.287)	-0.047 (0.000)	0.000 (0.958)	-0.042 (0.000)	-0.025 (0.004)
β	0.992 (0.000)	0.992 (0.000)	0.993 (0.000)	0.997 (0.000)	0.995 (0.000)	0.996 (0.000)
γ	0.107 (0.000)	0.117 (0.000)	0.101 (0.000)	0.107 (0.000)	0.101 (0.000)	0.102 (0.000)
v	4.960 (0.000)	4.310 (0.000)	12.285 (0.000)	8.292 (0.000)	9.371 (0.000)	9.521 (0.000)
			Panel C: Skewed Student's t			
μ	0.023 (0.262)	-0.034 (0.245)	-0.035 (0.270)	-0.032 (0.057)	-0.025 (0.312)	-0.031 (0.033)

φ	0.443 (0.000)	-0.277 (0.000)	-0.701 (0.000)	0.405 (0.000)	-0.825 (0.000)	-0.814 (0.000)
ϕ	-0.485 (0.000)	0.257 (0.000)	0.655 (0.000)	-0.463 (0.000)	0.781 (0.000)	0.782 (0.000)
ω	0.003 (0.129)	0.011 (0.000)	0.009 (0.000)	-0.000 (0.704)	0.004 (0.038)	0.006 (0.690)
α	0.009 (0.386)	0.010 (0.320)	-0.047 (0.000)	0.000 (0.971)	-0.043 (0.000)	-0.025 (0.003)
β	0.992 (0.000)	0.992 (0.000)	0.993 (0.000)	0.997 (0.000)	0.995 (0.000)	0.996 (0.000)
γ	0.107 (0.000)	0.118 (0.000)	0.103 (0.000)	0.109 (0.000)	0.104 (0.000)	0.103 (0.000)
ν	5.069 (0.000)	4.341 (0.000)	13.008 (0.000)	8.299 (0.000)	9.894 (0.000)	9.902 (0.000)
Skew-ξ	0.939 (0.000)	0.894 (0.000)	0.918 (0.000)	0.970 (0.000)	0.897 (0.000)	0.919 (0.000)
Panel D: GC						
μ	0.003 (0.000)	-0.003 (0.925)	-0.038 (0.493)	-0.031 (0.052)	-0.030 (0.534)	-0.087 (0.017)
φ	0.840 (0.000)	0.034 (0.000)	-0.726 (0.000)	0.491 (0.022)	-0.858 (0.000)	-0.880 (0.000)
ϕ	-0.853 (0.000)	-0.024 (0.000)	0.684 (0.000)	-0.527 (0.010)	0.818 (0.000)	0.855 (0.000)
ω	0.018 (0.000)	0.063 (0.000)	0.011 (0.002)	0.035 (0.000)	0.009 (0.003)	0.017 (0.000)
α	-0.005 (0.621)	-0.003 (0.848)	-0.056 (0.000)	-0.059 (0.004)	-0.049 (0.000)	-0.057 (0.000)
β	0.981 (0.000)	0.965 (0.000)	0.992 (0.000)	0.970 (0.000)	0.993 (0.000)	0.980 (0.000)
γ	0.055 (0.000)	0.056 (0.000)	0.020 (0.000)	0.046 (0.000)	0.028 (0.000)	0.033 (0.000)
d_3	-0.035 (0.004)	-0.042 (0.000)	-0.022 (0.034)	-0.032 (0.008)	-0.034 (0.002)	-0.026 (0.024)
d_4	0.166 (0.000)	0.244 (0.000)	0.111 (0.000)	0.408 (0.000)	0.116 (0.000)	0.228 (0.000)

μ, ϕ and φ are the parameters of the ARMA(1,1) model; ω, α, γ and β are the parameters of the EGARCH(1,1) model; ν and ξ are the parameters of the (skewed) Student's t; d_{3i} and d_{4i} are the parameters of the GC model. P-values for the t-test in parentheses.

The skewness parameter for Gold and Agricultural ETFs is close to one, when filtering the returns by an ARMA(1,1)-EGARCH(1,1) with innovations skewed-t distributed. This fact indicates that the distribution for these variables is close to the Student's t (as revealed by the similar loglikelihood and information criteria values in Table 14, Panel A and B), and the performance would be expected to be same for both distributions, when calculating risk measures for Gold ETF. In addition, the degrees of freedom parameter (ν) for Oil ETF is "relatively" high (12.28) when filtering the returns by an ARMA(1,1)-EGARCH(1,1) with innovations Student's t distributed. The higher the value of this shape parameter the less leptokurtic the distribution is (i.e. closer to the Gaussian pdf). This feature was also clear from the descriptive statistics (see Table 1), where Oil ETF presents the minimum excess kurtosis (2.47). It is worth noting that, after filtering Oil ETF returns by an ARMA(1,1)-EGARCH(1,1) model with skewed-t distribution for the innovations, the estimation for degrees of freedom is still relatively high (13) and skewness parameter close to one. The diagnostic tests for serial correlation (Ljung Box statistic) on the ARMA(1,1)-EGARCH(1,1) model (levels and squared) standardized residuals are displayed in Table 3, as well as the accuracy criteria (AIC, BIC and HQC) under the different distributional hypotheses.

Table 14. Diagnostics of ARMA(1,1)-EGARCH(1,1)

	Gold	**Silver**	**Oil**	**Agriculture**	**Energy**	**Broad**
			Panel A: Gaussian			
LL	-3548.78	-4738.13	-4648.64	-3323.15	-4146.57	-3497.14
AIC	3.131	4.179	4.100	2.933	3.658	3.086
BIC	3.149	4.197	4.118	2.950	3.675	3.103
HQC	3.138	4.185	4.106	2.939	3.664	3.092
LBSR	0.051	0.009	0.194	0.696 (0.404)	0.079	0.030
	(0.821)	(0.923)	(0.659)		(0.778)	(0.862)
LBSSR	7.829	10.31	1.679	0.485 (0.486)	1.006	0.303
	(0.005)	(0.001)	(0.195)		(0.316)	(0.582)
			Panel B: Student's t			
LL	-3458.40	-4647.04	-4632.68	-3287.33	-4117.24	-3468.68
AIC	3.053	4.099	4.087	2.902	3.633	3.062
BIC	3.073	4.120	4.107	2.922	3.653	3.082
HQC	3.060	4.107	4.094	2.909	3.640	3.069
LBSR	1.967	3.595	0.286	1.461 (0.227)	0.194	0.006
	(0.161)	(0.058)	(0.593)		(0.659)	(0.938)
LBSSR	33.83	30.40	1.758	0.720 (0.396)	0.930	0.222
	(0.000)	(0.000)	(0.185)		(0.335)	(0.637)
			Panel C: Skewed Student's t			
LL	-3455.90	-4638.37	-4628.84	-3286.86	-4111.05	-3464.98
AIC	3.051	4.093	4.084	2.903	3.628	3.059
BIC	3.074	4.115	4.107	2.925	3.651	3.082
HQC	3.060	4.101	4.093	2.911	3.637	3.068
LBSR	2.185	4.077	0.461	1.750 (0.186)	0.340	0.070
	(0.139)	(0.043)	(0.497)		(0.560)	(0.791)
LBSSR	33.70	30.18	1.552	0.624 (0.430)	0.753	0.257
	(0.000)	(0.000)	(0.213)		(0.385)	(0.612)
			Panel D: GC			
LL	-1389.89	-2593.33	-2562.65	-1278.65	-2045.83	-1434.25
AIC	1.232	2.292	2.265	1.134	1.809	1.271
BIC	1.255	2.314	2.287	1.157	1.832	1.294
HQC	1.240	2.300	2.273	1.142	1.818	1.279
LBSR	0.255	2.965	0.310	0.166 (0.683)	0.794	3.228
	(0.613)	(0.085)	(0.577)		(0.373)	(0.0724)
LBSSR	21.690	28.751	47.300	3.900 (0.048)	39.519	83.646
	(0.000)	(0.000)	(0.150)		(0.271)	(0.413)

LL: Loglikelihood, AIC: Akaike information criterion, BIC: Bayesian (Schwartz) information criterion, HQC: Hannan-Quinn information criterion, LBSR: Ljung Box Standardized Residuals, LBSSR: Ljung Box Squared Standardized Residuals (1 lag). P-values for testing H_0: no serial correlation in parentheses.

Accuracy criteria show that ARMA-EGARCH-Gaussian is outperformed by the other models for all series. Though the hypothesis of absence of serial correlation is rejected for squared standardized residuals in metal (Gold and Silver) ETFs, the ARMA-EGARCH model seems to be adequate for Commodity ETF returns, since their most important parameters to capture volatility persistence and leverage effects are significant.

4.2.2. Model performance for VaR and ES

a. Tests for 99%-VaR

According to Bernoulli Coverage test for VaR – see equation (18) – presented in Table 4, ARMA-EGARCH-skewed-t and ARMA-EGARCH-GC perform adequately for all Commodity ETFs, but the ARMA-EGARCH-GC fails in Agriculture ETF. ARMA-EGARCH-Gaussian and ARMA-EGARCH-t tend to underestimate risk, whereas ARMA-EGARCH-skewed-t seems to overestimate it. As seen in Table 15, ARMA-EGARCH-Gaussian and ARMA-EGARCH-t only perform reasonably well for Agriculture ETF returns according to the coverage test. Details about the relative performance tests of VaR at 99% (pairwise DM test) can be found in Appendix C. Results for relative performance shows the outperformance of non-Gaussian models with the exception of Oil ETF (as commented before, Oil ETF features the less leptokurtic pattern of all series), but ARMA-EGARCH-t performs poorly in most cases (all cases except for Silver and Oil ETFs). ARMA-EGARCH-GC is significantly better than ARMA-EGARCH-skewed-t for Silver and Oil ETFs, whereas ARMA-EGARCH-skewed-t performs more accurately than ARMA-EGARCH-GC for Energy ETF.

Table 15. Backtesting 99%-VaR for Commodity ETFs returns

Expected number of exceptions = 18 Data: December 2008 – January 2016						
1771 days	*Gold*	*Silver*	*Oil*	*Agriculture*	*Energy*	*Broad*
ARMA-EGARCH-Gaussian	43 (0.000)	38 (0.000)	35 (0.000)	24 (0.154)	38 (0.000)	33 (0.001)
ARMA-EGARCH-t	28 (0.023)	27 (0.039)	27 (0.039)	18 (0.945)	29 (0.013)	29 (0.013)
ARMA-EGARCH-skewed-t	25 (0.101)	15 (0.506)	15 (0.506)	21 (0.445)	17 (0.864)	15 (0.506)
ARMA-EGARCH-GC	21 (0.445)	21 (0.445)	16 (0.678)	5 (0.000)	24 (0.154)	17 (0.864)

P-values for Bernoulli Coverage test in parentheses.

b. Tests for 97.5%-VaR and 97.5%-ES

Table 16 displays the results of the Bernoulli Coverage test for VaR at 97.5%. They confirm the good performance of ARMA-EGARCH-skewed-t and ARMA-EGARCH-GC models. However, the ARMA-EGARCH-Gaussian for Silver, Agriculture and Broad ETFs are not rejected at 5% confidence. As expected, Gaussian model performs better when the confidence level of the VaR is not so high (e.g. 97.5%). The 97.5%-VaR relative performance (see results in Appendix D) supports the ARMA-EGARCH-GC and ARMA-EGARCH-skewed-t as the best models. ARMA-EGARCH-Gaussian (ARMA-EGARCH-t) only performs well for one (two) case(s).

Table 16. Backtesting 97.5%-VaR for Commodity ETFs returns

Expected number of exceptions = 44 Data: December 2008 – January 2016						
1771 days	*Gold*	*Silver*	*Oil*	*Agriculture*	*Energy*	*Broad*
ARMA-EGARCH-Gaussian	62 (0.011)	57 (0.063)	62 (0.011)	49 (0.479)	58 (0.046)	53 (0.197)
ARMA-EGARCH-t	62 (0.011)	56 (0.086)	59 (0.033)	45 (0.912)	56 (0.086)	53 (0.197)
ARMA-EGARCH-skewed-t	42 (0.727)	34 (0.103)	39 (0.412)	40 (0.508)	27 (0.005)	33 (0.072)
ARMA-EGARCH-GC	42 (0.727)	41 (0.614)	44 (0.967)	22 (0.000)	46 (0.794)	30 (0.021)

P-values for the Bernoulli Coverage test in parentheses.

Table 17 presents the results of the t-test – equations (23) and (24) – for model performance in terms of ES, the null hypothesis being zero mean of violation residuals. Hence, a model is good enough if there is no evidence to reject null hypothesis. The results of t-test for ES at 97.5% show that ARMA-EGARCH-skewed-t and ARMA-EGARCH-GC model the best (these models are never rejected) followed by ARMA-EGARCH-t (rejected in 4 occasions). AR-GARCH-Gaussian is always rejected when risk is measured by ES at 97.5%.

Table 17. T-test for 97.5%-ES for Commodity ETFs returns

Mean of violation residual is zero Data: December 2008 – January 2016						
1771 days	*Gold*	*Silver*	*Oil*	*Agriculture*	*Energy*	*Broad*
ARMA-EGARCH-Gaussian	4.443 (0.000)	3.455 (0.001)	3.004 (0.004)	2.203 (0.032)	3.539 (0.001)	3.410 (0.001)
ARMA-EGARCH-t	-4.586 (0.000)	-4.014 (0.000)	-1.292 **(0.201)**	-3.177 (0.003)	-1.785 (0.079)	-1.209 **(0.232)**
ARMA-EGARCH-skewed-t	1.756 **(0.087)**	1.206 **(0.237)**	0.342 **(0.734)**	0.951 **(0.347)**	1.561 **(0.130)**	1.541 **(0.133)**
ARMA-EGARCH-GC	1.855 **(0.071)**	1.856 **(0.071)**	1.179 **(0.245)**	-0.856 **(0.401)**	3.027 (0.004)	1.288 **(0.208)**

P-values for the t-test test in parentheses.

Relative performance (DM pairwise) test for ES is displayed in Appendix E. In this case, the best models are ARMA-EGARCH-t and ARMA-EGARCH-GC, followed by ARMA-

EGARCH-skewed-t. Gaussian model works better than Student's t and GC models for Oil and Energy ETFs respectively. Once more, this is not an unexpected result, since excess kurtosis in these two series are the lowest.

In summary, ARMA-EGARCH-GC outperforms the other models for most series and according to Bernoulli test, t-test and relative performance tests for VaR and ES estimations. Nevertheless, ARMA-EGARCH-skewed-t and ARMA-EGARCH-t seem to be the best models according to coverage test for 99%-VaR and relative performance for 97.5%-ES, respectively. ARMA-EGARCH-Gaussian performs very poorly according to both coverage and relative performance test for all series with the exception of the Oil ETFs series (the less leptokurtic one).

3.5 Portfolio risk quantification

The former section reveals the GC distribution as an accurate distribution to measure ETFs risk according to both VaR and ES. In this section we extend the analysis to the multivariate case by studying the relative performance of multivariate GC (henceforth, multivariate SNP distribution) and Gaussian distribution for ES backtesting of three bivariate portfolios of ETFs. The predicted risk measures for the portfolio return r_p at the time horizon t+1 and with confidence level α is given by

$$\mathrm{VaR}_{p,t+1}^{\alpha} = \hat{\mu}_{p,t+1} + \hat{\sigma}_{p,t+1}\hat{q}_{\alpha}(z_p), \tag{27}$$

$$\mathrm{ES}_{p,t+1}^{\alpha} = \hat{\mu}_{p,t+1} + \hat{\sigma}_{p,t+1}\widehat{ES}_{\alpha}(z_p), \tag{28}$$

where forecasted conditional mean and variance of the portfolio are computed as $\hat{\mu}_{p,t+1} = \sum_{i=1}^{n} w_i \hat{\mu}_{i,t+1}$ and $\hat{\sigma}_{p,t+1}^2 = \sum_{i=1}^{n} w_i w_j \hat{\sigma}_{ij,t+1}$, respectively, which are obtained through the forecasted mean for every asset i ($\hat{\mu}_{i,t+1}$) and the covariance of every couple of assets i and j ($\hat{\sigma}_{ij,t+1} = \hat{\rho}_{ij,t+1}\hat{\sigma}_{i,t+1}\hat{\sigma}_{j,t+1}$) and considering the weight of every asset i, denoted by w_i and satisfying $0 \leq w_i \leq 1$ and $\sum_{i=1}^{n} w_i = 1$. Note that $\hat{\rho}_{ij,t+1}$ accounts for the estimated conditional correlation coefficient between assets i and j and then $\hat{\sigma}_{ij,t+1} = \hat{\sigma}_{i,t+1}^2$ when $i = j$.

We assume that the marginal distribution of every individual asset follows the ARMA(1,1)-EGARCH(1,1) model in equations (1)-(4) with either Gaussian or GC innovations. In addition, an estimation for the conditional correlation $\hat{\rho}_{ij,t+1}$ is necessary to find the predicted volatility of portfolio returns. To this end, standard Dynamic Conditional Correlation (Gaussian-DCC) and semi-nonparametric DCC (SNP-DCC) models are employed in this section. The former model was introduced by Engle (2002), and the latter by Del Brio et al. (2011), which extends of the seminal Engle's (2002) work to non-Gaussian distributions. Next subsection reviews these models.

5.1. The Gaussian-DCC and SNP-DCC models

The multivariate SNP of vector $\boldsymbol{z}_t = (z_{1t}, z_{2t}, \dots, z_{nt})' \in \mathbb{R}^n$ with $z_{it} \sim GC(0,1)$ – i.e. distributed as in equation (8) – and conditional correlation matrix $\boldsymbol{R}_t$ (with general element $\{\rho_{ij}\}$) is characterized in terms of the following pdf:

$$\boldsymbol{f}_{SNP}(\boldsymbol{z}_t|\Omega_{t-1}) = (2\pi)^{-\frac{n}{2}}|\boldsymbol{R}_t|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\boldsymbol{z}_t'\boldsymbol{R}_t^{-1}\boldsymbol{z}_t\right\}\left[\sum_{i=1}^{n}\psi_i(x_{it})\right]\frac{1}{n} \quad (29)$$

where $\psi_i(x_{it}) = 1 + d_{3i}(x_{it}^3 - 3x_{it}) + d_{4i}\left(x_{it}^4 - 6x_{it}^2 + 3\right)$ and

$$\boldsymbol{x}_t = (x_{1t}, x_{2t}, \dots, x_{nt})' = \boldsymbol{R}_t^{-1/2}\boldsymbol{z}_t. \quad (30)$$

It is noteworthy that the multivariate SNP distribution collapses to the multivariate Gaussian distribution when $d_{3,i} = d_{4,i} = 0$ and then Gaussian-DCC is a particular case of SNP-DCC. The transformation in equation (30) is not unique, although this fact does not impact the estimates of the conditional correlations. For example, for the bivariate case ($n = 2$) and the eigenvalue decomposition this transformation yields

$$x_{1t} = \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_{12,t}}} + \frac{1}{\sqrt{1-\rho_{12,t}}}\right)z_{1t} + \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_{12,t}}} - \frac{1}{\sqrt{1-\rho_{12,t}}}\right)z_{2t} \quad (31)$$

$$x_{2t} = \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_{12,t}}} - \frac{1}{\sqrt{1-\rho_{12,t}}}\right)z_{1t} + \frac{1}{2}\left(\frac{1}{\sqrt{1+\rho_{12,t}}} + \frac{1}{\sqrt{1-\rho_{12,t}}}\right)z_{2t} \quad (32)$$

Then, the DCC model employed in this article can be formulated as:

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t(\boldsymbol{\phi}) + \boldsymbol{\varepsilon}_t \quad (33)$$

$$\boldsymbol{\varepsilon}_t \sim \boldsymbol{GC}(\boldsymbol{0}, \boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t) \quad (34)$$

$$\log\boldsymbol{D}_t^2 = diag\{\omega_i\} + diag\{\alpha_i\} \circ (|\boldsymbol{\varepsilon}_{t-1}| - \mathbb{E}[|\boldsymbol{\varepsilon}_{t-1}|]) + diag\{\gamma_i\} \circ \boldsymbol{\varepsilon}_{t-1} + diag\{\beta_i\} \circ \log\boldsymbol{D}_{t-1}^2 \quad (35)$$

$$\boldsymbol{z}_t = \boldsymbol{D}_t^{-1}\boldsymbol{\varepsilon}_t \quad (36)$$

$$\boldsymbol{Q}_t = \boldsymbol{S} \circ (\boldsymbol{\iota\iota}' - \boldsymbol{A} - \boldsymbol{B}) + \boldsymbol{A} \circ \boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}' + \boldsymbol{B} \circ \boldsymbol{Q}_{t-1} \quad (37)$$

$$\boldsymbol{R_t} = diag\{\boldsymbol{Q_t}\}^{-1/2}\boldsymbol{Q_t}diag\{\boldsymbol{Q_t}\}^{-1/2} \tag{38}$$

where $\boldsymbol{D_t^2}$ is the diagonal matrix of conditional variances with EGARCH dynamics[31] and $\boldsymbol{Q_t}$ the conditional covariance matrix of the DCC type – i.e. $\boldsymbol{\iota}$ is a vector of ones, the symbol $\circ$ represents the element-by-element multiplication operator (Hadamard product), matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{\iota\iota'} - \boldsymbol{A} - \boldsymbol{B}$ are positive definite matrices and $\boldsymbol{S}$ is the unconditional correlation matrix of $\boldsymbol{z_t}$.

It can be easily proved that the loglikelihood functions of both Gaussian-DCC and SNP-DCC models can be split in two terms, volatility and correlation, which allows to consistently estimate the density parameters in two steps: Firstly, conditional means and variances are estimated independently for every variable and, secondly, conditional correlations are estimated in the standardized (zero mean and unit variance) Gaussian and SNP distributions. In particular, the loglikelihood (LL) function in the second stage for SNP-DCC becomes (after deleting unnecessary constants)

$$\text{LL}(SNP) = -\frac{1}{2}\log|\boldsymbol{R}_t| - \frac{1}{2}\boldsymbol{z}_t'\boldsymbol{R}_t^{-1}\boldsymbol{z}_t + \sum_{t=1}^{T}\log\left\{\sum_{i=1}^{n}\psi_i(\boldsymbol{x}_{it})\right\}. \tag{39}$$

5.2. Application

For the sake of simplicity,[32] our application reduces to three bivariate (n = 2) and equally weighted ($w_i = 0.5, \forall i = 1{,}2$) portfolios: Portfolio A, formed by Gold and Silver ETFs; Portfolio B, formed by Gold and Oil ETFs; and Portfolio C, formed by Silver and Oil ETFs. Table 18 exhibits the main descriptive statistics for the portfolio returns, which are characterized by the same stylized facts than the individual commodity EFTs.

Table 18. Descriptive statistics of Commodity ETF Portfolios

Portfolios	**A: Gold-Silver**	**B: Gold-Oil**	**C: Silver-Oil**
Mean	0.0144	−0.0243	−0.343
Median	0.0592	0.0013	−0.0061
Standard deviation	1.6434 (25.98)	1.4132 (22.34)	1.8221 (28.81)
Min	−13.7779	−7.0676	−12.1956
Max	12.0853	7.4417	8.3828
Skewness	−0.7347	−0.1067	−0.4720
Excess Kurtosis	6.7643	2.4751	3.4321

[31] As far as we know this paper presents the first application of the SNP-DCC model with EGARCH innovations.
[32] Moreover, the other ETFs analyzed in this paper (Agriculture, Energy and Broad) track diversified commodity portfolios (see Appendix B).

Annual volatility in parentheses

Table 19 shows the estimated parameters for Gaussian-DCC (Panel A) and SNP-DCC (Panel B). The first two rows present the estimations for the DCC part, and results support the conditional correlation model. It must be noted our application incorporates an AR(1)-EGARCH(1,1). As expected, parameters for volatility part are similar to the estimations found in the univariate section (see Table 2). Furthermore, the parameters of the GC density for both dimensions are significant reflecting the outperformance of the SNP-DCC model.

Table 19. Estimates of DCC models

Panel A: Gaussian-DCC

Portfolios	A: Gold-Silver	B: Gold-Oil	C: Silver-Oil
A-DCC	0.041 (0.000)	0.044 (0.000)	0.024 (0.002)
B-DCC	0.937 (0.000)	0.909 (0.000)	0.950 (0.000)
ϕ_1	0.021 (0.338)	0.021 (0.338)	-0.015 (0.686)
ω_1	0.014 (0.002)	0.014 (0.002)	0.051 (0.000)
α_1	-0.012 (0.504)	-0.012 (0.504)	-0.016 (0.345)
δ_1	0.980 (0.000)	0.980 (0.000)	0.972 (0.000)
β_1	0.146 (0.000)	0.146 (0.000)	0.201 (0.023)
ϕ_2	-0.015 (0.686)	-0.042 (0.191)	-0.042 (0.191)
ω_2	0.051 (0.000)	0.013 (0.000)	0.013 (0.000)
α_2	-0.016 (0.345)	-0.052 (0.000)	-0.052 (0.000)
β_2	0.972 (0.000)	0.991 (0.000)	0.991 (0.000)
γ_2	0.201 (0.023)	0.102 (0.000)	0.102 (0.000)

Panel B: SNP-DCC

Portfolios	A: Gold-Silver	B: Gold-Oil	C: Silver-Oil
A-DCC	0.030 (0.000)	0.039 (0.000)	0.025 (0.001)
B-DCC	0.942 (0.000)	0.911 (0.000)	0.945 (0.000)
ϕ_1	0.003 (0.000)	0.003 (0.000)	-0.003 (0.925)
ω_1	0.016 (0.000)	0.016 (0.000)	0.061 (0.000)
α_1	-0.004 (0.621)	-0.004 (0.621)	-0.003 (0.848)
β_1	0.980 (0.000)	0.980 (0.000)	0.963 (0.000)

γ_1	0.053 (0.000)	0.053 (0.000)	0.055 (0.000)
ϕ_2	-0.003 (0.925)	-0.036 (0.490)	-0.036 (0.490)
ω_2	0.061 (0.000)	0.010 (0.000)	0.010 (0.000)
α_2	-0.003 (0.848)	-0.055 (0.000)	-0.055 (0.000)
β_2	0.963 (0.000)	0.990 (0.000)	0.990 (0.000)
γ_2	0.055 (0.000)	0.022 (0.000)	0.022 (0.000)
d_{31}	-0.008 (0.700)	-0.067 (0.003)	-0.073 (0.001)
d_{41}	0.101 (0.000)	0.099 (0.000)	0.116 (0.000)
d_{32}	-0.059 (0.005)	-0.048 (0.010)	-0.052 (0.007)
d_{42}	0.121 (0.000)	0.034 (0.000)	0.032 (0.000)

A-DCC and *B-DCC* are the parameters of the DCC model; ϕ_i is the parameter of the AR(1) model; ω_i, α_i, γ_i and β_i are the parameters of the EGARCH(1,1) model; d_{4i} and d_{4i} are the parameters of the GC model. P-values in parentheses.

Table 20 presents the results of t-test for ES at 97.5%. Once again, a model is good enough if the null hypothesis cannot be rejected. The SNP-DCC performs adequately in all cases, whereas Gaussian-DCC fails in the three analyzed portfolios.

Table 20. T-test for 97.5%-ES for Commodity ETFs portfolio returns

Mean of violation residual is zero Data: December 2008 – January 2016 (1771 days for backtesting)			
Portfolios	*A: Gold-Silver*	*B: Gold-Oil*	*C: Silver-Oil*
Gaussian-DCC	6.944 (0.000)	4.810 (0.000)	4.713 (0.000)
SNP-DCC	1.938 **(0.061)**	1.664 **(0.110)**	1.699 **(0.100)**

P-values for the t-test test in parentheses.

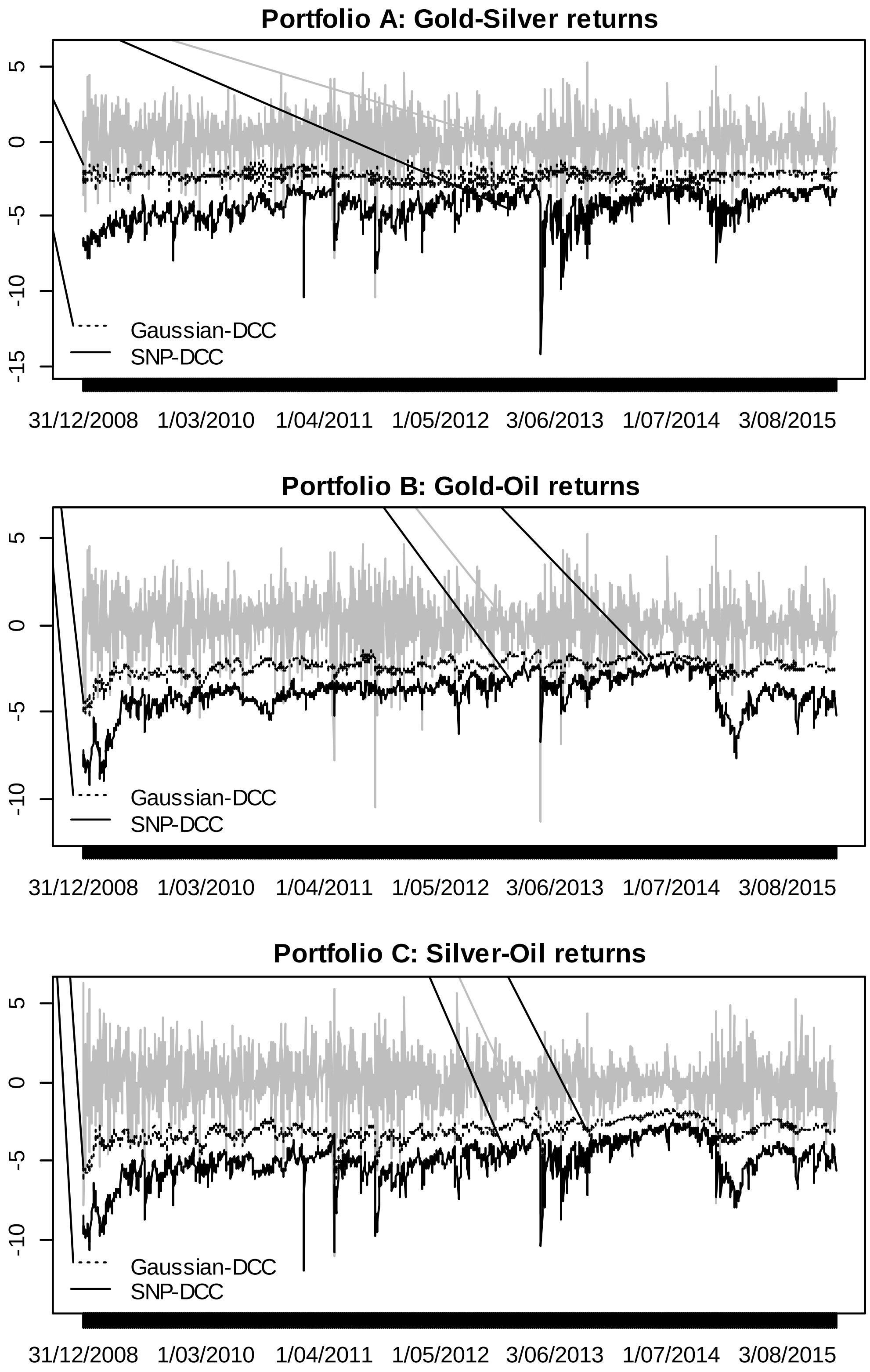
The results of relative performance test for ES are shown in Table 21. A pairwise Diebold Mariano test less than −1.96 indicates that SNP-DCC is preferred to Gaussian-DCC at 5% of significance. The results show the outperformance of SNP-DCC for all the analyzed portfolios.

Table 21. Pairwise Diebold Mariano test for 97.5%-ES

Portfolios	*A: Gold-Silver*	*B: Gold-Oil*	*C: Silver-Oil*
DM test	**-5.540** (0.000)	**-3.532** (0.000)	**-3.584** (0.000)

Finally, Figure 12 presents the comparison of 97.5%-ES estimated by Gaussian-DCC and SNP-DCC for the conditional correlations of the three analyzed portfolios. This figure illustrates the underperformance of the Gaussian-DCC for capturing extreme values and the more accuracy of the SNP-DCC for this purpose.

Figure 12. 97.5%-ES for portfolio of Commodity ETFs



3.6 Discussion

Since the global financial crisis, regulators and policy makers have been concerned about financial stability and an important instrument to achieve this goal is capital adequacy requirements. In this line, the Basel Committee published a consultative document (BCBS, 2013) on the Fundamental Review of Trading Book (FRTB). One of the main changes proposed in the FTRB is replacing 99%-VaR with a 97.5%-ES in the internal models approach for financial institutions. With the new rules, regulators estimate that banks would have to increase capital buffer around 40% in average to mitigate potential market risk losses in their trading books. Nevertheless, few studies are devoted to examine the impact for capital requirements with the new regulatory proposal. For instance, Kellner and Rösch (2016) find that ES is more sensitive towards regulatory arbitrage and model risk, especially under parameter misspecification. Table 22 presents the ratio 97.5%-ES to 99%-VaR in average for the backtesting period applied to the six commodity ETFs analyzed in this paper.

Table 22. Average ratio 97.5%-ES to 99%-VaR

Data: December 2008 – January 2016						
1771 days	**Gold**	**Silver**	**Oil**	**Agriculture**	**Energy**	**Broad**
Gaussian	1.005 (0.000)	1.005 (0.000)	1.005 (0.000)	1.005 (0.000)	1.005 (0.000)	1.005 (0.000)
Student's t	1.427 (0.294)	1.495 (0.340)	1.141 (0.227)	1.154 (0.060)	1.160 (0.104)	1.157 (0.102)
Skewed-t	1.160 (0.104)	1.160 (0.103)	1.160 (0.103)	1.160 (0.104)	1.161 (0.104)	1.161 (0.104)
GC	0.994 (0.004)	0.994 (0.022)	1.000 (0.004)	1.000 (0.006)	0.999 (0.004)	0.999 (0.005)

Standard deviation in parentheses.

As expected, the ratio 97.5%-ES/99%-VaR for the Gaussian case is 1, the reason for the Basel Committee to establish the 97.5% confidence level for ES. From the stylized facts of daily returns, empirical distribution of financial assets exhibits heavier tails than normal, and leptokurtic distributions such as Student's t and its skewed version have been employed. Results of Table 11 point to an average increment between 14% and 42% by replacing VaR with ES, if the financial institution assumes a Student's t risk model, and an average increment of 16% by using a skewed-t risk model. The ratio 97.5%-ES/99%-VaR for GC is also close to one as in the Gaussian case, since GC distribution is an expansion in terms of derivatives of the Gaussian pdf – and their corresponding Hermite polynomials, see equation (9). However, due to the fact that GC allows for skewness and kurtosis, capital requirements based on GC-

ES are higher than Gaussian-ES, and risk measures based on GC distribution perform well according to our results. The ratio ES to VaR for GC implies that 97.5%-ES behaves very similarly to 99%-VaR and studies have shown good backtesting performance for GC-VaR (Del Brio et al., 2014a,b). From the multivariate perspective used for analyzing portfolios, different results from calculating ES at 97.5% under a Gaussian or SNP distribution can be observed. The ratio SNP-ES/Gaussian-ES for Portfolio B (Gold-Oil) and C (Silver-Oil) is 1.66, and for Portfolio A (Gold-Silver) is 1.91. This evidence implies that risk is underestimated more than a half (for Portfolio A) if a financial institution employs the Gaussian model to buffer capital against potential losses in commodity ETF markets.

On the other hand, the high volatility in the commodity markets has been pressuring emerging economies, and the correlation of equity-commodities has increased with financialization of commodities (Basak and Pavlova, 2016). The results of this paper highlight the use of leptokurtic distribution to assessing risk, especially GC distribution, in order to mitigate potential losses when there is an exposure of commodities and other financial instruments.

ETFs have provided investors, portfolio managers and risk quantification area access to a variety of alternative strategies and investments with the advantages of liquidity transparency and cost effectiveness. In financial crises, the main concern for financial institutions is high volatility, and Volatility ETFs may be a solution for fluctuations in implied market volatility. Asset managers have been considering the inclusion of commodities in their portfolios, since commodities are negatively correlated to other financial assets in “relatively” calm periods or for market-neutral strategy reasons. One of the best ways to include commodities in portfolios is through ETFs due to the abovementioned advantages. However, in crisis periods correlations between asset returns tend to be positively high. Thus, this paper focuses on Commodity ETFs in order to quantify its risk especially in crisis periods.

3.7 Conclusions of Chapter 3

The Basel Committee’s fundamental review of trading book proposes to replace Value-at-Risk (VaR) at 99% by Expected Shortfall (ES) at 97.5% as the more accurate market risk measure. This proposal has initiated a controversial debate in the academy and financial industry about the appropriateness of such measure, mainly due to its troublesome backtesting implementation. On the other hand, there are straightforward methods to backtest VaR, although this risk measure fails to satisfy the subadditivity axiom, a desirable property to be a coherent risk measure. ES indeed satisfy this latter property but its lacks of elicitability opens the debate on whether ES might be backtestable. However, recent studies have shown that, under certain conditions, it is possible to implement performance tests in terms of ES jointly with VaR. This paper presents an application of these methods to different Commodity ETFs in order to shed some light on the risk assessment of different techniques, which cover parametric (Gaussian, Student's t and skewed-t), and semi-nonparametric (with two alternative specifications: the univariate and multivariate case). Particularly, as far as we know, the ES assessment of the latter is being tested for the first time.

Coverage test for backtesting 97.5%-VaR shows that ARMA-EGARCH-skewed-t performs satisfactorily for Commodity ETFs. Whereas, relative performance shows that ARMA-EGARCH-GC and ARMA-EGARCH-t are the best models for 97.5% VaR and 97.5% ES, which is the new confidence level proposed by Basel Accords. For portfolios of Commodity ETFs, the SNP-DCC is preferred to the Gaussian-DCC according to t-test and relative performance tests. Therefore, our study recommends the implementation of these techniques, particularly the one based on the Gram-Charlier expansion for which we provide a straightforward closed form for ES.

Future research will study of alternative semi-nonparametric methods for portfolio assessment – e.g. the SNP-DECO model by Ñíguez and Perote (2016) – and the applications of new risk measures such as median shortfall, which has been proven to be elicitable and satisfies a wide set of economic axioms (Kou and Peng, 2014). Spectral risk measures introduced by Acerbi (2002) appear to be other interesting avenues for research, and calculating confidence intervals of these risk measures on a specific model, since European Banking Authority urges the need for its calculation. Another future research will be focused on the analysis of interdependence in commodity market caused by financialization and contagion ES as per in Suh (2015). From the point of view of risk managers is also interesting to analyze the role of commodities (and commodity ETFs) in risk diversification, since the negative correlation between commodities and stocks has recently become strongly positive.

References

Acerbi, C. 2002. Spectral measures of risk: A coherent representation of subjective risk aversion. Journal of Banking and Finance 26, 1505-1518.

Acerbi, C., Tasche, D. 2002. On the coherence of expected shortfall. Journal of Banking and Finance 26, 1487-1583.

Aloui, C., Jammazi, R. 2015. Dependence and risk assessment for oil prices and exchange rate portfolios: A wavelet based approach. Physica A 436, 62-86.

Aloui, C., Mabrouk, S. 2010. Value-at-risk estimations of energy commodities via long memory, asymmetry and fat-tailed GARCH models. Energy Policy 38, 2326-2339.

Andriosopoulos, K., Nomikos, N. 2015. Risk management in the energy markets and Value-at-Risk modelling: a hybrid approach. The European Journal of Finance 21 (7), 548-574.

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.

Basak, S. Pavlova, A. 2016. A Model of Financialization of Commodities. The Journal of Finance 71 (4), 1511-1555.

BCBS (Basel Committee on Banking Supervision). 2013. Fundamental Review of the Trading Book: A Revised Market Risk Framework, Bank for International Settlements, Basel.

Bellini, F., Bignozzi, V. 2015. On elicitable risk measures. Quantitative Finance 15 (5), 725-733.

Bellini, F., Klar, B., Müller, A., Rosazza Gianin, E. 2014. Generalized quantiles as risk meausres. Insurance: Mathematics and Economics 54, 41-48.

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics 31, 307-327.

Chiu, Y.-C., Chuang, I.-Y., Lai, J.-Y. 2010. The performance of composite forecast models of value-at-risk in the energy market. Energy Economics 32, 423-431.

Cont, R., 2001. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. Quantitative Finance 1, 223-236.

Cont, R., Deguest, R., Scandolo, G. 2010. Robustness and sensitivity analysis of risk measurement procedures. Quantitative Finance 10, 593-606.

Del Brio, E. B., Perote J., 2012. Gram-Charlier Densities: Maximum Likelihood versus the Method of Moments. Insurance: Mathematics and Economics 51, 531–537.

Del Brio E.B., Níguez, T.M., Perote, J. 2011. Multivariate semi–nonparametric distributions with dynamic conditional correlations. International Journal of Forecasting 27, 347–364.

Del Brio, E.B., Mora-Valencia, A., Perote, J. (2014a). VaR performace during the subprime and sovereign debt crises: An application to emerging markets. Emerging Markets Review 20, 23–41.

Del Brio, E.B., Mora-Valencia, A., Perote, J. (2014b). Semi–nonparametric VaR forecasts for hedge funds during the recent crisis. Physica A 401, 330–343.

Diebold, F.X., Mariano, R.S. 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253-263.

Embrechts, P., Wang, B., Wang, R. 2015. Aggregation-robustness and model uncertainty of regulatory risk measures. Finance and Stochastics 19 (4), 763-790.

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.

Engle, R.F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom. Econometrica, 50, 987-1007.

Engle R.F., Kelly B. 2012. Dynamic Equicorrelation. Journal of Business and Economic Statistics 30, 212–228.

Fernández, C., Steel M.F.J., 1998. On Bayesian Modeling of Fat Tails and Skewness. Journal of the American Statistical Association 93, 359–371.

Fissler, T., Ziegel, J.F. 2016. Higher order elicitability and Osband's principle. Annals of Stastics, forthcoming.

Fissler, T., Ziegel, J.F., Gneiting, T. 2016. Expected Shortfall is jointly elicitable with value-at-risk: Implications for backtesting. Risk 29 (1) 58-61.

Gallant, A. R., Nychka, D. W., 1987. Semi-Nonparametric Maximum Likelihood Estimation. Econometrica 55, 363-390.

Giot, P., Laurent, S. 2003. Market risk in commodity markets: a VaR approach. Energy Economics 25, 433-457.

Gneiting, T. 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106, 746-762.

Hume, N., Sanderson, H. Sheppard, D. 2016, April 28. China futures gripped as speculators rush in. The Financial Times. Retrieved from http://www.ft.com.

Hung, J.-C., Lee, M.-C., Liu, H.-C. 2008. Estimation of value-at-risk for energy commodities via fat-tailed GARCH models. Energy Economics 30, 1173-1191.

Jarrow, R., Rudd A., 1982. Approximate Option Valuation for Arbitrary Stochastic Processes. Journal of Financial Economics 10, 347–369.

Jondeau, E., Rockinger, M. 2001. Gram-Charlier Densities. Journal of Economic Dynamics and Control 25, 1457-1483.

Jorion, P., 2006. Value at Risk: The New Benchmark for Managing Financial Risk, Third Edition, McGraw-Hill.

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 & Control 68, 45–63.

Kendall, M., Stuart, A., 1977. The Advanced Theory of Statistics. Griffin&Co, London.

Koch-Medina, P., Munari, C. 2016. Unexpected shortfalls of Expected Shortfall: Extreme default profiles and regulatory arbitrage. Journal of Banking and Finance 62, 141-151.

Kou, S., Peng, X. 2014. On the measurement of economic tail risk. Preprint arXiv: 1401.4787.

León, A., Mencía, J., Sentana, E., 2009. Parametric Properties of Semi-nonparametric Distributions. Journal of Business and Economic Statistics 27, 176-192.

Lu, X.F., Lai, K.K., Liang, L. 2014. Portfolio value-at-risk estimation in energy futures markets with time-varying copula-GARCH model. Annals of Operation Research 219, 333-357.

Mauleón, I., Perote, J. 2000. Testing Densities with Financial Data: An Empirical Comparison of the Edgeworth-Sargan Density to the Student's t. European Journal of Finance 6, 225-239.

McNeil, A.J., Frey, R., Embrechts, P. 2015. Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press (Revised Edition).

Nelson, D.B. 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347—370.

Newey, W.K., Powell, J.L. 1987. Asymmetric least squares estimation and testing. Econometrica 55, 819-847.

Ñíguez, T.M., Perote, J., 2012. Forecasting Heavy-Tailed Densities with Positive Edgeworth and Gram-Charlier Expansions. Oxford Bulletin of Economics and Statistics 74, 600-627.

Ñíguez, T.M., Perote, J., 2016. Multivariate Moments Expansion Density: Application of the Dynamic Equicorrelation Model. Journal of Banking and Finance 72, S216-S232.

Patton, A.J., Sheppard, K. 2009. Evaluating volatility and correlation forecasts. In T. Mikosch, J.-P. Kreiss, R. A. Davis, and T. G. Andersen, editors, Handbook of Financial Time Series, pages 801–838. Springer, Berlin.

Rockafellar, R.T., Uryasev, S. 2002. Conditional value-at-risk for general loss distributions. Journal of Banking and Finance 26 (7) 1443-1471.

Singleton, K.J. 2013. Investor Flows and the 2008 Boom/Bust in Oil Prices. Management Science 60 (2), 300-318.

Steen, M., Westgaard, S., Gjolberg, O. 2015. Commodity value-at-risk modeling: comparing RiskMetrics, historic simulation and quantile regression. Journal of Risk Model Validation 9 (2), 49-78.

Suh, S. 2015. Measuring sovereign risk contagion in the Eurozone. International Review of Economics and Finance 35 (2015) 45–65.

Tang, K., Xiong, W. 2012. Index Investment and Financialization of Commodities. Financial Analysts Journal 68 (6), 54-74.

Taylor, J.W. 2008. Estimating value at risk and expected shortfall using expectiles. Journal of Financial Econometrics 6 (2), 231-252.

Weber, S. 2006. Distribution-invariant risk measures, information, and dynamic consistency. Mathematical Finance (16) 419–441.

Yamai, Y., Yoshiba, T. 2005. Value-at-risk versus expected shortfall: A practical perspective. Journal of Banking and Finance 29, 997-1015.

Youssef, M., Belkacem, L., Mokni, K. 2015. Value-at-Risk estimation of energy commodities: A long-memory GARCH-EVT approach. Energy Economics 51, 99-110.

Ziegel, J.F. 2014. Coherence and elicitability. Mathematical Finance DOI: 10.1111/mafi.12080.

Appendix A. ES computation for GC

ES for the GC Type A distribution can be obtained as the expected probability of variable z_t conditioned on the fact that this variable had exceeded a given quantile $\hat{q}_\alpha(z_t) = \varphi^{-1}(\alpha)$. Therefore

$$EP\left[z_t/z_t < \varphi^{-1}(\alpha)\right] = \int_{-\infty}^{\varphi^{-1}(\alpha)} z_t f(z_t) dz_t$$

$$= \int_{-\infty}^{\varphi^{-1}(\alpha)} z_t \phi(z_t) \left[1 + \sum_{s=2}^{S} d_s H_s(z_t)\right] dz_t$$

$$= \int_{-\infty}^{\varphi^{-1}(\alpha)} z_t \phi(z_t) dz_t + \int_{-\infty}^{\varphi^{-1}(\alpha)} \sum_{s=2}^{S} d_s H_s(z_t)\, z_t \phi(z_t) dz_t$$

$$= \int_{-\infty}^{\varphi^{-1}(\alpha)} z_t \phi(z_t) dz_t + \sum_{s=2}^{S} \int_{-\infty}^{\varphi^{-1}(\alpha)} d_s H_s(z_t) z_t \phi(z_t)\, dz_t$$

$$= \int_{-\infty}^{\varphi^{-1}(\alpha)} z_t \phi(z_t) dz_t + \sum_{s=2}^{S} d_s \int_{-\infty}^{\varphi^{-1}(\alpha)} z_t H_s(z_t) \phi(z_t) dz_t$$

$$= -\phi\left(\varphi^{-1}(\alpha)\right) + \sum_{s=2}^{S} d_s \left[-H_s\left(\varphi^{-1}(\alpha)\right) \phi\left(\varphi^{-1}(\alpha)\right) + s \int_{-\infty}^{\varphi^{-1}(\alpha)} H_{s-1}(z_t) \phi(z_t) dz_t\right]$$

$$= -\phi\left(\varphi^{-1}(\alpha)\right) \left[1 + \sum_{s=2}^{S} d_s \left[H_s\left(\varphi^{-1}(\alpha)\right) + s H_{s-2}\left(\varphi^{-1}(\alpha)\right)\right]\right]$$

By direct application of Lemma 1 and Lemma 2 in Ñíguez and Perote (2012) for and taking into account that $H_1(z_t) = z_t$.

Consequently, for the GC expansion considered in Section 4 and $u = \varphi^{-1}(\alpha)$,

$$f(z_t) = [1 + d_3 H_3(z_t) + d_4 H_3(z_t) + d_6 H_6(z_t) + d_8 H_8(z_t)]\phi(z_t)$$

$$EP\left[z_t/z_t < u\right] = -\phi(u) - d_3\phi(u)[u^3 - 3u + 3u] - d_4\phi(u)[u^4 - 6u^2 + 3 + 4(u^2 - 1)]$$
$$- d_6\phi(u)[u^6 - 15u^4 + 45u^2 - 15 + 6u^4 - 36u^2 + 18]$$

$$-d_8\phi(u)[u^8 - 28u^6 + 21015u^4 - 420u^2 + 105 + 8(u^6 - 15u^4 + 45u^2 - 15)]$$

$$= -\phi(u) - d_3\phi(u)u^3 - d_4\phi(u)[u^4 - 2u^2 - 1] - d_6\phi(u)[u^6 - 9u^4 + 9u^2 + 3]$$
$$- d_8\phi(u)[u^8 - 20u^6 + 90u^4 - 60u^2 - 15]$$

Appendix B. Data description

ETF Commodity	Ticker	Description
Gold	GLD	SPDR Gold Shares is an investment fund incorporated in the USA. The investment objective of the Trust is for the Shares to reflect the performance of the price of gold bullion, less the Trust's expenses. The Trust holds gold and is expected from time to time to issue Baskets in exchange for deposits of gold and to distribute gold in connection with redemptions of Baskets.
Silver	SLV	iShares Silver Trust is a trust formed to invest in silver. The assets of the trust consist primarily of silver held by the custodian on behalf of the trust. The objective of the trust is for the shares to reflect the price of silver owned by the trust, less the trust's expenses and liabilities.
Oil	USO	United States Oil Fund LP is a Delaware limited partnership incorporated in the USA. The Fund's objective is to have changes in percentage terms of its unit's net asset value reflect the changes of the price of WTI Crude Oil delivered to Cushing, Oklahoma, as measured by changes in percentage terms of the price of the WTI Crude Oil futures contract on the NYMEX.
Agriculture	DBA	PowerShares DB Agriculture Fund is an exchange-traded fund incorporated in the USA. The Fund's objective is to reflect the performance of the DBIQ Diversified Agriculture Index Excess Return.
Energy	DBE	PowerShares DB Energy Fund is an exchange-traded fund incorporated in the USA. The Fund's objective is to track the DBIQ Optimum Yield Energy Index Excess Return.
Broad	DBC	PowerShares DB Commodity Index Tracking Fund is an investment fund incorporated in the USA. The Fund's objective is to reflect the performance of the DBIQ Optimum Yield Diversified Commodity Index Excess. The Fund invests in commodities such as Light, Sweet Crude Oil, Heating Oil, Aluminum, Gold, Corn and Wheat.

Source: Bloomberg LP.

Appendix C. Pairwise Diebold Mariano test for 99%-VaR

Model A → / Model B ↓	ARMA-EGARCH-t	ARMA-EGARCH-skewed-t	ARMA-EGARCH-GC
Panel C.1. Gold ETF returns			
ARMA-EGARCH-normal	**-2.346** (0.009)	-1.507 (0.145)	-1.912 (0.028)
ARMA-EGARCH-t		0.089 (0.535)	-1.161 (0.123)
ARMA-EGARCH- skewed-t			-0.931 (0.176)
Panel C.2. Silver ETF returns			
ARMA-EGARCH-normal	-1.600 (0.055)	4.304 (0.999)	-0.303 (0.381)
ARMA-EGARCH-t		4.151 (0.999)	0.071 (0.528)
ARMA-EGARCH- skewed-t			**-2.647** (0.004)
Panel C.3. Oil ETF returns			
ARMA-EGARCH-normal	1.311 (0.905)	3.828 (0.999)	**-2.129** (0.017)
ARMA-EGARCH-t		-1.311 (0.095)	-1.311 (0.095)
ARMA-EGARCH- skewed-t			**-4.363** (0.000)
Panel C.4. Broad ETF returns			
ARMA-EGARCH-normal	**-2.996** (0.001)	**-2.532** (0.006)	-1.682 (0.046)
ARMA-EGARCH-t		**-2.218** (0.013)	-0.509 (0.305)
ARMA-EGARCH- skewed-t			1.564 (0.941)
Panel C.5. Agriculture ETF returns			
ARMA-EGARCH-normal	-1.482 (0.069)	**-2.367** (0.009)	**-2.212** (0.013)
ARMA-EGARCH-t		-0.700 (0.242)	-1.130 (0.129)
ARMA-EGARCH- skewed-t			-0.944 (0.173)
Panel C.6. Energy ETF returns			
ARMA-EGARCH-normal	-1.826 (0.034)	-1.825 (0.034)	1.438 (0.925)
ARMA-EGARCH-t		1.697 (0.955)	2.620 (0.996)
ARMA-EGARCH- skewed-t			2.620 (0.996)

The table shows the Diebold-Mariano statistics for different models. Bold figures indicate Model A is preferred than Model B, figures in red indicate the opposite. Otherwise both models present the same performance. P-values in parentheses.

Appendix D. Pairwise Diebold Mariano test for 97.5%-VaR

Model A → / Model B ↓	ARMA-EGARCH-t	ARMA-EGARCH-skewed-t	ARMA-EGARCH-GC
Panel D.1. Gold ETF returns			
ARMA-EGARCH-normal	-0.268 (0.394)	-1.170 (0.121)	**-2.064** (0.019)
ARMA-EGARCH-t		-1.342 (0.089)	1.566 (0.059)
ARMA-EGARCH- skewed-t			-0.691 (0.245)
Panel D.2. Silver ETF returns			
ARMA-EGARCH-normal	0.633 (0.737)	5.172 (0.999)	-0.2285 (0.409)
ARMA-EGARCH-t		4.871 (0.999)	-0.356 (0.361)
ARMA-EGARCH- skewed-t			**-3.488** (0.000)
Panel D.3. Oil ETF returns			
ARMA-EGARCH-normal	1.309 (0.905)	4.997 (0.999)	**-2.685** (0.004)
ARMA-EGARCH-t		-1.309 (0.095)	-1.309 (0.095)
ARMA-EGARCH- skewed-t			**-5.619** (0.000)
Panel D.4. Broad ETF returns			
ARMA-EGARCH-normal	**-2.645** (0.004)	**-3.239** (0.001)	-1.777 (0.038)
ARMA-EGARCH-t		**-3.244** (0.001)	-1.035 (0.150)
ARMA-EGARCH- skewed-t			1.990 (0.977)
Panel D.5. Agriculture ETF returns			
ARMA-EGARCH-normal	-1.133 (0.129)	**-3.508** (0.000)	**-2.604** (0.005)
ARMA-EGARCH-t		**-2.752** (0.003)	-1.788 (0.037)
ARMA-EGARCH- skewed-t			0.305 (0.620)
Panel D.6. Energy ETF returns			
ARMA-EGARCH-normal	-1.555 (0.060)	-1.555 (0.060)	1.068 (0.857)
ARMA-EGARCH-t		1.738 (0.959)	1.866 (0.969)
ARMA-EGARCH- skewed-t			1.866 (0.969)

The table shows the Diebold-Mariano statistics for different models. Bold figures indicate Model A is preferred than Model B, figures in red indicate the opposite. Otherwise both models present the same performance. P-values in parentheses.

Appendix E. Pairwise Diebold Mariano test for 97.5%-ES

Model A → / Model B ↓	ARMA-EGARCH-t	ARMA-EGARCH-skewed-t	ARMA-EGARCH-GC
Panel E.1. Gold ETF returns			
ARMA-EGARCH-normal	**-2.451** (0.007)	**-2.166** (0.015)	-**1.993** (0.023)
ARMA-EGARCH-t		-1.683 (0.046)	0.855 (0.196)
ARMA-EGARCH- skewed-t			1.459 (0.927)
Panel E.2. Silver ETF returns			
ARMA-EGARCH-normal	-**2.186** (0.014)	0.989 (0.838)	**-1.961** (0.025)
ARMA-EGARCH-t		4.193 (0.999)	0.394 (0.637)
ARMA-EGARCH- skewed-t			**-4.143** (0.000)
Panel E.3. Oil ETF returns			
ARMA-EGARCH-normal	1.707 (0.956)	4.233 (0.999)	0.076 (0.530)
ARMA-EGARCH-t		2.415 (0.992)	-1.670 (0.047)
ARMA-EGARCH- skewed-t			-**3.977** (0.000)
Panel E.4. Broad ETF returns			
ARMA-EGARCH-normal	-0.933 (0.175)	**-2.114** (0.0173)	0.257 (0.602)
ARMA-EGARCH-t		-1.923 (0.027)	0.615 (0.731)
ARMA-EGARCH- skewed-t			2.028 (0.979)
Panel E.5. Agriculture ETF returns			
ARMA-EGARCH-normal	**-2.185** (0.014)	0.242 (0.596)	-1.163 (0.122)
ARMA-EGARCH-t		5.980 (0.999)	1.638 (0.949)
ARMA-EGARCH- skewed-t			-**2.177** (0.015)
Panel E.6. Energy ETF returns			
ARMA-EGARCH-normal	-1.320 (0.093)	-1.300 (0.097)	2.174 (0.985)
ARMA-EGARCH-t		1.697 (0.955)	2.410 (0.992)
ARMA-EGARCH- skewed-t			2.407 (0.992)

The table shows the Diebold-Mariano statistics for different models. Bold figures indicate Model A is preferred than Model B, figures in red indicate the opposite. Otherwise both models present the same performance. P-values in parentheses.

Chapter 4. A note on SMA vs. LDA-AMA: The dawning of a new regulation[33]

Abstract

A recent Basel Committee on Banking Supervision publication suggesting a switch from Advance Measurement Approach (AMA) to Standardized Measurement Approach (SMA), has generated debate in the financial industry and among academics regarding the new rules. This note presents a non-exhaustive review of the literature on operational risk quantification under a combination of the Loss Distribution Approach (LDA) model—the most commonly used approach to AMA models—and Extreme Value Theory (EVT). The literature review points out that Bayesian inference has provided solutions to different problems when modelling operational data. The main comments prepared by the financial industry in response to the new proposal and two recently published papers which analyze the impact of SMA are also summarized in the present document. Finally, the discussion section proposes an alternative solution, a single-loss approximation model (taking into account several severity and frequency distributions) with an appropriate risk measure under a Bayesian Model Averaging (BMA) setting as an intermediate solution to estimate operational risk capital, and its application will be the focus for future research.

Keywords: advanced measurement approach, loss distribution approach, standardized measurement approach, Bayesian methods

4.1 Introduction of Chapter 4

Operational risk is perhaps the most difficult risk to quantify. For a reliable assessment method, it is necessary to understand the nature of operational risk in accordance with its definition, its empirical characteristics and the risk factors of operational loss.

Operational risk is defined as: "the risk of a loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk" (BCBS, 2006, p. 144). Studies have found that the

[33] A version of this chapter has been published in the Journal of Operational Risk

main stylized facts (empirical characteristics) of operational loss are: (i) the existence of extremes and (right) skewed distributed, (ii) the variation of loss frequencies over time, (iii) the fact that loss severities are often infinite-mean,[34] and (iv) tail dependence (Moscadelli, 2004; McNeil et al., 2005; Nešlehová et al. 2006; Chavez-Demoulin, et al. 2006; Brechman et al., 2014). These empirical characteristics pose a challenge for risk managers and regulators in terms of estimating operational risk capital. Under Basel II, there are three possible methods to model operational risk: The Basic Indicator Approach (BIA), the Standard Approach (TSA) and the Advanced Measurement Approach (AMA). Inasmuch as AMA models, the Loss Distribution Approach (LDA) is the most commonly used by sophisticated banks.

The aim of the Basel Committee on Banking Supervision is to propose a simple method that captures the simplicity and comparability characteristics of a standardized approach, and that gathers the risk-sensitivity property of an advanced approach. Given the recent crises, a further purpose of the Basel Committee is to increase the minimum capital requirements, but not significantly so.

In March 2016, the Basel Committee (BCBS, 2016) published the consultative document on the Standardized Measurement Approach (SMA) as a means for calculating operational risk capital. This proposal has generated a debate in the financial industry and among academics. Some operational risk consultants argue that AMA models—which involve choice of severity, frequency models, and loss aggregation—are too complicated and that they do not make sense in terms of operational risk. The real reason for which AMA is considered complicated is that the nature of operational risk is not fully understood. Thus the Basel Committee wants to minimize the drawbacks of AMA by proposing a simple equation. On the other hand, a different section of the financial industry and a number of academics consider that the Basel Committee ignores the results of recent research, which indicates that tail operational losses are a source of systemic risk and a simple equation cannot adequately model the idiosyncratic and systemic components of operational risk. The SMA-averse sector agrees that new rules constitute a backward-looking measure, since they depend on historical data that is insensitive to current levels of risk. In addition, the main variable of SMA, the so-called "business indicator", depends only on size (gross income) and it is not related to a risk measure as is the

[34] Infinite-mean models are distributions which tail index is between 0 and 1. Therefore, it is not possible to quantify the expected value of the tail and risk measures such as expected shortfall are meaningless.

case for AMA models. This could provoke a "perverse" incentive to misreport losses. Moreover, the new proposal does not consider insurance hedging.

4.2 LDA-AMA Model

AMA allows banks to employ their internal models to assess capital charges and quantify operational risk at 99.9%-VaR in a 1-year horizon. LDA is the most commonly used approach by sophisticated banks and is a very well-known tool in insurance analytics. For the LDA, operational losses are classified into a matrix of eight business lines and seven event types. LDA-AMA includes four inputs: (1) internal loss data, (2) external loss data, (3) scenario analysis, and (4) business environment and internal control factors (BEICFs).

More recent references that cover the abovementioned inputs to model operational risk are Girling (2013), Cruz and Peters (2015), and Cavestany et al. (2015). A strand of the literature has proposed solutions to combine internal and external data, including Bayesian methods (see Table 1) and scaling models (Dahen and Dionne, 2010). For the fourth element (BEICFs), Dutta and Babbel (2013) propose a method that combines scenario analysis with historical loss data to assess the impact of each scenario on the total operational risk capital using the change of measure approach.

Table 23. Studies proposing Bayesian methods to combine internal and external data published in the Journal of Operational Risk

Authors	Title of the publication	Year
Shevchenko, P. Wütrich, M.	The structural modelling of operational risk via Bayesian inference: Combining loss data with expert opinions	2006
Lambrigger, D., Shevchenko, P., Wütrich, M.	The quantification of operational risk using internal data, relevant external data and expert opinion	2007
Gustaffson, J., Nielsen, J.P.	A mixing model for operational risk	2008
Peters, G., Shevchenko, P, Wütrich, M.	Dynamic operational risk: modelling dependence and combining different sources of information	2009
Agostini, A., Talamo, P., Vecchione, V.	Combining operational loss data with expert opinions through advanced credibility theory	2010

Several workshops and conferences have been organized in which practitioners and the academic community have discussed the challenges in measuring operational risk. For instance, the working paper "Statistical Issues in Financial Risk Modeling and Banking Regulation" which resulted from the workshop organized in 2009, found that combinations of LDA and Extreme Value Theory (EVT) were the most commonly used by the industry when calculating a high quantile (i.e. 99.9%-VaR); however, no best practice as such was established. Furthermore, the "optimal" threshold selection is still an open question under EVT, and the LDA model underestimates risk when compared to other approaches (Haubenstock and Hardin, 2003; Madigan, 2009). More details of the LDA model and its applications can be found in Frachot et al. (2001), Shevchenko (2010), among others.

4.3 Modelling the Tails

A big challenge in applying LDA is the severity distribution selection. There are a number of interesting articles in the Journal of Operational Risk that investigate several distributions (EVT distributions, (truncated) lognormal, Pareto, Champernowne, GB2, g-and-h), and how to model operational risk under LDA (single-loss approximation, fuzzy approach, convolution operator, fast Fourier transforms, Panjer recursions, copulas, Bayesian, robust and maximum entropy methods). Given the nature of operational losses, it seems that extreme value distributions are the most adequate. One of the first publications introducing EVT in operational risk was written by Medova (2000) who proposed the use of expected shortfall (ES) instead of Value-at-Risk (VaR), due to the fact that VaR can violate the subadditivity property. Several works have shown the good performance of EVT under operational loss data (de Fontnouvelle et al., 2003; de Fontnouvelle et al., 2004; Moscadelli, 2004; Han et al., 2015). Nevertheless, some conditions, especially in the data, must be satisfied in order to obtain consistent results in applying EVT under the operational risk framework. EVT accurately estimates high quantiles when there is sufficient data to calibrate the models (e.g. parameters in the Generalized Pareto Distribution (GPD) when applying the Peak Over Threshold (POT) methodology). Moreover, second order[35] information about the true data generating process complicates convergence to the GPD approach (Chavez-Demoulin et al., 2006). The latter may occur when the severities can be accurately modelled by a g-and-h distribution (Dutta and

[35] For more details of second-order regular variation theories and their relation with high-quantile estimation for operational risk, see for instance Degen and Embrechts (2008) and references therein.

Perry, 2006; Degen et al., 2006); however, other studies have shown similar and consistent results when comparing EVT and g-and-h calibration to operational losses (Jobst, 2007; Buch-Kromann, 2009). For other aspects to be considered when applying EVT, see for instance Diebold et al. (1998), Embrechts et al. (1997), Nešlehová et al. (2006), and Embrechts (2009). Robust statistics provide a plausible solution for scarce and contaminated data when estimating EVT distribution parameters of operational severities (Huber, 2008). Recently, Chavez-Demoulin et al. (2015) proposed a dynamic EVT approach which allows parameters to vary with certain covariates (business line, event type and time).

Nevertheless, the main critiques of LDA-AMA include the instability of capital charge (i.e. high variability in risk-weighted asset calculations), especially when there is a big loss, lack of forward-looking information, and inconsistent exposure estimates by using different parametric models to fit the losses (other critiques to AMA are found in Moosa, 2007; 2008). In particular, Zhou et al. (2016) and other authors referenced in Table 24, propose methods to quantify operational capital in a more accurate, robust and less volatile fashion, which considers the over-dispersion of the frequency process (Feria-Domínguez et al., 2015).

Table 24. Studies regarding robust estimation of operational risk published in the Journal of Operational Risk

Authors	Title of the publication	Year
Horbenko, N., Ruckdeschel, P., Bae, T.	Robust estimation of operational risk	2011
Opdyke, J.D., Cavallo, A.	Estimating operational risk capital: the challenges of truncation, the hazards of maximum likelihood estimation, and the promise of robust statistics	2012
Opdyke, J.D.	Estimating operational risk capital with greater accuracy, precision and robustness	2014

Non-parametric methods have also been proposed to quantify operational risk VaR, due to the advantage they present in terms of not assuming specific distributions (Buch-Kromann et al., 2007, Bolancé et al., 2012; Tursunalieva and Silvapulle, 2016), as well as right-truncated distributions which allow more stable estimations of capital requirements by employing AMA (Carrillo-Menéndez and Suárez, 2012).

The OpRisk North America and Europe conferences have also been actively gathering people in the industry to discuss the latest approaches to estimate capital buffer. Since the Basel Committee for Banking Supervision (BCBS) proposed that AMA should be replaced by SMA, the latest meetings have been deliberating the convenience of the implementation of the latter approach.

4.4 SMA

As in the cases for BIA and TSA, SMA only reports a capital charge at enterprise level and neglects the relevant information on loss events in different business lines and external operational loss. SMA comprises two main elements, the Business Indicator (BI) intended to reflect systematic risk, and the Loss Component (LC), which attempts to represent the idiosyncratic risk.

The BI is calculated as the sum of the 3-year average of (i) interest, lease and dividend component (ILDC), (ii) the service component (SC), and (iii) the financial component (FC). Banks are classified into five buckets depending on their BI value, and the BI factor ranges 0.11−0.29 in order to find the BI component (BIC) as indicated in Table 25.

Table 25. BI component assessment depending on BI values

Bucket	BI (€ bn)	BI component
1	0 – 1	0.11BI
2	1 – 3	110 m + 0.15(BI – 1 bn)
3	3 – 10	410 m + 0.19(BI – 3 bn)
4	10 – 30	1.74 bn + 0.23(BI – 10 bn)
5	30 to +∞	6.34 bn + 0.29(BI – 30 bn)

The constant values in the third column of Table 25 assure that the BI component is a piecewise linear function depending on the BI values. These values are obtained after data calibration collected in the Quantitative Impact Studies (QIS) by the Committee in 2015 and modification of its revision report to simple approaches (BCBS, 2014). As a result of the latter report, BI replaces Gross Income (GI) in the BIA and TSA methods, since BI variable is more risk sensitive according to the Basel Committee. The business lines and regulatory coefficients of TSA are also changed by five buckets and corresponding coefficients (after technical and cluster analyses) as presented in Table 26. Then, the BI factors in the SMA are in line with the coefficients initially proposed in BCBS (2014).

Table 26. Proposed coefficients per bucket (BCBS, 2014)

Bucket	BI (€ bn)	coefficient
1	0 – 0.1	10%
2	0.1 – 1	13%
3	1 – 3	17%
4	3 – 30	22%
5	30 to +∞	30%

Another important variable in SMA is the Internal Loss Multiplier (ILM), which depends on LC

$$ILM = \ln\left\{e^1 - 1 + \frac{LC}{BIC}\right\},$$

where

$$LC = 7\frac{1}{T}\sum_{t=1}^{T} S(t) + 7\frac{1}{T}\sum_{t=1}^{T} S(t)\mathrm{I}_{\{S(t)>10\mathrm{m}\}} + 5\frac{1}{T}\sum_{t=1}^{T} S(t)\mathrm{I}_{\{S(t)>100\mathrm{m}\}},$$

where $S(t)$ is the total annual operational loss and T is equal to 10 years, or 5 years if the bank does not have good quality loss data; $\mathrm{I}_{\{.\}}$ denotes the usual indicator function, which is equal to 1 if the condition inside the brackets holds, and zero otherwise. Thus, SMA suggests the use of historical 10-year internal loss data for calculating the ILM, whereas AMA requires 5 years of historical data. However, ILM is a backward-looking measure, which acts as a risk sensitivity factor whose purpose is to improve operational risk management.

Bringing the elements of SMA together, the regulatory capital is then calculated as

$$SMA = \begin{cases} BIC, & \text{if Bucket } 1, \\ 110 + ILM(BIC - 110), & \text{if Buckets } 2-5. \end{cases}$$

Despite the fact that the purpose of BCBS is not a significant increment in capital buffer, according to a study[36] by Operational Riskdata eXchange (ORX), there is an increase in the average of 61% capital under SMA compared with the current regulation. Another finding of the study is that SMA capital at a consolidated level can be much higher than the sum of subsidiary level capital. This may cause the decentralization of risk-management systems and banks would have the incentive to legally break up into their subsidiaries in order to report less regulatory capital; a problem faced when using a non-subadditive risk measure (see Section 6.1.1 of McNeil et al., 2005). Another criticism is that operational loss exhibits extremes, and ILM includes a division of LC by the BI component which is calculated as an average.

The comments on the BCBS consultative document on SMA were received by June 3, 2016. In general, the financial industry[37] disagrees with the proposed methodology due to; (i) a loss of investment in terms of resources to meet AMA requirements; (ii) a business indicator that does not capture conduct risk which is the main characteristic of operational risk; (iii) the needlessness of applying SMA to operational losses that can be modelled by AMA; (iv) the calibration of the approach using an incomplete QIS; (v) the fact that scrapping AMA could result in a loss of risk sensitivity; (vi) the possibility of SMA being very volatile, significantly increasing capital (since the new proposal is over-conservative and over-sensitive to BI size and large losses); (vii) multiple variations of capital tranches which may lead to less understandable and comparable capital requirements; (viii) inadequate incentives for risk management without BEICFs, among other criticisms.

The Journal of Operational Risk has been publishing outstanding technical papers regarding operational risk measurement from academics and practitioners, and two technical papers, which analyze the impact of SMA, have recently been published in the Journal.

Mignola et al. (2016) and Peters et al. (2016) ran several simulations to analyze the variability level of SMA capital under common distributions typically used in operational risk and consistent business indicators. The authors found that (i) a high variability of SMA capital (mainly explained by the variability of the BI Component) compared to current internal models,

[36] See The Operational Riskdata eXchange Association, Capital impact of the SMA, ORX benchmark of the proposed Standardized Measurement Approach 5 (May 18, 2016), available at https://www.orx.org/Lists/PublicDocuments/ORX%20Capital%20impact%20of%20the%20SMA.pdf

[37] Based on response documents from Institute and Faculty of Actuaries (IFoA), European Banking Federation (EBF), Canadian Bankers Association (CBA), British Bankers' Association (BBA), the International Swaps and Derivatives Association (ISDA), the International Association of Credit Portfolio Managers (IACPM), the Japan Financial Markets Council (JFMC), International Banking Federation (IBFed), the Institute of International Finance (IIF) and the Global Financial Markets Association (GFMA)

and (ii) that banks would be over-insured against operational loss under the new proposal when compared to AMA models. The simulations were performed using a compound Poisson process, where the severity follows a lognormal distribution, which is not precisely a *very* heavy-tailed distribution. It would be interesting to run similar simulations by employing distributions with a heavier tail than the lognormal distribution (Pareto, for instance) in order to model the "infinite-mean" characteristic of operational losses, and over-dispersed frequency distributions. However, the authors have found overwhelming results that discard two of main objectives of the SMA proposal: comparability and risk-sensitivity. Moreover, Peters et al. (2016) propose a way to standardize AMA (which consists mainly of a hybrid LDA model with factor regression components) instead of calling it off.

The main drawbacks of SMA and AMA models are described as follows. As is the case for AMA, SMA too is considered a backward-looking measure due to the fact that historic losses are employed to predict future losses. However, SMA, in contrast to AMA, is highly volatile, it is not related to a risk-based capital measure, and it does not consider insurance hedging. Since BI is related to a bank's income, this may provide financial institutions with incentives to report less income in order to shorten provisions for potential losses. There are challenges associated to modelling operational risk losses through LDA-AMA; for instance, the important fluctuations in regulatory capital for institutions with comparable operational risk exposures. Nevertheless, the literature recognizes that extreme loss is a stylized fact of operational risk, which can be better modelled using very well-known actuarial techniques. Moreover, AMA usually provides lower capital requirement levels than BIA and TSA creating incentives for a bank to switch to AMA models, even if they are more complex to use when calculating capital. In sum, both SMA and AMA present disadvantages given the trade-off between simplicity and sensitivity. The SMA is a simple formula, but it is not risk sensitive. However, the main source of these shortcomings is a lack of understanding of the operational risk drivers that did not allow the Committee to focus on a way to foster the advantages of AMA.

4.5 Discussion

Recent papers published in the Journal of Operational Risk show that SMA may lead to abnormal levels of capital and the main objectives of the regulator will not be satisfied, but LDA-AMA exhibits well-known shortcomings. This section introduces some ideas from the literature review to propose an intermediate solution in the future or a transition period.

The initial step in adequate operational risk management consist of an adequate control procedure for internal processes, people, systems and external events, which are the sources of operational loss. Some financial institutions have been (mega-)fined for their improper practices; however, such events continue to occur with increasing frequency, meaning that financial institutions' internal controls are failing and, as such, creating perverse incentives that can lead to conduct risk and, in turn, monetary loss. The literature has proposed that gaps in governance and risk culture need to be filled in order to prevent operational losses, see for instance Andersen et al. (2012, Section 5) for other mitigation strategies.

More research is needed on covariates or determinants of operational losses in order to understand its nature and determine a "reliable" business indicator. The literature identifies important variables such as firm-specific characteristics, board composition, the state of the economy, accounting and public disclosure standards, the constraints on the decision-making power of executives, the power of regulators, levels of corruption, and the quality of governance (Chernobai et al., 2011; Moosa, 2011; Cope et al., 2012; Wang and Hsu, 2013; Barakat, et al., 2014; Moosa, 2015; Li and Moosa, 2015). A sound understanding of the operational risk factors will allow banks to quantify risk and capital buffer more adequately in the future.

A solution proposed by a member of the financial industry involves the weighted sum of the internal model of a bank and SMA, where the weight assigned to the internal model is zero if the bank has a poor model. This could be useful for a transition period when a better solution is proposed.

Under frequency and severity distributions, an operational risk model should be (i) realistic, (ii) well-specified, (iii) flexible, and (iv) simple (Dutta and Perry, 2007). Thus, an intermediate solution is the OpCaR proposed by BCBS (2014) and based on works by Böcker and Klüppelberg (2005), and Böcker and Spritulla (2006). The OpCaR is a closed-form solution (aka the single-loss approximation) to quantify VaR by employing the quantile of the fitted severity distribution and the expected value of the assumed frequency process. This result is based on standard LDA model and subexponential severity distributions[38] under weak regularity conditions (see Theorem 1.3.9 of Embrechts et al., 1997 for more details). Though SMA has been calibrated with the OpCaR model, the proposal is to use the single-loss

[38] Examples of subexponential distributions are Pareto, log-gamma, g-and-h, Burr, modified Champernowne, lognormal, Weibull and Benktander type I and type II.

approximation with the internal loss data for each financial institution with an appropriate risk measure and an adequate model averaging as explained below.

Regarding risk measure quantification, median shortfall may be a good proposal, although it may not be coherent for all loss distributions. Expected shortfall is a coherent risk measure, which is very sensitive to extreme events and in infinite-mean models, it is not possible to quantify it. Median shortfall, on the other hand, is greater than VaR, but less than expected shortfall at the same confidence level (e.g. 99.9%), when the distribution of operational losses is right-skewed. Thus, it may be possible for the objective of non-significant increases in regulatory capital to be achieved. Moreover, median shortfall is a suitable and reliable risk measure for operational losses (Moscadelli, 2004). Nevertheless, if expected shortfall is the selected risk measure, the recent methodology proposed by Cirillo and Taleb (2016) may be useful to deal with infinite-mean models by employing what the authors call the dual distribution.

Concerning forward-looking quantification, since historical data is commonly employed to predict risk, Bayesian networks allow the combination of historical data (backward-looking) with expectations and opinions (forward-looking) to obtain posterior predictions of operational risk events, and to consider the correlation between losses in the different processes pertaining to financial institutions (Cornalba and Giudici, 2004; Aquaro, 2010; Sanford and Moosa, 2015; Yan and Wood, 2017). An excellent guide which illustrates how to apply Bayesian networks for operational risk is provided by Cowell et al. (2007).

Our proposed solution for forward-looking and instability capital aspects is a Bayesian Model Averaging (BMA) methodology. All models are subject to model uncertainty and it is usual to select one model from a variety of models by employing frequentist tests or Bayesian model selection (Peters and Sisson, 2006). BMA is the most effective way to deal with this problem and it has a better predictive ability than using a single model (Hoeting et al., 1999). The OpCaR methodology proposes that we should average VaR from different models, but Bayesian averaging is the most adequate manner to average plausible models, where the weight assigned to each forecast is given by the posterior probability of each model. The initial single-loss approximation model proposed by Böcker and Klüppelberg (2005) has been refined and can be tested (see Peters et al., 2013, and references therein). In what follows, we focus on two of the abovementioned ideas. One is median shortfall and the second concerns BMA.

5.1. Risk quantification

We calculate VaR, expected shortfall (ES) and median shortfall (MS) of operational risk data.

$$\mathrm{VaR}_\alpha = q_\alpha(F) = \inf\{l \in \mathbb{R}: F_L(l) \geq \alpha\},$$

$$\mathrm{ES}_\alpha = \mathrm{E}[L|L \geq \mathrm{VaR}_\alpha],$$

$$\mathrm{MS}_\alpha = \inf\{l \in \mathbb{R}: F_{\alpha,L}(l) \geq 1/2\},$$

where $q_\alpha(F)$ is the α-quantile of loss distribution F_L, and $F_{\alpha,L}$ denotes the α-tail distribution (with $\alpha = 0.999$ for operational risk). Our loss distributions are obtained by performing 100000 simulations given the parameters of frequency and severity distributions provided in Hess (2011), which is based on SAS OpRisk Global Data. The operational loss is given by:

$$L = \sum_{i=1}^{N} X_i,$$

where N follows a discrete distribution (frequency distribution), and X_i are positive independent and identically distributed random variables, which follow a continuous distribution (severity distribution). The author employs a Poisson distribution (with parameter λ) for frequency distribution, and evaluates lognormal (with parameters μ and σ) and gamma (with parameters k and θ) distributions for severity distribution in different business lines: Commercial Banking (CB), Corporate Finance (CF), Retail Banking (RB), Insurance (I), Trading and Sales (TS), Asset Management (AM), and Retail Brokerage (RB). The risk measures are calculated from the aggregate loss distributions: Poisson-Lognormal and Poisson-Gamma.

Table 27. VaR and shortfall risk measures for Poisson-Lognormal and Poisson-Gamma models

		CB	CF	RB	I	TS	AM	RB
Poi	λ	0.1233	0.1622	0.1525	0.1536	0.1495	0.1114	0.1476
LN	μ	1.6	2.73	0.12	1.52	1.37	1.85	1.09
	σ	2.37	2.25	2.50	2.26	2.97	2.50	2.05
	VaR	1529.65	4630.66	534.71	1343.08	6627.56	2506.37	437.93
	ES	4497.45	11348.57	1490.18	3421.67	24237.44	9189.78	1060.84
		(2.94)	(2.45)	(2.79)	(2.55)	(3.66)	(3.67)	(2.42)

	MS	2392.3	8626.47	966.86	2207.47	14243.31	4962.29	730.69
		(1.56)	(1.86)	(1.81)	(1.64)	(2.15)	(1.98)	(1.67)
Ga	k	0.44	0.44	0.53	0.55	0.47	0.45	0.57
	ϑ	379.1	573.5	135.9	131.9	566.3	488.5	99.5
	VaR	1805.45	2810.72	560.89	529.11	2526.82	2242	391.56
	ES	2068.73	3834.15	672.88	628.32	3080.10	2495.65	483.68
		(1.14)	(1.36)	(1.20)	(1.19)	(1.22)	(1.11)	(1.24)
	MS	1836.73	3881.17	601.33	567.26	3003.79	2276.60	496.77
		(1.02)	(1.38)	(1.07)	(1.07)	(1.19)	(1.01)	(1.27)

Ratio of shortfall measures (ES and MS) to 99.9%-VaR in parentheses. Poi stands for Poisson distribution, LN indicates lognormal distribution and Ga represents gamma distribution. VaR is assessed as the quantile of the aggregate loss at 99.9% confidence level, Expected Shortfall (ES) is calculated as the mean of losses given that the losses have exceeded VaR, whereas Median Shortfall (MS) is computed as the median of loss exceedances.

As seen in Table 27, most of the cases' expected shortfall is greater than median shortfall, since the tail of operational risk losses is generally right-skewed. The average ratio of ES to VaR is 2.95 (1.21) for lognormal (gamma) severity case, whereas the ratio MS to VaR is 1.81 (1.14). Other values can be obtained depending on the severity model employed, but MS could achieve the objective of non-significant increase with respect to the current regulation.

5.2. BMA

Though the use of BMA in finance is scarce, there are some interesting applications in exchange forecasting (Wright, 2008; Feldkircher et al., 2014), portfolio selection (Ando, 2009), and interest rates (Maltritz and Molchanov, 2013; Chua et al., 2013).

Let us assume that there are K models $M_1, \dots, M_K$; for instance: Poisson-Lognormal, and Poisson-Pareto, among others. Let θ_k be the vector of parameters for each model M_k; for example, λ, μ and σ for the Poisson-Lognormal model. The modeler has a prior belief that the data generating process is given by M_k, denoted by $P(M_k)$. This prior distribution is updated given observed (operational risk loss) data D, to obtain the posterior probability for model M_k:

$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_{l=1}^{K} P(D|M_l)P(M_l)},$$

where

$$P(D|M_k) = \int P(D|\theta_k, M_k)\, P(\theta_k|M_k) d\theta_k$$

is the marginal likelihood for model M_k, $P(\theta_k|M_k)$ is the prior density of θ_k under model M_k, and $P(D|\theta_k, M_k)$ is the likelihood. The integrals are not easily found analytically, but Bayesian computation methods can solve this problem. The model priors can be non-informative or real prior opinions driven by expert opinions or extracted from self-assessment questionnaires, as in the case of Fignini et al. (2015). The proposed quantity of interest Δ, is a risk measure that can be VaR, ES or MS. By knowing the posterior model probability $P(M_k|D)$, the risk measure is obtained as the posterior mean as follows:

$$\mathrm{E}(\Delta|D) = \sum_{k=1}^{K} \hat{\Delta}_k\, P(M_k|D),$$

where $\widehat{\Delta}_k = \mathrm{E}(\Delta|D, M_k)$, and the posterior model probability $P(M_k|D)$, can be seen as weights. Other model averaging methods can be employed such as the frequentist approach (weights depend on Akaike information criterion) and the predictive likelihood approach (see for instance Ando and Tsay, 2010; and the references therein). In particular, for the single-loss approximation, the quantity of interest Δ, is given by:

$$\mathrm{VaR}_\alpha = F^{-1}\left(1 - \frac{1-\alpha}{\mathrm{E}(N)}\right),$$

where F is the cumulative distribution function of the severity distribution for each model, $\mathrm{E}(N)$ is the expected value of the frequency distribution (λ, if the Poisson distribution is employed), and $\alpha \to 1$ is the confidence level. For several severity distributions, ES and MS can be analytically derived, and BMA would then be employed for these risk measures and to compare the results. The application of this proposal will be the focus for future study.

The ideas presented in this paper refer to strategies we consider convenient for a better understanding of operational risk and for finding the "almost" right model to quantify it. The debate surrounding SMA vs AMA models is just beginning and we propose a single-loss approximation model with the appropriate risk measure under a BMA setting as an alternative solution to estimate operational risk capital.

REFERENCES

Andersen, L.B., Häger, D., Maberg, S., Naess, M.B., Tungland, M. 2012. The financial crisis in an operational risk management context—A review of causes and influencing factors. Reliability Engineering & System Safety 105, 3-12. DOI: 10.1016/j.ress.2011.09.005

Ando, T. 2009. Bayesian portfolio selection using multifactor model and Bayesian predictive information criterion. International Journal of Forecasting 25, 550–566. DOI: http://dx.doi.org/10.1016/j.ijforecast.2009.01.005

Ando, T., Tsay, R. 2010. Predictive likelihood for Bayesian model selection and averaging. International Journal of Forecasting 26, 744–763. DOI: 10.1016/j.ijforecast.2009.08.001

Aquaro, V., Bardoscia, M., Belloti, R., Consiglio, A., De Carlo, F., Ferri, G. A Bayesian Networks approach to Operational Risk, Physica A 389, 1721-1728. DOI: 10.1016/j.physa.2009.12.043

Barakat, A., Chernobai, A., Wahrenburg, M. (2014). Information asymmetry around operational risk announcements. Journal of Banking & Finance 48, 152-179. DOI: 10.1016/j.jbankfin.2014.06.029

BCBS (2006). Basel Committee on Banking Supervision, International Convergence of Capital Measurement and Capital Standards: A Revised Framework Comprehensive Version. Available at http://www.bis.org/publ/bcbs128.pdf

BCBS (2014). Basel Committee on Banking Supervision, Operational risk – Revision to the simpler approaches. Available at http://www.bis.org/publ/bcbs291.pdf.

BCBS (2016). Basel Committee on Banking Supervision, Standardised Measurement Approach for operational risk. Available at http://www.bis.org/bcbs/publ/d355.htm.

Böcker, K., Klüppelberg, C. (2005) Operational VaR: a closed-form solution. RISK Magazine, December, 90-93.

Böcker, K., Spritulla, J. (2006). Operational VaR: meaningful means. RISK Magazine, December, 96-98.

Brechman, E., Czado, C., Paterlini, S. (2014). Flexible dependence modeling of operational risk losses and its impact on total capital requirements. Journal of Banking & Finance 40. 271-285. DOI: 10.1016/j.jbankfin.2013.11.040

Buch-Kroman, T. (2009). Comparison of tail performance of the Champernowne transformed kernel density estimator, the generalized Pareto distribution and the g-and-h distribution. The Journal of Operational Risk 4 (2), 43-67. DOI: 10.21314/JOP.2009.058

Buch-Kroman, T., Englund, M., Gustaffson, J., Nielsen, J.P., Thurin, F. (2007). Non-parametric estimation of operational risk losses adjusted for under-reporting. Scandinavian Actuarial Journal 4, 293-304. DOI: 10.1080/03461230701642471

Bulancé, C., Guillén, M., Gustaffson, J., Nielsen, J.P. (2012). Quantitative Operational Risk Models. Chapman and Hall/CRC.

Carrillo-Menéndez, S., Suárez, A. (2012). Robust quantification of the exposure to operational risk: Bringing economic sense to economic capital. Computers & Operations Research 39 (4), 792-804. DOI: http://dx.doi.org/10.1016/j.cor.2010.10.001

Cavestany, R., Boultwood, B., Escudero, L. (2015). Operational Risk Capital Models. Risk Books.

Chavez-Demoulin, V., Embrechts, P., Hofert, M. (2013). An extreme value approach for modeling operational risk losses depending on covariates. Journal of Risk and Insurance 83 (3), 735-776. DOI: 10.1111/jori.12059

Chavez-Demoulin, V., Embrechts, P., Nešlehová, J. (2006). Quantitative models for operational risk: extremes, dependence and aggregation. Journal of Banking & Finance 30(10), 2635–2658. DOI, 10.1016/j.jbankfin.2005.11.008

Chernobai, A., Jorion, P., Yu, F. (2011). The determinants of operational risk in U.S. financial institutions. Journal of Financial Quantitative Analysis 46, 1683–1725. DOI: 10.1017/S0022109011000500

Chua, C.L., Suardi, S., Tsiaplias, S. 2013. Predicting short-term interest rates using Bayesian model averaging: Evidence from weekly and high frequency data. International Journal of Forecasting 29, 442–455. DOI: 10.1016/j.ijforecast.2012.10.003

Cirillo, P., Taleb, N.N. (2016). Expected shortfall estimation for apparently infinite-mean models of operational risk. Quantitative Finance. DOI: 10.1080/14697688.2016.1162908

Cope, E.W., Piche, M.T., Walter, J.S. (2012). Macroenvironmental determinants of operational loss severity. Journal of Banking & Finance 36, 1362-1380. DOI: 10.1016/j.jbankfin.2011.11.022

Cornalba, C., Giudici, P. (2004). Statistical models for operational risk management. Physica A 338, 166-172. DOI: 10.1016/j.physa.2004.02.039

Cowell, R.G., Verrall, R.J., Yoon, Y.K. (2007). Modeling operational risk with Bayesian networks. The Journal of Risk and Insurance 74 (4), 795-827. DOI: 10.1111/j.1539-6975.2007.00235.x

Cruz, M., Peters, G., Shevchenko, P. (2015). Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk. Wiley: New Jersey.

Dahen, H., Dionne, G. (2010). Scaling models for the severity and frequency of external operational loss data. Journal of Banking & Finance 34, 1484-1496. DOI: http://dx.doi.org/10.1016/j.jbankfin.2009.08.017

de Fontnouvelle, P., DeJesus-Rueff, V., Jordan J., Rosengren, E. (2003). Using loss data to quantify operational risk. Availabe in www.bis.org/bcbs/events/wkshop0303/p04deforose.pdf.

de Fontnouvelle, P., Rosengren E., & Jordan J. (2004). Implications of alternative operational risk modelling techniques. Unpublished paper, Federal Reserve Bank of Boston.

Degen, M., Embrechts, P. (2008). EVT-based estimation of risk capital and convergence of high quantiles. Advanced in Applied Probability 40 (3): 696-715. DOI: http://dx.doi.org/10.1017/S0001867800002755

Degen, M., Embrechts, P., Lambrigger, D. (2007). The quantitative modeling of operational risk: Between g-h and EVT. ASTIN Bulletin, 37, 265-291. DOI: http://dx.doi.org/10.1017/S0001867800002755

Diebold, F. X., Schuermann, T., Stroughair, J. D. (1998). Pitfalls and opportunities in the use of extreme value theory in risk management. In A.P.N. Refenes, J.D. Moody, and A.N. Burgess (Eds.), Advances in computational finance (pp. 3-12). Amsterdam: Kluwer Academic Publishers.

Dutta, K.K, Babbel, D.F. (2013). Scenario analysis in the measurement of operational risk capital: A change of measure approach. Journal of Risk and Insurance 81 (2) 303-334. DOI: 10.1111/j.1539-6975.2012.01506.x

Dutta, K., Perry, J. (2007). A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital (Working Paper 06-13). Federal Reserve Bank of Boston.

Embrechts, P. (2009). Linear correlation and EVT: Properties and caveats. Journal of Financial Econometrics, 7(1), 30-39. DOI: 10.1093/jjfinec/nbn015

Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). Modelling extremal events for insurance and finance. Berlin: Springer.

Feldkircher, M., Horvath, R., Rusnak, M. 2014. Exchange market pressures during the financial crisis: A Bayesian model averaging evidence. Journal of International Money and Finance 40, 21–41. DOI: http://dx.doi.org/10.1016/j.jimonfin.2013.08.021

Feria-Domínguez, J.M., Jiménez-Rodríguez, E., Sholarin, O. (2015). Tackling the over-dispersion of operational risk: Implications on capital adequacy requirements. The North American Journal of Economics and Finance 31, 206-221. DOI: 10.1016/j.najef.2014.11.004

Figini, S., Gao, L. and Giudici, P. 2015. Bayesian operational risk models. The Journal of Operational Risk 10 (2), 45-60. DOI: 10.21314/JOP.2015.155

Frachot, A., Georges, P., & Roncali, T. (2001). Loss distribution approach for operational risk (Working Paper). Groupe de Recherche Opérationnelle, Crédit Lyonnais, France.

Girling, P. (2013). Operational risk management: a complete guide to a successful operational risk framework. New Jersey: Wiley.

Han, J., Wang, W., Wang, J. (2015). POT Model for Operational Risk: Experience with the Analysis of the Data Collected from Chinese Commercial Banks. China Economic Review 36, 325-340. DOI: 10.1016/j.chieco.2015.07.003

Haubenstock, M., & Hardin, L. (2003). The loss distribution approach. In C. Alexander (Ed.), Operational risk: Regulation, analysis and management (pp. 229-240). Upper Saddle River, NJ: Prentice Hall.

Hess, C. 2011. Can the single-loss approximation method compete with the standard Monte Carlo simulation technique? The Journal of Operational Risk 6 (2), 31-43. DOI: 10.21314/JOP.2011.091

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T. (1999). Bayesian Model Averaging: A tutorial. Statistical Sciences 14 (4), 382-417. DOI: 10.1214/ss/1009212519

Huber, S. (2008). (Non-)robustness of maximum likelihood estimators for operational risk severity distributions. Quantitative Finance 10 (8), 871-882. DOI: 10.1080/14697680903159240

Jobst, A. (2007). Operational Risk-The sting is still in the tail but the poison depends on the dose. The Journal of Operational Risk 2 (3), 3-59. DOI: https://dx.doi.org/10.5089/9781451868036.001

Li, L. Moosa, I. (2015). Operational risk, the legal system and governance indicators: a country-level analysis. Applied Economics 47 (20), 2053-2072. DOI: 10.1080/00036846.2014.1000533

Madigan, P. (2009). An operational model risk. RISK Magazine November, 77-79.

Maltritz, D., Molchanov, A. 2013. Analyzing determinants of bond yield spreads with Bayesian Model Averaging. Journal of Banking & Finance 37, 5275–5284. DOI: 10.1016/j.jbankfin.2013.07.007

McNeil, A.J., Frey, R., Embrechts, P. (2005). Quantitative risk management: Concepts, techniques and tools. Princenton: Princeton University Press.

Mignola, G., Ugoccioni, R., Cope, E. (2016). Comments on the BCBS proposal for a new standardized approach for operational risk. The Journal of Operational Risk 11 (3) 51-69.

Moosa, I. (2007). Operational risk: A survey. Financial Markets, Institutions & Instruments 16(4), 167-200. DOI: 10.1111/j.1468-0416.2007.00123.x

Moosa, I. (2008). A critique of the advanced measurement approach to regulatory capital against operational risk. Journal of Banking Regulation 9(3), 151-164. DOI: 10.1057/jbr.2008.7

Moosa, I. (2008). Operational risk as a function of the state of the economy. Economic Modelling 28, 2137-2142. DOI: http://dx.doi.org/10.1016/j.econmod.2011.05.011

Moosa, I. (2015). Governance indicators as determinants of operational risk. International Journal of Disclosure and Governance 12 (2), 132-143. DOI : 10.1057/jdg.2013.36

Moscadelli, M. (2004). The modelling of operational risk: Experience with the analyisis of the data collected by the Basel Committee. Banca d’Italia, Temi di discussione del Servizio Studi, 517, July.

Nešlehová, J., Chavez-Demoulin, V., & Embrechts, P., (2006). Infinite-mean models and the LDA for operational risk. Journal of Operational Risk, 1(1), 3-25. DOI: 10.21314/JOP.2006.001

Peters, G.W., Sisson, S.A. (2006). Bayesian inference, Monte Carlo sampling and operational risk. Journal of Operational Risk 1 (3) 27-50. DOI: 10.21314/JOP.2006.014

Peters, G.W., Shevchenko, P.V., Hassani, B., Chapelle, A. (2016). Should AMA be replaced with SMA for operational risk? The Journal of Operational Risk 11 (3), 1-49.

Peters, G.W., Targino, R.S., Shevchenko, P.V. (2013). Understanding Operational Risk Capital Approximations: First and Second Orders. Journal of Governance and Regulation 2 (3), 58-78. DOI: 10.22495/jgr_v2_i3_p6

Sanford, A., Moosa, I. (2015). Operational risk modelling and organizational learning in structured finance operations: A Bayesian network approach. The Journal of the Operational Research Society 66 (1), 86-115. DOI: 10.1057/jors.2013.49

Shevchenko, P.V. (2010). Implementing loss distribution approach for operational risk. Applied Stochastic Models in Business and Industry 26, 277-307. DOI: 10.1002/asmb.812

Tursunalieva, A., Silvapulle, P. (2016). Nonparametric estimation of operational value-at-risk (OpVaR). Insurance: Mathematics and Economics 69, 194-201. DOI: 10.1016/j.insmatheco.2016.05.010

Wang, T., Hsu, C. (2013). Board composition and operational risk events of financial institutions. Journal of Banking & Finance 37, 2042-2051. DOI: http://dx.doi.org/10.1016/j.jbankfin.2013.01.027

Wright, J.H. 2008. Bayesian model averaging and exchange rate forecasts. Journal of Econometrics 146, 329–341. DOI: http://dx.doi.org/10.1016/j.jeconom.2008.08.012

Yan, H., Wood, R. (2017). A structural model for estimating losses associated with the mis-selling of retail banking products. Journal of Operational Risk forthcoming.

Zhou, X., Durfee, A.V., Fabozzi, F.J. (2016). On stability of operational risk estimates by LDA: From causes to approaches. Journal of Banking & Finance 68, 266-278. DOI: 10.1016/j.jbankfin.2016.01.014

5. Conclusions of the Thesis

This section wraps the results up of this thesis. The aim of this thesis is to analyze performance of VaR and ES measures for market risk purposes and review the advantages and drawbacks of the new proposal for operational risk. The first chapter "Risk quantification in turmoil markets" analyzes several assets performance such as foreign currency (CHF/USD), Eurostoxx50, VIX index and a Commodity Index during global financial crisis under different distributional models for Value-at-Risk (VaR) quantification purposes. The results of this chapter show that the generalized Pareto distribution (GPD), α-stable and g-and-h distributions perform well for the analyzed data in our sample period according to VaR backtesting procedure. The second chapter "Quantifying Risk in Traditional Energy and Sustainable Investments" examines the risk performance for two diversified stock indexes, one of fossil fuel-related companies and another that includes sustainable companies from all sectors except these traditional energy companies. For Expected Shortfall (ES), it is employed the multinomial test and we conclude that this test is an appropriate and simple method to validate ES models as presented in this thesis. The third chapter "Risk quantification for Commodity ETFs: Backtesting Value-at-Risk and Expected Shortfall" analyzes the risk assessment of alternative methods for some univariate and portfolios of Commodity ETFs. Backtesting results for 97.5%-VaR shows that skewed-t performs satisfactorily for Commodity ETFs. According to ES validation tests, Gram-Charlier and Student-t are the best models for 97.5% VaR and 97.5% ES. For portfolios of Commodity ETFs, the SNP-DCC is preferred to the Gaussian-DCC according to t-test and relative performance tests. The last chapter "A note on SMA vs. LDA-AMA: The dawning of a new regulation" reviews the current and the proposed rules to quantify operational risk and proposes an adequate manner to "average" different amounts of OpRisk and proposes the median shortfall as an alternative to quantify operational risk.

Future research will be focused on analysis of non-modellable risk factors (NMRF) and implementation of the profit and loss attribution (PLA) test, which measures the accuracy of risk model estimates of profit and losses (P&L). In addition, other methods to test ES will be objective of study for market risk, as other methods to estimate operational risk capital in an appropriate way.

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea