SAMPLE SIZE, SKEWNESS AND LEVERAGE EFFECTS IN VALUE AT RISK AND EXPECTED SHORTFALL ESTIMATION

Laura García Jorcano

Número 26/2019
Cuadernos de Investigación SANFI

Editorial Universidad Cantabria

Laura García Jorcano. She holds a degree in Business Administration and Management, specializing in Finance, and a degree in Law from the Universidad de Zaragoza, and a M.Sc. in Banking and Quantitative Finance and a European Ph.D. in Banking and Quantitative Finance from the Universidad Complutense de Madrid. She is currently an assistant professor and researcher in the Financial Economics Area in the Department of Economic Analysis and Finance, Facultad de Ciencias Jurídicas y Sociales —Toledo—, in the Universidad de Castilla-La Mancha. Her fields of study are financial econometrics, quantitative finance, and financial economics. She has done several research works on financial risk measurement and she is currently working on the calculation of the probability unbiased Value at Risk under mixtures of normal distributions, on the forecasting of the Expected Shortfall with nonparametric methods based on expectiles, and in the measurement and forecasting of systemic risk.

SAMPLE SIZE, SKEWNESS AND LEVERAGE EFFECTS IN VALUE AT RISK AND EXPECTED SHORTFALL ESTIMATION

Colección Difunde #247

Cuadernos de Investigación SANFI Número 26/2019

SAMPLE SIZE, SKEWNESS AND LEVERAGE EFFECTS IN VALUE AT RISK AND EXPECTED SHORTFALL ESTIMATION

Laura García Jorcano

CANTABRIA CAMPUS INTERNACIONAL

Editorial Universidad Cantabria

García Jorcano, Laura, autor

Sample size, skewness and leverage effects in value at risk and expected shortfall estimation / Laura García Jorcano. – Santander : Editorial de la Universidad de Cantabria, [D.L. 2019]

158 páginas ; 24 cm. – (Difunde ; 247) (Cuadernos de investigación SANFI ; 26/2019)

D.L. SA. 806-2019. – ISBN 978-84-8102-912-3

1. Mercado financiero. 2. Gestión del riesgo-Modelos matemáticos. I. Fundación de la Universidad de Cantabria para el Estudio y la Investigación del Sector Financiero, entidad patrocinadora.

330.322

THEMA: KFFX

Promueve: Fundación de la Universidad de Cantabria para el Estudio y la Investigación del Sector Financiero (UCEIF)

Coordinadora: Mª Begoña Torre Olmo
Secretaria: Tamara Cantero Sánchez

Digitalización: Manuel Ángel Ortiz Velasco [emeaov]

ISBN: 978-84-8102-912-3

Depósito Legal: SA 806-2019

DOI: https://doi.org/10.22429/Euc2020.047

Imprenta Kadmos

Impreso en España. *Printed in Spain*

Santander Financial Institute (www.sanfi.es)

SANFI es el centro de referencia internacional en la generación, difusión y transferencia del conocimiento sobre el sector financiero, promovido por la UC y el Banco Santander a través de la Fundación UCEIF. Desde sus inicios, dirige actividades en áreas de formación, investigación y transferencia:

Máster en Banca y Mercados Financieros UC-Banco Santander

Constituye el eje nuclear de una formación altamente especializada, organizada desde la fundación en colaboración con el Banco Santander. Es Impartido en España, México, Marruecos, Chile y Brasil, dónde se están desarrollando la 24ª Edición, 21ª Edición, 13ª Edición y 3ª Edición respectivamente.

Formación In Company

SANFI potencia sus actividades para desarrollar la formación de profesionales del sector financiero, principalmente del propio Santander, destacando también su actuación dentro de otros programas, como el realizado con el Attijariwafa Bank.

Archivo Histórico del Banco Santander

Situado en la CPD del Santander en Solares, comprende la clasificación, catalogación, administración y custodia, así como la investigación y difusión de los propios fondos de Banco Santander como de otras entidades. Cabe destacar que posee más de 27.000 registros de fondo.

Educación Financiera

Finanzas para Mortales (www.finanzasparamortales.es). Elegido como el Mejor Proyecto de Educación Financiera 2018 por el Banco de España y la CNMV, está dirigido a fomentar la cultura financiera de la sociedad, a través de su plataforma online y sus sesiones presenciales gratuitas. Cuenta con más de 1.200 voluntarios procedentes en su mayoría de Banco Santander, distribuidos por los diferentes puntos de la geografía española. Desde 2015, han impartido más de 4.250 sesiones formativas, logrando acercar los conocimientos financieros a más de 10.000 beneficiarios de diversas instituciones entre las que destacan, colegios e institutos, Cruz Roja, Fundación del Secretariado Gitano, la ONCE, Fundación...

Investigación

- Atracción del Talento Internacional con programas de Becas y Ayudas para fomentar la realización de proyectos de investigación, especialmente de Jóvenes Investigadores, que fomenten el conocimiento de las metodologías y técnicas aplicables en el ejercicio de la actividad financiera, con especial interés en las realizadas por entidades bancarias, para favorecer al crecimiento y desarrollo económico de los países y al bienestar social.

- Premios Tesis Doctorales, cuyo fin es promover y reconocer la generación de conocimientos a través de actuaciones en el ámbito del doctorado que impulsen el estudio y la investigación en el sector financiero.

- Por último, la línea editorial en la que se enmarcan estos Cuadernos de Investigación, con el objetivo de poner a disposición de la sociedad, el conocimiento generado en torno al sector financiero fruto de todas las acciones desarrolladas en el ámbito del SANFI y especialmente, de los resultados de las Becas, Ayudas y Premios Tesis Doctorales.

ÍNDICE

II Summary

17 Chapter 1. Probability-unbiased VaR estimator

52 Chapter 2. Volatility specifications versus probability distributions in VaR estimation

111 Chapter 3. Testing ES estimation models: An extreme value theory approach

146 Bibliography

SUMMARY

The estimation of risk measures is an area of highest importance in the financial industry. Risk measures play a major role in the risk-management and in the computation of regulatory capital. The Basel III document has suggested to shift from Value at Risk (VaR) into Expected Shortfall (ES) as a risk measure and to consider stressed scenarios at a new confidence level of 97.5%. This change is motivated by the appealing theoretical properties of ES as a measure of risk and the poor properties of VaR. In particular, VaR fails to control for "tail risk". In this transition, the major challenge faced by financial institutions is the unavailability of simple tools for evaluation of ES forecasts (i.e. backtesting ES).

The objective of this thesis is to compare the performance of a variety of models for VaR and ES estimation for a collection of assets of different nature: stock indexes, individual stocks, bonds, exchange rates, and commodities. Throughout the thesis, by a VaR or an ES "model" is meant a given specification for conditional volatility, combined with an assumption on the probability distribution of return innovations.
Specifically, Chapter 1 considers the concept of unbiasedness in VaR estimation. Francioni and Herzog (2012) (FH) showed that there exists a bias correction for VaR when returns are Normally distributed. In this chapter the FH analysis is extended to the Student-t distribution as well as to Mixtures of two Normal distributions, using a bootstrapping algorithm proposed by FH. The use of the probability-unbiased VaR avoids the systematic underestimation of risk implied by the bias of standard VaR measures. The magnitude of the distortion that needs to be exerted on the quantile to move from the standard VaR to the probability-unbiased VaR depends on the sample size and on the distribution assumption on returns. Since financial returns usually have thick tails, the smaller the sample size and the lower the heaviness of the tail of the assumed distribution in estimation, the higher will be the distortion to be applied to achieve unbiasedness. This VaR adjustment allows us to work with small samples knowing that the estimated VaR will generally display a good performance. Furthermore, the results in the thesis show that

using a small sample may easily lead to more accurate VaR estimates than longer samples according to the Exceedance Probability and to the Observed Absolute Deviation per year (mean of the absolute differences between the expected number of exceedances and the number of observed exceedances). The good performance of the probability-unbiased VaR follows from the fact that a short sample size allows for capturing better the structural changes that arise over time in financial returns due to trading behavior.

Chapter 2 analyzes how the efficiency of VaR depends on the volatility specification and the assumption on the probability distribution for return innovations. This question is crucial for risk managers, since there are so many potential choices for volatility model and probability distributions that it would be very convenient to establish some priorities in modeling returns for risk estimation. We consider different conditional VaR models for assets of different nature, using symmetric and asymmetric probability distributions for the innovations and volatility models with and without leverage. We calculate VaR estimates following the parametric approach. The ability to explain sample return moments might be considered a natural condition to obtain a good VaR performance. However, even though significant effort is usually placed in selecting an appropriate combination of probability distribution and volatility specification in VaR estimation, the ability to explain sample return moments is seldom examined. After using simulation methods to calculate implied return moments from estimated models, we compare the implied levels of skewness and kurtosis of returns with the analogue sample moments.

We show that the ability to explain sample moments is in fact linked to performance in VaR estimation. Such performance is examined through standard tests: the unconditional coverage test of Kupiec (1995), the independence and conditional coverage tests of Christoffersen (1998), the Dynamic Quantile test of Engle and Manganelli (2004), as well as the loss functions proposed by Lopez (1998, 1999) and Sarma et al. (2003) and that of Giacomini and Komunjer (2005).

Relative to an ever increasing literature, we contribute in different ways:

1. considering a set of probability distributions that have recently been suggested to be appropriate for capturing the skewness and kurtosis of financial data, but whose performance for VaR estimation has not been compared yet on a common data set,
2. considering the APARCH and FGARCH volatility models with leverage that have also been recognized as being adequate for financial returns,
3. applying existing backtesting procedures for the different VaR models to a wide array of assets of different nature,
4. comparing the relevance of the assumed probability distribution for return innovations and the volatility specification for VaR performance,
5. introducing a dominance criterion to establish a ranking of models on the basis of their behavior under standard VaR validation tests, and
6. using the dominance criterion and the Model Confidence Set approach to search for robust conclusions on the preference of some probability distributions and volatility specifications.

Two clear results refer to issues that have been analyzed in previous research by a number of authors:

1. VaR models that assume asymmetric probability distributions for return innovations, like the skewed Student-t distribution, skewed Generalized Error distribution, Johnson SU distribution, and skewed Generalized-t distribution achieve better VaR performance than models with symmetric distributions, and
2. volatility models with leverage, like APARCH and FGARCH, show a better VaR performance than more standard GARCH and GJR-GARCH volatility specifications. Our analysis highlights other important issues.

A third result is that the shape and the skew of the assumed probability distribution for innovations are even more important for the performance of a Value at Risk model than including a leverage effect in volatility. This corroborates results by previous authors. We provide a thorough analysis of this issue by showing that the result holds for the wide set of assets we have considered:

a) the frequency of rejections of VaR tests in models that differ in their volatility specification is similar, while rejection frequencies among models with the same volatility specification but a different probability distribution for the innovations can differ very significantly,
b) changing the probability distribution in a VaR model affects the p-value of the statistic for VaR tests by a larger amount than changing the volatility specification, and
c) the dominance criterion we have introduced in this chapter establishes a clear ranking between models differing in their probability distribution, while the distinction between models that differ in their volatility specification is much less clear.

A fourth result deals with the fact that our estimates suggest that for a number of financial assets the true, unobserved volatility dynamics should not be specified in terms of the squared conditional standard deviation. Hence, models specified for the conditional variance are prone to produce biased results. Dealing with the power of the conditional standard deviation as a free parameter is an important feature of the APARCH/FGARCH volatility specifications that explains their better performance in validation tests of VaR forecasts.

A final result emerges from the consideration of the different criteria used in the chapter to choose among models for VaR estimation: the combination of APARCH or FGARCH volatility with a skewed Generalized Error, skewed Generalized-t or Johnson SU distributions seem to be have the best VaR performance for a wide array of assets of different nature.

In Chapter 3 we estimate the conditional Expected Shortfall based on the Extreme Value Theory (EVT) approach using asymmetric probability distributions for return innovations, and we analyze the accuracy of our estimates at 1- and 10-day horizons, before and during the 2008 financial crisis, using daily data. We take into account volatility clustering and leverage effects in return volatility by using the APARCH model under different probability distributions assumed for the standardized innovations: Gaussian, Student-t, skewed Student-t, skewed generalized error and Johnson SU and under EVT methods, following the two-step procedure of McNeil & Frey (2000). This two-step procedure fits a Generalized Pareto Distribution to the extreme values of the standardized

residuals generated by APARCH models. This two-step procedure fits a Generalized Pareto Distribution to the extreme values of the standardized residuals generated by APARCH models. Then, we compare the one-step-ahead out-of-sample ES forecast performance of all these models for different significance levels (α). Previously existing backtesting for ES have been shown to have serious limitations [as McNeil & Frey (2000) test, Berkowitz (2001) test, Kerkhof and Melenberg (2004) test and Wong (2008) test]. Such limitations are overcome by some recent ES backtesting proposals that we use for ES evaluation: the Righi & Ceretta (2013) test, two tests by Acerbi & Szekely (2014) that are straightforward but require simulation analysis (like the Righi & Ceretta test), the test of Graham & Pál (2014), which is an extension of the Lugannani-Rice approach of Wong, the quantile-space unconditional coverage test of Costanzino & Curran (2015) for the family of Spectral Risk Measures, of which ES is a member and, finally, the conditional test of Du & Escanciano (2016).
This chapter contributes to the literature in different ways:

i) comparing the performance of the standard parametric approach with two alternatives to ES forecasting that take into account volatility clustering and asymmetric returns: EVT and the semi-parametric Filtered Historical Simulation,
ii) comparing the results obtained under asymmetric probability distributions for return innovations with results under Normal and Student-t distributions,
iii) using the APARCH volatility specification because of its greater flexibility to represent the dynamics of conditional volatility (Garcia-Jorcano and Novales, 2017),
iv) forecasting VaR and ES over a 10-day horizon as in Basel capital requirements and test ES forecasting models at this horizon, an analysis that has seldom been considered in the financial literature, and
v) analyzing the accuracy of risk models for ES forecasting during pre-crisis and crisis periods as well as under different significance levels. To the best of our knowledge, this is the first time that a systematic test of ES forecasting models is done considering a variety of probability distributions and two alternatives to the standard parametric approach, like EVT and the semi-parametric FHS.

We obtain the following conclusions:

i) in standard conditional models fitted to the full distribution of return innovations we observe that asymmetric distributions play an important role in capturing tail risk at 1-day and 10-day horizons. This is because the stylized facts of financial returns such as volatility clusters, heavy tails and asymmetry are suitably captured by these asymmetric distributions.

ii) applying EVT to return innovations by modeling the tail with a GPD we obtain good ES forecasts at 1- and 10-day horizons regardless of the probability distribution used for returns. So, it looks as if considering just the return innovations in the tail of the distribution is more important than discriminating among probability distributions when forecasting ES.

iii) using Filtered Historical Simulation can be very useful. First, qualitative results under FHS in favor of the use of EVT in VaR and ES estimation are consistent with those obtained under the parametric approach, which is reassuring. Second, ES forecasts are much more similar for different probability distributions, as well as between forecasts from EVT-based models and non-EVT based models. That implies a considerable reduction in model risk.

iv) even during the crisis period, conditional EVT models are more accurate and reliable for predicting asset risk losses than conditional models that do not incorporate the EVT approach. However, during the crisis there is a systematic undervaluation of risk in both classes of models, with a number of violations above the theoretical one, suggesting that the models do not fully adapt to the occurrence of tail events. In general, p-values obtained in all tests during the pre-crisis period are higher than those obtained in the crisis period, suggesting a higher questioning of the models for ES forecasting over the crisis period.

CHAPTER I. PROBABILITY-UNBIASED VAR ESTIMATOR

1.1. Introduction

Value at Risk (VaR) has long been the most popular risk measure for the risk management of financial asset/portfolios. The Basel requirements for risk evaluation require the use of additional measures like Expected Shortfall, whose estimation is conditional on a previous VaR estimate. Hence, the ability to have good VaR estimates remains a central issue in risk management. The parametric approach to VaR estimation proceeds in two steps. First, the unknown parameters in the assumed probability distribution for portfolio returns are estimated from sample data by statistical methods. In the second step, these estimates are treated as the true parameter values and they are taken to the mathematical expression for VaR in the specific model considered to compute the desired distribution percentile. This is called a plug-in quantile estimator. It is well-known that this procedure is not efficient because the highly non-linear mapping from model parameters to the risk-measure introduces some biases. Statistical experiments show that such bias leads to a systematic underestimation of risk.

The quality of VaR estimates is controlled by backtesting. The tests to be conducted are usually given by the regulator (Basel Committee on Banking Supervision (BCBS), 2016). Even under the current emphasis on Expected Shortfall (ES) as the main current risk measure sanctioned by BCBS, backtesting is required only on VaR. There are two reasons for that: 1) a good VaR estimate is a needed condition for a precise ES estimate, and 2) the unavailability of simple tools for evaluation of ES forecasts. To backtest VaR the so-called failure rate procedure is often considered. This procedure focus on the rate of exceptions, i.e. ratio of scenarios in which the estimated capital reserve turns out to be insufficient. More precisely, given a data sample of size n, we start by estimating VaR at level α%. Then, we count how many times the actual return over the testing period exceeded the VaR estimate. Under a good

estimator we should observe that the relative frequency of exceptions should be close to α%.

The backtesting procedure crucially depends on the availability of historical data. This is most problematic if we only have available a small sample of historical data on a given asset or portfolio. There are several reasons why we have small samples: (i) the risk of the asset/portfolio under consideration may be determined by an instrument that has been issued recently, (ii) conditions for the evolution of some market may have changed and statistical analysis of historical data from the period preceding this change can not be expected to give a correct information about the probabilities of future changes, and (iii) the access to historical data for some instrument may be limited.

The question is: how can we estimate a quantile of an unknown probability distribution, if all we have is a small sample from this distribution?

The main goal of this chapter is to define a probability-unbiased VaR estimator in such a way that it behaves well under various backtesting procedures. We follow Francioni and Herzog, "Probability-unbiased Value-at-Risk estimators" (2012) in using probability unbiasedness as the criterion to search for good VaR estimates. VaR performance for each procedure is assessed by comparing the observed number of violations of the quantile estimator with the theoretical frequency. The VaR estimator should be unbiased regarding the relative frequency of violations of the quantile.

We use a non-parametric method (bootstrapping) introduced by Francioni and Herzog for the calculation of the *probability-unbiased* VaR estimator for the Normal distribution. These authors show how to change the probability level in the *plug-in* quantile estimator such that the resulting *plug-in* estimator is unbiased in probability. We extend their approach to other distributions such as Student-t and mixture of Normals. We also show how to use the parametric method to calculate the *probability-unbiased* VaR in the Normal case. This is possible because under Normality we can use the probability distributions of the parameter estimates to obtain a closed-form expression for the *probability-unbiased* VaR estimator.

We show that the use of *probability-unbiased* VaR avoids the systematic underestimation of risk implied by the bias of standard VaR measures in small samples. The magnitude of the distortion that needs to be exerted on the quantile to move from the *plug-in* VaR to the *probability-unbiased* VaR depends on the sample size and on the distribution assumption on returns. Our results suggest that using a small sample may easily lead to more accurate VaR estimates than a historical estimator based on long samples, according to VaR performance measures such that the Probability of Exceedance and the Observed Absolute Deviation per year. Short samples are more robust to the structural changes that may arise over time in financial returns due to trading behavior.

1.2. A review of literature

Estimation of risk measures is an area of highest importance in the financial industry since such measures play a major role in the risk-management and the computation of regulatory capital. For an in-depth treatment of the topic, see textbooks by McNeil, Frey and Embrechts (2005), and Alexander (2009). In particular, Embrechts and Hofert (2014) highlight that a major part of quantitative risk management is actually of a statistical nature. Statistical aspects in the estimation of risk measures have recently raised a lot of attention, see Acerbi and Szekely (2007), Davis (2014), Emmer et al. (2015), Du and Escanciano (2016), Costanzino and Curran (2015), Fissler et al. (2015) and Ziegel (2016). A careful analysis shows that in general risk estimators are biased, and they systematically underestimate risk.

Therefore, the occurrence of biases in risk estimation plays an important role in practice. The Basel III document suggests to change 10-day ahead Value at Risk at 99% confidence level by Expected Shortfall and to consider stressed scenarios where the risk level is set at 97.5%. Unfortunately, such a correction may reduce the bias only in the right scenarios. On the other hand, while the classical (statistical) concept of unbiasedness is always desirable from a theoretical point of view, it might be not prioritized by financial institutions or regulators, for whom backtests are the main source of estimation accuracy. Our goal is to obtain *probability-unbiased* estimators that perform the standard backtesting procedure proposed by Basel properly, i.e. with an expected

failure rate that is close to the theoretical VaR level α under non-Normal distributions.

Surprisingly, it turns out that the statistical properties of risk estimators have not yet been analyzed thoroughly. Schaller (2002) discussed the relevance of parameter uncertainty for VaR estimation under the assumption of Normally distributed data, proposing a correction to the standard VaR estimator. But the author restricts his attention to the uncertainty in the estimation of the variance without considering the uncertainty in mean estimation. Francioni and Herzog (2012) introduced the principle of probability unbiasedness and they derived the distribution of the VaR estimator distribution the parametric case for Normally distributed data. Their approach consists on computing an appropriate distortion for the desired significance level so that the resulting VaR estimate will be unbiased in probability. Additionally, they calculate approximate VaR confidence bands. Pitera and Schmidt (2016) propose a different methodology to obtain an unbiased VaR estimator under Normality. They propose a bootstrap algorithm to obtain unbiased estimators by distorting the estimated parameters of the distribution instead of the VaR significance level (α).

Moreover, as for any other statistical estimator based on a finite amount of data, the VaR estimator has a distribution that depends on the observed realization and the amount of data. This dependence on the number of observations has been recognized by others authors, such as Baysal and Staum (2008), and Chana and Peng (2006). In these two papers, the confidence bands for VaR were derived using Monte Carlo methods. Baysal and Staum (2008) introduce probability unbiasedness only with respect to the asymptotic distribution of the confidence intervals. Only asymptotically probability-unbiased confidence bands were obtained in both papers.

We adapt the VaR estimator and the confidence band of the VaR estimator such that both become probability unbiased, as Francioni and Herzog (2012) suggest for the case of the Normal distribution, extending their analysis to distributions different from Normal, such as Student-t and Mixtures of two Normal distributions.

1.3. Quantile or VaR estimator

As specified Francioni and Herzog (2012) we describe the concepts associated to the probability unbiased estimation of Value at Risk.
Let us suppose that X is an absolutely continuous random variable with distribution function F_θ, where θ is a parameter vector. The α quantile Q_α of X is defined as $Q_\alpha = F_\theta^{-1}(\alpha)$. By definition, the quantile has the property that $F_\theta(Q_\alpha) = \alpha$. This equation represents the intuitive concept of the quantile as a threshold that is exceeded with probability α. The quantile Q_α of the distribution of returns of a given financial asset or portfolio is known as the Value at Risk (VaR) at the level α or at the confidence level $1-\alpha$.

We assume that the parameter vector θ can be estimated by any method like Maximum Likelihood, Generalized Method of Moments or others in such a way that the observed data are well described. We will assume that estimator to be at least consistent.

In a general estimation setup, a *plug-in* estimator for a function $g(\theta)$ is an estimator obtained by replacing the parameter θ in the function by an estimate, that is

$$\widehat{g(\theta)} = g(\hat{\theta})$$

The quantile Q_α can be seen as a function of the parameter vector and the significance level, $Q_\alpha = g(\theta, \alpha)$.

The *plug-in* VaR estimator is the only method to estimate VaR under a parametric approach, that is $\widehat{VaR}_\alpha = \hat{Q}_\alpha = g(\hat{\theta}, \alpha)$.

We aim at estimating the risk of the future position where $\theta \in \Theta$ are unknown. If θ were known, we could directly compute the corresponding VaR as a function of θ, $g(\theta)$, specifically with F_θ, and we would not need to consider the family of VaR, $(g(\theta))_{\theta\in\Theta}$.

Our aim is to estimate Q_α in such a way that the estimator satisfies this probabilistic 'threshold property' in the mean for a F_θ–distributed random variable X for all θ, i.e.

$$\mathbb{E}_\theta\left[F_\theta\left(\hat{Q}_\alpha\right)\right] = \alpha$$

where $\mathbb{E}_\theta$ denotes the expectation operator under probability measure F_θ. This is a standard unbiasedness condition on the probability of exceeding the VaR estimate Q_α. That probability is usually checked by backtesting. Unbiasedness would imply that the VaR estimate Q_α will be exceeded with an expected probability equal to α.

Definition 1 *An estimator $g(\theta)$, obtained with sample observations $(X_1,\ldots,X_n)\sim F_\theta$ of $g(\theta)$, is said to be probability unbiased with respect to a random variable Z with distribution function F^Z, if*

$$F^Z\left(g(\theta)\right) = \mathbb{E}_\theta\left[F^Z\left(g\left(\hat{\theta}\right)\right)\right]$$

holds for all θ.

In the case of a quantile/VaR estimation where all $X_i \sim^{iid} F_\theta, i = 1,\ldots,n$, $g(\theta)$, is the α -quantile Q_α, Z is the next sample observation $Z = X_{n+1}$, and F^Z is the probability distribution from which the sample has been obtained. Hence, a probability-unbiased VaR estimator with respect to $Z = X_{n+1}$ must satisfy

$$\mathbb{E}_\theta\left[P\left(X_{n+1} < \hat{Q}_\alpha\right)\right] = \alpha \tag{1}$$

Unfortunately, under nonlinear mappings of the parameter vector θ, as it is the case of the quantile, the *plug-in* procedure generally introduces a small sample bias: $\mathbb{E}\left[P\left(X_{new} < \widehat{VaR}_\alpha\right)\right] \neq \alpha$. The reason is that it treats the estimated parameter vector as deterministic, even though θ is a random variable, a fact that must be incorporated into the estimation procedure in order to obtain probability-unbiasedness. As a consequence, the equation, $\hat{Q}_\alpha = F_{\hat{\theta}}^{-1}(\alpha)$ where $\hat{\theta}$ is an estimator of the parameter θ, is only true asymptotically, i.e. as the number of observations goes to infinity, provided the *plug-in* estimator is consistent

$$\widehat{VaR}_\alpha \equiv \hat{Q}_\alpha \xrightarrow{n\to\infty} VaR_\alpha \equiv Q_\alpha = F_\alpha^{-1}(\alpha)$$

almost surely for each $\theta\epsilon\Theta$, so that this asymptotically unbiased.

To obtain a probability-unbiased estimator for the quantile there are two approaches,

1. Replacing the level α by a suitable chosen level α_{pu} to modify the quantile of the estimated distribution. The VaR estimator will be

$$\hat{Q}_\alpha = \widehat{VaR}_\alpha = g\left(\hat{\theta}, \alpha_{pu}\right) = F_{\hat{\theta}}^{-1}(\alpha_{pu})$$

   where α_{pu} is chosen so that equation (1) is fulfilled.
   For example, if F is a Normal distribution, the VaR estimator can be written

$$\widehat{VaR}_\alpha = \hat{\mu} + \hat{\sigma} z_{\alpha_{pu}}$$

   where $\hat{\mu}$ and $\hat{\sigma}$ are the estimated mean and standard deviation, respectively, and $z_{\alpha_{pu}}$ is the inverse cumulative distribution function of the standard Normal(0,1) for α_{pu}.
2. Modifying the vector of estimated parameters $\hat{\theta}$ of the distribution F to $\hat{\theta}_{pu}$ when computing the *plug-in* estimator

$$\hat{Q}_\alpha = \widehat{VaR}_\alpha = g\left(\hat{\theta}_{pu}, \alpha\right) = F_{\hat{\theta}_{pu}}^{-1}(\alpha)$$

   If F is a Normal distribution: $\hat{\theta}_{pu} = (\hat{\mu}_{pu}, \hat{\sigma}_{pu})$, and the VaR estimator would then be written as follows

$$\widehat{VaR}_\alpha = \hat{\mu}_{pu} + \hat{\sigma}_{pu} z_\alpha$$

   On the other hand, the *plug-in* estimator, which has been used in the calculation of quantile / VaR is

$$\widehat{VaR}_\alpha = \hat{\mu} + \hat{\sigma} z_\alpha$$

In this chapter we follow the first of these two approaches to calculate the probability-unbiased VaR, and we use the second approach, whenever possible, to graph an approximation of the function F distorted by modifying the parameter $\hat{\theta}$. Thus, we will be computing a probability-unbiased estimator of VaR, that is, an estimator

$$\hat{Q}_\alpha = F_{\hat{\theta}}^{-1}(\alpha_{pu})$$

where α_{pu} is chosen so that equation (1) is fulfilled.

1.4. Parametric and Non-Parametric methods

We can use parametric or non-parametric methods to find α_{pu}. On the one hand, parametric methods or classical statistical methods have the basis for making inferences about the population in the theoretical sampling statistical distribution, whose parameters can be estimated from the observed sample. On the other hand, there are different procedures based on non-parametric methods. Those procedures generate samples from a set of observations constructing a sample distribution that can be used for parameter estimation and confidence intervals. Among them, the best known and most commonly used is the bootstrap method. The first mention of this method under this name is due to Efron (1979), although the same basic ideas came handling for at least a decade ago (Simon, 1969). Efron conceived the bootstrap method as an extension of "jackknife techniques", which usually consist in extracting samples ever constructed by removing one element of the original sample to assess the effect on certain statistical (Quenouille, 1949; Tukey, 1958; and Miller, 1974).

Unlike classical estimation methods, the bootstrap method does not make any distributional assumptions for the theoretical statistic. Instead, the distribution of the statistic is determined by simulating a large number of random samples constructed directly from observed data. That is, the original sample is used to generate new samples as a basis for estimating inductively the sampling distribution of the statistic, rather than deriving it from a theoretical distribution assumed a priori. This method has an immediate predecessor in the techniques of Monte Carlo simulation, consisting in extracting a large number of random samples from a known population to calculate from them the value of the statistic whose sampling distribution is intended to be estimated. However, in practice the population is not known and the information we have is a sample drawn from it.

Definition 2 *Bootstrapping (bootstrap) is a resampling method or algorithm that consists in generating a large number of resamples using sampling with replacement from an original random sample of size n that represents the population from which it was extracted. Each resample is the same size as the original random sample. The resamples serve as alternative population samples.*

According to the main idea of bootstrap, the procedure involves using the sample itself since we consider that it contains basic information about the population. Therefore, the suitability of this method will be greater when the sample contributes with more information about the population. A direct consequence is that the longer the sample size, the better the estimation about the sample distribution of a statistic. However, even with small samples, between ten and twenty observations, the bootstrap method can provide correct results (Bickel and Krieger, 1989) while being unsuitable for samples with less than five (Chernick, 1999).

1.5. Normal Distribution

The VaR calculated by the parametric approach for a Normal distribution is $Q_\alpha = \mu + \sigma z_\alpha$ where the parameter vector $\theta = (\mu, \sigma)$. To obtain the *plug-in* $\widehat{VaR}_\alpha$ the parameters of the Normal distribution are replaced by their Maximum Likelihood estimates $\hat{Q}_\alpha = \bar{x} + s z_\alpha$ where z_α is the inverse cumulative distribution function of a Normal(0,1), $\bar{x}$ is the sample mean and s is the sample standard deviation. These statistics are independent since the sample comes from a distribution $N(\mu, \sigma^2)$

$$\bar{x} = \frac{1}{n}\sum X_i$$

$$s = \sqrt{\frac{1}{n-1}\sum (X_i - \bar{x})^2}$$

For the Normal distribution, the statistical distributions are known, where the distribution of sample mean is a Normal distribution,

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

and the distribution of variance is calculated as follows,

If we start from a simple random sample with distribution $N(\mu, \sigma^2)$, then

$$\frac{n-1}{\sigma^2} s^2 \sim \chi^2_{n-1}$$

and therefore, the distribution of s^2 is

$$s^2 \sim \chi^2_{n-1}\left(\frac{n-1}{\sigma^2} s^2\right)\frac{n-1}{\sigma^2}$$

Note that the probability distribution of the two estimators depends on the size of the sample, *n*. The estimator $\bar{x}$ is unbiased, and s^2 is consiss tent (Casella and Berger, 2002).

To obtain $\mathbb{E}\left[P\left(X_{n+1} < \widehat{VaR}_\alpha\right)\right] = \alpha$, we have,

$$P(X_{n+1} < \bar{x} + sz_\alpha) = P\left(\frac{X_{n+1} - \mu}{\sigma} < \frac{\bar{x} + sz_\alpha - \mu}{\sigma}\right) = \Phi\left(\frac{\bar{x} - \mu}{\sigma} + \frac{s}{\sigma} z_\alpha\right)$$

so that,

$$\begin{aligned} &\mathbb{E}\left[P\left(X_{n+1} < \widehat{VaR}_\alpha\right)\right] = \mathbb{E}\left[\Phi\left(\frac{\bar{x} - \mu}{\sigma} + \frac{s}{\sigma} z_\alpha\right)\right] = \\ &= \iint \Phi\left(\frac{\bar{x} - \mu}{\sigma} + \frac{\sqrt{s^2}}{\sigma} z_\alpha\right) N\left(\bar{x}|\mu, \frac{\sigma^2}{n}\right) \chi^2_{n-1}\left(\frac{n-1}{\sigma^2} s^2\right)\frac{n-1}{\sigma^2} d\bar{x} ds^2 \end{aligned} \tag{2}$$

where we have calculated the expectation of a continuous function of two random variables

$$\mathbb{E}[g(\bar{x}, s^2)] = \iint g(\bar{x}, s^2) f_{\bar{x},s^2}\, d\bar{x} ds^2$$

where $f_{\bar{x},s^2}$ is the joint density function of two random variables. As $\bar{x}$ and s^2 are independent, the joint density function is just the product of the density functions of each variable,

$$N\left(\bar{x}|\mu,\frac{\sigma^2}{n}\right)\chi^2_{n-1}\left(\frac{n-1}{\sigma^2}s^2\right)\frac{n-1}{\sigma^2}$$

In (2) we do the following change of variable,

$$X=\frac{x-\mu}{\sigma}\quad d\bar{x}=\sigma dX$$
$$Y=\frac{n-1}{\sigma^2}s^2\quad ds^2=\frac{\sigma^2}{n-1}dY$$

The density function of X relates to that of the sample mean by

$$N\left(\bar{x}|\mu,\frac{\sigma^2}{n}\right)=\frac{\sqrt{n}}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2}=\frac{\sqrt{n}}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{n}{2}X^2}=\frac{1}{\sigma}N\left(X|0,\frac{1}{n}\right)$$

Therefore, the equation that defines α_{pu} with Maximum Likelihood estimator for the Normal distribution is

$$\iint\Phi\left(X+\sqrt{\frac{Y}{n-1}}z_{\alpha_{pu}}\right)N\left(X|0,\frac{1}{n}\right)\chi^2_{n-1}(Y)dXdY=\alpha \qquad (3)$$

Notice that equation (3) only depends on α and n, but it does not depend on θ, i.e. μ and σ. Non-dependence on θ arises under the Normal distribution because of its strong invariance structure. Being a location-scale distribution, we can reduce it to a standard Normal distribution that does not depend on these parameters. This property is important because estimation of the parameter θ make the VaR estimator to be just an approximation to the *probability-unbiased* VaR.

Therefore, the α_{pu} is unique for each sample size (n) and for each probability α and it does not vary from one sample to another of equal size because the function (3) does not depend on the distribution parameters. The VaR obtained with each of these α_{pu} will be *probability-unbiased*, that is, $\mathbb{E}\left[\widehat{VaR}\right]=VaR$.

To sum up, if the probability distribution from which we draw independent sample realizations belongs to the location-scale family, then we will be able to find an α_{pu} such that the VaR is unbiased.

Table 1 lists the probabilities α_{pu} obtained from equation (3), as a function of the sample size n and the value of α[1]. We can see that $\alpha_{pu} \to \alpha$ when $n \to \infty$. For instance, under the estimated probability distribution for a sample size n = 20, the 3.82% quantile has a 5% probability of being exceeded by a future observation drawn from the full distribution of returns. As we can see, for small sample sizes the estimated distribution function from a Normal sample is much heavier tailed than the Normal distribution associated to the plug-in estimator. As a consequence, the *plug-in* VaR estimator underestimates risk.

Table 2 presents the reverse question: What is the α associated to a given α_{pu}? Now, at 5% significance and n = 20, the plug-in VaR estimate would have a 6.25% probability of being exceeded by a future sample observation from the full distribution of returns. We observe that the differences are greater when we have small sample sizes and we can also observe that $\alpha_{pu} \to \alpha$ when $n \to \infty$.

Table I: Probabilities α_{pu}(%) to be used to obtain the probability-unbiased VaR_α for different values of α and n in the i.i.d. Normal distribution case

	α (%)			
n	0.5	1	5	10
10	0.033	0.154	2.727	7.345
15	0.105	0.336	3.445	8.239
20	0.169	0.463	3.821	8.683
25	0.217	0.552	4.051	8.948
50	0.340	0.757	4.520	9.476
100	0.415	0.874	4.759	9.738
150	0.442	0.915	4.839	9.826
200	0.456	0.936	4.879	9.869

1. The probabilities α_{pu} were obtained implicitly, by searching for the α_{pu} that make the double integral in (3) equal to α for a given sample size. Calculations were performed with Mathematica software 9.0.

Table 2: Shortfall probabilities α (%) that the next observation will be lower than the plug-in VaR estimate $Z\alpha_{pu}$ in the i.i.d. Normal distribution case

	α_{pu} (%)			
n	0.5	1	5	10
10	1.820	2.686	7.563	12.639
15	1.288	2.043	6.678	11.752
20	1.056	1.751	6.247	11.312
25	0.928	1.585	5.992	11.048
50	0.697	1.277	5.490	10.523
100	0.594	1.134	5.243	10.261
150	0.562	1.089	5.162	10.174
200	0.546	1.066	5.121	10.130

Figure 1 graphs the distortion function for different sample sizes (grey line). It corroborates the fact that, as we have more observations, the correction in the probability level is smaller and the distortion function converges to the identity (black line). This distortion function describes how probabilities need to be changed in the plug-in quantile estimator such that the plug-in estimator becomes probability-unbiased. Figure 2 shows the distortion of the quantiles of the standard Normal distribution function which describes how the plug-in estimate of the cumulative density function has to be changed for a given sample size n such that the estimate becomes probability-unbiased. In both figures, we only represent the left extreme quantiles, but it would be possible to enlarge the graph to represent the entire distribution. In Figure 2 we observe that for more extreme quantiles the distortion is greater, i.e. the differences between α and α_{pu} are larger. Also, α_{pu} is always lower than α, in other words, *probability-unbiased* VaR is greater (in absolute value) than *plug-in* VaR. The latter underestimates the extreme events and, therefore, is not an appropriate method to estimate risk measure with small samples.

Figure 1: Distortion function for the Normal distribution calculated with the parametric method. The diagonal (black line) represents no distortion

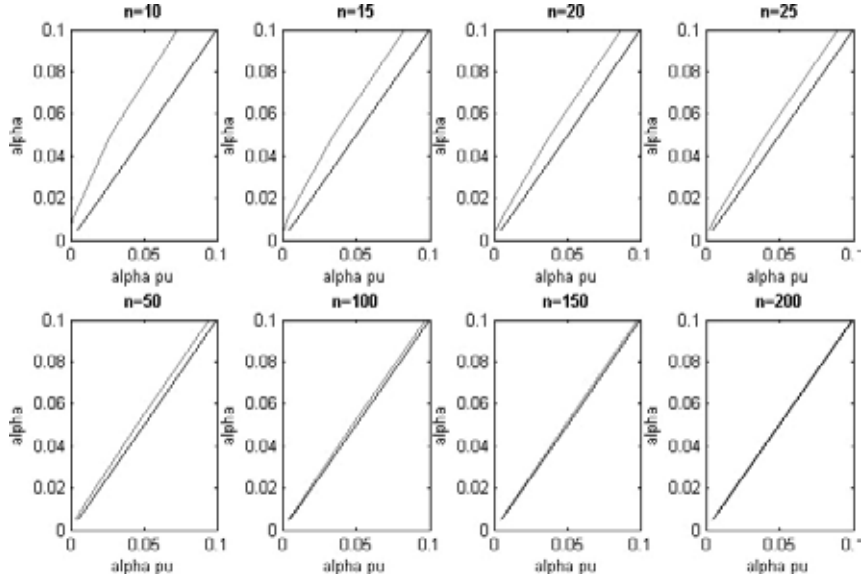
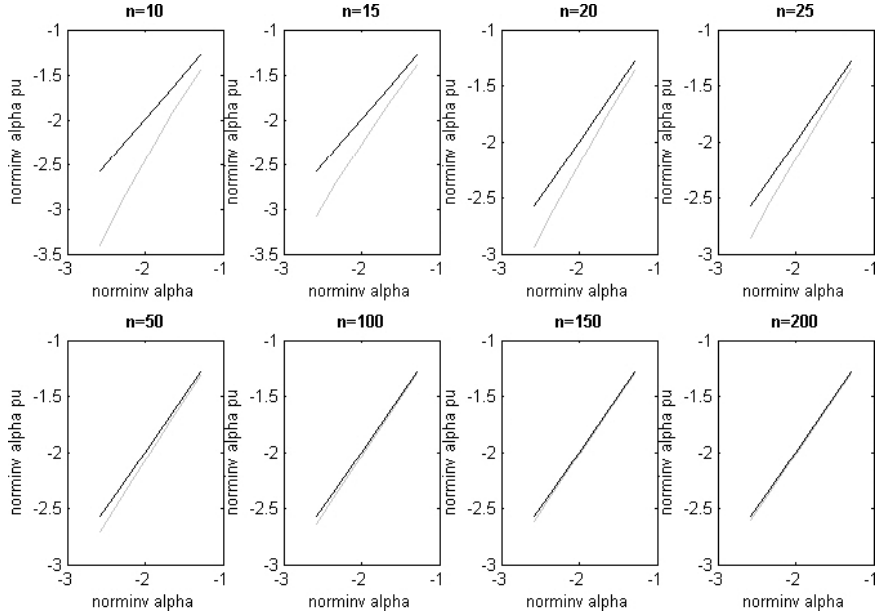


Figure 2: The quantiles of the Normal cdf versus the quantiles of the distorted Normal cdf calculated with the parametric method. The diagonal (black line) represents no distortion



1.5.1. Parametric probability-unbiased VaR estimator for a Normal distribution

We now turn to the estimation of VaR itself. We apply the first approach described in Section 1.3 to estimate the *probability-unbiased* VaR, which implies a modification of the quantile, replacing α by α_{pu}. Table 3 shows the *probability-unbiased* $\widehat{VaR}_\alpha$ (VaR_{pu}) and the *plug-in* $\widehat{VaR}_\alpha$ $(VaR_{plug-in})$ obtained for different sample sizes and α's. We can see that *plug-in* $\widehat{VaR}_\alpha$ underestimates risk, indicating smaller losses than we should really expect with $\alpha\%$ probability. Thus, for instance, for a random sample of size 25, the maximum expected loss with 95% probability or, equivalently, the minimum loss with a 5% is not 1.844, but 1.949.

The calculation of *probability-unbiased* VaR is particularly relevant for small sample sizes, when the difference in the estimation of VaR is higher than for large samples, for which the *probability-unbiased* $\widehat{VaR}_\alpha$ and the *plug-in* $\widehat{VaR}_\alpha$ are very similar.

Now, we follow the second approach described in Section 1.3, to obtain the *probability-unbiased* VaR estimator by calculating the standard deviation of $\hat{\sigma}_{pu}$ the distorted distribution function *F*.

If *F* is a Normal distribution, the *probability-unbiased* VaR estimator can be written in two alternative ways:

$$\widehat{VaR}_\alpha = \hat{\mu}_{pu} + \hat{\sigma}_{pu} z_\alpha = \hat{\mu} + \hat{\sigma} z_{\alpha_{pu}}$$

Table 3: Probability-unbiased $\widehat{VaR}_\alpha$ versus plug-in $\widehat{VaR}_\alpha$ in the case of Normal (0,1)

	VaR_{pu}				$VaR_{plug-in}$			
n	0.5	1	5	10	0.5	1	5	10
10	-3.227	-2.786	-1.768	-1.305	-2.409	-2.165	-1.496	-1.139
15	-2.796	-2.440	-1.570	-1.150	-2.309	-2.065	-1.400	-1.045
20	-3.206	-2.866	-2.011	-1.587	-2.839	-2.582	-1.880	-1.506
25	-3.117	-2.789	-1.949	-1.527	-2.825	-2.562	-1.844	-1.461
50	-2.925	-2.624	-1.827	-1.415	-2.783	-2.513	-1.775	-1.382
100	-2.546	-2.307	-1.666	-1.329	-2.488	-2.262	-1.645	-1.316
150	-2.621	-2.374	-1.705	-1.352	-2.581	-2.342	-1.690	-1.343
200	-2.393	-2.165	-1.547	-1.220	-2.365	-2.143	-1.537	-1.213

that illustrate the two equivalent approaches to *probability-unbiased* VaR estimation: either we distort the quantile and use the estimated parameters or we maintain the original quantile while distorting the estimated parameters. This equation also shows that once we have calculated α_{pu} we can obtain $\sigma z_{\alpha_{pu}}$ and viceversa.

Since the mean of the distribution is very low in high frequency returns and it is estimated with very low precision, we can consider it to be the same for the distorted distribution as for the original distribution, i.e. $\hat{\mu}_{pu} = \hat{\mu}$. Then, we will calculate the standard deviation $\hat{\sigma}_{pu}$ of the distorted distribution function implicitly so that the previous equation holds. That standard deviation will be different for every α and for each sample size (n) because $\widehat{VaR}_{\alpha}$ also changes with α and with n.

Table 4 shows the $\hat{\sigma}_{pu}$ values obtained for different α and n. Notice that $\hat{\sigma}_{pu}$ is greater for small sample sizes suggesting the heavier tails of the distorted distribution. For a given sample size, we obtain larger differences between $\hat{\sigma}$ and $\hat{\sigma}_{pu}$ for the more extreme quantiles. For a given α, we obtain greater differences between $\hat{\sigma}$ and $\hat{\sigma}_{pu}$ for small sample sizes. As the sample size increases the $\hat{\sigma}_{pu}$'s move closer to the sample standard deviation ($\hat{\sigma} = s$) for any α and, therefore, closer to the population standard deviation, 1. At $n = 200$ we see a distortion produced by estimating the mean. Had we set $\hat{\mu} = 0$ when calculating the parametric VaR estimate, we would get $\hat{\sigma}_{pu}$ converging to $\hat{\sigma}$ and hence, to σ, the population standard deviation, which is equal to 1, as the sample size increases.

Figure 3 shows the true density function of a random variable N(0,1) (solid line), the density function of the Normal distribution with the parameters estimated from a random sample of size 15 extracted from a N(0,1) (dashed line), and the density function of the distorted estimated distribution function using the $\hat{\sigma}_{pu}$ estimate (dotted line). We can see that the distorted distribution function has heavier tails, which should allow for a better fit to most asset returns. The *probability-unbiased* $\widehat{VaR}$ (black circle) indicates higher losses than *plug-in* $\widehat{VaR}$ (black triangle). In other words, the *plug-in* estimator underestimates risk. This will generally be the case with small size samples. Besides, the smaller the sample size the greater the correction or adjustment needed on the probability distribution.

Table 4: Estimated standard deviations for the distorted distribution function

	α(%)			
n	0.5	1	5	10
10	1.299	1.248	1.147	1.111
15	1.165	1.137	1.079	1.058
20	1.172	1.152	1.109	1.093
25	1.167	1.151	1.118	1.105
50	1.138	1.130	1.114	1.108
100	0.928	0.925	0.919	0.916
150	0.972	0.970	0.966	0.964
200	0.901	0.899	0.8965	0.895

Figure 3: The true N(0,1) pdf (solid line), the *plug-in* pdf (dashed line) and the pdf of the *unbiased* cdf (dotted line). On the horizontal axis the data points for the true $\widehat{VaR}_{5\%}$ (black square), the *plug-in* $\widehat{VaR}_{5\%}$ (black triangle) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (black circle) are plotted

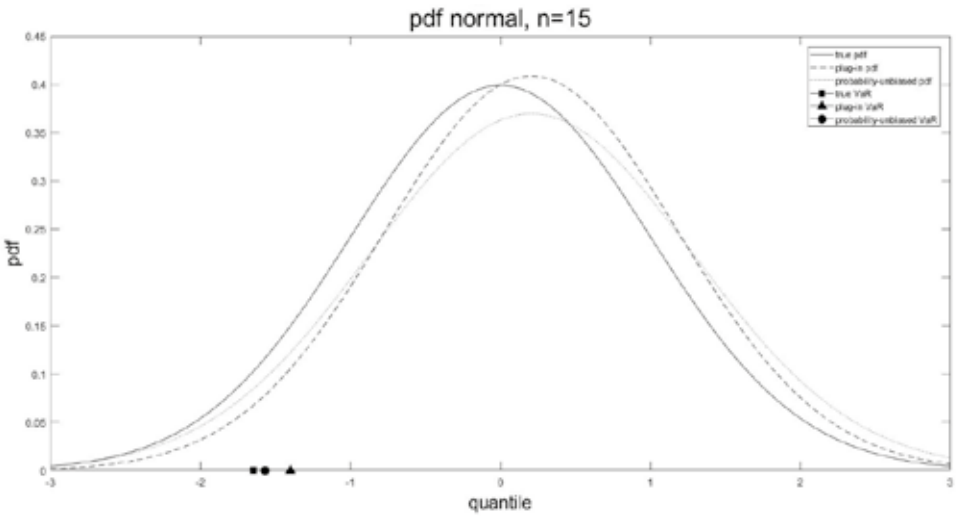
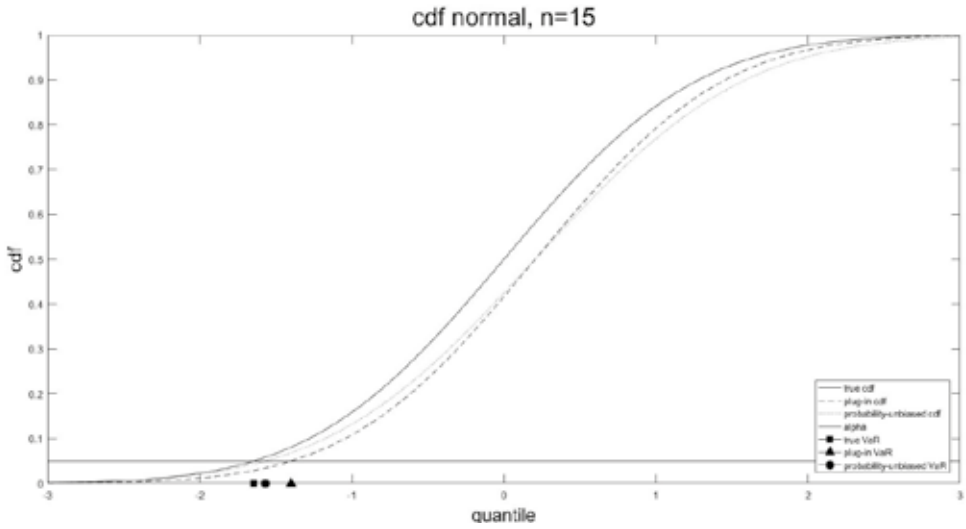


Figure 4 shows the cumulative distribution function of N(0,1) (solid line), the *plug-in* cumulative distribution function (dashed line) and the *unbiased* cumulative distribution function (dotted line). It also displays VaR estimates at 5% significance level. As Figure 3, these representations are based on the estimates obtained from a random sample of size 15. The smaller the sample size the larger the distortion in the *plug-in* distribution function.

Figure 4: The true N(0,1) cdf (solid line), the *plug-in* cdf (dashed line) and the *unbiased* cdf (dotted line). On the horizontal axis the data points for the true $\widehat{VaR}_{5\%}$ (black square), the *plug-in* $\widehat{VaR}_{5\%}$ (black triangle) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (black circle) are plotted



Figures 5 and 6 show pdf's and cdf's, respectively, for different sample sizes. We also show the *plug-in* $\widehat{VaR}_{5\%}$ and the *probability-unbiased* $\widehat{VaR}_{5\%}$. These Figures show the convergence of the *plug-in* distribution and the *probability-unbiased* distribution to the true distribution as the sample size increases.

Figure 5: Enlarged left tail of the true N(0,1) pdf (solid line), the *plug-in* pdf (dashed line) and the pdf of the *unbiased* cdf (dotted line) for different sample sizes. On the horizontal axis the data points for the true $\widehat{VaR}_{5\%}$ (black square), the *plug-in* $\widehat{VaR}_{5\%}$ (black triangle) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (black circle) are plotted

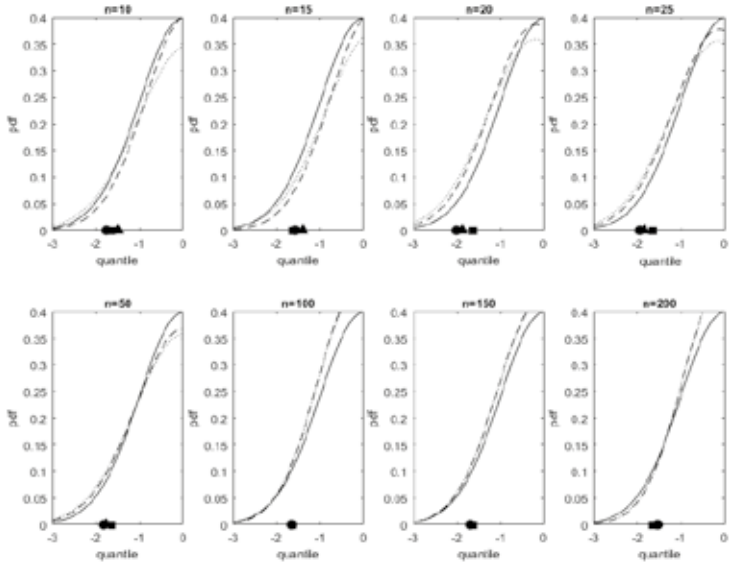
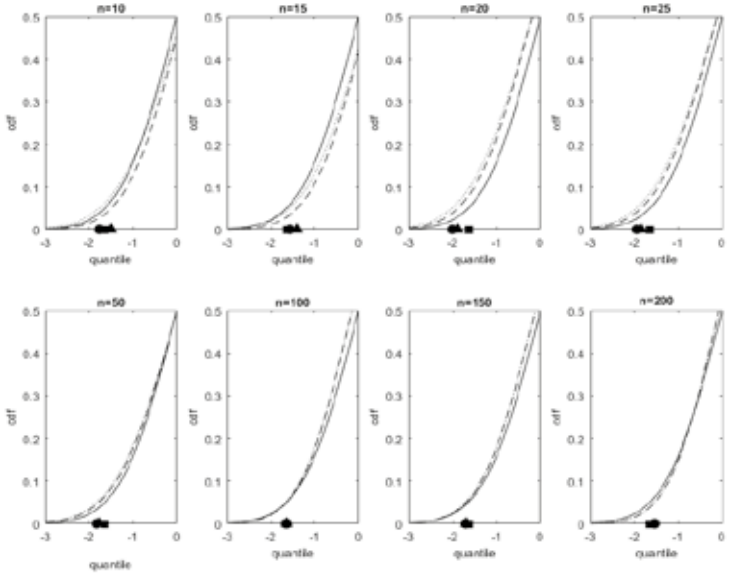


Figure 6: Enlarged left tail of the true N(0,1) cdf (solid line), the *plug-in* cdf (dashed line) and the *unbiased* cdf (dotted line) for different sample sizes. On the horizontal axis the data points for the true $\widehat{VaR}_{5\%}$ (black square), the *plug-in* $\widehat{VaR}_{5\%}$ (black triangle) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (black circle) are plotted



1.5.2. A comparison of parametric estimates of probability-unbiased VaR and plug-in VaR under Normality

We now compare the exceedance probabilities obtained from *probability-unbiased* VaR and *plug-in* VaR. We simulate the estimation of the *plug-in* VaR estimator and the *probability- unbiased* VaR estimator and calculate the exceedance probabilities. According to the argument in the previous section, we expect to obtain a number of exceedances for *probability-unbiased* VaR to be close to the theoretical α regardless of the sample size considered.

The Monte-Carlo exercise with S simulations is performed using the following steps:

1. Set the counter of the simulation $s = 0$.
2. Increment the counter of the simulation $s = s + 1$.
3. Simulate normally distributed data with $n + 1$ observations and predefined values of μ and σ.
4. Estimate the mean and the standard deviation based on the first n data points.
5. Calculate the *plug-in* VaR estimator. Calculate the *probability-unbiased* VaR estimator using the α_{pu} from Table 1 for the μ and σ estimates in step 4.
6. Check if the $n + 1$ data point is smaller than the *plug-in* $\widehat{VaR}$ and the *probability- unbiased* $\widehat{VaR}$ estimators. When the data point is smaller (VaR exceedance) record a 1, otherwise a 0.
7. Return to step 2 while $s < S$.
8. Calculate the exceedance probability by summing the recorded values and dividing them by the number of simulations S.

The results in Table 5 for samples of size 10, 15, 20 and 25, for $\widehat{VaR}_{1\%}$ and $\widehat{VaR}_{5\%}$, with $\mu = 0$, $\sigma = 1$ and $S = 100\,000$, show that the probability of a VaR exceedance from the *probability-unbiased* estimator is close to the theoretical values of 1% and 5%.

However, the probability of a VaR exceedance for the *plug-in* VaR differs than the theoretical probability. This confirms the results presented in Table 1. As the sample size increases, the probability of an excess from the *plug-in* VaR estimator calculated from the simulations approaches the theoretical value. For the *probability-unbiased* VaR estimator, probability remains similar to the theoretical probability for all sample sizes.

Table 5: The shortfall probabilities α% with which the next observation is always lower than the plug-in $\widehat{VaR}$ and the probability-unbiased $\widehat{VaR}$ in the Monte-Carlo simulation for the Normal distribution

n	1% plug-in	1% pu	5% plug-in	5% pu
10	2.674	0.985	7.567	5.015
15	2.063	0.953	6.748	5.047
20	1.825	1.028	6.268	5.038
25	1.608	0.999	5.978	4.930

1.5.3. Non-parametric (Bootstrapping) estimation of probability-unbiased VaR

In the previous section we have calculated α_{pu} analytically, corroborating that the exceedance probabilities obtained from *probability-unbiased* VaR are closer to theoretical α than those obtained from *plug-in* VaR for all sample sizes. We lack a closed-form solution for unbiased estimators for other distributions, such as Student-t and Mixture of two Normals, because the statistical distributions of the sample parameters are unknown, so that the bootstrap algorithm proposed by FH is extremely useful. In this subsection, we use that algorithm to calculate α_{pu} for Normal distributions, as FH propose, and we compare the results obtained with this method and the analytical method.

The algorithm proposed by FH when sampling from a Normal distribution replaces the level α by a suitably chosen level α_{pu} so as to minimize the average distance between the bootstrapped estimators and α. The α_{pu} obtained through resampling will change with the sample size (n), the significance level α and the observed values in the sample. The algorithm approximates the α_{pu} level and achieves an approximation to the *probability-unbiased* VaR through a modification of the significance level. The change from α to α_{pu} corrects the fact that we do not observe infinite realizations. For a large number of observations the *plug-in* estimator and the *probability-unbiased* estimator become very similar. The *plug-in* estimator has good properties only asymptotically, while the *probability-unbiased* estimator is a good estimator even in short samples.

Suppose we have a random sample of size n drawn from a distribution F_θ. Then, we generate B resamples of the same size n. These resamples are obtained by sampling with replacement. The steps to be performed are:

1. From observed values $X_1, ..., X_n \sim^{i.i.d.} F_\theta$
2. Estimate $\theta = \theta(X_1, ..., X_n)$
3. For i=1:B
   Samples $X_1, ..., X_n$ from $F_{\hat{\theta}}$[2]
   Estimate $\hat{\theta}_i^*$

Find the α_{pu} that minimizes the following objective function

$$\alpha_{pu} = argmin_\gamma \left| \frac{1}{B} \sum_{i=1}^{B} F_{\hat{\theta}} \left(F_{\hat{\theta}_i^*}^{-1}(\gamma) \right) - \alpha \right| \tag{4}$$

The level of α_{pu} is chosen so that equation (1) is satisfied. Substituting α for α_{pu} obtain the *probability-unbiased* VaR estimator.

We start with a random sample of size n generated from a Normal distribution with mean 0 and standard deviation 1. From this original random sample we obtain 10 000 resamples of size n. As we increase the sample size, the Maximum Likelihood estimates of mean and standard deviation of the original random sample, μ_y and σ_y, tend to the population average ($\mu_x = 0$) and the population standard deviation ($\sigma_x = 1$). For each resample we estimate the mean and the standard deviation, obtaining 10 000 means and 10 000 standard deviations. These estimates are used to find the α_{pu} that minimizes the objective function (4).

We have shown that for a Normal distribution it is possible to obtain a closed-form solution for probability-unbiased VaR. But for other

2. Notice that the samples must be resamples of the original sample (our observed values) for three reasons: 1) if the samples were generated from the distribution function estimated with the original sample, we would obtain from each resample very different values of $\hat{\theta}_i^*$, especially with small sample sizes, and we would have to draw many samples to obtain suitable results. Indeed, even extracting 100 000 samples from F_θ we have not obtained the expected results, and α_{pu} does not tend to α when n tends to ∞, 2) we have just one random sample, possibly of small size, and we cannot use classical statistical inference to find the sampling distribution because we do not know the parameters of the population distribution and we cannot take the estimated parameters as population parameters. Therefore, to find the sampling distribution, at least approximately, we create many resamples by repeatedly sampling with replacement from the original random sample. Each resample has the same size as the original random sample, and 3) a bootstrap algorithm is based on a large number of new samples obtained by sampling from the original sample, not by simulation.

distributions for which the probability distributions of estimated parameters are either unknown or they are difficult to obtain, we suggest using this bootstrap algorithm. In particular, we will use below the FH algorithm to calculate probability-unbiased VaR for Student-t distributions as well as for a mixture of two Normal distributions.

1.6. Student-t Distribution

1.6.1. Probability-unbiased VaR estimator for a Student-t distribution

We assume that we have a finite-short sample from a Student-t distribution function (F_θ). Following Francioni and Herzog, the VaR estimator is a modification on the α-quantile from the estimated distribution. We replace α by α_{pu} thereby taking a quantile of the estimated probability distribution different from α.

Hence, if F is a t-Student distribution, the VaR estimator can be written:

$$\widehat{VaR}_\alpha = t^{-1}(\alpha_{pu})$$

where α_{pu} is chosen so that the equation $\mathbb{E}_\theta[P\,(X_{n+1}) < \hat{Q}_\alpha)] = \alpha$ is satisfied. The α_{pu} approximation is obtained by a bootstrap algorithm. The change of α for α_{pu} corrects for the fact that do not observe infinite realizations. The *probability-unbiased* VaR estimator can be obtained for any sample size, including small sample sizes, while the estimator *plug-in* is only *probability-unbiased* when $n \to \infty$.

We start with a random sample of size n generated from a Student-t distribution with 2 degrees of freedom. This is the original random sample from which we will generate 10 000 resamples of sample size n. The parameter to be estimated in this distribution is the number of degrees of freedom. We use a method of moments estimator: $\nu = \frac{2\sigma^2}{\sigma^2-1}$ because of its simplicity although it requires that the distribution has a variance greater than 1[3]. For each resample, we estimate the number of degrees of

3. An alternative estimator might combine two different GMM estimators based on the variance and the kurtosis of the sample. However, often the number of estimated degrees of freedom is below 4, not allowing for the use of the kurtosis estimator.

freedom, which is then used to find the α_{pu} value solving the previously established equation (4).

Table 6 contains probabilities α_{pu} for the different values of α and n in the i.i.d. Studen-t distribution case with 2 degrees of freedom. This table shows that $\alpha_{pu} \to \alpha$ as n $\to \infty$. Comparing Table 1, probabilities α_{pu} obtained from closed-formed solution for Normal distribution case, with Table 6, we observe the convergence under the Student-t distribution is faster than under the Normal, and for small sample sizes we obtain an α_{pu} closer to the theoretical α than under the Normal distribution. This is because the higher kurtosis of the Student-t distribution makes more likely the occurrence of extreme events, so that the correction needed on α is smaller.

Table 6: Probabilities α_{pu}(%) to be used to obtain a probability-unbiased VaRα for different values of α and n in the i.i.d. Student-t distribution case

	α (%)			
n	**0.5**	**1**	**5**	**10**
10	0.071	0.332	3.865	8.903
15	0.227	0.639	4.470	9.490
20	0.357	0.798	4.663	9.668
25	0.363	0.804	4.656	9.654
50	0.373	0.821	4.701	9.706
100	0.389	0.833	4.784	9.776
150	0.450	0.926	4.862	9.858
200	0.445	0.919	4.855	9.853

Table 7 lists $\widehat{VaR}_{\alpha}$ *probability-unbiased*, ($\widehat{VaR}_{pu}$), and $\widehat{VaR}_{\alpha}$ *plug-in* (the usual VaR estimate) obtained for different sample sizes and α's. This table shows that the *plug-in* $\widehat{VaR}_{\alpha}$ underestimates risk for any probability level α%. As with the Normal distribution, the calculation of *probability-unbiased* VaR is especially relevant for small sample sizes although in this case, differences between both VaR estimates are larger.

Table 7: Probability-unbiased $\widehat{VaR}_\alpha$ versus plug-in $\widehat{VaR}_\alpha$ in the case of Student-t distribution

n	0.5	1	5	10	0.5	1	5	10
10	-17.102	-8.949	-2.975	-1.885	-7.515	-5.564	-2.609	-1.752
15	-14.275	-8.533	-3.078	-1.940	-9.650	-6.808	-2.887	-1.872
20	-7.928	-5.716	-2.612	-1.750	-6.919	-5.205	-2.522	-1.714
25	-9.530	-6.635	-2.815	-1.838	-8.247	-5.997	-2.709	-1.796
50	-7.437	-5.448	-2.556	-1.727	-6.634	-5.031	-2.479	-1.695
100	-8.666	-6.198	-2.736	-1.805	-7.762	-5.711	-2.643	-1.767
150	-8.372	-6.054	-2.715	-1.798	-7.990	-5.846	-2.674	-1.781
200	-8.512	-6.128	-2.728	-1.804	-8.078	-5.898	-2.686	-1.786

1.7. Mixture of two Normal distributions

1.7.1. Probability-unbiased VaR estimator for Mixtures of Normal distributions

As an example, we consider a mixture of Normal distributions with different mean and different standard deviation: N(-5,10) and N(0,1) with mixing parameter p = 0.1. With so different Normal distributions and a small p, the resulting mixture can capture potential extreme data much better than a Normal distribution, which may provide a better fit to some of the statistical characteristics observed in asset returns. Table 8 shows moments for sample sizes 100, 200, 300 and 400. As a result of this mixing, we obtain a distribution having a smaller mean, greater deviation and largest kurtosis than the second Normal distribution. The quantiles of a mixture distribution do not accept a closed form solution but rather, they require solving an implicit equation. Therefore, to calculate the VaR we cannot use the parametric approach. We then need to work with samples larger than in the case of Normal and t-Student distributions because both *plug-in* VaR and *probability-unbiased* VaR are now calculated as a sample percentile. For example, if we want to compute the 1% percentile, we must compute it from a sample of considerable size to avoid that it might fall outside the data range. For instance, the *prctile* function of MatLab would return the first value of the sample, in spite of the fact that the first value might be significantly larger than the 1% percentile.

Table 8: First four moments for samples of size 100, 200, 300 and 400 of a mixture distribution of N(-5,10) and N(0,1) with a mixing parameter p = 0.1

	MIXTURE of N(-5,10) and N(0,1)			
n	μ_{mix}	σ_{mix}	$skewness_{mix}$	$kurtosis_{mix}$
100	-0.5274	3.8400	-3.0260	18.8137
200	-0.5415	3.4229	-4.9035	34.4905
300	-0.2653	2.7198	-2.1651	29.9570
400	-0.2644	3.4468	-2.6797	27.9013

Table 9 shows the α_{pu} calculated from the bootstrap method to estimate the probability-unbiased VaR. We again see that the 1% and 5% plug-in VaR underestimate risk.

Table 9: Probabilities α_{pu}(%) needed to obtain probability-unbiased $\widehat{VaR}_{\alpha}$ for samples of size 100, 200, 300 and 400 of a mixture distribution of a Normal(-5,10) and of a Normal(0,1) with mixing parameter p = 0.1. We also show the associated probability-unbiased VaR and plug-in VaR for α = 1% and α = 5%

	α = 1%			α = 5%		
n	α_{pu}	VaR_{pu}	$VaR_{plug-in}$	α_{pu}	VaR_{pu}	$VaR_{plug-in}$
100	0.5806	-23.2413	-19.0515	4.6457	-6.6337	-5.5756
200	0.7963	-18.4699	-17.7663	4.6833	-2.2543	-1.8864
300	0.8333	-12.6246	-12.4064	4.8733	-2.1380	-2.1126
400	0.8677	-20.5952	-18.8131	4.8821	-2.7855	-2.6125

1.8. Empirical application and comparison with other VaR models

In this section we follow McNeil et al. (2005, chapter 2.3.6) to test different VaR estimation methods using the last 1000 data observations from a portfolio that invests 30% in the Financial Times 100 Shares Index (FTSE 100), 40% in the Standard & Poor's 500 (S&P 500) and 30% in Swiss Market Index (SMI) between 1992 and 2003. We consider the application of methods belonging to the general categories of variance-covariance and historical simulation methods to the portfolio of an

investor in international equity indexes. The investor is assumed to have domestic currency pound sterling (GBP) and to invest in FTSE 100, S&P 500 and SMI. The investor thus has currency exposure to US dollars (USD) and Swiss francs (CHF) and the value of portfolio is influenced by five risk factors (three log index values and two log exchange rates). We standardize the total portfolio value V_t in sterling to be one

$$V_t = 0.3 \cdot x_1 + 0.4 \cdot (x_2 + x_4) + 0.3 \cdot (x_3 + x_5)$$

where x_1, x_2 and x_3 represent log-returns on the three indexes and x_4 and x_5 are log-returns on the GBP/USD and GBP/CHF exchanges rates, respectively.

The return series show little serial correlation, to the point that it is safe to treat returns as being i.i.d.

The VaR estimation methods considered are,

- VC: The standard unconditional variance-covariance method assuming multivariate Gaussian risk-factor changes.
- HS: The standard unconditional historical simulation method.
- VC-t: The standard unconditional variance-covariance method assuming multivariate Student-t risk-factor changes.
- HS-GARCH: A conditional version of the historical simulation method in which GARCH(1,1) models with a constant conditional mean term and Gaussian innovations are fitted to the historically simulated losses to estimate the volatility of the next day's loss.
- VC-MGARCH: A conditional version of the variance-covariance method in which a multivariate GARCH model (a first-order constant conditional correlation model) with multivariate Normal innovations is used to estimate the conditional covariance matrix of the next day's risk-factor changes.
- HS-EWMA: A conditional method, similar to HS-GARCH, in which the EWMA method is used to estimate the conditional covariance matrix of the next day's risk- factor changes.
- VC-EWMA: A similar method to VC-MGARCH but a multivariate version of the EWMA method is used to estimate the conditional covariance matrix of the next day's risk-factor changes.

- HS-GARCH-t: A similar method to HS-GARCH but Student-t innovations are assumed in the GARCH model.
- VC-MGARCH-t: A similar method to VC-MGARCH but multivariate Student-t innovations are used in the MGARCH model.
- HS-CONDEVT: A conditional method using a combination of GARCH modeling and EVT (extreme value theory).

The characteristics of the distortion of α allows for efficiently estimating the VaR quantile from a short amount of data to capture the clusters in the data. This is relevant because extreme returns appear in clusters (McNeil et al., 2005) and if we use the i.i.d. model with long windows we will be likely to underestimate risk. VaR estimates would then change very slowly, being unable to capture changes that may occur in the market as soon as they happen.

On the contrary, calculation of the probability-unbiased VaR estimator from short rolling windows under the i.i.d. approach has some advantages (Francioni and Herzog, 2012): i) only a few data points are needed to obtain a very good VaR estimate and ii) this approach outperforms other alternatives that need many data points to calibrate the model, e.g. EWMA, GARCH, ... (at least 1 000 data are necessary to calibrate the models, McNeil et al., 2005). In fact, we have confirmed in previous sections results by Francioni and Herzog showing that the standard plug-in VaR estimates of a Normal population is biased. We have also described their suggestion to distort the significance level α so that the resulting VaR estimate is unbiased, and we have extended their results to other probability distributions.

We now calculate α_{pu} values to obtain the *probability-unbiased* VaR estimator in a rolling window of size n. We start with the simpler case of the Normal distribution, a member of the location-scale family. Under Normality there is no dependence on any additional parameter and we can use α_{pu} values provided in Table 1 (calculated by closed-form). The Student-t has the number of degrees of freedom as an additional parameter, which we estimate by Maximum Likelihood first from portfolio return data to subsequently calculate α_{pu} values[4]. A similar procedure is followed for mixtures of two Normals, starting with the GMM estimation

4. The GMM parameter estimates for returns from this portfolio are $\mu_1 = -0.0020$, $\mu_2 = -0.0011$, $\sigma_1 = -0.0020$, $\sigma_2 = -0.0214$ and $p = 0.2487$, very different from those used in the simulation exercise.

of the five additional parameters, μ_1, μ_2, σ_1, σ_2, and p, using portfolio returns for then calculate α_{pu} values. We follow the algorithm proposed by FH previously presented in the simulation exercises to obtain α_{pu} values needed to calculate *probability-unbiased* VaR in each window of size n. We propose mixtures of Normals as a more realistic distributions to further improve the results obtained under the Normal distribution.

Table 10 shows the number of annual VaR exceedances for the i.i.d. rolling window 5% VaR estimator under Normal and Student-t distributions, for different sample sizes. Every year, that number is relatively close to its expected value of 13 (aprox. 5% 260 days)[5]. The results in Table 10 suggest that under a Student-t distribution, windows with n = 20, 25 and 50 data points are generally outperformed by the shorter n = 15 window. In particular, for 1993, 1994, 2000 and 2001 VaR is poorly estimated under the Student-t distribution, with too many violations of the 95% VaR estimates. In general, 2000, 2001 and 2002 were the most difficult years to use in prediction for most models, since returns became very volatile, with many extreme losses. In the case of the Normal distribution, n = 15 and n = 25 perform better than n = 50.

For the mixture of two Normal distributions, the window with n = 100 outperforms longer window sizes especially during 1996, 1997 and 1998. We again work with samples larger than in the case of Normal and Student-t distributions for reasons explained above.

When comparing the performance of the different models for VaR estimation we look at estimates from 1996 through 2003, because that is the time period considered by McNeil et al. (2005) with whom we want to compare our results. We use rolling windows and calculate *probability-unbiased* VaR estimator in each window. On the contrary, the models considered by McNeil et al. (2005) consider the full period (1996-2003) to calculate *plug-in* VaR estimates. For the performance analysis, two different quantities are calculated: i) the overall exceedance probability, defined as the number of observed exceedances in the period divided by the number of data points, ii) the Observed Absolute Deviation per year (OAD), used by McNeil et al. (2005), which was introduced by Francioni and Herzog as the mean of the absolute difference between the expected number of exceedances (i.e. 3 for 1% VaR and 13 for 5% VaR) and the number of observed exceedances.

5. Except for 1992, when we lost n data observations due to the rolling window.

Table 10: Number of 5% VaR exceedances per year for the i.i.d. rolling window model with window length n. Absolute differences between the expected number of exceedances (13 per year) and the number of observed exceedances is reported in parentheses

Normal	Year											
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
n = 10	5 (8)	5 (8)	8 (5)	8 (5)	8 (5)	6 (7)	5 (8)	4 (9)	9 (4)	3 (10)	3 (10)	3 (10)
n = 15	7 (6)	7 (6)	8 (5)	11 (2)	14 (1)	9 (4)	11 (2)	7 (6)	12 (1)	7 (6)	8 (5)	6 (7)
n = 20	7 (6)	14 (1)	11 (2)	8 (5)	12 (1)	10 (3)	15 (2)	9 (4)	14 (1)	14 (1)	6 (7)	8 (5)
n = 25	7 (6)	18 (5)	12 (1)	10 (3)	13 (0)	10 (3)	13 (0)	12 (1)	13 (0)	13 (0)	9 (4)	6 (7)
n = 50	7 (6)	13 (0)	17 (4)	11 (2)	15 (2)	12 (1)	14 (1)	11 (2)	15 (2)	18 (5)	11 (2)	9 (4)
Student-t	Year											
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
n = 10	8 (5)	12 (1)	10 (3)	13 (0)	13 (0)	9 (4)	10 (3)	14 (1)	17 (4)	10 (3)	11 (2)	10 (3)
n = 15	8 (5)	15 (2)	12 (1)	12 (1)	14 (2)	10 (3)	15 (2)	12 (1)	15 (2)	16 (3)	13 (0)	10 (3)
n = 20	9 (4)	19 (6)	16 (3)	10 (3)	15 (2)	14 (1)	17 (4)	12 (1)	17 (4)	18 (5)	14 (1)	8 (5)
n = 25	10 (3)	19 (6)	15 (2)	11 (2)	16 (3)	13 (0)	18 (5)	13 (0)	17 (4)	18 (5)	12 (1)	9 (4)
n = 50	9 (4)	13 (0)	18 (5)	11 (2)	15 (2)	13 (0)	16 (3)	12 (1)	17 (4)	19 (6)	12 (1)	9 (4)
Mixture	Year											
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
n = 100	7 (6)	13 (0)	15 (2)	9 (4)	15 (2)	14 (1)	14 (1)	8 (5)	19 (6)	15 (2)	19 (6)	8 (5)
n = 200	1 (12)	12 (1)	15 (2)	5 (8)	17 (4)	16 (3)	19 (6)	6 (7)	14 (1)	15 (2)	16 (3)	2 (11)
n = 300	0 (13)	7 (6)	20 (7)	4 (9)	22 (9)	23 (10)	19 (6)	4 (9)	17 (4)	16 (3)	20 (7)	3 (10)
n = 400	0 (13)	4 (9)	19 (6)	6 (7)	21 (8)	26 (13)	20 (7)	5 (8)	13 (0)	19 (6)	22 (9)	4 (9)

Table 11 clearly shows that the i.i.d. model with rolling window outperforms the other models with respect to the overall exceedance probability and OAD[6]. For the 1% probability-unbiased VaR, the Normal model

6. Values for the overall exceedance probability and OAD for standard models in the lower half of Tables 11 and 12 are taken from McNeil et al. (2005).

with window of length 25, the Student-t model with window of length 15 and the mixture with window of length 100 show the best overall probabilistic properties. At 5% significance, Table 12 shows that the i.i.d. Student-t model with rolling window of size 15 and the Normal model with rolling window of size 50 outperform the other models with respect to OAD and to the overall exceedance probability, respectively. The mixture with a window of length 100 is the best within models with mixture distribution and outperforms in VaR estimation many of the methods proposed by McNeil et al. (2005).

Table II: Historical 1% VaR exceedance probabilities of the various models and historical Observed Absolute Deviation (OAD) per year

Model	Exc. Prob. (%)	OAD
N i.i.d. n = 15	0.16	2.50
N i.i.d. n = 25	0.87	1.33
N i.i.d. n = 50	1.43	1.42
ST i.i.d. n = 15	0.83	1.25
ST i.i.d. n = 25	1.32	1.33
ST i.i.d. n = 50	1.65	1.50
NM i.i.d. n = 100	1.02	1.33
NM i.i.d. n = 200	1.29	2.00
NM i.i.d. n = 300	1.38	2.42
NM i.i.d. n = 400	1.43	2.33
VC	3.03	5.55
HS	2.02	3.00
VC-t	2.35	3.87
HS-GARCH	2.26	2.87
VC-MGARCH	2.31	2.87
HS-EWMA	2.07	2.62
VC-EWMA	2.02	2.62
HS-GARCH-t	1.68	1.62
VC-MGARCH-t	1.44	3.12
HS-CONDEVT	1.35	1.25

Table 12: Historical 5% VaR exceedance probabilities of the various models and historical Observed Absolute Deviation (OAD) per year

Model	Exc. Prob. (%)	OAD
N i.i.d. n = 15	3.43	4.16
N i.i.d. n = 25	4.38	2.41
N i.i.d. n = 50	4.97	2.42
ST i.i.d. n = 15	4.87	1.92
ST i.i.d. n = 25	5.5	2.83
ST i.i.d. n = 50	5.32	2.50
NM i.i.d. n = 100	5.15	2.92
NM i.i.d. n = 200	4.71	4.17
NM i.i.d. n = 300	5.48	6.50
NM i.i.d. n = 400	5.82	6.25
VC	7.36	7.88
HS	7.65	8.12
VC-t	8.46	10.25
HS-GARCH	6.11	3.38
VC-MGARCH	6.64	4.50
HS-EWMA	6.2	3.62
VC-EWMA	5.92	3.38
HS-GARCH-t	6.34	3.75
VC-MGARCH-t	6.93	5.50
HS-CONDEVT	5.77	2.75

In general, the i.i.d. approach with short windows incorporates too little information to produce good VaR estimates, whereas with large windows VaR estimates are too static and they do not adapt to new information fast enough. However, the i.i.d. approach to compute the 1% and 5% probability-unbiased VaR outperforms the alternative models considered by McNeil et al., which are more complex and use the *plug-in* VaR estimator. Figures 7 and 8 show a plot of the probability-unbiased VaR estimates of the i.i.d. Student-t model with a rolling window size of 20 and 200 data points, respectively. The rolling window with 20 data points clearly reacts extremely fast to new data, with occasional large jumps in the VaR estimate.

Figure 7: Portfolio log-returns from 1992 to 2003 and i.i.d. VaR estimates for different α based on a Student-t rolling window model with a window length of 20 observations

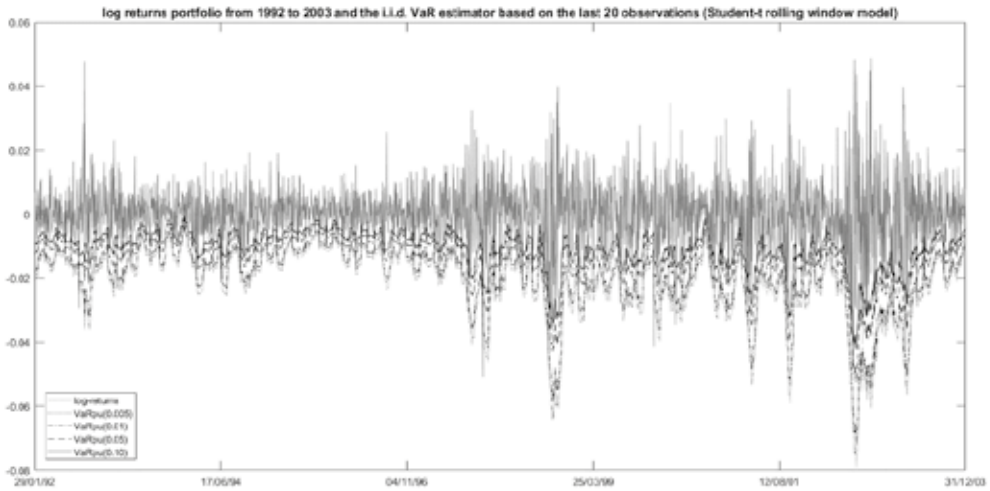
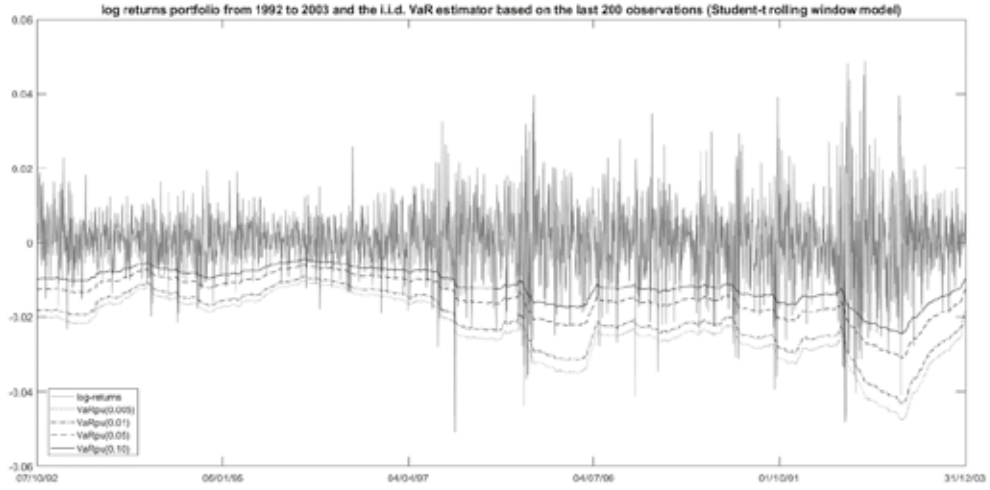


Figure 8: Portfolio log-returns from 1992 to 2003 and i.i.d. VaR estimates for different α based on a Student-t rolling window model with window length of 200 observations



1.9. Conclusions

Francioni and Herzog (2012) (FH) proposed a standard resampling bootstrap algorithm to estimate a probability unbiased VaR in the case of Normal returns. The main idea is to replace the level α by a suitable chosen level α_{pu} which minimizes the averaged distance of the bootstrapped estimators to α. In other words, their strategy consisted on modifying the desired significance level α to obtain α_{pu} in such a way that we obtain an unbiased estimate of VaR at the original significance level. Their analysis suggested that VaR estimates based on short samples may have a good performance for Normal distributions, often beating standard VaR estimates based on long samples.

We explore the properties of the *probability-unbiased* VaR proposed by FH as an interesting alternative to *plug-in* VaR when working with short samples and a small significance level α. It is then when the *probability-unbiased* VaR differs more from *plug-in* VaR. We extend work by FH to Student-t distributions and mixtures of Normal distributions. Our results suggest that for a variety of distributions the *plug-in* VaR estimator underestimates risk for a given range of probabilities (α) when estimated from short samples. The smaller the sample size, the greater the underestimation of risk by the *plug-in* VaR estimator. The range of probabilities for which *plug-in* VaR underestimates risk depends on the sample size and on the assumed probability distribution for returns. In all these cases the *probability-unbiased* VaR performs better.

In the Gaussian case we can use the parametric approach to estimate VaR in closed form. For other cases we use an appropriate bootstrapping algorithm suggested by FH. We show that the performance of the *probability-unbiased* estimators for small sample sizes is surprisingly good also for Student-t distributions as well as for mixtures of Normals. The reason is that the shorter the period, the more uniform will be the sample. Besides, the conditional volatility will not change much over a short sample, making the sample almost i.i.d.. The difference between α_{pu} and α is larger for a Normal sample than for a Student-t distribution. For a Mixture of Normals, the difference depends on the mixing parameters.

We also estimate *probability-unbiased* confidence intervals for the VaR estimator. For the three distributions (Normal, Student-t and mixture of

Normals) the *probability-unbiased* confidence interval is shifted to the left, relative to the standard confidence interval calculated using the *plug-in* VaR estimator. Once again, the leftward shift of the *probability-unbiased* confidence interval is due to the fact that most simulated VaR values fall to the left of the *VaR* estimate. Hence, a symmetric confidence interval would not be appropriate. The findings in this chapter suggest that the unbiased VaR estimator is a valuable tool for the practice of risk management.

CHAPTER 2. VOLATILITY SPECIFICATIONS VERSUS PROBABILITY DISTRIBUTIONS IN VAR ESTIMATION

2.1. Introduction

A traditional discussion in risk measurement analysis has been whether volatility models that incorporate a leverage effect, with negative innovations having a larger impact on volatility than positive innovations of the same size, lead to better Value at Risk (VaR) forecasts. A second modeling issue refers to whether asymmetric probability distributions for return innovations lead to an improved VaR model[7].

The goal of this chapter is to examine the relative importance of the two issues, the volatility specification and the assumption on the probability distribution of return innovations, for the efficiency of VaR forecasts. The question is crucial for risk managers, since there are so many potential choices for volatility model and probability distributions that it would be very convenient to establish some priorities in modeling returns for risk estimation. To that end, we have performed an extensive analysis of VaR forecasts in assets of different nature, using symmetric and asymmetric probability distributions for the innovations on volatility models with and without leverage. Additionally, we want to make some progress in characterizing the more appropriate probability distributions and volatility specifications to be used for innovations in financial returns.

We consider three general volatility specifications with leverage, GJR-GARCH, APARCH and FGARCH as well as the standard symmetric GARCH model as benchmark. The FGARCH model includes as special cases many other volatility specifications, like the symmetric GARCH, GJR-GARCH and APARCH. It is, in fact, a nested family of GARCH- type

7. Along the chapter we refer to a VaR forecasting model as a combination of a probability distribution and a volatility specification for return innovations.

models, thereby allowing for testing how simpler models fit the data. Besides, the FGARCH and APARCH models take the power on the conditional standard deviation of the innovations as a free parameter. That way, they provide more flexibility to the dynamics of volatility, allowing for shifts and rotations in the news impact curve. The two types of asymmetry in volatility produced by shifts and rotations are distinct, and they should not be treated as substitutes for each other (Hentschel, 1995).

As probability distributions for the innovations we compare the performance of the skewed Student-t distribution and skewed Generalized Error distribution as introduced in Fernandez and Steel (1998), the unbounded Johnson S_U distribution, skewed Generalized-t distribution (Theodossiou, 1998) and Generalized Hyperbolic skew Student-t distribution (Aas and Haff, 2006), with the Normal and symmetric Student-t distributions as benchmark. An interesting feature of our work is the consideration of a variety of assets of different nature: stock market indexes, individual stocks, interest rates, commodity prices and exchange rates.

A novel approach of our analysis is to use standard statistical tests to examine the extent to which the estimated probability distributions fit the distribution of empirical return innovations. Additionally, each estimated combination of volatility specification and probability distribution for return innovations determines the distribution of returns themselves. We use simulation methods to analyze whether our estimated models fit the main characteristics of return distributions. These should be expected to be two natural conditions for the good VaR performance of a model. But, in spite of the fact that significant effort is generally placed in selecting an appropriate probability distribution and volatility model, the ability of estimates to explain sample moments is seldom examined.

We calculate VaR forecasts following the parametric approach. An AR(1) was estimated for daily returns in all cases. The performance of VaR forecasts is examined through standard tests: the unconditional coverage test of Kupiec (1995), the independence and conditional coverage tests of Christoffersen (1998), the Dynamic Quantile test of Engle and Manganelli (2004), as well as the evaluation of the Asymmetric Linear Tick loss function (AlTick) proposed by Giacomini and Komunjer (2005).

The combination of 19 assets, 7 probability distributions, 4 volatility specifications and 4 backtests of VaR leads to an extensive set of results that need to be summarized in a search for some consistent conclusions. One of the contributions of this chapter is to follow a diverse strategy to examine test results in search of robust patterns that might suggest some preferred VaR model specifications. We proceed along several lines: *i)* comparing the number of realized and theoretical violations of VaR for the alternative models across the set of assets, *ii)* comparing the p-values of VaR tests over the alternative models, *iii*) applying a Dominance criterion we introduce in this chapter to the alternative VaR models considered, *iv*) following the Model Confidence Set approach to select the most preferred models. Such multiple strategy for summarizing the information allows us to draw some clear-cut conclusions on the benefits of the alternative models.

Our results suggest that the important assumption for VaR performance is that of the probability distribution of the innovations, with the choice of volatility model playing a secondary role. Indeed, validation tests for VaR forecasts yield very similar results for a given probability distribution as we change the volatility model. On the contrary, test results drastically change for a given volatility model when we change the assumption on the probability distribution of the innovations. In fact, the main difference arises when we move from symmetric to asymmetric probability distributions for the innovations, a result consistent with work by Gerlach et al. (2011) and Dendramis et al. (2014), among others. The unbounded Johnson distribution, the skew Generalized-t distribution and the skewed Generalized Error distributions seem to dominate other asymmetric distributions, like the skewed Student-t and the Generalized Hyperbolic skewed Student-t. Symmetric distributions are clearly inappropriate. Furthermore, FGARCH and APARCH volatility specifications dominate other alternatives. Indeed, our results suggest that the standard deviation, rather than the variance, should often be used to model volatility dynamics.

Relative to the ability to reproduce sample moments, different volatility models with the same probability distribution for the innovations fit sample moments similarly. On the other hand, while it is obviously true that asymmetric distributions are needed to explain the skewness in returns, symmetric and asymmetric probability distributions, imposed

on the same volatility model lead to minor differences in kurtosis. The ability of estimated models to fit the empirical distributions of returns and returns innovations seems in fact, a necessary condition for a good VaR performance.

2.2. A review of literature

Among parametric methods for VaR estimation, some authors have analyzed the improvement on VaR estimation provided by volatility models with leverage. Giot and Laurent (2003a) estimated daily VaR for stock indexes using different volatility models. They stated that more complex models like APARCH performed better than RiskMetrics or GARCH specifications (for a comparison of volatility models in VaR estimation see also El Babsiri and Zakoian, 2001). Angelidis et al. (2004) show that volatility models with leverage fare better than symmetric specifications, as they capture more efficiently the characteristics of the underlying series and provide better VaR forecasts since they perform better in the low probability regions that VaR tries to measure (see also Ane, 2006). McMillan and Kambouroudis (2009) provide evidence on the performance of alternative VaR models for a large number of individual stocks and exchange rates. They conclude that the APARCH model should be preferred for more extreme VaR forecasts, while the RiskMetrics model seems to be adequate at more moderate significance levels. In their work, RiskMetrics seems adequate in providing volatility forecasts for most Asian markets; however, the APARCH model is superior in obtaining forecasts for the G7 markets, as well as for other European markets and for the larger Asia markets.

Given the widespread evidence on the skewness of the distribution of asset returns, analyzing whether the assumption of an asymmetric distribution of return innovations leads to more efficient VaR forecasts is a second methodological issue of interest. Based on the influence of leverage effects on the accuracy of VaR forecasts, Brooks and Persand (2003) concluded that models which do not allow for asymmetries either in the unconditional distribution of returns or in the volatility specification underestimate the true VaR. Giot and Laurent (2003a) used daily data for stock market indexes and individual stocks, showing that models that rely on a symmetric density for return innovations underperform with

respect to skewed density models that require modeling both the left and right tails of the distribution of returns. Lee and Su (2015) estimate VaR for eight stock market indexes from Europe and Asia by a parametric GARCH approach as well as by the semi-parametric approach of Hull and White. The only asymmetric distribution they consider, the skewed Generalized-t, is shown to have a better VaR forecasting performance than the Student-t, with the Normal distribution being the last in the ranking, according to the unconditional coverage test of Kupiec and two different loss functions.

Corlu et al. (2016) investigate the ability of five alternative probability distributions to represent the behavior of daily equity index returns over the period 1979-2014: the skewed Student-t distribution, the generalized lambda distribution, the Johnson system of distributions, the normal inverse Gaussian distribution, and the g-and-h distribution. The explanatory power of the alternative distributions is tested using in-sample Value at Risk (VaR) failure rates. Their focus is on the unconditional distribution of equity returns, not on conditional distributions. They find that the generalized lambda distribution is a prominent alternative for modeling the behavior of daily equity index returns.

More recently, some papers have jointly examined the performance of both, the variance specification and the probability distribution of return innovations in VaR estimation. Gerlach et al. (2011) examine the performance of a wide class of volatility models: RiskMetrics, asymmetric GARCH, IGARCH, GJR-GARCH and EGARCH, under four alternative probability distributions: Gaussian, Student-t, Generalized Error Distribution and skewed Student-t in VaR forecasting at 1% and 5% significance in different time periods (pre-crisis, crisis-GFC and post-crisis) incorporating parameter uncertainty through a Bayesian approach. Results are varied and hard to summarize, but their evidence suggests a preference for asymmetric probability distributions for the innovations of the return process. Giot and Laurent (2003b) analyze daily returns on commodities fitting ARCH and APARCH models under a skewed Student-t probability distribution for the innovations, and using Riskmetrics as a benchmark. While the skewed Student-t APARCH model performs best in all cases, it is unclear whether the forecasting gain is enough to dominate over the computationally simpler skewed Student-t ARCH model. Bubak (2008), Tu et al. (2008), Kang and Yoon (2009) and

Diamandis et al. (2011), analyze Eastern and Central European stock markets, Asian stock markets, Asian emerging markets and developed and emerging markets, respectively. Comparing a wide range of univariate conditional variance models, they show that models that incorporate an asymmetric distribution for return innovations tend to perform better than models with a symmetric distribution, in terms of both in-sample and out-of-sample (one-day-ahead) VaR forecasts. Dendramis et al. (2014) show that the VaR performance of alternative parametric models like EGARCH or the Markov regime-switching model is enhanced when combined with asymmetric probability distributions for return innovations. Tang and Shieh (2006) and Mabrouk and Saadi (2012) include Fractionally Integrated time varying GARCH models designed to capture not only volatility clustering, but also long memory in asset return volatility. Both papers consider three probability distributions, Normal, Student-t and skew Student-t. Tang and Shieh (2006) consider FIGARCH and HYGARCH (Hyperbolic GARCH) models, showing that for the three stock index futures considered, HYGARCH models with skewed Student-t distribution perform better based on the Kupiec LR tests. Mabrouk and Saadi (2012) conclude that the skewed Student-t FIAPARCH model outperforms the alternative GARCH and HYGARCH models because it can simultaneously account for fat tails, asymmetry, volatility clustering and long memory. However, given that the VaR forecasts required by the Basel accords are short run, the inclusion of long-memory is expected not to make any fundamental difference [see for example So and Yu (2006)]. So, the need to consider asymmetric probability distributions for return innovations seems to be well established at this point. Recently, Leccadito et al. (2014) have compared the performance of a variety of volatility specifications and asymmetric distributions using multilevel VaR tests that apply independence and conditional coverage tests at different confidence levels.

As in the latter group of papers, we also examined the performance of both, the variance specification and the probability distribution of return innovations in VaR estimation. We consider a complex and flexible volatility model proposed by Hentschel (1995), FGARCH, which is an omnibus model that subsumes some of the most popular GARCH models. To the best of our knowledge, there are no papers examining the performance of this model for VaR forecasting. Besides, we introduce

distributions rarely used in the literature on VaR performance, such as the skewed Generalized Error Distribution [Fernandez and Steel (1998)], Johnson SU distribution [Johnson (1949)], skewed Generalized-t [Theodossiou (1998)] and Generalized Hyperbolic skew Student-t distribution (GHST) [Aas and Haff (2006)]. In the VaR literature, Johnson distributions are suggested in Zangari (1996), Mina and Ulmer (1999), in RiskMetrics Technical Document (1996) and Choi (2001), which examines empirically a GARCH model with Johnson innovations. Simonato (2011) documents the performance of the Johnson system relative to closely competing approaches, such as the Gram-Charlier and Cornish-Fisher approximations. He considers the case of Expected Shortfall computation without performing a backtesting analysis, just comparing the moments of the distributions and root-mean-squared errors. The GHST distribution has hardly been employed in financial applications because its estimation is computationally demanding. Nakajima and Omi (2012) use GHST distribution to perform a Bayesian analysis of a stochastic volatility model. Among multivariate applications, Hu (2005) Multivariate Generalized Hyperbolic Distribution using the EM algorithm. Paolella and Polak (2015) also use the Generalized Hyperbolic distribution in a context of multivariate time series.

Relative to this ever increasing literature, we contribute in different ways: *i*) considering a set of probability distributions that have recently been suggested to be appropriate for capturing the skewness and kurtosis of financial data, but whose performance for VaR estimation has not been compared yet on a common dataset, *ii*) considering the APARCH and FGARCH volatility models with leverage that have also been recognized as being adequate for financial returns, *iii*) applying existing backtesting procedures for the different VaR models to a wide array of assets of different nature, *iv*) comparing the relevance of the assumed probability distribution for return innovations and the volatility specification for VaR performance, *v*) introducing a dominance criterion to establish a ranking of models on the basis of their behavior under standard VaR validation tests, *vi*) using the dominance criterion and the Model Confidence Set approach to search for robust conclusions on the preference of some probability distributions and volatility specifications.

2.3. Volatility models and probability distributions

Let x_i, for $t = 1, \ldots, T$, be a time series of asset returns. It is convenient to break down the complete characterization of x_i into three components: *i)* the conditional mean, μ_i, *ii)* the conditional variance, which contains a scale parameter that measures the dispersion of the distribution, σ_i^2 and *iii)* the shape parameters, which determine the form of a conditional distribution (e.g., skewness, kurtosis) within a general family of distributions. Thus, we may write

$$x_t = \mu_t(\theta) + \varepsilon_t \qquad \mu_t(\theta) = \mathbb{E}[x_t|\mathcal{F}_{t-1}] = \mu(\theta, \mathcal{F}_{t-1}) \qquad \varepsilon_t = \sigma_t(\theta) z_t$$
$$\sigma_t^2(\theta) = \mathbb{E}[(x_t - \mu_t)^2|\mathcal{F}_{t-1}] = \sigma^2(\theta, \mathcal{F}_{t-1}) \qquad z_t \sim f(z_t|\theta)$$

The standardized innovation, $z_t = \big(x_t - \mu_t(\theta)\big)/\sigma_t(\theta)$ has zero mean and a unit variance. It follows a conditional distribution f with shape parameters that capture the possible asymmetry and fat-tailedness of returns, except in the case of the Normal distribution. Vector θ contains all the parameters associated with the conditional mean and variance and the conditional distribution.

An AR(1) model for the conditional mean return is sufficient to produce serially uncorrelated innovations for all assets. We consider three general volatility models with leverage, GJR-GARCH, APARCH and FGARCH with a standard symmetric GARCH model as benchmark. As probability distributions for the innovations we compare the performance of skewed Student-t, skewed Generalized Error, unbounded Johnson S_U, skewed Generalized-t and Generalized Hyperbolic skew Student-t distributions, with the Normal and symmetric Student-t distributions as benchmark.

In all models we jointly estimate by maximum likelihood the parameters in the equation for the mean return, the equation for its conditional variance and the probability distribution for the return innovations. The exception is the skewed Generalized-t distribution, for which we use a two-step estimation method because of the numerical difficulty of estimating all parameters jointly[8].

8. In that case, we first estimated the AR(1)-GARCH conditional mean-volatility model assuming a Generalized Error distribution (GED) for the innovations, as suggested by Bali and Theodossiou (2007). The parameters of the skewed Generalized-t distribution (SGT) were estimated in a second stage using the standardized returns $\left(\frac{r_t - \phi_0 - \phi_1 r_{t-1}}{\sigma_t} = \frac{\varepsilon_t}{\sigma_t}\right)$ obtained in the first step.

2.3.1. Volatility models

The conditional variance of GARCH(p,q) model (Bollerslev, 1986) is used as a benchmark, i.e.

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

where $\omega > 0, \alpha_i, \beta_j \geq 0, \sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_j < 1$.

The standard GARCH model captures the existence of volatility clustering but is unable to express the leverage effect, since it assumes that positive and negative error terms have the same effect on volatility. To incorporate asymmetric effects on volatility from positive and negative surprises, Glosten, Jagannathan and Runkle (1993) proposed a GJR-GARCH(p,q) model, adding the negative impact of leverage in the conditional variance equation. This model incorporates positive and negative shocks on the conditional variance asymmetrically via the use of the indicator function $I(\varepsilon_{t-i} \leq 0)$, so that the variance equation becomes,

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} [\alpha_i \varepsilon_{t-i}^2 + \gamma_i I(\varepsilon_{t-i} \leq 0) \varepsilon_{t-i}^2] + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

The volatility effect of a unit negative shock is $\alpha_i + \gamma_i$ while the effect of a unit positive shock is α_i. A positive value of γ_i indicates that a negative innovation generates greater volatility than a positive innovation of equal size, and on the contrary for a negative value of γ_i.

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

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

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

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

The FGARCH model (Family GARCH) of Hentschel (1995) is an omnibus model which subsumes some of the most popular GARCH models. It is similar to the APARCH model, but more general, since it allows the decomposition of the residuals in the conditional variance equation to be driven by different powers for z_t and σ_t. It also allows for both shifts and rotations in the news impact curve, where the shift is the main source of asymmetry for small shocks while rotation drives the asymmetry for large shocks.

$$\sigma_t^{\lambda} = \omega + \sum_{i=1}^{q} \alpha_i \sigma_{t-1}^{\lambda} f^{\delta}(z_{t-i}) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^{\lambda}$$

where $f^{\delta}(z_{t-i}) = (|z_{t-i} - \eta_{2i}| - \eta_{1i}(z_{t-i} - \eta_{2i}))^{\delta}$.

Positivity of $f^{\delta}(z_{t-i})$ is guaranteed when $|\eta_1| \leq 1$, which ensures that neither arm of the rotated absolute value function crosses the abscissa. The parameter η_2, however, is unrestricted in size and sign. The magnitude and direction of a shift in the news impact curve are controlled by the parameter η_2 while the magnitude and direction of a rotation in the news impact curve are controlled by the parameter η_1. Other GARCH models only permit either a shift or a rotation, but not both. Allowing for shifts in the news impact curve, the FGARCH model is more flexible than previous models, being able to capture asymmetries in volatility even in the presence of small shocks.

2.3.2. Probability distributions

To account for the excess skewness and kurtosis typical of financial data, the parametric volatility models presented in the previous section can be combined with skewed and leptokurtic distributions for return innovations. The skewed Student-t by Fernandez and Steel and Lambert and Laurent (2001)[9] is

$$f(z|\xi,\nu)=\frac{2}{\xi+\frac{1}{\xi}}s\{g[\xi(sz+m)|\nu]I_{(-\infty,0)}(z+m/s)+g[(sz+m)/\xi|\nu]I_{[0,\infty)}(z+m/s)\} \quad (5)$$

where $g(\cdot\,|\nu)$ is the symmetric (unit variance) Student-t density and ξ is the skewness parameter[10]; m and s^2 are, respectively the mean and the variance of the non-standardized skewed Student-t and are defined as,

$$\mathbb{E}(\varepsilon|\xi)=M_1(\xi-\xi^{-1})\equiv m$$
$$\mathbb{V}(\varepsilon|\xi)=(M_2-M_1^2)(\xi^2+\xi^{-2})+2M_1^2-M_2\equiv s^2$$

where $M_r=2\int_0^\infty s^r g(s)ds$ is the absolute moments generating function. Note that when $\xi=1$ and $\nu=+\infty$ we get the skewness and the kurtosis of the Gaussian density. When $\xi=1$ and $\nu>2$ we have the skewness and the kurtosis of the (standardized) Student-t distribution.

An alternative distribution for return innovations which can capture skewness and kurtosis can be based on the Generalized Error Distribution (GED) by Nelson (1991). According to Lambert and Laurent the innovation process z_t is said to follow a (standardized) skewed Generalized error distribution, $SGED(0,1,\xi,\kappa)$, if

$$f(z|\xi,\kappa)=\frac{2}{\xi+\frac{1}{\xi}}s\{g[\xi(sz+m)|\kappa]I_{(-\infty,0)}(z+m/s)+g[(sz+m)/\xi|\kappa]I_{[0,\infty)}(z+m/s)\}$$

9. Lambert and Laurent (2001) and Giot and Laurent (2003a) have shown that for various financial daily returns, it is realistic to assume that standardized innovations $\hat{z}_t$ follows a skewed Student-t distribution.

10. The skewness parameter $\xi>0$ is defined such that the ratio of probability masses above and below the mean is

$$\frac{Prob(z\geq 0|\xi)}{Prob(z<0|\xi)}=\xi^2$$

where $g(\cdot|\kappa)$ is the symmetric (unit variance) Generalized Error distribution, ξ is the skewness parameter, κ representing the shape parameter and $\Gamma(\cdot)$ is the gamma function. Mean (m) and standard deviation (s) are calculated in the same way as in the case of skewed Student-t distribution. As κ increases the density gets flatter and flatter while in the limit, as $\kappa \to \infty$, the distribution tends toward the uniform distribution. Special cases are the Normal when $\kappa = 2$ and the Laplace distribution when $\kappa = 1$. For $\kappa > 2$ the distribution is platykurtic and for $\kappa < 2$ it is leptokurtic.

Another alternative is the Johnson S_U distribution. It was one of the distributions derived by Johnson (1949) based on translating the Normal distribution by certain functions. Letting $Z \sim N(0,1)$, the standard Normal distribution, the random variable Y has the Johnson system of frequency curves if it is a transformation of Z by $Z = \gamma + \delta g((Y - \xi)/\lambda)$. The form of the resulting distribution depends on the choice of function g. When $g(u) = sinh^{-1}(u)$, the distribution is unbounded, called the Johnson S_U distribution. The parameters of the distribution are ξ, $\lambda > 0$, γ, $\delta > 0$.

We use a parametrization[11] of the original Johnson S_U distribution, so that parameters ξ and λ are the mean and the standard deviation of the distribution. The parameter γ determines the skewness of the distribution with $\gamma > 0$ indicating positive skewness and $\gamma < 0$ negative skewness. The parameter δ determines the kurtosis of the distribution. δ should be positive and most likely in the region above 1.

The pdf of the Johnson's S_U, denoted here as $JSU(\xi, \lambda, \gamma, \delta)$, is defined by

$$f_Y(y) = \frac{\delta}{c\lambda} \frac{1}{\sqrt{(r^2+1)}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right]$$

11. This parametrization is used by R rugarch package, which we use for estimating the parameters of our models.

where

$$z = -\gamma + \delta sinh^{-1}(r) = -\gamma + \delta \log\left[r + (r^2+1)^{1/2}\right]$$

$$r = \frac{y - (\xi + c\lambda\omega^{\frac{1}{2}} sinh\Omega)}{c\lambda}$$

$$c = \left\{\frac{1}{2}(\omega - 1)[\omega cosh 2\Omega + 1]\right\}^{-1/2}$$

where $\omega = \exp(\delta^{-2})$ and $\Omega = -\gamma/\delta$. Note that $Z \sim N(0,1)$. Here $\mathbb{E}(Y) = \xi$ and $\mathbb{V}ar(Y) = \lambda^2$.

A very flexible distribution is the skewed Generalized-t distribution proposed by Theodossiou (1998). They developed a skewed version of the Generalized-t distribution introduced by McDonald and Newey (1988).

The skewed Generalized-t distribution has the probability density function

$$f(x|\mu,\sigma,\lambda,p,q) = \frac{p}{2\nu\sigma q^{\frac{1}{p}} B\left(\frac{1}{p},q\right)\left(\frac{|x-\mu+m|^p}{q(\nu\sigma)^p(\lambda sign(x-\mu+m)+1)^p}+1\right)^{\frac{1}{p}+q}}$$

where

$$m = \frac{2\nu\sigma\lambda q^{\frac{1}{p}} B\left(\frac{2}{p}, q-\frac{1}{p}\right)}{B\left(\frac{1}{p},q\right)}$$

$$\nu = q^{-\frac{1}{p}}\left[(3\lambda^2+1)\frac{B\left(\frac{3}{p},q-\frac{2}{p}\right)}{B\left(\frac{1}{p},q\right)} - 4\lambda^2\left(\frac{B\left(\frac{2}{p},q-\frac{1}{p}\right)}{B\left(\frac{1}{p},q\right)}\right)^2\right]^{-\frac{1}{2}}$$

where $B(\cdot)$ is the beta function, and μ, σ, λ, p and q are the location, scale, skewness, peakedness and tail-thickness parameters, respectively. Note that the parameters have the following restrictions $\sigma > 0$, $-1 < \lambda < 1$, $p > 0$ and $q > 0$. The skewness parameter λ controls the rate of descent of the density around $x = 0$. The parameters p and q control the height and tails of the density, respectively. The parameter q has the degrees of freedom interpretation in case $\lambda = 0$ and $p = 2$.

More complex and novel are the distributions belonging to the generalized hyperbolic family. An special case of this family is the Generalized Hyperbolic skew Student-t distribution proposed by Aas and Haff (2006). This distribution has the important property that one tail has polynomial and the other exponential behavior. Further, it is the only subclass of the Generalize Hyperbolic family of distribution having this property. This is an alternative for modeling the empirical distribution of financial returns. It is often skewed, having one heavy and one semiheavy or more Gaussian-like tail. The skew extensions to the Student-t distribution, like that of Fernandez and Steel, have two tails behaving as polynomials. This means that they fit heavy-tailed data well, but they do not handle substantial skewness, since that requires one heavy tail and one nonheavy tail.

The probability density function of the Generalized Hyperbolic skew Student-t is given by

$$f_X(x) = \frac{2^{\frac{1-\nu}{2}}\delta^{\nu}|\beta|^{\frac{\nu+1}{2}}\mathrm{K}_{\frac{\nu+1}{2}}\left(\sqrt{\beta^2(\delta^2+(x-\mu)^2)}\right)\exp\left(\beta(x-\mu)\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}\left(\sqrt{\delta^2+(x-\mu)^2}\right)^{\frac{\nu+1}{2}}} \qquad \beta \neq 0$$

and

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\delta\Gamma\left(\frac{\nu}{2}\right)}\left[1+\frac{(x-\mu)^2}{\delta^2}\right]^{-(\nu+1)/2} \qquad \beta = 0$$

where $\mathrm{K}_{\nu}(x) \sim \sqrt{\frac{\pi}{2x}}\exp(-x)$ for $x \to \pm\infty$ is the modified Bessel function (Abramowitz and Stegun, 1972), μ, δ, β and ν determine the location, scale, skew and shape parameters, respectively.

When $\beta = 0$ the density $f_x(x)$ can be recognized as that of noncentral Student-t distribution with v degrees of freedom, expectation μ and variance $\delta^2/(v-2)$.

2.4. The data

We work with daily percentage returns on five groups of assets of different nature over the sample period 1/4/2000-12/31/2015 (4173 observations). Daily returns are computed as 100 times the first difference of log prices, i.e. $100[ln(P_{t+1}) - ln(P_t)]$%. The financial assets considered are: stock market indexes: IBEX 35 (€), NASDAQ 100 ($), FTSE 100 (£) and NIKKEI 225 (¥); individual stocks: IBM ($), SAN (€), AXA (€) and BP (£); interest rates: IRS 5Y (€), interest rate of GERMAN BOND 10Y (€) and interest rate of US BOND 10Y($); commodity prices CRUDE OIL BRENT ($ per barrel), NATURAL GAS ($ per Million British Thermal Units), GOLD ($ per Troy Ounce) and SILVER (Cents $ per Troy Ounce) and exchange rates EUR/USD (€), GBP/USD (£), JPY/USD (¥) and AUD/USD (Australian $). The data were extracted from Datastream.

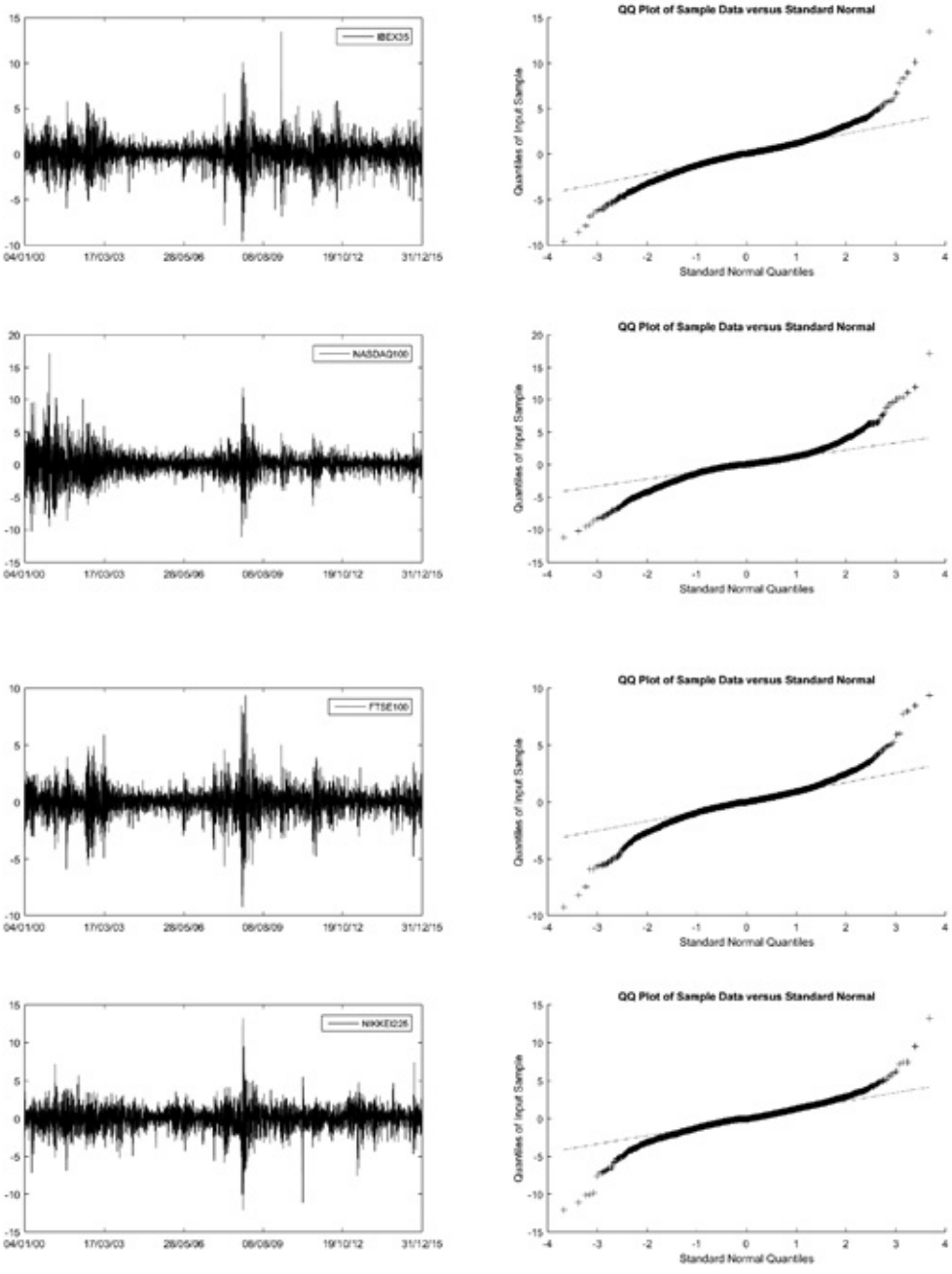
Table 13 reports descriptive statistics for daily returns. All the assets have mean and median returns close to zero. Returns on interest rates are obtained as log changes in the price of implicit zero coupon bonds having the value of an interest rate as a yield. In terms of standard deviation, the sample range is higher for AUD/USD (18.7), IRS (18.0) and US BOND (17.1) and lower for JPY/USD (13.2), EUR/USD (13.4), SILVER (13.8) and the interest rate on the GERMAN BOND (13.9). The unconditional standard deviation is relatively similar for assets in the same class, except for commodities, where GAS (4.19) and OIL BRENT (2.28) are more volatile than GOLD (1.13) and SILVER (1.93). NASDAQ is more volatile than other stock market indexes and AXA is the most volatile stock. The $US exchange rate for the Australian dollar has higher standard deviation than the one for the euro, British pound or Yen. AUD/USD, SILVER, GOLD and NIKKEI have significant negative skewness, while GAS, AXA, JPY/USD and NASDAQ have high positive skewness. For all the assets considered the kurtosis is high, implying that the return distributions have much thicker tails than the Normal distribution. Kurtosis is especially large for AUD/USD, GAS, IBM and AXA while EUR/USD, while the interest rate of the GERMAN BOND and the JPY/USD exchange rate have lower kurtosis. Together with a large sample size, these values for skewness and kurtosis lead to a very large Jarque-Bera statistic, rejecting the assumption of Normality in all cases.

Table 13: Descriptive statistics for daily percentage returns. Mean and median returns are in basis points. SD denotes the standard deviation, and J-B is the Jarque-Bera statistic to test for Normality. Sample: 01/04/2000-12/31/2015

	Mean (bps.)	Median (bps.)	Max	Min	S.D.	Skewness	Kurtosis	J-B
IBEX	-0.47	2.89	13.48	-9.58	1.49	0.08	7.93	4234.84
NASDAQ	0.46	3.68	17.20	-11.11	1.85	0.19	9.62	7652.53
FTSE	-0.25	0	9.38	-9.26	1.21	-0.16	9.36	7042.80
NIKKEI	0.01	0	13.23	-12.11	1.50	-0.41	9.72	7979.58
IBM	0.42	0	12.26	-16.89	1.66	-0.07	11.63	12947.74
SAN	1.01	0	20.87	-15.19	2.19	0.15	9.11	6515.50
AXA	0.55	0	19.78	-20.35	2.67	0.27	10.09	8790.79
BP	-1.35	0	10.58	-14.04	1.71	-0.13	7.81	4041.28
IRS	0.55	0.48	1.92	-1.86	0.21	-0.28	8.53	5367.17
GER BOND	1.11	0.97	3.39	-2.33	0.41	-0.09	5.97	1536.83
US BOND	0.98	0.96	4.53	-5.57	0.59	-0.22	7.96	4307.77
BRENT	0.98	0	17.97	-18.72	2.28	-0.19	8.26	4831.81
GAS	0.01	0	37.81	-28.90	4.19	0.56	12.81	16946.14
GOLD	3.10	0.01	6.86	-10.16	1.13	-0.41	8.81	5991.49
SILVER	2.26	0	13.66	-12.98	1.93	-0.57	8.62	5724.23
EUR/USD	0.16	0	4.62	-3.84	0.63	0.14	5.48	1091.11
GBP/USD	-0.20	0	4.43	-3.88	0.57	-0.04	7.27	3170.80
JPY/USD	-0.41	-0.99	4.61	-3.71	0.63	0.27	6.96	2779.74
AUD/USD	0.23	1.86	6.70	-8.83	0.83	-0.82	15.13	26058.43

Figure 9 displays daily percentage returns of each stock market indexes. It is clear from the graph that large price changes tend to also be followed by large changes, and small changes tend to follow small changes. Such volatility clustering is a property of asset prices that each index seems to exhibit. This graphical evidence is an indication of the presence of ARCH effect in our daily returns series that should be accounted for when estimating Value at Risk. Figure 9 also displays QQ-plot of each index against the Normal distribution. These QQ-plot show that all returns distributions exhibit fat tails and also fat tails are not symmetric.

Figure 9: Stock market indexes daily percentage returns and QQ-plot against the Normal distribution



2.5. Parameter estimates

To perform a VaR analysis we estimate four volatility models: GARCH, GJR-GARCH, APARCH and FGARCH under each of the different probability distributions assumed for the innovations: Normal, Student-t, skewed Student-t, skewed Generalized Error, unbounded Johnson S_U, skewed Generalized-t and Generalized Hyperbolic skew Student-t distributions. An AR(1) model was specified for the conditional mean return in all cases. Most computations were performed with the rugarch package (version 1.3-4) of R software (version 3.1.1), designed for the estimation and forecast of various univariate ARCH-type models. The exception is the estimation of models under the skewed Generalized-t and Generalized Hyperbolic skew Student-t distributions for which we used the sgt package (version 2.0) and the SkewHyperbolic package (version 0.3-2), respectively.

The Ljung-Box Q statistic for five lags computed on the standardized residuals does not show evidence of autocorrelation at 1% significance level except for GAS. But for one lag, GAS does not show autocorrelation at 1%, inasmuch as the p-values of the Q statistics are 0.0899, 0.2621, 0.2440, 0.0452, 0.2288, 0.0447 and 0.4053 for N-, ST-, SKST-, SGED-, JSU-, SGT- and GHST-APARCH models, respectively. The same statistic computed with nine lags on the squared standardized residuals is not significant at 1% except for IBEX, SAN, IRS, GERMAN BOND, OIL, GOLD and SILVER. If we consider one lag, we obtain a Q statistic not significant at 1% significance level for IBEX and SAN but it remains significant for the remaining assets. A significant statistic indicates a possible problem with this model. In the lower panels of these tables we present the log-likelihood values of the four volatility models (GARCH, GJR-GARCH, APARCH and FGARCH). Their similarity suggests that the implied volatility specifications are very similar. The autoregressive effect in volatility is strong, with a β_1 -parameter generally above 0.90, suggesting strong memory effects. The range of β_1 is [0.88, 0.97] where the minimum is obtained for GAS and the maximum is obtained for EUR/USD. The coefficient γ_1 is positive and statistically significant for most series, indicating the existence of a leverage effect for negative returns in the conditional variance specification. Estimates of γ_1 are close to 1 for IBEX, NASDAQ and FTSE (in the GJR-GARCH model we also obtain an α_1 (parameter close to 0). Compared to estimates for other

assets these values are very high, suggesting that only negative shocks contribute to volatility. We also obtain a γ_1 estimate close to 1 in the APARCH model (equivalently α_1 close to 0 in GJR-GARCH) with other indexes not considered in this chapter such as CAC 40, DAX 30 and S&P 500 for this same sample period. We obtain the same parameter estimates for these models and indexes using MatLab, R, Eviews and Gretl. The coefficient γ_1 is negative and statistically significant for interest rates with some models, GOLD, SILVER and JPY/USD, indicating that a positive shock generates greater volatility than a negative shock of equal size.

It is also important that estimates of ξ in the skewed Student-t and skewed Generalized Error are less than 1 for most assets, suggesting the convenience of incorporating negative asymmetric features in the probability distribution in order to model innovations appropriately. A similar consideration applies to the skewness parameter γ of the Johnson S_U, λ of the skewed Generalized-t and β of Generalized Hyperbolic skew Student-t, which in these cases the skewness parameters have negative sign. We obtain positive skewness with GAS and GOLD with some models, EUR/USD and JPY/USD. According to kurtosis, the estimates of ν (Student-t and skewed Student-t) and δ (Johnson S_U) are between 1.35 and 12.50, capturing the heavy tails of the distribution. The smallest values are obtained with Johnson S_U. The kurtosis parameters κ and p of skewed Generalized Error and skewed Generalized-t, respectively, measure the peakness of the distribution. For most assets and with most models, we obtain values lower than 2 indicating that the distribution is leptokurtic. Note that skewed Generalized-t have two parameters related to kurtosis, p and q. The parameters p and q control the peak and the tails of density, respectively. And the parameter q only has the degrees of freedom interpretation in case $\lambda = 0$ and $p = 2$. We obtain high q values accompanied with low p values for some assets, indicating in these cases that the kurtosis is mainly due to higher peak, rather than thicker tails of the distribution. Finally, δ takes values between 0.95 and 2.33, being significantly different from 2 in most cases[12]. Our estimates of the APARCH model for the different asset classes (not shown in the

12. This result is in line with those of Taylor (1986), Schwert (1990) and Ding et al. (1993) who indicate that there is substantially more correlation among absolute returns than among squared returns, a reflection of the 'long memory' of high-frequency financial returns.

tables) suggest that, contrary to standard practice, we should model the conditional standard deviation for stock market indexes, individual stocks and metals, the conditional variance ($\delta = 2$) for interest rates, and a value between conditional standard deviation and variance ($\delta = 1.5$) for energy commodities and exchange rates. In summary, these results indicate the need for a model featuring a negative leverage effect in the equation for conditional volatility (conditional asymmetry) combined with an asymmetric distribution for the underlying error term (unconditional asymmetry) when representing stock market data. Furthermore, that equation should be specified for the right power of the conditional standard deviation.

Figure 10 displays, for each stock market index, histograms and QQ-plots against theoretical quantiles for estimated standardized residuals ($\hat{z}_t$) of the SKST-APARCH model. We can observe that standardized innovations show, indeed, fat tails and negative skewness.

Figure 10: Histograms and QQ-plots of standardized innovations from SKST-APARCH model for stock market indexes against the skewed Student-t distribution

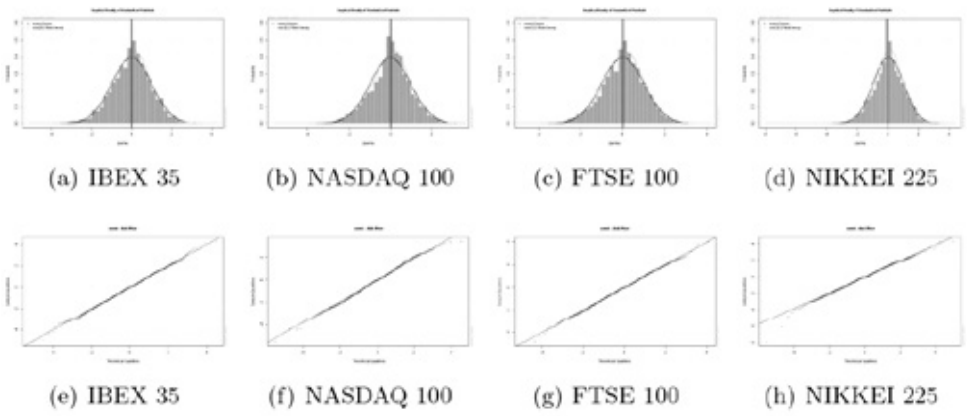
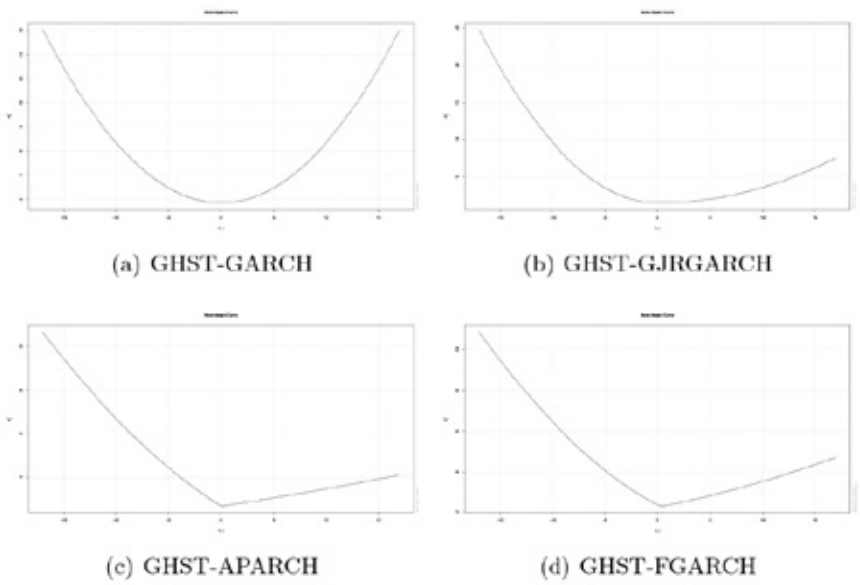


Figure 11 displays the news impact curves of different volatility models for IBM. We can observe that GARCH and GJRGARCH models are based on the variance equation, while APARCH and FGARCH models introduce the Box-Cox transformation in the conditional standard deviation, and the free parameter (δ in APARCH and λ in FGARCH) determines the shape of the transformation. For IBM the value of this parameter is

δ = 1.01 and λ = 1.10 for the GHST-APARCH and GHST-FGARCH model, respectively. These parameters are significantly different from zero and two, but not from one. Furthermore, FGARCH model permit not only rotations, like APARCH model, but also shifts of the news impact curve. As can be seen from Figure 11 d), the asymmetry caused by the shift η_2 = 0.20 is most pronounced for small shocks. For extremely large shocks, the asymmetric effect becomes a negligible part of the total response. On the other hand, the rotated news impact curve of Figure 11 c), γ = 0.61 maintains the hypothesis that a zero shock results in the smallest increase of conditional variance. Additionally, the size of the asymmetric effect of small shocks is very small in absolute terms. The estimates of γ in APARCH model imply that negative shocks result in higher volatility than equally large positive shocks, which is in accordance with the "leverage effect". In Figure 11 d) the shift η_2 = 0.20 and rotation η_1 = 0.42 are combined in one news impact curve. Both parameters are significant. By appropriately shifting and rotating the news impact curve, it is possible to have asymmetry for small shocks, a roughly symmetric response for moderate shocks, and asymmetry for very large shocks.

Figure II: News impact curves of different volatility model for IBM



(a) GHST-GARCH (b) GHST-GJRGARCH

(c) GHST-APARCH (d) GHST-FGARCH

2.6. Fitting the data

VaR models are usually evaluated according to the performance of their VaR estimates using appropriate testing procedures. However, the ability of a VaR model to reproduce the main characteristics of return data is hardly ever examined. A possible justification for such inattention is the argument that good VaR estimates have to do just with the quality of the fit to the tails of the distribution of returns. A good overall fit might not be all that interesting because it might be obtained at the expense of not fitting so well the distribution tails. However, the fit to the tail of return distribution is usually not examined either. The fact is that it is unclear whether a good overall fit of the return distribution helps to produce good VaR estimates or whether it should be enough to care about the fit to the tail of the distribution, and we want to throw some light into that question. In particular, if fitting the tail distribution is what matters, that might explain why the type of models considered in extreme value theory tend to beat other alternatives in VaR estimation.

We examine in this section the extent to which each model fits the return data, and we will later check whether the models with a better overall fit lead to better VaR estimates. We start by checking the extent to which each model fits the likelihood of return data. After that, we examine the ability of each model to fit the main sample moments of returns. To evaluate the fit to the distribution of returns Monte Carlo simulation is needed, as explained below.

2.6.1. Likelihood ratio tests

Models with FGARCH volatility, combined with JSU and SGED distributions for stock market indexes, with SKST and SGED distributions for individual stocks, with JSU and GHST distributions for interest rates, with SGED for commodities and with SGED and JSU for exchange rates, often achieve the highest log-likelihood. Likelihood ratio tests in Table 14 show a superiority of the FGARCH specification over the APARCH, GJR-GARCH and the symmetric GARCH specifications for stock market indexes. In all comparisons in the table, the more restricted model appears to the left. At 5% significance, the test clearly favors the APARCH model against the GJRGARCH model and the FGARCH model against the APARCH. Indeed,

for stock market indexes, individual stocks and commodities the FGARCH model is preferred to the APARCH model whereas for interest rates and exchange rates the APARCH model is preferred. Overall, FGARCH and APARCH are the best models according to this criterion.

Table 14: Likelihood ratio tests of volatility specifications for stock market indexes

Test statistic	IBEX	NASDAQ	FTSE	NIKKEI
N-GARCH vs N-APARCH	205.186	138.348	195.112	78.296
N-GJRGARCH vs N-APARCH	30.148	11.934	15.908	9.164
N-APARCH vs N-FGARCH	21.816	52.104	37.834	47.568
ST-GARCH vs ST-APARCH	159.688	119.084	248.592	79.222
ST-GJRGARCH vs ST-APARCH	25.748	15.412	18.558	17.354
ST-APARCH vs ST-FGARCH	8.762	27.646	32.422	44.818
SKST-GARCH vs SKST-APARCH	167.376	134.806	186.022	77.49
SKST-GJRGARCH vs SKST-APARCH	26.388	18.518	19.916	17.154
SKST-APARCH vs SKST-FGARCH	11.064	38.588	33.902	46.58
SGED-GARCH vs SGED-APARCH	163.090	123.216	137.098	66.682
SGED-GJRGARCH vs SGED-APARCH	24.716	15.794	**-19.958**	13.778
SGED-APARCH vs SGED-FGARCH	12.574	19.094	30.93	40.332
JSU-GARCH vs JSU-APARCH	166.902	135.970	184.646	74.574
JSU-GJRGARCH vs JSU-APARCH	25.992	19.584	19.006	16.460
JSU-APARCH vs JSU-FGARCH	**1.216**	14.958	33.778	47.004
SGT-GARCH vs SGT-APARCH	154.116	108.794	157.816	66.148
SGT-GJRGARCH vs SGT-APARCH	24.028	12.516	15.348	13.442
SGT-APARCH vs SGT-FGARCH	10.89	27.402	30.93	38.618
GHST-GARCH vs GHST-APARCH	168.844	148.648	180.538	71.766
GHST-GJRGARCH vs GHST-APARCH	34.134	60.512	19.006	16.460
GHST-APARCH vs GHST-FGARCH	**-6.826**	13.926	20.554	57.444

Note: The null hypothesis is rejected, except where indicated by boldface

2.6.2. Fitting standardized innovations

2.6.2.1. Fitting the empirical distribution of return innovations

Table 15 reports the results obtained when comparing the empirical distribution of estimated innovations to the theoretical distribution used in estimation for the four stock market indexes.[13] We use the Kolmogorov-Smirnov (KS) test (Kolmogorov, 1933, Smirnov, 1939 and Massey, 1951), which quantifies the distance between the empirical distribution function of standardized innovation and the cumulative distribution function of the reference distribution, and the Chi-square (Chi2) test (Pearson, 1900) applied to a partition of the return data range into 10 bins.[14] The null distribution of these statistics is calculated under the null hypothesis that the sample is drawn from the reference distribution. These tests suggest that models with an asymmetric distribution for the innovations are to be preferred. Test statistics also tend to be smaller for the APARCH and FGARCH volatility specifications.

According to the KS test, models with N distributions fits the data well 11 out of 76 cases (4 volatility models by 19 assets), ST fits the data well in 53 cases, SKST in 54, SGED in 59, JSU in 47, SGT in 62 and GHST in 47 cases. Regarding volatility models, distributions with GARCH model fit the data well 79 out of 133 cases (7 probability distributions by 19 assets), GJRGARCH and APARCH fit the data well in 85 cases and FGARCH in 84 cases. According to the Chi2 test, models with N distributions fits the data well 1 out of 76 cases, ST fits the data well in 20 cases, SKST and SGED in 32, JSU and SGT in 30 and GHST in 18 cases. Respect to volatility models, distributions with GARCH, APARCH and FGARCH models fit the data well 40 out of 133 cases and GJRGARCH in 43 cases. To sum up, the SGED and SGT are preferred to fit the innovations and GJRGARCH and APARCH to model the volatility.

13. Results for other assets are available on request.
14. The number of bins affects the results of the Pearson test.

Table 15: Goodness-of-fit tests for standardized innovations of stock market indexes. Figures in parentheses denote p-values

	IBEX35		NASDAQ100		FTSE100		NIKKEI225	
	KS	Chi2	KS	Chi2	KS	Chi2	KS	Chi2
N-GARCH	0.039 (0.000)	243110 (0.000)	0.043 (0.000)	86.329 (0.000)	0.032 (0.000)	107.345 (0.000)	0.051 (0.000)	243267 (0.000)
ST-GARCH	0.022 (0.038)	21.348 (0.011)	0.028 (0.002)	17.048 (0.048)	0.020 (0.083)	37.191 (0.000)	0.039 (0.000)	21.667 (0.010)
SKST-GARCH	0.027 (0.005)	9.892 (0.359)	0.030 (0.001)	7.344 (0.601)	0.031 (0.001)	15.078 (0.089)	0.038 (0.000)	13.744 (0.132)
SGED-GARCH	0.022 (0.035)	50.690 (0.000)	0.023 (0.027)	5.633 (0.776)	0.027 (0.005)	17.152 (0.046)	0.026 (0.006)	39.174 (0.000)
JSU-GARCH	0.026 (0.006)	10.010 (0.349)	0.030 (0.001)	6.336 (0.706)	0.031 (0.001)	13.381 (0.146)	0.041 (0.000)	14.011 (0.122)
SGT-GARCH	0.029 (0.001)	23.392 (0.005)	0.028 (0.003)	6.001 (0.740)	0.034 (0.000)	17.480 (0.042)	0.028 (0.003)	39.408 (0.000)
GHST-GARCH	0.021 (0.056)	8.799 (0.456)	0.028 (0.003)	7.665 (0.568)	0.019 (0.099)	22.349 (0.008)	0.037 (0.000)	10.983 (0.277)
N-GJRGARCH	0.034 (0.000)	228.860 (0.000)	0.047 (0.000)	39.537 (0.000)	0.039 (0.000)	117.613 (0.000)	0.048 (0.000)	971626 (0.000)
ST-GJRGARCH	0.024 (0.020)	49.752 (0.000)	0.031 (0.001)	40.861 (0.000)	0.029 (0.002)	57.236 (0.000)	0.038 (0.000)	73.794 (0.000)
SKST-GJRGARCH	0.018 (0.129)	14.610 (0.102)	0.022 (0.041)	12.553 (0.184)	0.016 (0.222)	8.206 (0.514)	0.036 (0.000)	43.630 (0.000)
SGED-GJRGARCH	0.015 (0.338)	21.714 (0.010)	0.017 (0.166)	17.948 (0.036)	0.016 (0.262)	9.942 (0.355)	0.030 (0.001)	127.937 (0.000)
JSU-GJRGARCH	0.017 (0.155)	14.262 (0.113)	0.020 (0.072)	12.947 (0.165)	0.016 (0.248)	5.923 (0.748)	0.039 (0.000)	38.568 (0.000)
SGT-GJRGARCH	0.021 (0.046)	20.689 (0.014)	0.025 (0.012)	20.488 (0.015)	0.026 (0.008)	11.848 (0.222)	0.029 (0.002)	133.807 (0.003)
GHST-GJRGARCH	0.027 (0.004)	18.724 (0.028)	0.033 (0.000)	12.947 (0.165)	0.028 (0.003)	28.488 (0.001)	0.035 (0.000)	24.784 (0.000)
N-APARCH	0.036 (0.000)	248.980 (0.000)	0.047 (0.000)	141.086 (0.000)	0.042 (0.000)	111.868 (0.000)	0.048 (0.000)	243023 (0.000)
ST-APARCH	0.026 (0.006)	48.544 (0.000)	0.030 (0.001)	29.656 (0.001)	0.031 (0.001)	44.470 (0.000)	0.038 (0.000)	47.912 (0.000)

	IBEX35		NASDAQ100		FTSE100		NIKKEI225	
	KS	Chi2	KS	Chi2	KS	Chi2	KS	Chi2
SKST-APARCH	0.019	15.015	0.020	2.589	0.019	3.208	0.037	20.405
	(0.093)	(0.091)	(0.062)	(0.978)	(0.110)	(0.956)	(0.000)	(0.016)
SGED-APARCH	0.019	24.748	0.017	3.676	0.015	5.960	0.031	54.437
	(0.114)	(0.003)	(0.162)	(0.931)	(0.275)	(0.744)	(0.001)	(0.000)
JSU-APARCH	0.020	16.025	0.020	1.759	0.018	3.576	0.038	15.731
	(0.081)	(0.066)	(0.064)	(0.995)	(0.120)	(0.937)	(0.000)	(0.073)
SGT-APARCH	0.019	21.066	0.026	5.237	0.025	6.753	0.029	60.521
	(0.094)	(0.012)	(0.008)	(0.813)	(0.013)	(0.663)	(0.002)	(0.000)
GHST-APARCH	0.030	22.105	0.031	8.823	0.030	21.973	0.037	20.293
	(0.001)	(0.009)	(0.001)	(0.454)	(0.001)	(0.009)	(0.000)	(0.016)
N-FGARCH	0.037	132.260	0.047	97.099	0.042	106.330	0.051	971730
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ST-FGARCH	0.027	14.633	0.025	32.188	0.033	44.745	0.037	60.110
	(0.004)	(0.102)	(0.011)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SKST-FGARCH	0.019	4.929	0.022	15.5601	0.019	2.967	0.034	27.787
	(0.102)	(0.840)	(0.035)	(0.077)	(0.088)	(0.966)	(0.000)	(0.001)
SGED-FGARCH	0.018	10.078	0.023	14.149	0.017	2.206	0.032	110.135
	(0.142)	(0.344)	(0.029)	(0.117)	(0.158)	(0.988)	(0.000)	(0.000)
JSU-FGARCH	0.019	4.467	0.026	12.654	0.020	2.393	0.035	22.448
	(0.097)	(0.878)	(0.007)	(0.179)	(0.082)	(0.984)	(0.000)	(0.008)
SGT-FGARCH	0.018	7.566	0.028	14.121	0.023	3.179	0.027	121.760
	(0.123)	(0.578)	(0.002)	(0.118)	(0.022)	(0.957)	(0.005)	(0.000)
GHST-FGARCH	0.026	15.856	0.032	33.125	0.032	25.915	0.037	25.226
	(0.008)	(0.070)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.003)

To compare the adequacy of the different distributions we can also employ out-of-sample density forecasts, as proposed by Diebold, Gunther and Tay (1998) (DGT). Let $f_i(y_i|\Omega_i)_{i=1}^m$ be a sequence of m one-step-ahead density forecasts produce by a given model, where Ω is the conditioning information set, and $p_i(y_i|\Omega_i)_{i=1}^m$ the sequence of densities defining the Data Generating Process y_i (which is never observed). The null hypothesis is H_0: $f_i(y_i|\Omega_i)_{i=1}^m = p_i(y_i|\Omega_i)_{i=1}^m$. DGT use the fact that under null hypothesis, the probability integral transform $\zeta_i = \int_{-\infty}^{0} f_i(t)dt$ is i.i.d. with a Uniform(0,1) distribution. To check H_0 , they propose to use an independence test for i.i.d. U(0,1). The i.i.d.-ness property of

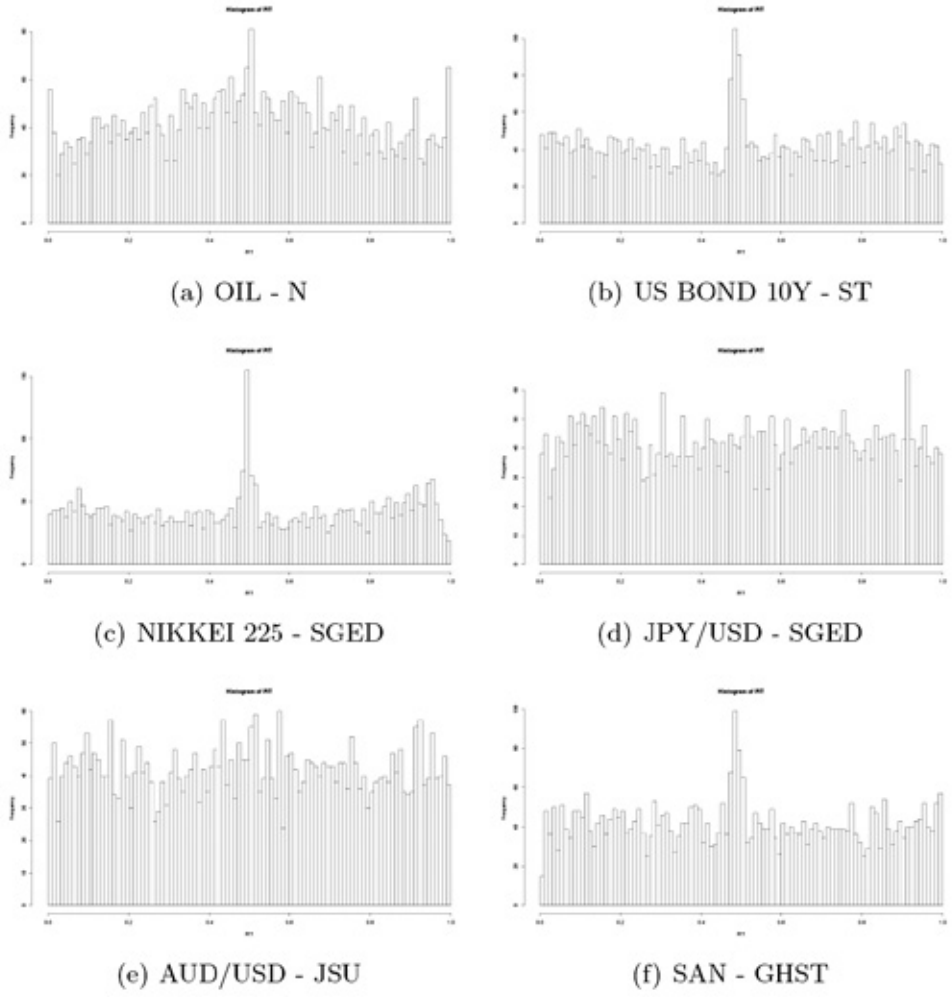
ζ_i can be evaluated by plotting a histogram of ζ_i. A humped shape of the ζ-histogram would indicate that the issued forecasts are too narrow and that the tails of the true density are not accounted for. On the other hand, a U-shape of the histogram would suggest that the model issues forecasts that either under- or overestimate too frequently [Bauwens, Giot, Grammig and Veredas (2000)].

Figures 12 a) and 12 b) show a sample of such histograms for the assets in our data set. The humped shape of the histograms shows that symmetrical distributions are not suitable to model the OIL and US BOND 10Y returns. Figures 12 c) and 12 d) show that the skewed Generalized Error distribution is not suitable for NIKKEI 225. It is appropriate for JPY/USD because its probability integral transform is Uniformly distributed. In 12 e) the Johnson S_U distribution is also appropriate for AUD/USD. Figure 12 f) shows that the assumption of a Generalized Hyperbolic skew Student-t for the innovation is not appropriate for SAN. These results are consistent with the goodness-of-fit tests previously carried out. For the rest of assets, the results are similar, the symmetrical distributions and the Generalized Hyperbolic skew Student-t for the innovations are not appropriate for most of the assets whereas skewed Student-t, skewed Generalized Error and Johnson S_U are suitable.

2.6.2.2. Fitting the sample moments of return innovations

For a given asset, the innovations change with the estimated model, so we compare the theoretical moments of a given probability distribution with the sample moments for the standardized innovations for that model. In fact, however, sample moments for innovations are similar across models, showing a near zero mean and a unit variance in all models, as expected. But that is also the case under all the estimated probability distributions, so it makes sense to focus on the ability of each estimated distribution to fit the sample skewness and kurtosis of standardized innovations. Table 16 compares the theoretical value of skewness and kurtosis from the estimated probability distribution with the similar sample moments of the standardized innovations calculating the absolute differences between these both values for stock market indices. Obviously, the Normal and the symmetric Student distribution do not produce any skewness. This is

Figure 12: ζ-histograms (100 cells) for 4173 one-step-ahead forecasts. We assume different distributions with AR(1)-FGARCH(1,1) model for different assets



(a) OIL - N

(b) US BOND 10Y - ST

(c) NIKKEI 225 - SGED

(d) JPY/USD - SGED

(e) AUD/USD - JSU

(f) SAN - GHST

a limitation of these distributions since skewness and kurtosis are present in standardized innovations. For most assets, the skewed t-Student distribution produces negative skewness, although not as much as it is observed in the data. The unbounded Johnson distribution achieves a higher level of negative skewness, often being close to that observed in the data. The GHST distribution does not fit innovation moments very well, especially overestimating the degree of negative skewness. Indeed,

the GHST distribution usually produces the maximum absolute difference between the theoretical and the sample skewness in most stock market indexes, individual stocks and exchange rates. The GHST distribution has been proposed as being suitable for assets with high skewness and heavy-tailed (Aas & Haff, 2006) and the assets we consider do not have high skewness. In fact, only the standardized innovations in SILVER and AUD/USD have a negative high skewness and in these two cases models with GHST produce the best fit to sample skewness. Additionally, asymmetric probability distributions are unable to reproduce the positive skewness shown by a few return innovations, such as those in IRS5Y and GAS.

On the other hand, the GHST distribution can explain the high kurtosis often observed in our standardized innovations, except when used with a GJRGARCH volatility for stock market indexes or when used with APARCH and FGARCH specifications for IRS5Y. The symmetric and the skewed Student-t distributions explain the level of kurtosis observed in the data[15], while the Johnson distribution generally implies higher kurtosis than it is observed in the data[16].

In fact, for skewness the results are concentrated in the SKST distribution, it fits skewness best in 8 of the 19 cases. For kurtosis results are not so concentrated: ST (for 5 assets), SKST (4), SGT (6) y SGED (4) fit kurtosis best. Pulling together the fit of both moments, the SKST distribution performs best in 12 out of the 38 cases, followed by SGT and SGED with 7 cases. The FGARCH specification fits skewness best in 7 assets, while the GARCH specification fits kurtosis best in 8 assets. Overall, the SGT and SKST distributions with GARCH, GJR-GARCH and FGARCH do better in capturing skewness and kurtosis of the standardized innovations than other combinations.

15. The theoretical kurtosis for the Student-t distribution has been calculated as $K = \frac{6}{\nu-4} + 3$. For GOLD and SILVER, the Student-t distribution for some volatility models produces negative kurtosis because we have obtained in the estimation a number of degrees of freedom (ν) less than 4.

16. For GOLD and SILVER, as well as for IBM and BP, the unbounded Johnson distribution produces extremely high kurtosis because the estimated kurtosis parameter (δ) of the Johnson distribution is close to 1. As $\delta > \infty$ the distribution approaches the Normal density function and we obtain a kurtosis = 3.

Table 16: Absolute differences between standardized innovation moments and theoretical moments for stock market indexes. Bold figures show the lowest value for each asset

	IBEX35		NASDAQ100		FTSE100		NIKKEI 225	
	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
N-GARCH	0.263	1.327	0.235	0.900	0.318	0.748	0.377	1.334
ST-GARCH	0.284	0.149	0.245	0.651	0.317	0.488	0.398	0.278
SKST-GARCH	0.079	0.045	0.048	0.555	0.065	0.313	0.197	0.127
SGED-GARCH	0.104	0.420	0.101	0.206	0.102	0.063	0.341	**0.074**
JSU-GARCH	0.070	2.529	0.145	3.869	0.062	1.020	0.076	3.273
SGT-GARCH	0.114	0.254	0.123	0.222	0.118	0.082	0.351	0.105
GHST-GARCH	0.345	1.303	0.416	2.406	0.210	0.333	0.249	1.558
N-GJRGARCH	0.260	0.952	0.275	0.734	0.332	0.633	0.366	1.499
ST-GJRGARCH	0.269	0.159	0.288	0.476	0.332	0.196	0.384	0.267
SKST-GJRGARCH	0.032	0.119	0.030	0.421	**0.047**	0.115	0.198	0.339
SGED-GJRGARCH	0.055	0.187	0.070	0.188	0.066	0.011	0.316	0.268
JSU-GJRGARCH	0.025	0.952	0.069	2.094	0.171	0.165	0.032	1.718
SGT-GJRGARCH	0.063	0.084	0.096	0.203	0.081	0.010	0.336	0.264
GHST-GJRGARCH	0.501	4.861	0.843	18.402	0.304	2.140	0.731	16.204
N-APARCH	0.251	0.916	0.307	0.773	0.338	0.685	0.365	1.577
ST-APARCH	0.258	0.159	0.332	0.420	0.346	0.085	0.395	0.451
SKST-APARCH	0.019	0.130	0.061	0.323	0.052	0.033	0.206	0.511
SGED-APARCH	0.043	0.172	0.094	**0.108**	0.067	0.088	0.313	0.461
JSU-APARCH	0.027	0.852	**0.023**	1.778	0.176	0.097	**0.001**	2.072
SGT-APARCH	0.049	0.058	0.118	0.127	0.083	0.089	0.343	0.458
GHST-APARCH	0.391	2.119	0.334	2.779	0.199	0.498	0.277	1.444
N-FGARCH	0.232	0.763	0.299	0.677	0.338	0.649	0.385	1.632
ST-FGARCH	0.250	0.197	0.297	0.204	0.349	0.038	0.415	0.513
SKST-FGARCH	**0.003**	0.184	0.053	0.352	0.055	**0.002**	0.216	0.596
SGED-FGARCH	0.027	0.082	0.122	0.131	0.065	0.099	0.318	0.588
JSU-FGARCH	0.035	0.745	0.185	0.140	0.201	0.021	0.051	1.326
SGT-FGARCH	0.035	**0.011**	0.106	0.168	0.080	0.096	0.353	0.582
GHST-FGARCH	0.291	2.414	0.313	2.303	0.080	1.123	0.218	0.818

2.6.3. Fitting observed returns

2.6.3.1. Fitting the empirical distribution of asset returns

How about the ability of each estimated model to fit sample return moments? Unfortunately, except in cases when returns do not show any stochastic structure, it is not easy to derive the moments of asset returns from the estimated probability distribution for return innovations. Hence, we characterize the implied probability distribution for returns by simulation. Taking random draws for the estimated probability distribution for innovations, we generate 1000 time series for returns with the same length as our data set. For each simulation we apply the two-sample KS test (Kolmogorov, 1933, Smirnov, 1939 and Massey, 1951) and the Chi2 test (Pearson, 1900) to compute the failure rates of the respective null hypotheses.

The KS test quantifies the distance between the empirical distribution function of observed returns and the one obtained from each simulated time series. The KS test statistic is:

$$D = sup_x \left| F_{1,n}(x) - F_{2,n'}(x) \right|$$

where sup_x is the supremum of the set of distances between two empirical distributions, $F_{1,n}$ and $F_{2,n'}$. The null hypothesis is rejected at level α if $D > c(\alpha)\sqrt{\frac{n+n'}{nn'}}$ where n and n' are the sizes of first and second sample respectively. The value of $c(\alpha)$ is given in the table below for each level of α,

α	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

Table 17 reports the failure rates of the KS and Chi2 null hypothesis at confidence level 99%. The models with lower failure rate in either the KS and the Chi2 tests are the SGED distribution with GJRGARCH, APARCH or FGARCH volatility specifications, and the SKST and JSU distributions with APARCH and FGARCH specifications, respectively. Hence, we observe again the preference for asymmetric distributions and volatility models with leverage.

Table 17: Goodness-of-fit tests for observed returns of stock market indexes. Figures denote the fail rates for each model

confidence level = 0.99	IBEX35		NASDAQ100		FTSE100		NIKKEI225	
fail rate	KS	Chi2	KS	Chi2	KS	Chi2	KS	Chi2
N-GARCH	0.821	0.985	0.993	1.000	0.367	0.846	1.000	0.999
ST-GARCH	0.069	0.978	0.391	0.999	0.100	0.676	0.798	0.997
SKST-GARCH	0.202	0.714	0.575	0.994	0.551	0.423	0.836	0.773
SGED-GARCH	0.110	0.594	0.155	0.991	0.519	0.461	0.460	0.979
JSU-GARCH	0.208	0.713	0.549	0.993	0.555	0.436	0.855	0.542
SGT-GARCH	0.324	0.669	0.332	0.992	0.800	0.428	0.465	0.994
GHST-GARCH	0.067	0.889	0.416	0.998	0.067	0.447	0.845	0.811
N-GJRGARCH	0.593	0.997	0.993	1.000	0.721	0.970	0.997	0.999
ST-GJRGARCH	0.149	0.995	0.420	1.000	0.221	0.887	0.806	0.997
SKST-GJRGARCH	0.029	0.775	0.185	0.995	0.054	0.433	0.789	0.818
SGED-GJRGARCH	0.009	0.668	0.051	0.993	0.056	0.450	0.521	0.993
JSU-GJRGARCH	0.025	0.803	0.164	0.996	0.050	0.459	0.841	0.637
SGT-GJRGARCH	0.074	0.648	0.150	0.990	0.270	0.421	0.461	0.991
GHST-GJRGARCH	0.223	0.974	0.617	0.996	0.218	0.705	0.923	0.804
N-APARCH	0.672	0.999	0.992	1.000	0.819	0.993	0.999	0.998
ST-APARCH	0.250	0.994	0.407	1.000	0.313	0.943	0.791	1.000
SKST-APARCH	0.032	0.812	0.177	0.993	0.040	0.527	0.747	0.826
SGED-APARCH	0.024	0.706	0.058	0.989	0.034	0.583	0.565	0.989
JSU-APARCH	0.037	0.818	0.154	0.996	0.037	0.551	0.762	0.708
SGT-APARCH	0.045	0.657	0.170	0.983	0.219	0.515	0.473	0.996
GHST-APARCH	0.318	0.984	0.463	0.999	0.302	0.789	0.808	0.938
N-FGARCH	0.735	0.998	0.984	1.000	0.858	0.983	1.000	1.000
ST-FGARCH	0.305	0.999	0.285	1.000	0.423	0.939	0.768	0.999
SKST-FGARCH	0.033	0.831	0.158	0.998	0.046	0.453	0.714	0.825
SGED-FGARCH	0.030	0.746	0.142	0.999	0.025	0.452	0.607	0.991
JSU-FGARCH	0.033	0.864	0.338	1.000	0.047	0.500	0.700	0.700
SGT-FGARCH	0.040	0.670	0.336	0.998	0.173	0.405	0.441	0.989
GHST-FGARCH	0.257	0.998	0.455	1.000	0.344	0.861	0.802	0.959

2.6.3.2. Fitting the sample moments of asset returns

In addition to the fit to the whole distribution, we now examine the ability of each combination of volatility specification and probability distribution to fit the main moments of observed returns: sample mean, standard deviation, skewness, kurtosis, maximum, minimum and the observed range. To that end, we assign to each model the average value for each of these moments over the set of 1000 simulations, to be compared with their sample return analogues. Table 18 presents sample return moments for stock market indices together with a summary of the average simulated return moments over probability distributions and volatility specifications[17]. Column 1 in Table 18 displays sample moments, while column 2 shows the median value of the average simulated moments across all models (28 in total). The remaining columns show median values of moments across subsets of models[18]. The first panel, from third to ninth column, considers median values of moments across alternative volatility specifications, for a given probability distribution for return innovations. The second panel, from tenth to thirteenth column, presents median values of simulated moments across probability distributions, for a given volatility specification. We also compute the mean absolute difference between the average moments obtained by simulation and the analogue sample moments (mean, standard deviation, skewness, kurtosis, maximum, minimum and the observed range). The last row displays the median value of these absolute differences. Finally, we take the range[19] of MAE values across the set of volatility specifications or across the set of probability distributions, as shown in the last two columns.

The first panel shows that for most assets all probability distributions explain the standard deviations of return data similarly, with the Normal and Student-t distributions doing somewhat better than the rest. The Johnson S_U distribution approximates very well the level of skewness in returns and skewed Generalized Error distribution does better than other distributions to approximate the level of kurtosis. We conclude that the Normal distribution performs well on this account for stock market

17. Results for other assets are available on request.
18. Remember that for each model we take the average value of each moment over 1000 simulations.
19. The difference between the highest and the lowest MAE values.

indexes because it fits very well the second moment but not because it fits well the higher order moments, i.e. the third and fourth moment. In the second panel, the differences between volatility specifications are small compared with differences between probability distributions but APARCH and FGARCH models fit standard deviation better than another volatility models, GJRGARCH and FGARCH volatilities seem to fit skewness best, while APARCH and FGARCH fit kurtosis best.

Summarizing, all the probability distributions other than the Normal produce levels of kurtosis as high as those in the return data, but they fall short of explaining the negative skewness observed in some market returns. They also fall a bit short of reproducing the maximum returns. However, they tend to produce a minimum that is higher in absolute value than the one for observed returns. Consequently, the range of values implied by the estimated models is just a bit narrower than that observed in return data for all assets.

According to the median MAE, the Normal distribution is the preferred one for 2 assets, the symmetric Student-t is the best for 4 assets, the skewed Student-t for 3, the skewed Generalized Error for 2, the Johnson S_U for 4, the skewed Generalized-t for 1 and Generalized Hyperbolic skew Student-t for 3 assets. In terms of volatility models, the standard GARCH is the preferred volatility specifications for 4 assets, the GJR-GARCH model for 1, the APARCH model for 6 and the FGARCH model is the best for 8 assets. So, from this point of view, it looks as if the FGARCH and APARCH volatility specifications and the symmetric Student-t and the Johnson S_U probability distribution should be preferred.

If we exclude from consideration the ability to reproduce the maximum and minimum observed returns the Normal distribution is the preferred one for 2 assets, the symmetric Student-t is the best for 2 assets, the skewed Student-t for 2, the skewed Generalized Error for 5, the Johnson S_U for 3, the skewed Generalized-t for 2 and Generalized Hyperbolic skew Student-t for 3 assets. In terms of volatility models, the standard GARCH is the preferred volatility specifications for 5 assets, the GJR-GARCH model for 1, the APARCH model for 7 and the FGARCH model is the best for 6 assets. Again, the APARCH and FGARCH volatility models perform better than GARCH and GJR-GARCH, but now the skewed Generalized Error distribution is the preferred one.

Table 18: Empirical moments vs sample moments for stock market indexes

		Median over	Median over probability distributions						Median over volatility models					Ranges over distributions (left)	
	Sample	all models	N	ST	SKST	SGED	JSU	SGT	GHST	GARCH	GJRGARCH	APARCH	FGARCH	and over volatility models (right)	
IBEX35															
Mean	-0.005	0.018	0.007	0.028	0.011	0.006	0.012	0.029	0.009	0.046	0.015	0.007	0.004	0.023	0.042
Standard deviation	1.495	1.743	1.592	1.474	1.874	1.685	1.743	2.004	2.400	2.049	1.750	1.516	1.674	0.926	0.532
Skewness	0.083	-0.193	0.042	0.064	-0.272	-0.224	-0.318	-0.078	-0.740	-0.244	-0.182	-0.235	-0.079	0.804	0.165
Kurtosis	7.932	12.364	6.875	12.419	14.043	12.206	13.732	10.226	16.545	12.860	13.755	9.423	13.217	9.670	4.333
Maximum	13.484	13.675	10.356	12.234	15.004	13.348	14.102	14.953	16.584	15.095	13.989	10.735	14.316	6.228	4.360
Minimum	-9.586	-13.746	-9.976	-11.114	-15.613	-13.746	-15.138	-14.938	-21.414	-16.437	-13.701	-11.861	-13.792	11.438	4.576
Range	23.070	27.584	20.332	23.348	30.866	27.208	29.557	29.891	37.448	31.533	27.690	22.596	27.791	17.115	8.936
	Median MAE		**0.853**	**1.237**	**2.483**	**1.580**	**2.174**	**1.630**	**4.130**	**2.459**	**1.820**	**0.953**	**2.001**	**3.277**	**1.505**
NASDAQ100															
Mean	0.005	0.052	0.027	0.060	0.032	0.042	0.049	0.059	0.037	0.070	0.033	0.031	0.055	0.034	0.039
Standard deviation	1.848	1.759	1.591	1.420	1.828	1.714	1.789	1.955	2.676	1.876	1.939	1.744	1.245	1.257	0.694
Skewness	0.192	-0.097	0.062	0.066	-0.227	-0.177	-0.297	-0.043	-0.788	-0.199	-0.140	-0.219	0.044	0.854	0.263
Kurtosis	9.623	11.838	6.833	11.691	15.384	11.343	12.643	9.933	33.514	11.669	15.312	12.087	7.035	26.681	8.277
Maximum	17.203	13.741	9.866	11.939	15.234	12.849	13.616	15.024	19.485	13.598	15.853	13.885	10.031	9.618	5.822
Minimum	-11.115	-14.403	-9.292	-10.803	-15.034	-13.247	-15.093	-14.524	-24.307	-14.978	-15.248	-14.469	-7.301	15.015	7.946
Range	28.318	28.319	19.158	22.742	30.079	26.096	28.709	29.548	42.950	28.326	31.101	29.092	19.883	23.792	11.218
	Median MAE		**2.134**	**1.507**	**2.296**	**1.739**	**2.431**	**1.508**	**7.888**	**1.783**	**2.534**	**1.845**	**2.896**	**6.380**	**1.113**

		Median over		Median over probability distributions						Median over volatility models				Ranges over distributions (left)	
	Sample	all models	N	ST	SKST	SGED	JSU	SGT	GHST	GARCH	GJRGARCH	APARCH	FGARCH	and over volatility models (right)	
FTSE100															
Mean	-0.003	-0.003	-0.006	0.010	-0.005	-0.009	-0.005	0.010	-0.008	0.028	-0.003	-0.008	-0.011	0.019	0.039
Standard deviation	1.210	1.346	1.221	1.237	1.321	1.244	1.279	1.665	1.762	1.287	1.326	1.231	1.551	0.541	0.320
Skewness	-0.161	-0.277	0.065	0.067	-0.325	-0.277	-0.404	-0.094	-0.596	-0.290	-0.261	-0.305	-0.265	0.663	0.044
Kurtosis	9.356	11.624	7.223	11.786	13.472	12.302	13.579	10.420	21.043	11.269	14.362	9.968	17.074	13.820	7.106
Maximum	9.384	10.687	8.027	9.726	10.406	9.813	10.133	12.133	14.531	9.083	10.786	8.984	13.586	6.504	4.602
Minimum	-9.266	-10.675	-7.673	-9.437	-10.967	-10.049	-11.161	-12.544	-16.806	-10.260	-10.751	-10.189	-14.170	9.132	3.981
Range	18.650	21.293	15.700	19.164	21.337	19.863	21.294	24.677	31.337	19.343	21.537	19.229	27.293	15.637	8.064
	Median MAE		**0.888**	**0.926**	**1.213**	**0.856**	**1.294**	**1.349**	**4.219**	**0.868**	**1.384**	**0.539**	**2.941**	**3.363**	**2.402**
NIKKEI225															
Mean	0.000	0.012	0.012	0.012	0.012	-0.001	0.012	0.012	-0.011	0.012	0.012	0.012	0.002	0.023	0.010
Standard deviation	1.499	1.561	1.533	1.502	1.506	1.529	1.500	2.182	1.693	1.545	1.482	1.484	1.678	0.681	0.195
Skewness	-0.410	-0.064	0.014	0.024	-0.218	-0.064	-0.336	0.000	-0.751	-0.066	-0.037	-0.061	-0.092	0.775	0.055
Kurtosis	9.725	8.711	4.825	8.581	8.987	9.273	8.898	8.565	13.131	8.672	9.301	7.444	11.667	8.306	4.223
Maximum	13.235	10.939	8.372	10.785	10.597	11.194	10.177	15.646	11.342	10.992	10.886	10.076	12.773	7.275	2.697
Minimum	-12.120	-11.439	-8.158	-10.532	-11.305	-11.072	-11.567	-15.496	-15.983	-11.422	-11.189	-10.693	-14.054	7.825	3.361
Range	25.354	22.376	16.530	21.318	21.902	22.266	21.744	31.142	27.325	22.309	22.074	20.357	27.491	14.612	7.134
	Median MAE		**2.367**	**0.947**	**0.798**	**0.991**	**0.821**	**1.341**	**1.618**	**0.983**	**1.076**	**1.323**	**1.259**	**1.570**	**0.340**

Interestingly enough, the last two columns show that median values of the simulated statistics for different volatility specifications are more similar among them than median values for the alternative probability distributions. This suggests again that the assumption we can make on the probability distribution of return innovations may be more important to fit return data than the assumption on the volatility specification.

2.7. VaR Performance

We now analyze the VaR performance of our estimated models restricting our attention to the left tail of the distribution and the 1% significance level. The choice of the 1% level is a compromise between trying to capture extreme events and trying to avoid a too low number of exceptions. Results for alternative significance levels are available from the authors upon request. Considering the left tail is not a trivial choice, since results for both tails may differ significantly for asymmetric return distributions. In all cases we present out-of-sample VaR forecasts over the last five years in the sample: 2011-2015 (1260 data observations). Every day we compute 1-day ahead 1% VaR, reestimating each model every 50 days. The latter choice tries to reduce the computational cost as well as avoiding frequent parameter variation that might be due in part to just noise.

We estimate the one-step ahead VaR parametrically as $VaR_{\alpha,t} = \mu_t(\theta) + \sigma_t(\theta)F^{-1}(\alpha|\theta)$, where $\mu_t(\theta)$ represents the conditional mean, $\sigma_t(\theta)$ is the conditional standard deviation and $F^{-1}(\alpha|\theta)$ denotes the corresponding quantile of the distribution of the standardized innovations z_t at a given α% significance. After that, we examine the performance of VaR models through standard tests: the unconditional coverage test of Kupiec (1995), the independence and conditional coverage tests of Christoffersen (1998), the Dynamic Quantile test of Engle and Manganelli (2004), as well as by evaluating the Asymmetric Linear Tick loss function (AlTick) proposed by Giacomini and Komunjer (2005). For a comprehensive review on VaR forecasting and backtesting, see Nieto and Ruiz (2015).

The unconditional coverage test introduced by Kupiec (1995) is based on the number of violations, i.e. the number of times (T_1) that returns

exceed the predicted VaR over a period T for a given significance level. If the VaR model is correctly specified, the failure rate $\left(\hat{\pi} = \frac{T_1}{T}\right)$ should be equal to the pre-specified VaR level (α). The null hypothesis $H0: \pi = \alpha$ is evaluated through a likelihood ratio test:

$$LR_{uc} = -2ln\left(\frac{L(\Pi_\alpha)}{L(\Pi)}\right) = -2ln\left(\frac{(1-\alpha)^{T_0}\alpha^{T_1}}{(1-\hat{\pi})^{T_0}\hat{\pi}^{T_1}}\right) \xrightarrow{T\to\infty} \chi_1^2$$

where $T_0 = T - T_1$.

Two other tests by Christoffersen (1998) examine whether VaR exceedances are independent. We consider two states of nature each period: state 0 if the return does not fall below VaR: $r_t < VaR_\alpha$, and state 1, if $r_t < VaR_\alpha$. For the alternative hypothesis of VaR inefficiency, it is assumed that the process of violations $I_t(\alpha)$, where $I_t(\alpha) = 1$ if $r_t < VaR_\alpha$ and $I_t(\alpha) = 0$ otherwise, can be modeled as a Markov chain with $\pi_{ij} = \Pr[I_t(\alpha) = j | I_{t-1}(\alpha) = i]$. Let us then denote by T_{ij} the number of observations in state j after having been in state i in the previous period and define $T_0 = T_{00} + T_{10}$ and $T_1 = T_{11} + T_{01}$. The two probabilities of a VaR excess (state 1), conditional on the state of the previous period π_{01} and π_{11} are estimated by $\hat{\pi}_{01} = T_{01}/(T_{00} + T_{01})$ and $\hat{\pi}_{11} = T_{11}/(T_{10} + T_{11})$. Under the null hypothesis of independence of VaR exceedances: $\pi_{01} = \pi_{11} = \pi = (T_{11} + T_{01})/T$, the likelihood function is $L(\hat{\Pi}) = (1-\hat{\pi})^{T_0}\hat{\pi}^{T_1}$.

The likelihood under the alternative hypothesis is: $L(\hat{\Pi}_1) = (1-\hat{\pi}_{01})^{T_{00}}\hat{\pi}_{01}^{T_{01}}(1-\hat{\pi}_{11})^{T_{10}}\hat{\pi}_{11}^{T_{11}}$.

The independence test of Christoffersen (1998) is a test of the hypothesis of serial independence in VaR exceedances against a first-order Markov dependence. The likelihood ratio LR_{ind} statistic is: $LR_{ind} = -2\ln(L(\hat{\Pi})/L(\hat{\Pi}_1))$ with a distribution χ_1^2. The conditional coverage test is based on the likelihood ratio statistic, $LR_{cc} = -2\ln(L(\Pi_\alpha)/L(\hat{\Pi}_1)) = LR_{uc} + LR_{ind}$, which is asymptotically distributed χ_2^2.

While the conditional coverage test is easy to use, it is rather limited for two main reasons, *i*) the independence is tested against a very particular form of alternative dependence structure that does not take into account a dependence of order higher than one, *ii*) the use of a Markov chain

only considers the influence of past violations $I_t(\alpha)$ and not the influence of any other exogenous variable. The Dynamic Quantile Test proposed by Engle and Manganelli (2004) overcomes these two drawbacks of the conditional coverage test. These authors suggest using a linear regression model that links current violations to past violations. Let us define the auxiliary variable: $Hit_t(\alpha) = I_t(\alpha) - \alpha$, so that $Hit_t(\alpha) = 1 - \alpha$ if $r_t < VaR_{t|t-1}(\alpha)$ and $Hit_t(\alpha) = -\alpha$ otherwise. The null hypothesis of this test is that the sequence of hits (Hit_t) is uncorrelated with any variable that belongs to the information set Ω_{t-1} available when the VaR was calculated and it has a mean value of zero, which implies that the hits are not autocorrelated. The Dynamic Quantile test is a Wald test of the null hypothesis that all slopes in the regression model,

$$Hit_t(\alpha) = \delta_0 + \sum_{i=1}^{p} \delta_i Hit_{t-1} + \sum_{j=1}^{q} \delta_{p+j} X_j + \epsilon_t$$

are zero, where X_j are explanatory variables contained in Ω_{t-1}. The test statistic has an asymptotic χ^2_{p+q+1} distribution. In our implementation of the test, we use $p = 5$ and $q = 1$ (where $X_1 = VaR(\alpha)$) as proposed Engle and Manganelli (2004). By doing so, we are testing whether the probability of an exception depends on the level of VaR.

To evaluate the consequences of a VaR exceedance, we use the Asymmetric Linear Tick loss function (AlTick) proposed by Giacomini and Komunjer (2005), which takes into account the magnitude of the implicit cost associated with VaR forecasting errors. Hence, it takes into consideration not only the returns that exceed the VaR, but also the opportunity cost produced by an overestimation of VaR. When there are not exceptions, the loss function penalizes for the excess capital retained:

$$L_\alpha(e_{t+1}) = \begin{matrix} (\alpha - 1)e_{t+1} & if\ e_{t+1} < 0 \\ \alpha e_{t+1} & if\ e_{t+1} \geq 0 \end{matrix}$$

where $e_{t+1} = r_{t+1} - VaR_{t+1}$. Giacomini and Komunjer use the asymmetric linear loss function with α equal to the significance level used to forecast VaR. The AlTick function can be seen as the implicit loss function

whenever the object of interest is a forecast of a particular quantile of the conditional distribution of returns. That way, a VaR model is preferable if it has a lower average value of the loss function.

The different combinations of probability distributions and volatility specifications, applied to each of the 19 assets considered, yield a large number of VaR tests and it is hard to summarize so much information in order to achieve some clear-cut conclusion on the adequacy of each model.

Some authors compare VaR methodologies using a two-stage selection process. This approach proposed by Sarma et al. (2003) consists in removing in a first stage those methods or models that fail to pass statistical accuracy tests (backtesting) like those described above. The VaR models selected in this stage are then compared in a second stage on the basis of loss functions. Even though this two-stage selection approach helps in selecting a smaller set of competing models, it could fail to identify suitable models because they might have been removed in the first stage. Indeed, a model may be rejected in the first stage because of failing to pass a given test at a specific confidence level, despite producing a smaller loss than another one that has been judged to be statistically appropriate in the first stage. In the extreme case when we identify a single model as appropriate in the first stage, we would be making a decision based on statistical accuracy tests without taking into account the size of the losses beyond the VaR. Under that approach the VaR accuracy tests resemble more a decision-making process than an evaluation using loss functions.

Instead, we will proceed in the next section along four lines: i) the frequency of rejections of a given model when applying each test to the set of assets, ii) how often the p-value of a given test decreases when switching between two models differing in either the probability distribution or the volatility specification, iii) selecting the preferred models by a concept of Dominance among VaR models we introduce below, iv) implementing a Model Confidence Set approach to select the preferred VaR models for each asset. This approach is based on the use of a specific loss function. The first three criteria are based on properties of the tests for validation of VaR forecasts, while the fourth criterion deals with the size of the sample returns exceeding the estimated VaR.

2.7.1. Frequency of violations

For each asset, we calculate the number of observed violations of VaR forecasts, the statistic and p-value of each test for each combination of volatility model and probability distribution for the innovations[20]. Naturally, violation rates close to $\alpha = 0.01$ (13 violations) are desirable. Further, under the Basel Accord, models that over-estimate risk are preferable to those that under-estimate risk levels. In our case, less than 20 violations of VaR would define the 'green zone', between 20 and 50 violations would correspond to the 'yellow zone' and the 'red zone' would be defined by more than 50 violations[21]. In fact, falling inside the green zone is not necessarily a good thing if the number of violations of VaR is too low, since then the bank would be taking an excessive opportunity cost of capital.

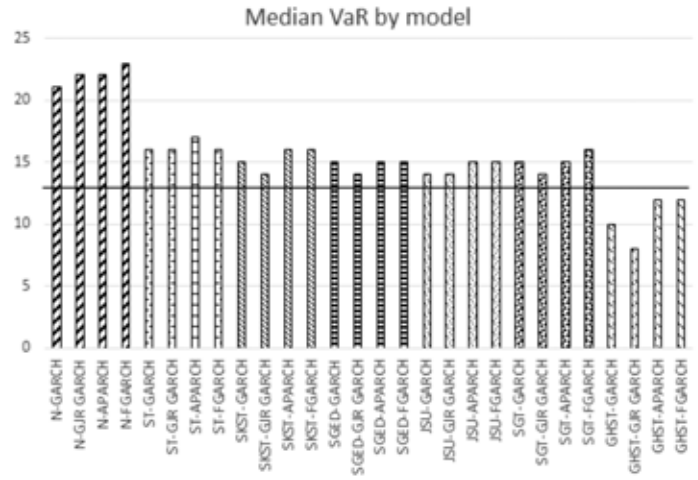
We never observed a model to fall in the red zone for any asset. The expected number of violations (13) falls in the green zone, so a good model should be in that zone. Across the 76 VaR analysis performed (4 volatility specifications and 19 assets) models under the Normal distribution fell in the green zone 26 times out of 76 (34%), 55 times under the Student-t distribution (72%), 72 times under SKST (95%), 69 times under SGED (91%), 75 times under JSU (99%), 73 times under SGT (96%) and 70 times under GHST (92%). All the other models fell in the yellow zone. The Normal distribution falls too often in the yellow zone. The frequency of the Student-t distribution to produce a model in the green zone was not very high either. All other probability distributions lead frequently to models in the green zone.

Figure 13 shows the median number of VaR violations for each combination of probability distribution and volatility specification. The Normal distribution leads to the largest median number of violations (22)

20. These results are available from the author upon request.

21. In terms of Basel Accord, based on a sample of 250 observations, if the number of exceptions is less than, or equal to 4 (the green zone), the test results are consistent with an accurate model and the possibility of erroneously accepting an inaccurate model is low. At the other extreme, if there are 10 or more exceptions (the red zone), the test results are extremely unlikely to have resulted from an accurate model, and the probability of erroneously rejecting an accurate model on this basis is remote. In between these two cases we have the yellow zone, where the backtesting results could be consistent with either accurate or inaccurate models, and the supervisor should encourage a bank to present additional information about its model before taking action. We have applied to these thresholds a scale factor based on our sample size of 1260 observations.

Figure 13: Median number of VaR violations for each model over the set of 19 assets



across the 76 models (4 volatility specifications and 19 assets). Since the expected number of violations is 13, the Normal distribution clearly underestimates the level of risk. The GHST distribution produces the lowest median number of violations (10), with a clear overestimation of risk. All the other probability distributions have a median number of violations around 15, with a slight underestimation of risk that is more evident for the Student-t distribution. We can say that except by the Normal and GHST distributions, all other distributions perform well. Being more specific, the median frequency of violations is 1.75% for models with Normal innovations, 1.27% for Student-t innovations, 1.19% for skewed Student-t, skewed Generalized Error and skewed Generalized-t innovations, 1.11% for Johnson S_U innovations and 0.79% for Generalized Hyperbolic skew Student-t innovations. According to the frequency of violations, the unbounded Johnson S_U distribution shows the best behavior among the asymmetric probability distributions. The performance of GHST might be acceptable under some criteria, although it would lead to an excessive opportunity cost of capital.

Differences among volatility specifications are much smaller. Models with a GARCH specification fell 114 times out of 133 cases (7 probability distributions and 19 assets) in the green zone (86%), 109 times for the GJR-GARCH (82%), 108 times for APARCH (81%) and 109 times for FGARCH (82%) out of 133 VaR analysis. The median number of violations was 15,

15, 16 and 16, respectively, very similar across volatility specifications. The frequency of violations for all volatility specifications is 1.19% for GARCH and GJR-GARCH, and 1.27% for APARCH and FGARCH models. This result already suggests the need to be careful when choosing an appropriate probability distribution for return innovations. Selecting the best volatility specification is also important, but the consequences of not making the right choice do not seem to be so crucial.

It is also interesting to examine the performance by asset type. Most models tend to overestimate risk in energy commodities (OIL and GAS). The median number of violations over the set of 28 models (7 probability distributions and 4 volatility specifications) is 7 for OIL and 5 for GAS (see Figure 14). A similar result is obtained for the GBP/USD and AUD/USD exchange rates, with a median number of 10 violations in both cases, which is not the case for the two other exchange rates[22]. But the general result is that more often than not, models tend to underestimate risk in all assets, with a number of violations above the expected value of 13. Underestimation is especially evident in the non-industrial metals (GOLD and SILVER) and some Spanish stock market variables (SAN and IBEX).

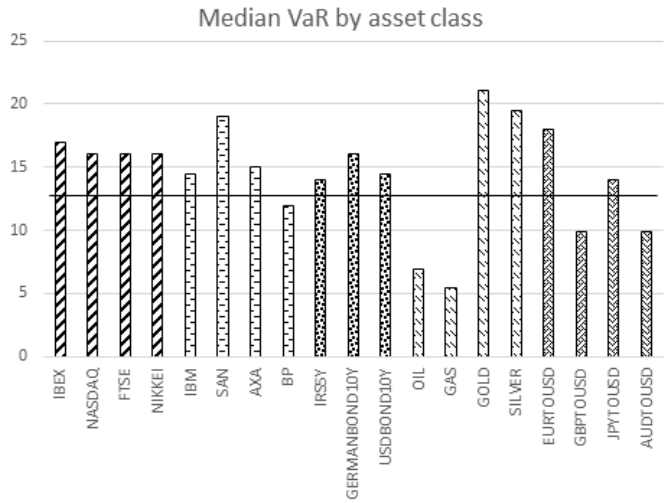
2.7.2. Switching between models

For 19 assets, we have a total of 216 tests performed under each probability distribution, and 378 tests under each volatility specification[23]. They produce a large amount of information, and we need to design ways to summarize that information in order to be able to draw some conclusion on the relative merits of each probability distribution and each volatility specification. This is what we do in the next sections.

We start by comparing, for each of the four VaR tests described above (Kupiec, independence, conditional coverage and Dynamic Quantile tests), the *p*-values of the test statistics for models that differ in either the probability distribution for the innovations or in the specification of

22. The median number of violations is also below 13 for BP, but it is so close to that target that we have to consider the difference as sampling error.
23. Notice that the independence and the conditional coverage tests not always can be applied.

Figure 14: Median number of VaR violations for each asset over the set of 28 models



volatility dynamics. In these tests the null hypothesis is H0: the VaR model is 'appropriate', in some sense that is specific to each test. As the probability of finding a similar sample with a more contrary evidence to H0, the *p*-value gives us a numerical indication on how favorable our sample to H0 is. Hence, when comparing any two VaR forecasting models, we should prefer the one with a higher *p*-value in VaR validation tests. To summarize the results of this analysis, Table 19 displays the number of cases in which the *p*-value of the test statistic increases or decreases when we change either the probability distribution or by the specification of the volatility model. We cannot make any formal testing, but by comparing *p*-values, we are searching for patterns that might suggest that a particular model is preferred over a given alternative.

If we consider all the possible specifications sharing the same probability distribution for return innovations, we see that switching from a Normal to a Student-t distribution for return innovations increases the *p*-value of VaR tests in 160 out of a total of 216 comparisons, suggesting in those cases a more accurate VaR model[24]. Even though the test statistics are obviously subject to sampling error, that frequency of increases

24. The number of possible comparisons arises from applying all the VaR tests to all the assets. The difference between this number and the sum of increases and decreases in the *p*-value is the number of cases in which the *p*-value of the test statistic does not change.

in *p*-value suggests that, as expected, the Student-t distribution is generally more appropriate than the Normal distribution to represent financial returns. Switching from the symmetric to the skewed Student-t distribution achieves a further increase in *p*-value in 114 comparisons, while decreasing in 75 cases. Moving from the asymmetric Student-t to the unbounded Johnson distribution achieves an increase in 91 cases while decreasing in 55 cases. Switching from the asymmetric Student-t (SKST) to other asymmetric distributions (SGT, JSU, SGED), the *p*-value increases more often than otherwise. On the contrary, if we switch from the SKST, SGED, JSU or SGT distributions to the GHST distribution, the opposite happens, with the *p*-value usually decreasing. Hence, we consider the SKST, SGED, JSU and SGT distributions to be preferable to GHST. Between these asymmetric distributions, switching to JSU or SGT leads to an increase in *p*-value in a greater number of cases.

Among volatility models, switching from the symmetric GARCH to GJR-GARCH increases the *p*-value of the statistic in 176 out of 378 comparisons. The *p*-value increases in 131 cases when switching from GJR-GARCH to APARCH, but it decreases in 167 cases.

On the other hand, if we move from the APARCH to the FGARCH model, the *p*-value increases in 151 out of 378 cases, decreasing in 128 cases. Overall, the FGARCH model seems to be the preferable volatility specification. Percent differences between the number of cases in which the value of the test statistic increases or decreases when switching between volatility models are much smaller than the ones obtained when switching between two probability distributions. This suggests again that, according to the performance of the models for VaR estimation, the specification of the volatility dynamics is not as important as the choice of probability distribution for the innovation in returns.

2.7.3. Dominance among VaR models

In the previous sections we have used four backtesting tests for VaR performance: the unconditional likelihood-ratio test, the independence test, the conditional coverage test, and the dynamic quantile test, and each test has been run for a variety of models and assets. In this section we evaluate the adequacy of the different models considered for VaR forecasting by comparing the specific situations in which each model has been rejected by each test.

Table 19: Number of cases in which the *p*-value of the test statistic increases or decreases when changing the probability distribution or the volatility model for all assets

For each test, the left (right) column shows the number of cases when the *p*-value increases (decreases) when switching between probability distributions (upper panel) or between volatility models (lower panel). The last two columns show the results when aggregating results for the four tests. LR_{uc} denotes the unconditional coverage test of Kupiec and LR_{ind} and LR_{cc} are the independence and the conditional coverage tests of Christoffersen, respectively. DQT denotes the Dynamic Quantile test. Rows with bold figures show the total number of tests run.

	LR_{uc}		LR_{ind}		LR_{cc}		*DQT*		Total	
Total number of statistics	**76**		**32**		**32**		**76**		**216**	
Increases/Decreases	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓
N → ST	64	12	8	24	30	2	58	18	160	56
ST → SKST	45	11	9	20	21	7	39	37	114	75
SKST → JSU	25	6	4	15	16	4	46	30	91	55
SKST → SGT	14	15	6	12	16	2	49	27	85	56
SKST → GHST	33	37	9	19	11	16	32	44	85	116
SKST → SGED	17	16	5	15	14	6	47	29	83	66
SGED → JSU	28	13	8	9	9	8	41	35	86	65
SGED → SGT	6	11	7	2	6	3	52	24	71	40
SGED → GHST	29	38	9	16	8	17	29	47	75	118
JSU → SGT	10	30	12	6	9	9	43	33	74	78
JSU → GHST	22	43	8	17	8	18	25	51	63	129
SGT → GHST	29	29	6	16	2	20	29	47	66	112
Total number of statistics	**133**		**56**		**56**		**133**		**378**	
Increases/Decreases	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓
GARCH → GJRGARCH	46	56	36	19	35	21	59	74	176	170
GJRGARCH → APARCH	32	50	25	16	16	26	58	75	131	167
APARCH → FGARCH	34	44	21	13	18	16	78	55	151	128

We base our analysis in a new concept of dominance between VaR models we introduce in this chapter[25]. Let us consider k tests that we apply to m alternative models to represent the dynamics of n assets. Each model is estimated for each of the n assets and subject to the k tests.

Definition 1. *We say that model M1 is dominated by model M2 if i) M1 has been rejected in at least as many cases as M2, and ii) whenever M2 is rejected by a test, M1 is also rejected.*

This definition introduces a transitive relationship among VaR models, although it is too strong to be satisfied in practice. So, we also consider the weaker concept of p-dominance.

Definition 2. *Given a confidence level between 0 and 1, we say that model M1 is p-dominated by model M2 if i) M1 has been rejected in at least as many cases as M2, and ii) in a percentage of at least p of the cases when M2 is rejected by a test, M1 is also rejected.*

In the special case $p = 1$ we have the first dominance criterion above. Unfortunately, for $p < 1$, p-dominance is not a transitive relationship. Notice that p does not need to be related to the confidence level at which VaR validation tests are implemented. We would expect p to be around .90 in most practical applications.

The interesting feature of this dominance criterion is that it compares any two model specifications across all the statistical tests and assets thereby allowing us to achieve some robust results. The criterion could accommodate different weights for each test depending on the relevance we want to assign them. The dominance criterion would then use the number of rejections in each test, weighted by relevance. An interesting possibility would consist of assigning a larger weight to tests having a larger ability to discriminate among models. Weights could also be chosen as a bounded function of the size of the test rejection, either in terms of the test statistic or the p-value of the test.

25. Sener et al. (2012) introduce a ranking model and a complementary predictive ability test statistic to investigate the forecasting performances of different Value at Risk (VaR) methods. The increasing literature on competitions among a wide array of alternative forecasting models is stimulating a well needed literature on this issue.

The dominance criterion could also be used to choose among forecasting models that are required to satisfy some condition to be considered acceptable. For instance, if competing models are used over a number of periods to forecast a given variable, and there is a maximum forecast error that is acceptable, the dominance relationship would be based on the number of periods in which each model exceeds that error threshold.

Table 20 contains the information needed to establish dominance comparisons. The upper panel corresponds to implementing the VaR validation tests at 99% confidence, while the lower panel has been obtained with test results implemented at 95% confidence. In each panel, the upper part compares the rejections of models using probability distributions D1 (left) and D2 (right) when combined with all the volatility specifications. The lower part compares the rejections of models made up with volatility specifications M1 (left) and M2 (right) when combined with all the probability distributions. The first two columns of each panel in Table 20 show the number of cases when the two probability distributions or the two volatility specifications listed in the first column have been rejected by the data when applying the unconditional coverage tests of Kupiec. The third column displays the percentage of rejections of D2 (M2) that were also rejections of D1 (M1). We will conclude that the probability distribution (or the volatility specification) with the lower number of rejections dominates the competitor when this percentage is below a pre-specified threshold for p. The following three columns refer to the independence tests, and the next columns come from the conditional coverage test and the Dynamic Quantile test. The final three columns aggregate the number of rejections across tests. For instance, if we take a threshold of .90 for p-dominance, the independence test of Kupiec rejected 36 models made up with the Normal distribution and just 7 models with the Student-t distribution. Besides, those 7 models rejected under the Student-t distribution were also rejected under a Normal distribution. Hence, the Student-t distribution dominates the Normal distribution according to this test. The independence test rejected 7 models made up with either the Normal or the Student-t distributions. In 5 of the 7 rejections under a Student-t distribution for return innovations the model was also rejected under a Normal distribution. That ratio is 5/7=0.714, so that we could not conclude that models with a Student-t distribution for return innovations dominate models with a Normal distribution according to the independence test.

Table 20: Dominance between VaR models

The upper panel shows results from tests implemented at 1% significance, while the lower panel shows results from tests at the 5% significance level. $n1$ is the number of tests in which H0 is rejected when D1 (M1) is specified as distribution (volatility model) for the different assets, $n2$ is the number of tests in which H0 is rejected when D2 (M2) is the probability distribution (volatility model) for the different assets and p is the proportion of times that H0 is rejected with both D2 (M2) and D1 (M1). Rows with bold figures show the total number of tests run.

Confidence level 99%	LR_{uc}			LR_{ind}			LR_{cc}			*DQT*			TOTAL		
Total number statistics	**76**			**32**			**32**			**76**			**216**		
D1 → D2	n1	n2	p	n1	n2	p	n1	n2	p	n1	n2	p	n1	n2	p
N → ST	36	7	1	7	7	0.714	25	13	1	44	29	1	112	56	0.964
ST → SKST	7	0	1	7	6	0.833	13	7	1	29	21	1	56	34	0.971
SKST → JSU	0	0	1	6	4	1	7	4	1	21	21	0.952	34	29	0.966
SKST → SGT	0	1	0	6	5	1	7	6	1	21	21	0.952	34	33	0.939
SGED → SKST	1	0	1	6	6	0.833	7	7	0.857	22	21	1	36	34	0.941
SGED → JSU	1	0	1	6	4	1	7	4	1	22	21	1	36	29	1
SGED → SGT	1	1	1	6	5	1	7	6	1	22	21	1	36	33	1
SGT → JSU	1	0	1	5	4	1	6	4	1	21	21	1	33	29	1
GHST → SKST	9	0	1	7	6	0.667	9	7	1	24	21	0.762	49	34	0.794
GHST → SGED	9	1	1	7	6	1	9	7	1	24	22	0.727	49	36	0.833
GHST → JSU	9	0	1	7	4	1	9	4	1	24	21	0.714	49	29	0.793
GHST → SGT	9	1	1	7	5	1	9	6	1	24	21	0.714	49	33	0.818
Total number statistics	**133**			**56**			**56**			**133**			**378**		
M1 → M2	n1	n2	p	n1	n2	p	n1	n2	p	n1	n2	p	n1	n2	p
GARCH → GJRGARCH	10	12	0.833	9	9	0.778	16	12	0.917	48	45	0.844	83	78	0.846
GJRGARCH → APARCH	12	14	0.714	9	13	0.615	12	23	0.609	45	46	0.739	78	96	0.688
APARCH → FGARCH	14	18	0.722	13	11	0.818	23	20	0.800	46	43	0.930	96	92	0.848

Confidence level 95%	LR_{uc}			LR_{ind}			LR_{cc}			DQT			TOTAL		
Total number statistics	76			32			32			76			216		
D1 → D2	n1	n2	p	n1	n2	p	n1	n2	p	n1	n2	p	n1	n2	p
N → ST	50	23	0.826	13	13	0.769	32	23	1	52	35	1	147	94	0.926
ST → SKST	23	6	1	13	16	0.813	23	18	1	35	27	0.963	94	67	0.940
SKST → JSU	6	3	1	16	17	0.941	18	17	1	27	25	0.960	67	62	0.968
SKST → SGT	6	6	1	16	16	0.875	18	17	1	27	28	0.964	67	67	0.955
SGED → SKST	8	6	0.833	17	16	1	18	18	1	28	27	1	71	67	0.985
SGED → JSU	8	3	1	17	17	0.941	18	17	1	28	25	1	71	62	0.984
SGED → SGT	8	6	1	17	16	1	18	17	1	28	28	0.964	71	67	0.985
SGT → JSU	8	3	1	16	17	0.882	17	17	0.941	28	25	1	67	62	0.866
GHST → SKST	21	6	0.833	17	16	0.938	19	18	0.944	30	27	0.926	87	67	0.925
GHST → SGED	21	8	0.875	17	17	0.882	19	18	0.944	30	28	0.929	87	71	0.901
GHST → JSU	21	3	1	17	17	0.882	19	17	0.941	30	25	1	87	62	0.952
GHST → SGT	21	6	1	17	16	0.875	19	17	0.941	30	28	0.929	87	67	0.925
Total number statistics	133			56			56			133			378		
M1 → M2	n1	n2	p	n1	n2	p	n1	n2	p	n1	n2	p	n1	n2	p
GARCH → GJRGARCH	23	30	0.700	32	30	0.867	40	37	0.865	56	51	0.863	151	148	0.831
GJRGARCH → APARCH	30	33	0.758	30	25	0.960	37	36	0.972	51	59	0.797	148	153	0.856
APARCH → FGARCH	33	31	0.968	25	22	0.955	36	32	1	59	59	0.915	153	144	0.951

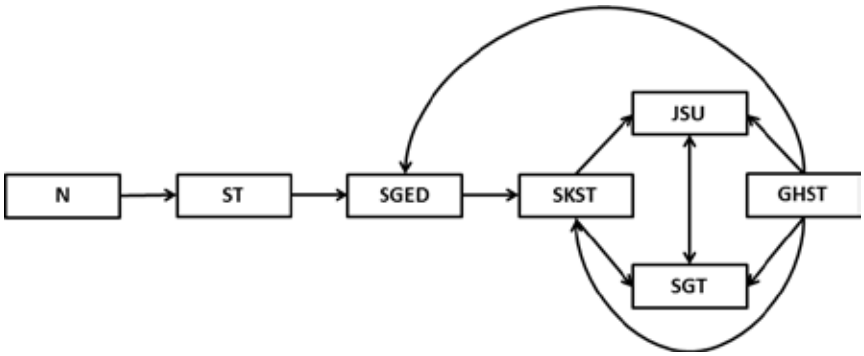
The number of pairwise comparisons between probability distributions or between volatility specifications is very high because they could be made in both directions, so we show in Table 20 the more interesting ones. For instance, we do not explicitly show the comparisons between the Normal distribution and asymmetric distributions because the latter always dominate. Similarly, we do not show pairwise comparisons between Student-t and any asymmetric distribution other than the skewed Student-t (SKST) because the skewed Student-t tends to *p*-dominate the

standard Student-t, and the majority of asymmetric distributions *p*-dominate in turn the skewed Student-t distribution[26].

Taking into account the aggregate results across the four tests we can summarize the comparisons at α = 95% as in the Figure 15:

Figure 15: Dominance relationship among probability distributions from aggregate results across the four tests at α = 95%

Each arrow head points to a model that dominates the model where the arrow originates. A two-headed arrow indicates two models that do not dominate each other.



No matter whether we take α = 99% or α = 95%, the Normal, Student-t, SKST and SGED distributions are dominated by other alternatives, specially JSU and SGT. We observe that JSU and SGT distributions seem to dominate all others, while not being dominated by each other. According to this dominance criterion the GHST distribution is judged again not to be appropriate for VaR estimation, since it is dominated by the rest of asymmetric distributions. The Normal distribution is also dominated by all other distributions.

At α = 99% there is not a clear dominance ordering between volatility specifications. For α = 95% the FGARCH specification seems to dominate but, once again, differences are not as clear as when comparing probability distributions.

26. Even though *p*-dominance is not transitive it seems safe to focus on the models that tend to be *p*-dominant.

A preference for APARCH and FGARCH models against standard GARCH and GJRGARCH has been a constant throughout our analysis up to this point. So, a robust conclusion is the need to incorporate a leverage effect in volatility and, possibly more important, the convenience to model standard deviations, rather than variances. The preference for asymmetric probability distributions in Table 20 is also consistent with results in Table 19 when comparing *p*-values of the test statistics. Both analyses are based on the same information, but they use it in a very different fashion. Nothing guarantees that the conclusions on the preferred probability distributions should be the same in both analyses. On the contrary, this coincidence should be seen as a proof of the robustness of such preference.

2.7.4. Model Confidence Sets

We calculate the values of the AlTick loss function using percent returns for different models and assets[27]. With a few exceptions, including a leverage effect in volatility reduces the loss function with independence of the assumption on the probability distribution of innovations. There is also a noticeable reduction in the value of the loss function when we move from symmetric to asymmetric distributions. Bold figures in each column show the VaR model that achieves the lowest value of the loss function for each asset. It is striking that the probability distributions that perform well according to other criteria do not do well according to the criterion of minimizing the loss function[28]. Something similar happens with the APARCH model. This suggests that loss functions should not be used by themselves. Different loss functions should be expected to yield different results, and there are not clear criteria to prefer one function versus another. Besides, we cannot say anything about whether differences are statistically significant and, sometimes, they are small. As we have done with the results of backtesting, we prefer to embed the evaluation of the AlTick loss function into a more complete approach to model selection that can provide us with some robust evidence on the performance of alternative VaR models.

27. These values are available from the author upon request.

28. In 5 cases, the minimum loss for each asset is achieved by a SGED and the Normal distributions, in 4 cases by SGT, in 3 cases by the GHST distribution and in 1 case by JSU and SKST. The FGARCH model achieves the minimum loss in 12 cases, the GARCH in 3 cases, APARCH and the GJRGARCH model in 2 cases.

The availability of several model specifications being able to adequately describe the unobserved data generating process (DGP) opens the question of selecting the 'best fitting model' according to a given optimality criterion. Recently, significant effort has been placed on developing testing procedures being able to deliver the 'best fitting' models among a set of alternatives. One of the first proposals was Diebold & Mariano (1995), but it is not applicable when the forecasts come from nested models or when the forecasts are calculated from semiparametric or non- parametric methods (Giacomini & Komunjer, 2005). This has been overcome by the Reality Check (RC) approach of White (2000), the Stepwise Multiple Testing procedure of Romano and Wolf (2005), the Superior Predictive Ability (SPA) test of Hansen and Lunde (2005), the Conditional test of Giacomini and White (2006), and the Model Confidence Set (MCS) procedure developed by Hansen, Lunde and Nason (2011). All these approaches are relevant from an empirical point of view, especially when the set of competing alternatives is large.

We implement the Model Confidence Set (MCS) procedure developed by Hansen, Lunde and Nason (2011) to discriminate among models. The MCS is a general approach to model selection that it does not assume knowledge of the correct specification. Furthermore, it does not require that the "true" model must be available as one of the competing models. This approach considers that all models have the same possibility of being correct and it compares them with each other. Another advantage of MCS is that it does not discard a model unless it is found to be significantly inferior relative to other models[29]. It is an appealing method to use when comparing a set of forecasting models because in practice it often cannot be ruled out that two or more competing models are equally good, being then members of the Set of Superior Models (SSM). In this sense, the MCS approach may be preferred over methods that search for a single model to be selected as the "best model".

The MCS procedure consists of a sequence of tests to construct the 'Set of Superior Models' (SSM). The MCS is a sequential testing procedure that eliminates at each step the worst model, until the hypothesis of

29. In this respect it is clearly different from the two-stage approach to model selection we described at the beginning of this Section.

Equal Predictive Ability (EPA) is not rejected for any of the models in the current SSM. On the other hand, each element in the SSM is characterized as having better predictive ability than models not in the set. The SSM has an interpretation similar to a confidence interval for a parameter in the sense that, with a given level of confidence, the SSM contains the best model. The EPA test statistic is evaluated under a given loss function, so that it is possible to test models on various aspects depending on the chosen loss function. The possibility of user supplied loss functions provides enough flexibility to the procedure that can be used to test competing models with respect to different dimensions. This is in common with Diebold & Mariano (1995), although we are here somewhat more specific in comparing whether the number and size of VaR violations are different across models. We apply the EPA tests using the AlTick loss function, not just the difference between observed and predicted returns, but results with other functions might be different.

Formally, the loss function $\ell_{i,t}$ associated to the i-th model $\ell_{i,t} = \ell(Y_t, \hat{Y}_{i,t})$ measures the cost associated to the difference between the observation at time t, Y_t, and $\hat{Y}_{it}$ the output of model i at time t. The MCS procedure starts from an initial set of models $\hat{M}_0$ of dimension m made up by all combinations of probability distributions and volatility specification considered in previous sections. Then, for a given confidence level $1-\alpha$, we obtain a smaller set, the superior set of models, SSM, $\hat{M}^*_{1-\alpha}$ of dimension $m^* \leq m$. Let us denote by d_{ij} the loss differential between models i and j,

$$d_{ij,t} = \ell_{i,t} - \ell_{j,t} \quad i,j = 1,\dots,m, \quad t = 1,\dots,T$$

The EPA hypothesis for a given set of models M can be formulated:

$$H_{0,M}: c_{ij} = 0, \qquad for\ all \quad i,j = 1,\dots,m$$

$$H_{1,M}: c_{ij} \neq 0, \qquad for\ some \quad i,j = 1,\dots,m$$

where $c_{ij} = \mathbb{E}(d_{ij})$ is assumed to be finite and not time dependent. This hypothesis can be tested using the test statistic [Hansen et al. (2011)],

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}} \quad for\ i,j \in M$$

where $\bar{d}_{ij} = n^{-1}\sum_{t=1}^{n} d_{ij,t}$ measures the relative sample loss between the i-th and j-th models, with $\widehat{var}(\bar{d}_{ij})$ is a bootstrapped estimate of $var(\bar{d}_{ij})$. Following Hansen et al. (2011) we calculate the bootstrapped variances by a block-bootstrap procedure. The block-bootstrap is the most general method to improve the accuracy of bootstrap for time series data. By dividing the data into several blocks, it can preserve the original time series dependency structure within a block. To that end, we divide the full time series (1260 data observations) into overlapping blocks of length k. The accuracy of the block- bootstrap is sensitive to the choice of block length, and the optimal block length depends on the sample size, the data generating process, and the statistic considered[30]. The block length p is usually estimated as the maximum number of significant parameters obtained by fitting an AR(p) process to all the d_{ij} terms. Since financial returns exhibit little linear autocorrelation, an AR(1) is enough to capture the dependence structure, so that we take $p = 1$ and resample individual observations. Using a block length of 2 would not change significantly the characterization of the MCS.

As discussed in Hansen et al. (2011) the EPA null hypothesis maps naturally into the statistic,

$$T_{R,M} = \max_{i,j \in M} \left| t_{ij} \right|$$

30. See Goncalves and White (2004, 2005), Künsch (1989), Liu and Singh (1992), and Politis and Romano (1994). Details about the implemented bootstrap procedure can be found in White (2000), Kilian (1999), Clark and McCracken (2001), Hansen et al. (2003), Hansen and Lunde (2005), Hansen et al. (2011) and Bernardi et al. (2016).

Since the asymptotic distributions of this test statistic is nonstandard, the relevant distribution under the null hypothesis needs to be estimated using a bootstrap procedure similar to that used to estimate $var(d_{ij})$.

Table 21 reports the frequency by which each probability distribution and each volatility specification enter into the Superior Set of Models for each asset using the AlTick loss function[31]. Tests are performed at the 90% confidence level, using a block-bootstrap procedure of 10000 resamples with a block length of 1. The table shows that for some assets, like NASDAQ 100, FTSE 100, EUR/SD and JPY/USD, the SSM includes a variety of distributions and volatility specifications. That indicates that the one-step ahead 1% VaR forecasting performance of the competing combinations of probability distribution and volatility specification is relatively similar, suggesting that for these assets the use of simple models for VaR forecasting may be justified.

The SGT, JSU, SGED and GHST distributions are the ones that enter most often into the MCS of the set of assets considered. Among the volatility models, FGARCH and APARCH seem to describe quite well the behavior of financial time series, although the symmetric GARCH also enters into the MCS quite often. Concerning the distribution specifications, we observe that the MCS confirms the common finding that the Normal distribution provides a poor description of the behavior of financial time series. Under the AlTick loss, the skewed Generalized-t and skewed Generalized Error distributions perform better than the Generalized Hyperbolic skew Student-t. Definitely, the Normal, Student-t and skewed Student-t distributions do not seem to be appropriate for VaR forecasting, at least for the wide set of financial assets considered in this chapter.

31. We believe that the opportunity cost of overestimating VaR is non-trivial. The AlTick loss function not only penalizes underestimation but also risk overestimation, because of the excess capital retained, and therefore we prefer it over other loss functions, such as those proposed by Lopez (1998, 1999) and Sarma et al. (2003) which only penalize risk underestimation. However, it would be worthwhile to explore other loss functions that might focus on different characteristics of VaR forecasts.

Table 21: Number of times that each probability distribution and volatility model enter into the Superior Set of models for each asset

Bold figures in the last column of the lower panel are aggregates for each probability distribution or volatility model. Bold figures in the last row of each panel display aggregates for each asset.

AlTick	IBEX	NASDAQ	FTSE	NIKKEI	IBM	SAN	AXA	BP	IRS	GER BOND
					Models					
GARCH	0	0	0	0	1	0	0	0	1	2
GJRGARCH	0	2	2	1	0	0	0	0	0	0
APARCH	2	5	5	0	0	0	0	2	0	0
FGARCH	3	4	3	2	0	1	1	0	0	3
					Distributions					
N	0	0	0	0	0	0	0	1	0	0
ST	0	1	0	0	0	0	0	0	0	0
SKST	0	2	2	0	0	0	0	0	0	0
SGED	2	2	2	2	1	0	1	1	0	1
JSU	0	2	3	0	0	1	0	0	0	2
SGT	2	3	3	1	0	0	0	0	1	1
GHST	1	1	0	0	0	0	0	0	0	1
Total Number	**5**	**11**	**10**	**3**	**1**	**1**	**1**	**2**	**1**	**5**

AlTick	US BOND	BRENT	GAS	GOLD	SILVER	EUR/ USD	GBP/ USD	JPY/ USD	AUD/ USD	TOTAL
					Models					
GARCH	6	0	0	1	4	2	0	3	0	**20**
GJRGARCH	0	0	0	0	1	2	1	0	0	**9**
APARCH	0	0	0	0	0	2	0	4	2	**22**
FGARCH	1	1	1	0	1	6	0	4	1	**32**
					Distributions					
N	0	1	1	0	0	0	1	0	2	**6**
ST	1	0	0	0	0	2	0	3	1	**8**
SKST	1	0	0	0	1	1	0	0	0	**7**
SGED	1	0	0	0	0	1	0	3	0	**17**
JSU	1	0	0	0	2	1	0	0	0	**12**
SGT	1	0	0	0	1	4	0	3	0	**20**
GHST	2	0	0	1	2	3	0	2	0	**13**
Total Number	**7**	**1**	**1**	**1**	**6**	**12**	**1**	**11**	**3**	

2.8. Conclusions

This chapter extends previous work on the forecasting performance of alternative VaR models by considering four volatility specifications: GARCH, GJR-GARCH, APARCH and FGARCH and a set of asymmetric probability distributions: skewed Student-t, skewed Generalized Error, unbounded Johnson, skewed Generalized-t and Generalized Hyperbolic skew Student-t distributions, some of them being relatively new to the financial literature. Standard symmetric distributions and GARCH models without leverage are also used as a benchmark. Our sample of daily data for assets of different nature for the January 2000-December 2015 period covers the recent financial crisis of 2007-2009.

Two clear results refer to issues that have been analyzed in previous research by a number of authors: *i*) VaR models that assume asymmetric probability distributions for return innovations, like the skewed Student-t distribution, skewed Generalized Error distribution, Johnson SU distribution, and skewed Generalized-t distribution achieve better VaR performance than models with symmetric distributions, *ii*) volatility models with leverage, like APARCH and FGARCH, show a better VaR performance than more standard GARCH and GJR-GARCH volatility specifications.

Our analysis highlights other important issues. A third result is that the shape and the skew of the assumed probability distribution for innovations are even more important for the performance of a Value at Risk model than including a leverage effect in volatility. This corroborates results by other authors (Lopez and Walter, 2000, Angelidis and Degiannakis, 2006 and Braione and Scholtes, 2016). We provide a thorough analysis of this issue by showing that the result holds for the wide set of assets we have considered: *i*) the frequency of rejections of VaR tests in models that differ in their volatility specification is similar, while rejection frequencies among models with the same volatility specification but a different probability distribution for the innovations can differ very significantly, *ii*) changing the probability distribution in a VaR model affects the *p*-value of the statistic for VaR tests by a larger amount than changing the volatility specification, *iii*) the dominance criterion we have introduced in this chapter establishes a clear ranking between models differing in their probability distribution, while the distinction between models that differ in their volatility specification is much less clear.

A fourth result deals with the fact that our estimates suggest that for a number of financial assets the true, unobserved volatility dynamics should not be specified in terms of the squared conditional standard deviation. Hence, models specified for the conditional variance are prone to produce biased results. Dealing with the power of the conditional standard deviation as a free parameter is an important feature of the APARCH/FGARCH volatility specifications which explains their better performance in validation tests of VaR forecasts.

Fifth, our analysis suggests that, as expected, a good fit of the moments of the distribution of returns usually leads to a good VaR performance. The MAE calculated over estimates for the four first moments selects the combination of a skewed Generalized Error distribution and an APARCH/FGARCH volatility specification as the best model to reproduce the skewness and kurtosis in asset returns.

According to VaR performance, switching to a Johnson SU or a skewed Generalized-t distribution tends to increase the *p*-value of VaR validation tests. In terms of the dominance criterion among VaR models we have introduced in this chapter, the unbounded Johnson and skewed Generalized-t dominate other asymmetric distributions like the skewed Student-t, the Generalized Hyperbolic skew Student-t and the skewed Generalized error distribution, as well as the symmetric distributions like Student-t and Normal. The skewed Generalized-t and skewed Generalized Error distributions perform better than the other distributions in terms of the Model Confidence Set procedure. According to all these analyses, FGARCH seems the preferred model to capture the volatility of financial time series, with APARCH as a close second. In summary, the combination of APARCH or FGARCH volatility with a skewed Generalized Error, skewed Generalized-t or unbounded Johnson SU distributions seem to be have the best VaR performance for a wide array of assets of different nature.

This evidence has been obtained trying to get broad and robust conclusions over the set of assets considered. But it could be the case that alternative VaR models provided different VaR performance for distinct asset classes. We have just a few assets of each class, which may explain the disparate results that are likely to arise if we repeat the analysis in the chapter by asset classes in our sample. But this is clearly an important issue that deserves being considered for further research.

CHAPTER 3. TESTING ES ESTIMATION MODELS: AN EXTREME VALUE THEORY APPROACH

3.1. Introduction

The Basel Committee on Banking Supervision (BIS) has recently chose Expected Shortfall (ES) as the market risk measure to be used for banking regulation purposes, replacing Value at Risk (VaR). The change is motivated by the superior properties of ES as a measure of risk, since it is based on information on the whole tail of the distribution of returns. The main drawback with the use of ES for risk regulation is the unavailability of simple tools for evaluation of ES forecasts (i.e. backtesting ES). In fact, the Basel Committee backed down on requiring the backtesting of ES. A debate started by Gneiting led many to believe that ES could not be backtested because it was not "elicitable". That point was settled recently by Fissler, Ziegel and Gneiting (2015) and by Acerbi and Szekely (2014), who demonstrated that lack of elicitability is not an impediment to backtesting ES. The latest Basel consultative document of January 2016, however, proposed to calculate risk and capital using ES, but to conduct backtesting only on VaR. VaR backtests are applied comparing whether the observed percentage of outcomes covered by the risk measure is consistent with the intended level of coverage. However, it is important that the capital reserve indicated by the VaR calculation could be tested, and the hypothesis that the level of reserves is adequate could be subject to a valid statistical test.

There is not much work evaluating and comparing the performance of ES forecasting models using recently introduced ES backtesting. Alexander and Sheedy (2008) develop a two-stage methodology for conducting stress tests whereby an initial shock event is linked to the probability of its occurrence. Working with three major currency pairs they found that results compared favorably with the traditional historical scenario stress testing approach. Jalal and Rockinger (2008) use a circular block bootstrap to take adequately into account the possible dependency among exceedances.

Applying the two-step procedure of McNeil and Frey (2000), they found that ES forecasts captured actual shortfalls satisfactorily. Ergün and Jun (2010) show that the Autoregressive Conditional Density (ARCD) model of Hansen (1994) with a time-varying conditional skewness parameter seems to provide more ES forecasts, beating forecasts from other GARCH-based models as well as from the extreme value theory (EVT) approach. Wong et al. (2012) compare ES forecasting models using the saddlepoint backtest proposed by Wong (2008). Righi and Ceretta (2015) evaluate unconditional, conditional and quantile/expectile regression-based models for ES forecasting using the ES backtest proposed by McNeil and Frey (2000) and a proposed test based on the standard deviation of returns beyond VaR. Clift, Costanzino and Curran (2016) apply three approaches recently proposed in the literature for backtesting ES by Wong (2008), Acerbi & Szekely (2014) and Costanzino & Curran (2015), but they only consider a GARCH volatility specification and a Normal distribution for ES forecasting. In these papers there is some indication on the benefits of using asymmetric probability distributions and EVT for ES forecasting[32].

We estimate VaR and ES at 1-day and 10-day horizons using standard conditional models as well as an EVT approach. For the latter, we use the two-step algorithm proposed by McNeil and Frey (2000) that fits a Generalized Pareto distribution to the extreme values of the standardized residuals generated by a given conditional volatility model. In both analysis we use asymmetric probability distributions for return innovations that are relatively new to the financial literature, and we analyze the accuracy of our estimates before and during the 2008 financial crisis using daily data. We take into account volatility clustering and leverage effects in return volatility by using the APARCH model (Ding, Granger and Engle, 1993) under different probability distributions for the standardized innovations: Gaussian, Student-t, skewed Student-t [Fernandez and Steel (1998)], skewed Generalized Error [Fernandez and Steel (1998)] and Johnson S_U [Johnson (1949)]. Then, we compare the out-of-sample 1-day and 10-day ahead ES forecast performance of all these models. For ES evaluation, we use the most recent ES backtesting proposals, which overcome the limitations of previous tests [McNeil &

32. Other studies have VaR as their primary measure of interest, leaving ES to a second level, such as Zhou (2012), Degiannakis, Floros and Dent (2013) and Tolikas (2014), where no extensive focus is placed on ES forecasting patterns.

Frey (2000), Berkowitz (2001), Kerkhof and Melenberg (2004) and Wong (2008)]. These are the test of Righi & Ceretta (2013), the first two tests of Acerbi & Szekely (2014) that are straightforward but require simulation analysis (as the Righi & Ceretta test) to compute critical values and p-values, the test of Graham & Pál (2014) which is an extension of the Lugannani-Rice approach of Wong (2008), the quantile- space unconditional coverage test of Costanzino & Curran (2015) for the family of Spectral Risk Measures of which ES is a member, and the conditional test of Du & Escanciano (2016). The last two tests can be thought of as the continuous limit of the Emmer, Kratz & Tasche (2015) idea in that they are joint tests of a continuum of VaR levels.

EVT has rarely been implemented beyond a one-day horizon when forecasting the ES of financial assets, even though there are several economic and practical reasons for computing long-term risk measures. Risk horizons longer than one day are particularly important for risk liquidity management, for long term strategic asset allocation as well as to compute capital requirements. Besides, the Basel Committee obliges banks to compute their level of risk over a 10-day horizon. The difficulty resides in getting enough homogeneous data on 10-day returns over non-overlapping periods. That explains the extended use of the scaling law, whose use is also proposed in the Basel Committee supervision documents. We get around this limitation by using Filtered Historical Simulation (FHS) to obtain time series of 10-day returns and we estimate 10-day ES by applying the same methodologies as for 1-day ahead ES forecasting.

To sum up, this work contributes to the literature in four ways. First, we compare the performance of the standard parametric approach with two alternatives to ES forecasting that take into account volatility clustering and asymmetric returns: EVT and the semi- parametric Filtered Historical Simulation. Second, we compare the results obtained under asymmetric probability distributions for return innovations with results under Normal and Student-t distributions. Third, we use the APARCH volatility specification because of its greater flexibility to represent the dynamics of conditional volatility (Garcia-Jorcano and Novales, 2017). Fourth, we forecast VaR and ES over a 10-day horizon as in Basel capital requirements and test ES forecasting models at this horizon, an analysis that has seldom been considered in the financial literature. Finally, we

examine the accuracy of risk models for ES forecasting during pre-crisis and crisis periods as well as under different significance levels. To the best of our knowledge, this is the first time that a systematic test of ES forecasting models is done considering a variety of probability distributions and two alternatives to the standard parametric approach, like EVT and the semi-parametric FHS.

3.2. Review of Literature

The quantiles of the distribution of returns (VaR) can be estimated by extreme value theory (EVT), which models the tails of the distribution of returns without making any specific assumption concerning the center of the distribution (Rocco, 2014). The tail index parameter in EVT can be estimated nonparametrically without assuming any particular model for the tail. There are many estimators that can be used to accomplish this task, such as Hill estimator (Hill, 1975) and Pickands estimator (Pickands, 1975).

For the estimation of the tail index parameters in EVT there are also two parametric approaches based on classical methods such as maximum likelihood. The first parametric approach is Block Maxima (BM) based on the Generalized Extreme Value (GEV), which divides the sample into m subsamples of n observations each and picks the maximum of each subsample; see for example Longin (2000), Diebold, Schuermann and Stroughair (2000). The second EVT parametric approach is the Peak Over Threshold (POT) based on the Generalized Pareto Distribution, according to which any observations that exceed a given high threshold, u, are modeled separately from non-extreme observations. McNeil and Frey (2000) show that the EVT method based on the Generalized Pareto distribution yields quantile estimates that are more stable than those from the Hill estimator. When working with threshold exceedances the choice of cut-off between the central part of the distribution and the tails may have severe consequences for risk estimates. If the threshold is chosen too low VaR forecasts will be biased and the asymptotic limit theorems will not apply. If the threshold is too large VaR forecasts will have large standard deviations due to the limited number of sample observations over the threshold.

An alternative to the unconditional approach is to calculate the conditional quantile. Under a parametric approach, the usual option to estimate the conditional quantile is assuming a particular distribution for return innovations. The most popular parametric distribution for standardized returns is Gaussian and Student-t distributions, and the Skewed Student-t distribution of Hansen (1994). An alternative leptokurtic and asymmetric distribution that has been considered in this context is the Skewed-Generalized-t (SGT) distribution proposed by Theodossiou (1998). The SGT distribution has the attractive feature of encompassing most of the distributions that are usually assumed for standardized returns, such as Gaussian, Generalized Error Distribution (GED), Student-t and Skewed Student-t distributions, for example. Recently, Ergen (2015) has considered the Skewed-t distribution proposed by Azzalini and Capitanio (2003) and Aas and Haff (2006) propose the use of the Generalized Hyperbolic Skew Student-t distribution for unconditional and conditional VaR forecasting.

Another possibility is to estimate the conditional quantile using the EVT approach. Danielsson and de Vries (2000) and McNeil and Frey (2000) suggest estimating the quantiles of return innovations by applying EVT to the standardized returns, which are i.i.d. if the conditional mean and variance are specified correctly. Chan and Gray (2006) introduce a description of the conditional EVT and its application to the forecasting of the VaR of daily electricity prices. McNeil and Frey (2000) propose filtering returns by estimating a GARCH model, then applying EVT to the tails of the empirical distribution of innovations while bootstrapping to the central part of the distribution. They verify that the General Pareto distribution of EVT results in better estimates for ES than the Gaussian model. Jalal and Rockinger (2008) show that this procedure appears to perform a remarkable job when combined with a well-chosen threshold estimation, such as that in Gonzalo and Olmo (2004).

Following non-parametric methods, innovation quantiles can be estimated using bootstrapping, which does not need to assume any particular probability distribution (Ruiz and Pascual, 2002). In particular, Barone-Adesi, Giannopoulus and Vosper (1999, 2002) propose a bootstrap method known as filtered historical simulation (FHS), which is based on the idea of using random draws with replacement from the standardized residuals. Bootstrap procedures have the advantage that they allow

for the construction of confidence intervals for VaR estimates. Pascual, Ruiz and Romo (2006) propose a bootstrap procedure that allows for the incorporation of parameter uncertainty. Kourouma et al. (2011) compare unconditional and conditional historical simulation and EVT in VaR and ES forecasting. They conclude that conditional EVT model is more accurate and reliable for VaR forecasting, according to the rate of violations and Wald, Kupiec and Christoffersen tests, and for ES forecasting, according to an ES test proposed by them that is based on the average difference between realized returns and the predicted ES.

As regards ES, in spite of its advantages as a measure of risk it is still less used than VaR. However, the Basel committee (2016) has recently placed a stronger emphasis on ES and backtesting ES is clearly in the future agenda for capital requirements at financial institutions. The problem is that backtesting ES is much harder than backtesting VaR.

Recently, some ES backtesting procedures have been developed, like the residual approach introduced by McNeil and Frey (2000), the censored Gaussian approach proposed by Berkowitz (2001), the functional delta approach of Kerkhof and Melenberg (2004), and the saddlepoint technique introduced by Wong (2008). While Berkowitz's censored Gaussian approach and Kerkhof & Melenberg's functional delta method rely on large samples for convergence to the required limiting distributions, the saddlepoint techniques proposed by Wong are accurate and have reasonable test power even if the sample size is small. The saddlepoint technique makes use of a small sample asymptotic method that involves higher order moments of the underlying distribution and is able to approximate to a very high degree of accuracy the required tail probability even for very small sample sizes. But this test has a few disadvantages, such as the Gaussian distribution assumption and the full distribution conditional standard deviation that is used as the dispersion measure.

However, these approaches present some drawbacks. The backtest of McNeil and Frey (2000), Berkowitz (2001) and Kerkhof and Melenberg (2004) rely on asymptotic test statistics that might be inaccurate when the sample size is small, and this could penalize financial institutions because of an incorrect forecasting of ES. Further, these tests compute the required p-value based on the full sample size rather than conditional on the number of exceptions. The test proposed by Wong (2008) is

robust to these questions, making it possible to detect failure of a risk model based on just one or two exceptions before any more data is observed. Nonetheless, the Wong (2008) backtest has some disadvantages, such as the Gaussian distribution assumption, and the use of the full distribution conditional standard deviation as a dispersion measure.

To overcome these limitations, Emmer, Kratz & Tasche (2015) propose a new ES backtest based on a simple linear approximation, in which the ES estimate is obtained as the average of quantiles at different VaR levels. The ES estimate is considered acceptable if all the VaR estimates pass Kupiec test. Also, the test proposed by Righi & Ceretta (2013) verifies whether the average observed deviations from the ES for those returns below VaR is zero. In this test returns are standardized using the mean and standard deviation from the distribution of returns truncated to the left of VaR. Later, Acerbi & Szekely (2014) introduced three model-free, non-parametric backtesting methodologies for ES that are shown to be more powerful than the Basel VaR test. Graham & Pál (2014) generalized Wong's result in a tractable and intuitive manner to allow for any VaR modeling, and therefore distributional, approach. Costanzino & Curran (2015) developed a methodology that can be used to backtest any spectral risk measure, including ES. It is based on the idea that ES is an average of a continuum of VaR levels. They introduce an unconditional ES backtest similar to the unconditional VaR backtest of Kupiec, to test whether the average cumulative violation is equal to $\alpha/2$. Later, Du & Escanciano (2016) proposed backtesting for ES based on cumulative violations, which are the natural analogue of the commonly used conditional backtest for VaR, extending the results obtained by Costanzino & Curran (2015).

Several papers have considered a comparison between alternative ES forecasting models: Kourouma et al. (2011) introduce a validation test for ES models and use it to compare unconditional and conditional ES forecasting models at 1-, 5- and 10-day horizons. As conditional model they specify a GJR-GARCH under a Normal distribution for return innovations. Wong et al. (2012) compare conditional models with GARCH and APARCH volatility specifications and Normal, Student-t, skew Student-t and EVT combined with a Normal distribution using the ES validation tests introduced in Wong (2008, 2010). Righi & Ceretta (2015) analyze a much richer variety of alternative models and methods for ES

forecasting using the McNeil & Frey (2000) test and the Righi & Ceretta (2015) test. Clift et al. (2016) consider the Wong test (2010), Costanzino & Curran test (2015) and Acerbi & Szekely tests (2014) but use simple specifications as illustration: a constant volatility model and a GARCH model under Normality.

We use five different approaches for evaluation of ES estimates, with six methods overall. The test of Righi & Ceretta (2015) and the first two tests of Acerbi & Szekely (2014) are straightforward, but require simulations, the test of Graham & Pál (2014), which is an extension of the Lugannani-Rice approach of Wong (2010), the quantile-space unconditional coverage test of Costanzino & Curran (2015) for the family of Spectral Risk Measures, of which ES is a member and, finally, the conditional test of Du & Escanciano (2016). The last two tests can be thought of as the continuous limit of the Emmer, Kratz & Tasche (2015) idea in that it is a joint test of a continuum of VaR levels.

3.3. Background

3.3.1. Standard Risk Measures

Value at Risk (VaR) is a simple risk measure that tells us what loss will be exceeded only a small percentage of times in the next k trading days ($100\alpha\%$). Thus, given the log-return r_{t+k} of a portfolio in period $t+k$, VaR at a level α is defined as $\Pr(r_{t+k} < VaR_{t+k}^{\alpha}) = \alpha$. For simplicity, let us assume at we are predicting the VaR at some level α for 1-day ahead returns, $r_{t+1} = \mu_{t+1} + \sigma_{t+1}z_{t+1}$ where μ_{t+1} is the conditional mean return in period $t+1$, σ_{t+1}^2, is the conditional variance and z_{t+1} represents the white noise time series of return innovations, which will follow a given probability distribution. From the VaR_{t+1}^{α} definition we have, $\Pr(z_{t+1} < (VaR_{t+1}^{\alpha} - \mu_{t+1})/\sigma_{t+1}) = \alpha$, which amounts to $F((VaR_{t+1}^{\alpha} - \mu_{t+1})/\sigma_{t+1}) = \alpha$, or

$$VaR_{t+1}^{\alpha} = \mu_{t+1} + \sigma_{t+1}F^{-1}(\alpha) \tag{6}$$

where F denotes the probability distribution function of return innovations z_t. Given the drawbacks of VaR as a risk measure, it is convenient

to compute the ES, which accounts for the magnitude of large losses as well as their occurring probability. The ES is defined from VaR as $ES_{t+k}^{\alpha} = \mathbb{E}_{t+k}[r_{t+k}|r_{t+k} < VaR_{t+k}^{\alpha}]$ and tells us the expected value of the loss k-days ahead, conditional on it being worse than the VaR. As VaR is usually negative for low α values, the expectation of returns below VaR will also be negative. For the 1-day ahead ES we have $ES_{t+1}^{\alpha} = \mathbb{E}_{t+1}[r_{t+1}|r_{t+1} < VaR_{t+1}^{\alpha}] = \mu_{t+1} + \sigma_{t+1}\mathbb{E}_{t+1}[z_{t+1}|z_{t+1} < (VaR_{t+1}^{\alpha} - \mu_{t+1})/\sigma_{t+1}]$. Finally, using (6) we get,

$$ES_{t+1}^{\alpha} = \mu_{t+1} + \sigma_{t+1}\mathbb{E}_{t+1}[z_{t+1}|z_{t+1} < F^{-1}(\alpha)] \tag{7}$$

If we assume the existence of an absolutely continuous cdf F, ES is defined as

$$\mathbb{E}_{t+1}[z_{t+1}|z_{t+1} < F^{-1}(\alpha)] = \frac{1}{\alpha}\int_0^{\alpha} F^{-1}(s)ds = \frac{1}{\alpha}\int_{-\infty}^{F^{-1}(\alpha)} rf(r)dr$$

3.3.2. Estimating risk: Conditional models for the full distribution

We now define VaR and ES in conditional models. For that purpose, we consider that R is a stationary process with a fully parametric location-scale specification based on the expectation, dispersion and random components: $r_t = \mu_t + \sigma_t z_t$, where for period t, r_t is the return of an asset, μ_t is the conditional mean (location), σ_t the conditional standard deviation (scale) and z_t represents a zero location and unit scale innovation white noise series, which can assume many probability distribution functions F. Under this specification, the risk measures become,

$$VaR_t^{\alpha} = \mu_t + \sigma_t F^{-1}(\alpha)$$

$$ES_t^{\alpha} = \mu_t + \sigma_t\left(\frac{1}{\alpha}\int_0^{\alpha} F^{-1}(s)ds\right)$$

$$SD_t^{\alpha} = \left[\sigma_t^2\frac{1}{\alpha}\int_0^{\alpha}\left(F^{-1}(s) - \left(\frac{1}{\alpha}\int_0^{\alpha} F^{-1}(s)ds\right)\right)^2 ds\right]^{1/2}$$

The SD measure in the last expression is the dispersion around the expected value truncated by the VaR. This will be considered for ES backtesting of Righi & Ceretta.

3.3.3. Estimating risk: Conditional models for extreme events

The alternative approach is to consider only extreme events, precisely those captured by risk measures. In this regard, the Extreme Value Theory (EVT) is concerned with the distribution of the smallest order statistics and it considers only the tail of the distribution of returns. For further reference, see Longin (2005). Although EVT is interesting for risk modeling, the stylized facts make the *iid* assumption inappropriate for most financial data. To solve this issue, one should apply the EVT analysis to the filtered residuals z_t of a previously estimated model, as proposed by Diebold, Schuermann and Stroughair (2000) and McNeil and Frey (2000). This is possible because under a correct model specification, the filtered residuals will be approximately *iid*, an assumption of EVT modeling.

Under the EVT approach, VaR and ES are modeled using the concept of threshold exceedance. The peaks-over-threshold (POT) models developed around this concept center on the analysis of the Generalized Pareto distribution, which may be understood as a limiting tail distribution for a wide variety of commonly studied continuous distributions. The POT is the typical approach used in finance. Under the *iid* assumption, let us consider the distribution function of excesses $Y = u - Z$ over a high, fixed threshold u, $F_u(y) = P(Y = u - Z \leq y | Z < u) = [F(u) - F(u - y)]/[F(u)], y \geq 0$[33]. Pickands (1975) shows that the Generalized Pareto distribution (GPD) arises naturally as the limit distribution of the scaled excesses of identical and independently distributed (*iid*) random variables over high thresholds. We say that excesses from a given threshold follow a General Pareto distribution $Y = u - Z \sim GPD(\xi, \beta)$ if

33. Notice that we focus on the lower tail of the data, and we have adapted all formulations accordingly. The choice of u is subject to a trade-off: very high u leads to an estimator with large variance, while low u induces bias. The choice of u is the most important implementation issue.

$$F_u(y) \approx GPD_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\beta}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\beta}\right), & \xi = 0 \end{cases}$$

$GPD_{\xi,\beta}(y)$ has support $y \geq 0$ if $\xi \geq 0$ and $0 \leq y \leq -\beta/\xi$ if $\xi < 0$ where $\beta > 0$ is a scale parameter and ξ is the tail shape parameter, which is crucial because it governs the tail behavior of $GPD_{\xi,\beta}(y)$. The case $\xi > 0$ corresponds to heavy-tailed distributions whose tails decay like power functions, such as Pareto, Student-t, Cauchy, Burr, loggamma and Fréchet distributions. In this case, the tail index parameter equal to $1/\xi$ corresponds to, for example, the degrees of freedom of the Student-t distribution. The case $\xi = 0$ corresponds to distributions like Normal, exponential, gamma and lognormal, whose tails essentially decay exponentially. The final group of distributions are short-tailed distributions ($\xi < 0$) with a finite right endpoint, such as the uniform and beta distributions.

The implied assumption is that the tail of the underlying distribution begins at the threshold u. From our sample of T data a random number T_u will exceed this threshold. If we assume that the T_u excesses over the threshold are *iid* with exact GPD distribution, Smith (1987) has shown that maximum likelihood estimates $\hat{\xi} = \hat{\xi}_N$ and $\hat{\beta} = \hat{\beta}_N$ of the GPD parameters ξ and β are consistent and asymptotically normal as $T_u \to \infty$, provided $\xi > -1/2$. Under the weaker assumption that the excesses are *iid* from a $F_u(y)$ which is only approximately GPD he also obtains asymptotic normality results for ξ and β.

Consider now the following equality for points $z < u$ in the left tail of F:

$$F(z) = F(u) - F_u(u - z)F(u) = F(u)(1 - F_u(u - z))$$

If we estimate the first term on the right-hand side of the equation using the proportion of tail data T_u/T, and if we estimate the second term by approximating the excess distribution with a generalized Pareto distribution fitted by maximum likelihood, we get the tail estimator

$$\hat{F}_Z(z) = \frac{T_u}{T}\left(1 + \hat{\xi}\frac{u - z}{\beta}\right)^{-1/\hat{\xi}}$$

It is very important to note that the distribution F of the conditional model and the distribution $GPD_{\xi,\nu}$ for $\{z\}$ over threshold u are not linked. Thus, it is possible to use any conditional model to filter the data before applying EVT to z_T. In our analysis we assume a variety of asymmetric distributions for F that give rise to different conditional EVT estimates. Under the EVT approach, the risk measures are obtained,

$$VaR_t^\alpha = \mu_t + \sigma_t\left(u + \frac{\beta}{\xi}\left[1 - \left(\frac{\alpha}{T_u/T}\right)^{-\xi}\right]\right)$$

$$ES_t^\alpha = \mu_t + \sigma_t\left(\frac{1}{\alpha}\int_0^\alpha F_{z,u}^{-1}(s)ds\right) = \mu_t + \sigma_t\left(\frac{F_{z,u}^{-1}(\alpha)}{1-\xi} - \left(\frac{\beta + \xi u}{1-\xi}\right)\right)$$

$$SD_t^\alpha = \left[\sigma_t^2\frac{1}{\alpha}\int_0^\alpha\left(F_{z,u}^{-1}(s) - \left(\frac{1}{\alpha}\int_0^\alpha F_{z,u}^{-1}(s)ds\right)\right)^2 ds\right]^{1/2}$$

Summarizing, McNeil & Frey proceed as follows: In the first step, they filter the dependence in the time series of returns by computing the residual of a GARCH-type model, which should be *iid* if the GARCH-type model correctly fits the data. In the second step, they model the extreme behavior of the residual using the tail approach explained above. Finally, in order to produce a VaR forecast of original returns, they trace back the steps by first producing the α-quantile forecast for the GARCH-type filtered residuals and transforming the α-quantile forecast for the original returns using the conditional forecast at the required horizon.

It is worth emphasizing that the GARCH-EVT approach incorporates the two ingredients required for an accurate evaluation of the conditional VaR, i.e. a model for the dynamics of the first and second return moments, and an appropriate model for the conditional distribution of returns. An obvious improvement of this approach as compared to the unconditional EVT is that incorporates in VaR forecasting changes in expected return and volatility. For instance, if we assume a change in

volatility over the recent period, the GARCH-EVT is able to incorporate this new feature in its VaR evaluation, whereas the unconditional EVT would remain stuck at the average level of volatility over the estimation sample.

McNeil & Frey (2000) also perform a backtesting experiment in which they compare the performance of various methods to correctly reproduce the quantiles of several asset returns. They show that the GARCH-EVT performs much better than unconditional EVT, suggesting that the ability to capture changes in volatility is crucial for VaR computation.

3.3.4. Estimating risk: Filtered Historical Simulation

The standard historical approach is often limited to the 1-day horizon because of the lack of enough historical data to use non-overlapping h-day returns. Using overlapping h-day returns would distort the tail behavior of the return distributions leading to significant error in VaR and ES forecasts at extreme quantiles. An alternative for VaR and ES forecastings at risk horizons longer than one-day is Filtered Historical Simulation (FHS). Barone-Adesi et al. (1998, 1999) extend the idea of volatility adjustment to multi-step historical simulation, using overlapping data in a way that does not create blunt tails for the h-day portfolio return distribution. Their idea is to apply a statistical bootstrap to the residuals of a parametric dynamic model of returns, to simulate log returns on each day over the risk horizon. Typically, the model will incorporate a specification of the GARCH family for volatility dynamics. The filtering involved in the FHS approach allows for h-day return distributions to be generated from overlapping samples, since the bootstrap allows for increasing the number of observations used for building the h-day return distribution.

FHS is in fact a hybrid method combining some attractive features of both historical and Monte Carlo VaR models. The advantages of FHS approach are 1) it captures current market conditions by means of the volatility dynamics, 2) no assumptions need to be made on the distribution of the return innovations and 3) the method allows for the computation of any risk measure at any investment horizon of interest because we can generate as many h-day returns as we like.

Suppose that at a time s, we want to simulate returns for the next h days. We select $\{z^*_{s+1,}z^*_{s+2}, \dots, z^*_{s+h}\}$ at random with replacement (statistical bootstrap) from the set of standardized innovations from our model $\{\hat{z}_1, \hat{z}_2, \dots, \hat{z}_s\}$ after filtering out APARCH and AR models. We use the APARCH model to simulate future returns for dates $t = s+1, s+2, \dots, s+h$:

$$\sigma^*_t = (\hat{\omega} + \hat{\alpha}_1(|\varepsilon^*_{t-1}| - \hat{\gamma}_1\varepsilon^*_{t-1})^{\hat{\delta}} + \hat{\beta}_1(\sigma^*_{t-1})^{\hat{\delta}})^{1/\hat{\delta}} \tag{8}$$

$$\varepsilon^*_t = z^*_t\sigma^*_t \tag{9}$$

$$r^*_t = \hat{\phi}_0 + \hat{\phi}_1 r^*_{t-1} + \varepsilon^*_t \tag{10}$$

The algorithm contains the following steps,

(i) Select $\{z^*_{s+1,}z^*_{s+2}, \dots, z^*_{s+h}\}$ drawing randomly with replacement from $\{\hat{z}_1, \hat{z}_2, \dots, \hat{z}_s\}$.

(ii) Take as initial values the last estimates: $\sigma^*_s = \hat{\sigma}_s, \varepsilon^*_s = \hat{\varepsilon}_s, r^*_t = r_t$.

(iii) Set up for $t = s+1, s+2, \dots, s+h$,

- Plug σ^*_{t-1} and ε^*_{t-1} in equation (8) to get σ^*_t.
- Plug z^*_t (from step (i)) and σ^*_t in equation (9) to get ε^*_t.
- Plug r^*_{t-1} and ε^*_t in equation (10) to get r^*_t .
- Then the simulated log return over h days ($r^*_{s:h}$) is the sum $r^*_{s+1} + r^*_{s+2} + \cdots + r^*_{s+h}$

(iv) Repeating this procedure N times yields N simulated h-day returns, $r^*_{i,s:h}$, $i = 1, 2, \dots, N$.

We compute h-day ahead VaR and ES forecasts as

$$VaR^{\alpha}_{s:h} = Percentile\{r^*_{i,s:h}, i = 1, \dots, N; 100\alpha\}$$

$$ES^{\alpha}_{s:h} = (N\alpha)^{-1}\sum_{i=1}^{N}\left(r^*_{i,s:h}\mathbb{I}_{\{r^*_{i,s:h} < VaR^{\alpha}_{s:h}\}}\right)$$

where $\mathbb{I}$ is the indicator function that assumes value 1 if the h-day return $r^*_{i,s:h}$ is lower than VaR and 0 otherwise. Thus, the ES is just the mean of the simulated returns below VaR. Finally,

$$SD^{\alpha}_{s:h} = \left\{(N\alpha)^{-1}\sum_{i=1}^{N}\left[\left(r^*_{i,s:h}\mathbb{I}_{\{r^*_{i,s:h}<VaR^{\alpha}_{s:h}\}}\right) - ES^{\alpha}_{s:h}\right]\right\}^{1/2}$$

and, thus SD is just the standard deviation around the ES, considering only the values below VaR.

We use an expanding window to estimate the model, starting with the 2915 observations from the 10/2/2000-12/2/2011 period. Each day we add a new observation, estimate the models and apply the algorithm to generate N = 5000 h-day ahead return simulations from which we compute forecasts for VaR and ES. The forecasting exercise extends over 1260 days, the last five years in our sample, 12/5/2011-9/30/2016, obtaining daily forecasts of h-day ahead of the VaR, ES and SD risk measures.

Following the McNeil & Frey (2000) proposal, under the EVT approach we estimate the ξ and β parameters by fitting the Generalized Pareto distribution (GPD) to the left tail of standardized return innovations. We generate N = 5000 simulations for the h-day ahead return $r^*_{s:h}$ using a combination of bootstrapping in-sample residuals from the fitted models (i.e., FHS) and GPD simulation. We apply the following algorithm, which was also proposed independently by Danielsson and de Vries (2000),

(i) Use bootstrapping to randomly sample from the standardized innovations for each future period and for each of the N trajectories.

(ii) If a selected innovation z^* is below the threshold (u), we draw a realization y from the previously estimated GPD($\hat{\xi}$, $\hat{\beta}$). The value y is taken as the excess below the threshold u, i.e., the numerical value of the innovation to be used in simulation will be: $z^* = u - y$.

(iii) Otherwise, return standardized innovations themselves.

(iv) Finally, we trace back from simulated standardized innovations to recover the returns and we end up with N sequences of hypothetical daily returns for day $s + 1$ through day $s + h$. From these, we calculate the hypothetical h-day returns as $r^*_{s:h} = \sum_{h=1}^{H} r_{i,s+h}$ for $i = 1, 2, \ldots, N$ „ and we can calculate the h-day VaR, h-day ES and h-day SD, as described above.

(v) We repeat this procedure for $s + 1$, $s + 2$, $s + 3$, ..., $s + 1259$, to cover the out-of-sample period.

3.4. Data and Estimation Models

We work with daily percentage returns on assets over the sample period 10/2/2000 - 9/30/2016 (4175 sample observations). Daily returns are computed as 100 times the difference of the log prices, i.e. $100[ln(P_{t+1}) - ln(P_t)]\%$. The financial assets considered are: International Business Machines [IBM] ($), Banco Santander [SAN] (€), AXA [AXA] (€) and BP [BP] (£). The data were extracted from Datastream.

Table 22 reports descriptive statistics for the daily percentage return series. All of them have a mean close to zero. Median returns are zero. SAN has the largest total range ($max - min$) and BP has the smallest range. The unconditional standard deviation (S.D.) is around 2, with AXA having the ighest and IBM the lowest one for IBM. All assets have negative skewness, except AXA. For all assets considered the kurtosis statistic is large, implying that the distributions of those returns have much thicker tails than the Normal distribution. Accordingly, the Jarque-Bera statistic (J-B) is statistically significant, rejecting the assumption of Normality in all cases.

Table 22: Descriptive statistics for daily percent returns

Sample: 10/2/2000 - 9/30/2016 (4175 daily observations). Mean and median returns in basis points. S.D. is the standard deviation, J-B is the Jarque-Bera test statistic.

	Mean (bps)	Median (bps)	Max	Min	S.D.	Skewness	Kurtosis	J-B
IBM	0.83	0	11.35	-16.89	1.58	-0.22	12.33	15194.87
SAN	1.56	0	20.88	-22.17	2.26	-0.07	10.50	9793.17
AXA	1.47	0	19.78	-20.35	2.69	0.19	10.24	9155.81
BP	-0.69	0	10.58	-14.04	1.69	-0.19	8.01	4390.88

To perform an ES analysis, we estimate the APARCH volatility model (Ding, Granger and Engle, 1993) under the different probability distributions for return innovations: Gaussian, Student-t, skewed Student-t, skewed Generalized Error and Johnson S_U. An AR(1) model was considered for the conditional mean return, which is sufficient to produce serially uncorrelated innovations[34]. The APARCH model is particularly successful in capturing the heteroscedasticity exhibited by the data due to the power of the conditional standard deviation is a free parameter, which provides more flexibility to the dynamics of volatility.

For a given return series $r_1, \cdots, r_T$, the model adopted is

$$r_t = \phi_0 + \phi_1 r_{t-1} + \varepsilon_t \qquad \varepsilon_t = \sigma_t z_t \qquad t = 1, 2, \ldots, T$$

$$\sigma_t^\delta = \omega + \alpha_1(|\varepsilon_{t-i}| - \gamma_1 \varepsilon_{t-i})^\delta + \beta_1(\sigma_{t-j})^\delta$$

where $\omega, \alpha_1, \gamma_1, \beta_1$ and δ are parameters to be estimated. The parameter γ_1 reflects the leverage effect $(-1 < \gamma_1 < 1)$. A positive (resp. negative) value of γ_1 means that past negative (resp. positive) shocks have a deeper impact on current conditional volatility than past positive (resp. negative) shocks. The parameter δ plays the role of a Box-Cox transformation of $\sigma_t(\delta > 0)$. .

In EVT implementation we use 10% of the data as the threshold excess. For the conditional models, a filter is necessary to model the conditional mean and the variance of the data. Thus, we estimate the AR(1)-APARCH(1,1) model described above, in which *z* represents a *F* distributed white noise series with unit variance. As explained previously, we set *F* to be Gaussian, Student-t, skewed Student-t, skewed Generalized Error and Johnson S_U distributions. In all models we jointly estimate by maximum likelihood the parameters in the equation for the mean return, the equation for its conditional standard deviation and the probability distribution for the return innovations. In addition, through the usual diagnostics performed on the standardized residuals and their squared values, we assess that returns

34. All computations were performed with the *rugarch* package (version 1.3-4) of R software (version 3.1.1) designed for the estimation and forecast of various univariate ARCH-type models. In the estimation of EVT models, we use *ismev* (version 1.41) and *evir* (version 1.7-3) packages.

are properly filtered. Based on this filtering, the conditional models are estimated as described in the previous subsections.

Table 23 presents the results of the estimation by the maximum likelihood method of the Generalized Pareto distribution parameters jointly with the respective parameters of the probability distribution of the innovations and those of the model AR(1)- APARCH(1,1), for a given threshold u , for each asset. For all asset returns, the estimated tail index ξ of the Generalized Pareto distribution is positive. Then, the left tail of the GPD distribution is fat and the probability of occurrence of extreme losses is higher than predicted by the Normal distribution. The estimated tail indexes of IBM and SAN are higher than those of AXA and BP, reflecting a thicker left tail of the return distribution.

Table 24 shows estimated parameters for the EVT-JSU-AR(1)-APARCH(1,1) model under a JSU distribution for all assets[35]. The autoregressive effect in the volatility specification is strong, with β_1 around 0.93, suggesting strong memory effects. The estimated γ_1 coefficient is positive and statistically significant at 10% in all cases, indicating the existence of a leverage effect for negative returns in the conditional volatility specification. It is also important that the skewness parameter in the Johnson S_U is less than 0 for all assets, suggesting the convenience of incorporating negative asymmetry to model innovations appropriately, although this parameter is not significant at 5% for IBM and at 10% for BP. The shape parameter is low, implying high kurtosis. Finally, δ takes values between 1.04 and 1.09, being significantly different from 2. This result suggests that, instead of modeling the conditional variance, we should model the conditional standard deviation, as it has been pointed out for a variety of assets by Garcia-Jorcano and Novales (2017).

The maximum likelihood estimates of the Generalized Pareto distribution parameters for IBM are $\left(\hat{\xi}, \hat{\beta}\right) = (0.39, 0.51)$, with standard errors of 0.12 and 0.07 respectively. Figure 16 shows a well-defined likelihood profile for this asset with a maximum log-likelihood of -91.877 reached for $\hat{\xi} = 0.39$. Thus, the model we have fitted is essentially a very heavy-tailed, infinite-variance model.

35. Estimation results for alternative models are available from the author upon request.

Table 23: Parameter estimates for the Generalized Pareto Distribution using daily returns

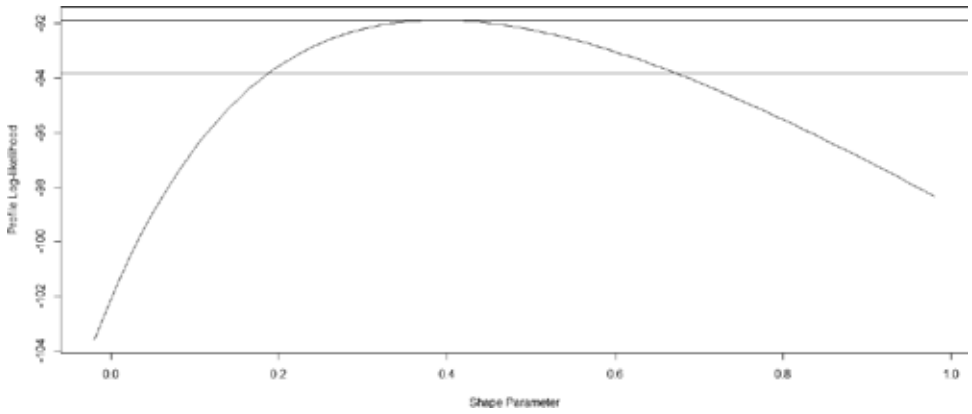
Sample: 10/2/2000 - 9/30/2016. u is the threshold, ξ is the shape parameter, β is the scale parameter. (ξ) and (β) correspond to the standard error of the shape and scale parameters, respectively.

	Daily	u	ξ	β	(ξ)	(β)
N	IBM	-1.041	0.392	0.493	0.121	0.072
	SAN	-1.239	0.240	0.522	0.103	0.070
	AXA	-1.139	0.048	0.712	0.067	0.079
	BP	-1.131	0.055	0.642	0.086	0.079
ST	IBM	-1.061	0.391	0.514	0.120	0.075
	SAN	-1.249	0.235	0.534	0.101	0.071
	AXA	-1.175	0.059	0.697	0.070	0.079
	BP	-1.158	0.072	0.635	0.089	0.080
SKST	IBM	-1.051	0.390	0.514	0.120	0.075
	SAN	-1.235	0.229	0.539	0.100	0.071
	AXA	-1.159	0.057	0.702	0.069	0.079
	BP	-1.154	0.078	0.627	0.090	0.079
SGED	IBM	-1.037	0.376	0.524	0.118	0.076
	SAN	-1.233	0.225	0.542	0.100	0.072
	AXA	-1.152	0.055	0.705	0.069	0.079
	BP	-1.145	0.072	0.628	0.089	0.079
JSU	IBM	-1.053	0.392	0.516	0.121	0.075
	SAN	-1.236	0.230	0.539	0.100	0.071
	AXA	-1.157	0.057	0.703	0.069	0.079
	BP	-1.152	0.074	0.631	0.089	0.079

Table 24: Parameter estimates of AR(I)-APARCH(I,I)-JSU model for individual stocks, fitting GPD to the residuals

	IBM			SAN			AXA			BP		
	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
φ_0	-0.00050	0.01760	0.978	-0.01736	0.02345	0.459	-0.02234	0.02857	0.434	-0.01607	0.02108	0.446
φ_1	-0.02048	0.01442	0.156	-0.00478	0.01590	0.764	0.02099	0.01532	0.171	-0.00856	0.01575	0.587
ω	0.02108	0.02664	0.429	0.02506	0.00670	0.000	0.03250	0.00136	0.000	0.02574	0.01955	0.188
α_1	0.07614	0.05961	0.202	0.06690	0.01386	0.000	0.06009	0.00476	0.000	0.06995	0.02945	0.018
β_1	0.92882	0.06487	0.000	0.93626	0.01428	0.000	0.93809	0.00138	0.000	0.92991	0.03718	0.000
γ_1	0.50884	0.29529	0.085	0.86807	0.15866	0.000	0.95065	0.02665	0.000	0.57416	0.28027	0.041
δ	1.07512	0.27157	0.000	1.04276	0.13396	0.000	1.06576	0.10144	0.000	1.09268	0.39636	0.006
skew	-0.09240	0.04808	0.055	-0.25352	0.07497	0.001	-0.27503	0.09489	0.004	-0.11002	0.07202	0.127
shape	1.52764	0.06332	0.000	2.00085	0.12402	0.000	2.29264	0.16404	0.000	1.98823	0.12293	0.000
ξ	0.39168	0.12073	0.001	0.22972	0.10037	0.001	0.05659	0.06883	0.411	0.07416	0.08915	0.406
β	0.51558	0.07523	0.000	0.53888	0.07147	0.000	0.70277	0.07903	0.000	0.63068	0.07938	0.000

Figure 16: Likelihood profile for ζ-parameter from the threshold excess model applied to filtered residuals of IBM under JSU-EVT model



We consider the tail of the IBM return distribution as defined by a threshold u = 1.0533, which leaves us with 126 exceedances (10% of 1260 data points). Figure 17 shows the fitted GPD model for the excess distribution, $F_u(y)$ where $y = z - u$, superimposed on points plotted at empirical estimates of excess probabilities for each loss (126 losses)[36]. Notice the good correspondence between the empirical estimates and the GPD curve. Under the EVT approach the filtered residuals from all models considered show a very similar fit to the GPD curve, especially when the filtered residuals come from asymmetric distributions. Figure 18 shows the estimation tail probabilities on logarithmic axes. The points on the graph are the 126 threshold exceedances and are plotted at y-values corresponding to the tail of the empirical distribution function. The smooth curve running through the points is the tail estimator (defined for the right tail):

$$1 - \hat{F}(z) = \frac{T_u}{T}\left(1 + \hat{\xi}\frac{z-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}$$

36. Figures 17 and 18 show the right tail, considering losses as positive numbers.

Figure 17: Empirical distribution of threshold excesses for IBM filtered residuals under EVT-AR(1)-APARCH(1,1)-JSU model versus the fitted GPD

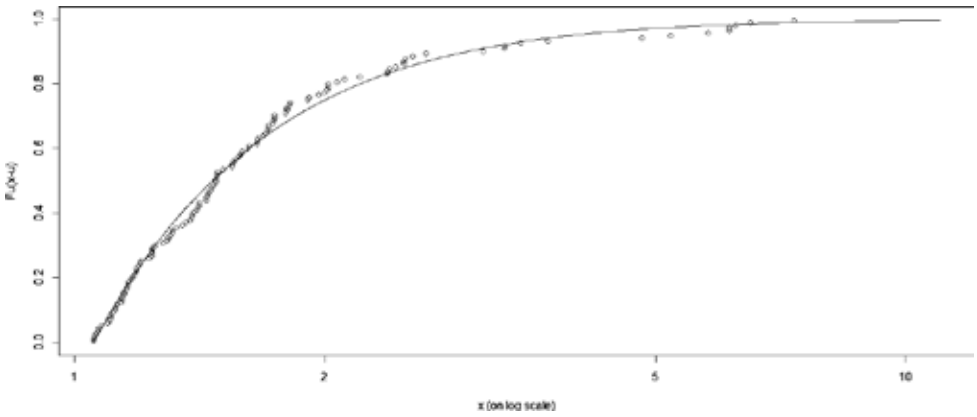
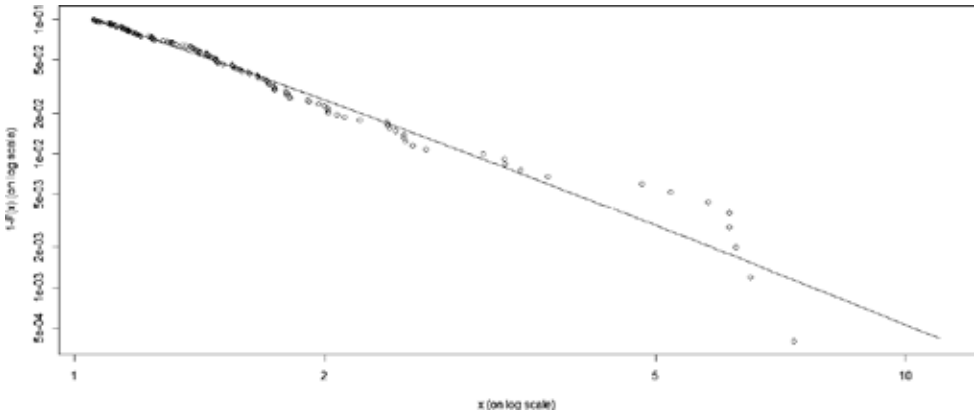


Figure 18: The smooth curve through the points shows the estimated tail of filtered residuals for IBM under AR(1)-APARCH(1,1)-JSU model using the tail estimator. Points are plotted at empirical tail probabilities calculated from the empirical distribution function



3.5. Evaluating 1-day ES forecasts

3.5.1. ES forecasts under the parametric approach

In this section we present the results from VaR and ES forecasts following a standard time-varying parametric approach. We restrict our attention to the left tail of the distribution and the 1%, 2.5% and 5% significance levels. We compute recursive ES forecasts from an expanding

window. First, each model is estimated using 2915 daily observations from the 10/2/2000-12/2/2011 sample period. After that, we increase the initial sample by one data point each day until the end of 2016, to compute 1-day ahead VaR and ES forecasts over five years: 2012-2016 (1260 data observations). In this forecasting period models are estimated every 50 days. This choice tries to reduce the computational cost while avoiding frequent parameter variation due in part to pure noise.

Table 25 displays descriptive statistics for returns for in-sample (10/2/2000-12/2/2011) and out-of-sample (12/5/2011-9/30/2016) periods. Skewness is negative, except for SAN and AXA in the in-sample period. Likewise, kurtosis is higher than 3 for the four stocks in both periods. We are thus confronted with fat tail distributions and the Jarque-Bera statistic clearly rejects the null hypothesis of a Normal distribution. VaR and ES forecasts based on the assumption of a Normal distribution of returns are therefore inappropriate, so we will compute them under a non-Normal framework using alternative distributions as well as relying on a different approach, like EVT. Focusing on the behavior of the left tail of these leptokurtic distributions seems justified as it should allow for a better estimation of extreme variations in financial returns.

We forecast both risk measures, not only with the full distribution but also using only extreme events as explained previously. Figure 19 shows IBM daily percentage returns (1260 data) together with out-of-sample $VaR_{1\%}$ and $VaR_{5\%}$ forecasts from an AR(1) model for returns with a JSU-APARCH(1,1) model for return innovations. Such forecasts are compared with those obtained by applying Extreme Value Theory (EVT), fitting a GPD density to the tail of the distribution. The differences in VaR calculated with the two models are small for the 5% quantile but they become more important for the 1% quantile. VaR forecasts under EVT indicate higher losses than predicted VaR without the use of EVT. Figure 20 shows $ES_{1\%}$ and $ES_{5\%}$ forecasts obtained with EVT and without EVT. We can see that the forecast of average losses exceeding VaR under the GPD distribution in the EVT approach is greater than the one obtained from a JSU distribution in the non-EVT approach, especially for the more extreme quantiles.

Table 25: Descriptive statistics for log-returns (%) of individual stocks over the in-sample (10/2/2000-12/2/2011) and out-of-sample (12/5/2011-9/30/2016) periods. JB stat. is the Jarque-Bera test statistic

	In-Sample				Out-of-Sample			
Daily	IBM	SAN	AXA	BP	IBM	SAN	AXA	BP
Observations	2915	2915	2915	2915	1260	1260	1260	1260
Mean (bps.)	1.79	-2.18	-3.94	-0.89	-1.41	-0.13	4.26	-0.26
Median (bps.)	0.00	0.00	0.00	0.00	0.00	1.94	11.04	0.00
St. Dev	1.74	2.31	2.96	1.80	1.18	2.13	1.93	1.43
Skewness	-0.12	0.27	0.31	-0.19	-1.00	-1.09	-0.72	-0.16
Kurtosis	11.47	9.09	9.33	7.87	9.77	14.79	9.10	6.84
Maximum	11.35	20.88	19.78	10.58	4.91	10.14	7.28	6.93
10 percentile	-1.73	-2.63	-3.07	-1.96	-1.24	-2.43	-2.09	-1.60
5 percentile	-2.66	-3.67	-4.60	-2.73	-1.73	-3.38	-3.22	-2.31
1 percentile	-5.13	-6.68	-8.46	-5.32	-3.59	-4.97	-4.96	-3.59
Minimum	-16.89	-12.72	-20.35	-14.04	-8.64	-22.17	-16.82	-9.08
JB stat.	8724.16	4548.12	4918.84	2902.54	2614.85	7549.73	2063.42	780.63

Figure 19: IBM daily percent returns and $VaR_{1\%}$ and $VaR_{5\%}$ forecasts with the full sample as well as using only extreme values

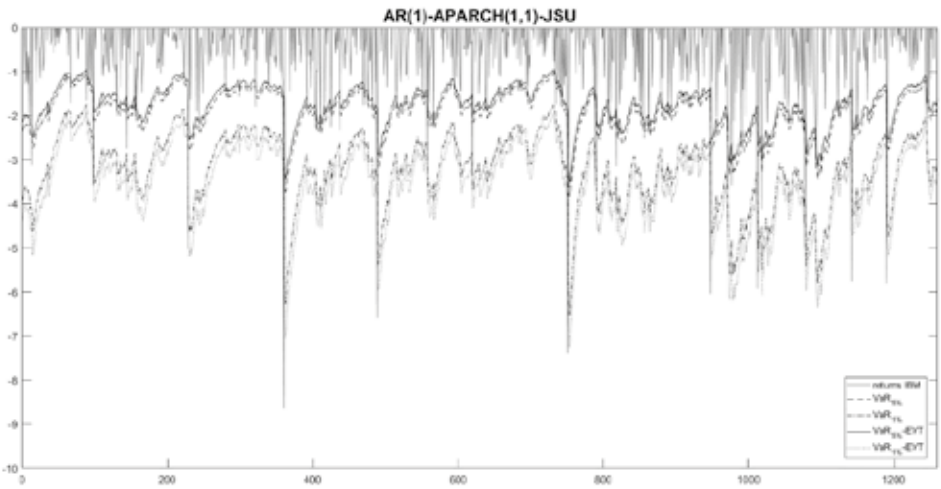
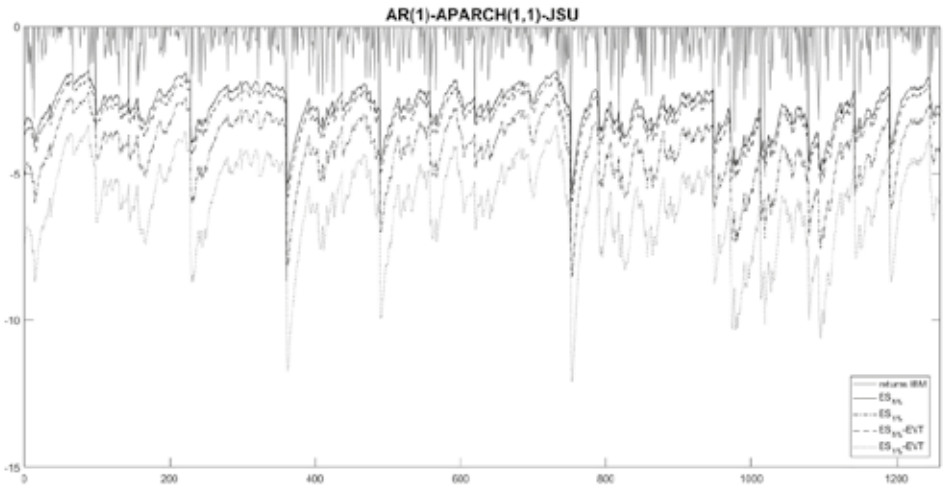


Figure 20: IBM daily percent returns and $ES_{1\%}$ and $ES_{5\%}$ forecasts with the full sample as well as using only extreme values



We now examine our forecasts for the complete out-of-sample period (5 years, 1260 data). Since we generate time series of VaR and ES forecasts, we present a summary of results over the 5-year period. Table 26 presents the average of out-of-sample 1-day ES forecasts ($\overline{ES}$), the violations ratio (Viol) of the underlying VaR and the backtesting results for the different models for IBM[37]. Our discussion here is focused on the general patterns that appear in these estimation results. The average ES forecasts from conditional EVT-based models can be seen to be "more negative" than forecasts from conditional models not based on EVT. As shown in Figure 20, differences on ES forecasts at 1% significance level are larger than those at 5% significance level.

It seems desirable that a good ES model may have a violation ratio close to the theoretical one. Indeed, as we will see below, some backtesting tests for ES are based on this comparison. Conditional EVT-based models tend to yield a violation ratio very close to the theoretical one. Departures from the theoretical violation ratio are larger for models not using EVT, especially under the Normal and Student-t distributions for return innovations. In general, the violations ratio suggests that conditional

37. Results for the rest of individual stocks assets are available from the author upon request.

EVT-based models forecast the VaR quantile correctly, corroborating Kuester, Mittnik and Paolella (2006), who attest to the superiority of this approach. On the other hand, conditional ES models not based on EVT that incorporate heavy-tailed distributions also perform well, corroborating Mabrouk and Saadi (2012). But we will show below that EVT-based models not only show an accurate violation ratio, but they also have a good performance in ES backtesting. On the other hand, non-EVT based models have a violation ratio higher than expected and they show a worse ES forecasting performance than EVT-based models.

If we focus on the conditional models not based on EVT, all tests show that models with asymmetric distributions for return innovations produce better ES forecasts. If we take higher p-values as an indication of how well the model fulfills the condition established in the null hypothesis, then the JSU distribution can be seen as showing the best performance in ES forecasting for the set of four stocks. On the other hand, the Z_1 test by Acerbi & Szekely and the tests by Costanzino & Curran and Du & Escanciano do not discriminate among asymmetric distributions.

Under the null hypothesis Acerbi & Szekely the number of theoretical VaR breaches is $\mathbb{E}_{H0}[N_T] = T\alpha$, where N_T is the indicator of VaR breaches. The relationship between the two test statistics of Acerbi & Szekely is: $Z_2 = (1 + Z_1)N_T/T\alpha - 1$. This shows that while Z_1, being just an average taken over excesses themselves, is insensitive to an excessive number of exceptions, Z_2 depends on that number through the ratio $N_T/T\alpha$. This is why, when the number of violations exceeds the theoretical level, p-values for the Z_2-test are lower than for the Z_1 test. An ES model will pass the Z_2 test when not only the magnitude but also the frequency of the excesses is statistically equal to the expected one[38].

At the 1% significance level, p-values of Acerbi & Szekely and Graham & Pál tests for the conditional models not based on EVT theory are very close to 0. In these cases, we obtain positive realized values for Z_1 and Z_2, instead of them being equal to zero. In short, we reject H_0 because of risk undervaluation. For these three tests we observe large differences in

38. Acerbi & Szekely (2014) show that the $Z2$ test is more powerful than the $Z1$ test when the null and alternative hypothesis differ in volatility, while $Z1$ is more powerful than $Z2$ in the case of different tail indexes.

p-values between conditional models based on EVT and non-EVT based conditional models in favor of the former, which seem to produce better risk forecasts. The Graham & Pál test discriminates against the Normal and Student-t distribution for almost all significance levels for the four stocks, but only for the non-EVT based ES models.

We indicate in boldface the *p*-values of the Righi & Ceretta, Acerbi & Szekely and Graham & Pál tests when we have obtained statistics with opposite sign to the one embedded in the alternative hypothesis. That essentially arises for EVT-based models. In the Righi & Ceretta test we have $H_0: \mathbb{E}(BT_t) = 0$, where BT_t is the statistic of the test which estimate the expected loss and its dispersion through the ES and SD, against $H_1: \mathbb{E}(BT_t) < 0$ but with some models we obtain $\mathbb{E}(BT_t) > 0$, reflecting that most excesses fall between VaR and ES, not beyond ES, especially under the EVT approach. The first test by Acerbi & Szekely specifies $H_0: \mathbb{E}(Z_1) = 0$ against $H_1: \mathbb{E}(Z_1) > 0$ and the second one, $H_0: \mathbb{E}(Z_2) = 0$ against $H_1: \mathbb{E}(Z_2) > 0$. However, with some models, especially models based on EVT, we obtain $\mathbb{E}(Z_1) < 0$ and $\mathbb{E}(Z_2) < 0$, respectively. In the first test that means that the average of realized excesses is lower in absolute value than the predicted ES. In the second test, it means that not only the average excess but also the number of excesses is lower than expected. Finally, in the Graham & Pál test we have $H_0: TR_\alpha = TR^0$ against $H_0: TR_\alpha < TR^0$ where TR^0 is equal to $-\alpha$ under the exponential assumption. The null hypothesis is rejected if the realized value of the sample statistic $\widehat{TR}_\alpha$ is significantly lower than the theoretical level of tail risk TR^0. If we obtain $TR_\alpha = TR^0$, we will say that the risk model captures tail risk sufficiently, or that it provides enough risk coverage, although risk may then be overvalued[39]. When that happens, the logarithmic difference between the probability of an excess and the significance level for VaR (α) follows a distribution with thicker tails than the exponential distribution. When the forecast CDF is a correct estimate of the real and unobservable P&L distribution, such probability differences follow an exponential distribution[40].

39. For more details of this tests, see Graham & Pál (2014).
40. That amounts to return violations, in probability terms, following a Uniform (0,1) distribution.

Table 26: Mean *ES* forecasts *($\overline{ES}$)*, violation ratio (Viol) and backtesting results (p-values) for ES forecasts for IBM

BT_T is the test of Righi & Ceretta (2015), Z_1 and Z_2 are the tests of Acerbi & Szekely (2014), *TR* is the test of Graham & Pál (2014), and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional (*lags* = 1 and *lags* = 5) tests of Costanzino & Curran (2015) and Du & Escanciano (2016). *p*-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified under the alternative hypothesis.

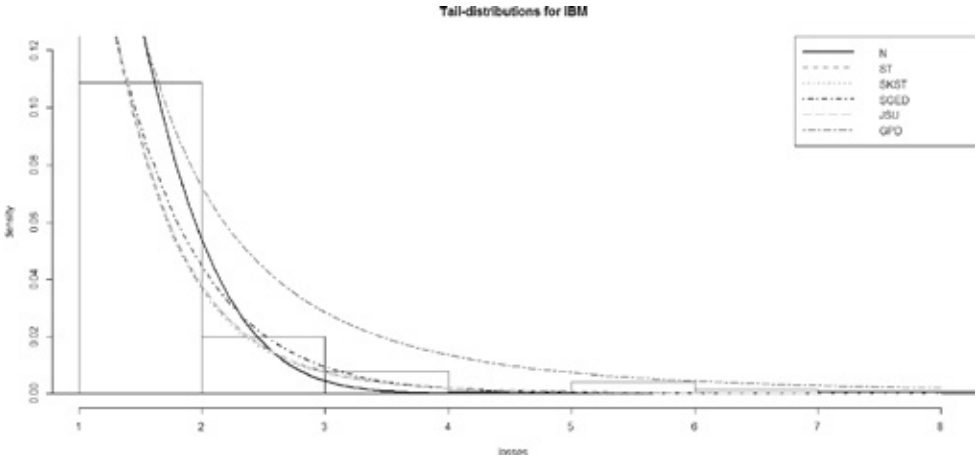
IBM			1% significance level						
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	*TR*	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.249	0.014	0.00	0.01	0.01	0.00	0.00	0.65	0.97
ST	-4.205	0.010	0.07	0.00	0.00	0.00	0.01	0.72	0.99
SKST	-4.266	0.010	0.07	0.00	0.00	0.00	0.02	0.73	0.99
SGED	-3.921	0.010	0.02	0.00	0.00	0.00	0.01	0.72	0.99
JSU	-4.206	0.010	0.06	0.00	0.00	0.00	0.02	0.73	0.99
N-EVT	-5.931	0.010	**0.38**	**0.96**	**1.00**	0.30	0.26	0.76	0.99
ST-EVT	-6.059	0.010	**0.35**	**0.92**	**0.99**	0.30	0.26	0.77	0.99
SKST-EVT	-6.050	0.010	**0.35**	**0.93**	**1.00**	0.30	0.26	0.77	0.99
SGED-EVT	-5.923	0.010	**0.34**	**0.89**	**0.98**	0.29	0.25	0.77	0.99
JSU-EVT	-6.052	0.010	**0.34**	**0.90**	**1.00**	0.30	0.26	0.76	0.99
IBM			2.5% significance level						
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	*TR*	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.847	0.022	0.01	0.01	0.03	0.00	0.13	0.79	0.84
ST	-3.305	0.024	0.13	0.03	0.05	0.01	0.29	0.86	0.89
SKST	-3.350	0.021	0.12	0.01	0.08	0.03	0.39	1.00	0.91
SGED	-3.255	0.018	0.05	0.01	0.32	0.00	0.48	0.63	0.91
JSU	-3.352	0.019	0.10	0.01	0.32	0.02	0.46	0.81	0.92
N-EVT	-4.042	0.022	**0.36**	**0.79**	**0.95**	0.49	0.35	0.70	0.91
ST-EVT	-4.122	0.022	0.31	0.68	0.94	0.47	0.38	1.00	0.93
SKST-EVT	-4.116	0.024	**0.34**	**0.73**	**0.96**	0.47	0.38	0.99	0.93
SGED-EVT	-4.056	0.022	0.32	0.66	0.92	0.48	0.35	0.69	0.91
JSU-EVT	-4.116	0.023	0.32	0.68	**0.99**	0.47	0.38	1.00	0.93

IBM	5% significance level								
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.510	0.037	0.05	0.04	0.49	0.00	0.17	0.37	0.55
ST	-2.699	0.044	0.18	0.13	0.40	0.09	0.33	0.11	0.41
SKST	-2.734	0.044	0.20	0.11	0.39	0.16	0.22	0.11	0.40
SGED	-2.738	0.037	0.15	0.07	0.75	0.02	0.06	0.36	0.58
JSU	-2.752	0.040	0.18	0.11	0.67	0.15	0.14	0.13	0.42
N-EVT	-3.003	0.053	**0.48**	**0.69**	**0.90**	0.52	0.41	0.41	0.67
ST-EVT	-3.055	0.049	**0.40**	**0.70**	**0.97**	0.54	0.37	0.13	0.46
SKST-EVT	-3.050	0.050	**0.41**	**0.76**	**0.96**	0.54	0.37	0.13	0.46
SGED-EVT	-3.015	0.053	**0.44**	**0.69**	**0.96**	0.54	0.36	0.23	0.59
JSU-EVT	-3.050	0.051	**0.41**	**0.63**	**0.93**	0.54	0.37	0.13	0.45

Bold figures in the tables signal a frequent overvaluation of risk for EVT-based ES models. In them, the number of violations does not depart much from the theoretical value, reflecting good VaR forecasts. But the sign of the test statistic is contrary to that in the null hypothesis, showing an overvaluation of ES that would imply too high a level of required capital. Such overvaluation will not be detected by one-sided tests. However, the absolute value of the statistic is generally very small, suggesting that the estimation error may be statistically acceptable. The possible overvaluation of risk can be seen in Figure 21 as it shows the tail probability distributions estimated for IBM. Other assets show a similar picture. Curve lines show the estimated tail probabilities, while the rectangles display observed relative frequencies. Estimated parameters for each distribution are shown in parenthesis. We observe that most probability distributions other than GPD tend to undervalue the weight of extreme returns. Such undervaluation is especially obvious for the Normal distribution. On the contrary, the GPD is suitable to appropriately capture tail risk, and it avoids underestimating extreme risks, although at the price of slight overvaluation of the risk of medium range losses.

Figure 21: Estimated tail-distributions for IBM. N is the Normal distribution

ST is the Student-t (4.67), SKST is the Skewed Student-t (0.97, 4.69), SGED is the Skewed Generalized Error (0.99, 1.15) and JSU is the Johnson SU (-0.092, 1.53) distribution. Numerical estimates for parameters in brackets.



By and large, we have obtained that for our sample of stocks, conditional EVT-based models not only produce better VaR forecasts, but also, they yield the best results in ES forecasts according to different ES backtests. In many cases, we obtain p-values close to 1 with EVT-based models. The success of EVT models for ES forecasting corroborates Marinelli *et al.* (2007), Jalal and Rockinger (2008) and Wong *et al.* (2012). However, we must bear in mind that the Righi & Ceretta, Acerbi & Szekely and Graham & Pal tests are one-sided by nature and they focus on risk undervaluation. Therefore, in those tests risk overvaluation does not lead to a rejection of the null hypothesis, and that seems to be often the case in ES forecasting with EVT-based models.

3.5.2. ES forecasts under Filtered Historical Simulation

We evaluate the performance results of 1-day ahead out-of-sample ES forecasts from FHS using the test of Righi & Ceretta and the two tests of Acerbi & Szekely because they are suitable for non-parametric VaR and ES forecasts. Table 27 shows average ES forecasts ($\overline{ES}$), the violations

Table 27: Mean *ES* forecasts *($\overline{ES}$)*, violation ratio (Viol) and backtesting results (p- values) for ES forecasts for IBM and for 1-day returns calculated by Filtered Historical Simulations (FHS)

BT_T is the test of Righi & Ceretta (2015) and Z_1 and Z_2 are the tests of Acerbi & Szekely (2014). p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified under the alternative hypothesis.

IBM			1% significance level		
FHS	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-4.553	0.011	0.15	0.02	0.02
ST	-4.629	0.011	0.11	0.00	0.00
SKST	-4.627	0.010	0.10	0.02	0.02
SGED	-4.577	0.010	0.13	0.02	0.02
JSU	-4.619	0.011	0.15	0.00	0.12
N-EVT	-4.490	0.010	**1.00**	**0.90**	**0.97**
ST-EVT	-4.592	0.011	0.12	0.14	0.97
SKST-EVT	-4.575	0.011	0.12	0.10	0.97
SGED-EVT	-4.519	0.011	0.09	0.08	1.00
JSU-EVT	-4.573	0.011	0.12	0.00	0.97
IBM			2.5% significance level		
FHS	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-3.474	0.021	0.19	0.02	0.29
ST	-3.486	0.021	0.14	0.03	0.26
SKST	-3.485	0.022	0.16	0.01	0.13
SGED	-3.471	0.021	0.19	0.03	0.30
JSU	-3.481	0.022	0.18	0.03	0.45
N-EVT	-3.471	0.021	**1.00**	**0.80**	**0.87**
ST-EVT	-3.492	0.021	0.14	0.37	0.94
SKST-EVT	-3.481	0.022	0.17	0.42	0.88
SGED-EVT	-3.466	0.021	0.14	0.31	0.93
JSU-EVT	-3.482	0.022	0.17	0.16	0.88
IBM			5% significance level		
FHS	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-2.791	0.042	0.27	0.14	**0.68**
ST	-2.784	0.044	0.21	0.13	0.50
SKST	-2.782	0.043	0.22	0.05	0.43
SGED	-2.780	0.043	0.26	0.10	0.51
JSU	-2.781	0.044	0.24	0.12	0.45
N-EVT	-2.792	0.042	**1.00**	**0.75**	**0.75**
ST-EVT	-2.792	0.044	0.25	0.44	0.74
SKST-EVT	-2.784	0.045	0.26	0.42	0.80
SGED-EVT	-2.784	0.044	0.25	0.39	0.82
JSU-EVT	-2.786	0.044	0.25	0.52	0.79

ratio of the underlying VaR and backtesting results[41]. A comparison with Table 26 shows that (i) conditional EVT-based models do not always present "more negative" average ES values than conditional models not based on EVT. Besides, average ES values over the out-of-sample period (5 years, 1260 data) are now more similar among models than under the parametric approach. This observation is important because it amounts to a reduction in Model Risk, i.e., in the uncertainty that arises on the true value of VaR and ES due to the availability of forecasts coming from a variety of alternative models, (ii) average ES forecasts under FHS are closer to those obtained under the parametric approach for non EVT-based models than for EVT-based models, (iii) regarding VaR violation rates, there is some tendency to undervalue risk at the 1% significance level (more violations than expected) and overvalue risk at the 5% significance level (less violations than expected), (iv) unlike Table 26, EVT-based models do not yield a lower violation rate than non-EVT based models, (v) models not based on EVT seem again unsuitable in terms of ES forecasts, being rejected by Acerbi & Szekely Z_1 and Z_2 tests for $ES_{1\%}$ and $ES_{2.5\%}$.

Less discrimination is obtained at 5% significance level. For instance, at that level, all models display a good ES performance for BP at 10% significance, although the Z_2 test suggests that ES is possibly overvalued, and (vi) overvaluation of risk as signaled by a sign of the test statistic contrary to H_1 in the one-tail tests is much less frequent than under the parametric approach.

The conclusions obtained when applying ES backtests under the parametric and FHS approaches are similar, which is reassuring. Differences between conditional models based on EVT and not based on EVT are more evident under the parametric approach, because the power and flexibility of conditional volatility models is diluted by historical simulation. The dilution depends on the number of realizations or paths generated from the standardized residuals from the first step estimation.

41. Results for the rest of individual stocks assets are available from the author upon request.

3.6. Conclusions

In spite of the substantial theoretical evidence documenting the superiority of Expected Shortfall (ES) over VaR as a measure of risk, financial institutions and regulators have only recently embraced ES as an alternative to VaR for financial risk management. One of the major obstacles in this transition has been the unavailability of simple tools for the evaluation of ES forecasts. While the Basel rules for VaR tests are based on counting the number of exceptions, assessing the adequacy of an ES model requires the consideration of the size of tail losses beyond the VaR boundary. Different approaches have been proposed in the literature for ES backtesting in the last few years but, to the best of our knowledge, this paper provides the first extensive comparison of a variety of alternative ES backtesting procedures.

We use daily market closing prices for 10/2/2000 to 9/30/2016 on IBM, Santander, AXA and BP, and we consider some flexible families of asymmetric distributions for as- set returns that include more standard probability distributions as special cases. Normal and Student-t distributions are considered as a benchmark for comparison. Given the evidence put forward in Garcia-Jorcano and Novales (2017) we use an APARCH volatility specification for all assets. We are initially interested in exploring which probability distribution seems more appropriate to model asset returns in order to get good ES forecasts. Following the standard risk management methodology, once we estimate the dynamics of returns and the parameters of the probability distribution for return innovations, we forecast returns and volatility and apply a parametric approach to forecast VaR and ES. After that, we use a variety of tests recently proposed for ES model validation.

As the true temporal dependency of financial returns is a complex issue, the standard approach to risk management can be improved by considering a two-step procedure that applies Extreme Value Theory (EVT): First, filtering the returns through a more or less complex GARCH model and second, estimating an extreme value theory type of density for the tail of the distribution of return innovations, using their assumed *iid* structure. This two-step procedure was proposed by McNeil & Frey (2000) and it leads to a significant improvement, since VaR and ES

forecasts then incorporate changes in expected returns and volatility over time. So, in the application of EVT we first estimate a dynamic model for returns and their volatility under a given probability distribution. After that, we fit a Generalized Pareto Distribution for return innovations once we have filtered autocorrelations and GARCH effects. As in the standard approach, we then forecast VaR and ES at different significance levels at 1- and 10-day horizons and compare the results with those obtained under the standard parametric approach.

In standard conditional models fitted to the full distribution of return innovations we observe that asymmetric distributions play an important role in capturing tail risk. This is because some stylized facts of financial returns such as volatility clusters, heavy tails and asymmetry are collected suitably by these asymmetric distributions. When we apply EVT to return innovations by modeling the tail with a GPD we obtain good ES forecasts regardless of the probability distribution used for returns. So, it looks as if considering just the return innovations in the tail of the distribution is more important than discriminating among probability distributions when forecasting ES. Besides, each combination of APARCH volatility and probability distribution under the EVT approach dominates the similar specification under the standard approach fitted to the full distribution. Conditional EVT models turn out to be more accurate and reliable than standard conditional models not based on EVT both, for forecasting VaR and for predicting losses beyond VaR.

We have also shown that using Filtered Historical Simulation can be very useful. First, qualitative results under FHS are very similar to those obtained under the parametric approach, which is reassuring. EVT-based models dominate non-EVT based models for forecasting both VaR and ES, and asymmetric probability distributions yield more accurate ES forecasts. Second, ES forecasts are much more similar for different probability distributions, and also between forecasts from EVT-based models and non-EVT based models. That implies a considerable reduction in model risk, i.e., the uncertainty in ES forecasting because of having alternative model specifications. Given the extreme importance of these forecasts for capital requirements at financial institutions, reducing model risk is a central issue in tail risk estimation.
The ES tests we consider focus on a possible undervaluation of risk, except for Costanzino & Curran and Du & Escanciano tests which are

two-tailed tests. We have pointed out that in some cases backtesting does not reject the model specification because the sample evidence is against both the null and the alternative hypothesis. In other words, some ES models are not rejected in spite of the fact that they overvalue risk, albeit by a small amount in most cases. When using ES to build the institution's reserves to cover potential losses in times of crisis, the undervaluation may be fatal, but overvaluation will lead to inefficient use of capital. This is a relevant consideration that should be taken into account for ES model validation.

A final remark from this research relates to the possible weak power of currently available tests for ES forecasting. Other than showing a clear preference for an EVT approach as well as a rejection of symmetric probability distributions for return innovations, none of the tests we have considered are able to discriminate much among alternative probability distributions. However, the recommendation to use FHS under an EVT specification for ES forecasting is a clear conclusion of this research.

BIBLIOGRAPHY

1. Aas, K., & Haff, I.H. (2006). The generalized hyperbolic skew student's t-distribution. *Journal of Financial Econometrics*, 4(2), 275-309.
2. Abad, P., & Benito, S. (2012). A detail comparison of value at risk estimates. *Mathematics and Computers in Simulation*, 94, 258-276.
3. Abad, P., Benito, S., & Lopez, C. (2015). Role of the loss function in the VaR comparison. *Journal of Risk Model Validation*, 9(1), 1-19.
4. Abad, P., Benito, S., Lopez, C., & Sanchez-Granero, M.A. (2016). Evaluating the performance of the skewed distributions to forecast value-at-risk in the global financial crisis. *Journal of Risk*, 18(5), 1-28.
5. Abramowitz, M., & Stegun, I. A. (1972). Handbook of mathematical functions with formulas, graphs, and mathematical tables (Vol. 9). Dover: New York.
6. Acerbi, C., & Szekely, B. (2014). Backtesting Expected Shortfall. Publication of MSCI.
7. Acerbi, C., & Tasche, D. (2002). On the coherence of Expected Shortfall. *Journal of Banking and Finance*, 26, 1487-1503.
8. Andersen, T., Bollerslev, T., Diebold, F., & Labys, P. (2003). Modelling and forecasting realized volatility. *Econometrica*, 71, 529-629.
9. Ane, T. (2006). An analysis of the flexibility of Asymmetric Power GARCH models. *Computational Statistics and Data Analysis*, 51, 1293-1311.
10. Angelidis, T., Benos, A., & Degiannakis, S. (2007). A robust VaR model under different time periods and weighting schemes. *Review of Quantitative Finance and Accounting*, 28, 187-201.
11. Angelidis, T., & Degiannakis, S. (2006). Backtesting VaR models: An Expected Shortfall approach. http://dx.doi.org/10.2139/ssrn.898473
12. Alexander, C. (2009). Market risk analysis, Value at Risk models. John Wiley & Sons, Vol. 4.
13. Alexander, C., & Sheedy, E. (2008). Developing a stress testing framework based on market risk models. *Journal of Banking and Finance*, 32, 2220-2236.

14. Artzner, P., Delbaen, F., Eber, J.M., & Heath, M. (1999). Coherent measures of risks. *Mathematical Finance*, 9, 203-228.
15. Azzalini, A., & Capitanio, A. (2003). An asymmetric Generalization of Gaussian and Laplace Laws. *Journal of the Royal Statistical Society*, Series B, 65, 367-389.
16. Bali, T.G., & Theodossiou, P. (2007). A conditional-SGT-VaR approach with alternative GARCH model. *Annals of Operations Research*, 151, 241-267.
17. Barndorff-Nielsen, O.E. (1997). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian Journal of Statistics*, 24, 1-14.
18. Barone-Adesi, G., Bourgoin, F., & Giannopoulus, K. (1998). Don't look back. *Risk*, 11, 100-103.
19. Barone-Adesi, G., Giannopoulus, K., & Vosper, L. (1999). VaR without correlations for portfolios of derivative securities. *Journal of Futures Markets*, 19, 583-602.
20. Barone-Adesi, G., Giannopoulos, K., & Vosper, L. (2002). Backtesting derivative portfolios with filtered historical simulation (FHS). *European Financial Management*, 8(1), 31-58.
21. Basel Committee on Banking Supervision (2016). Standards: Minimum Capital requirements for market risk. Bank for International Settlements.
22. Bauwens, L., Giot, P., Grammig, J., & Veredas, D. (2004). A comparison of financial duration models via density forecasts. *International Journal of Forecasting*, 20, 589-609.
23. Baysal, R. E., & Staum, J. (2008). Empirical likelihood for value-at-risk and expected shortfall. *The Journal of Risk*, 11(1), 3-32.
24. Bellini, F., Klar, B., Müller, A., & Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54, 41-48.
25. Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Review of Financial Studies*, 14, 371-405.
26. Berkowitz, J., Christoffersen, P.F., & Pelletier, D. (2011). Evaluating Value-at-Risk models with desk-level data. *Management Science*, 57(12), 2213-2227.
27. Berkowitz, J., & O'Brien, J. (2002). How accurate are Value-at-Risk models at commercial banks? *Journal of Finance*, 1093-1111.

28. Bernardi, M., Catania, L., & Petrella, L. (2016). Are news important to predict the Value-at-Risk? *The European Journal of Finance*.
29. Bickel, P.J., & Krieger, A.M. (1989). Confidence bands for a distribution function using the bootstrap. *Journal of the American Statistical Association*, 84, 95-100.
30. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.
31. Braione, M., & Scholtes, N.K. (2016). Forecasting Value-at-Risk under different distributional Assumptions. *Econometrics*, 4(3).
32. Brooks, C., & Persand, G. (2003). Volatility forecasting for risk management. *Journal of Forecasting*, 22, 1-22.
33. Bubak, V. (2008). Value-at-Risk on Central and Eastern European stock markets: An empirical investigation using GARCH models. Institute of Economics Studies, Charles University, 18. https://www.econstor.eu/handle/10419/83344
34. Caporin, M. (2008). Evaluating Value-at-Risk measures in the presence of long memory conditional volatility. *The Journal of Risk*, 10(3), 79-110.
35. Carver, L., 2013. Mooted VaR substitute cannot be backtested, says top quant. *Risk*, March, 8.
36. Casella, G., & Berger, R. (2002). Statistical Inference. *Duxbury Advanced Series*.
37. Chan, K.F., & Gray, P. (2006). Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting*, 22(2), 283-300.
38. Chana, N.H., & Peng, S.J.D.L. (2006). Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics*, 137, 556-576.
39. Chen, J.M. (2014). Measuring market risk under the basel accords: VaR, stressed VaR, and Expected Shortfall. *Aestimatio, The IEB International Journal of Finance*, 8, 184-201.
40. Chernick, M.R. (1999). Bootstrap methods: A practitioner's guide. John Wiley & Sons.
41. Christoffersen, P. (1998). Evaluating internal forecasting. *International Economic Review*, 39, 841-862.
42. Choi, P., & Nam, K. (2008). Asymmetric and leptokurtic distribution for heteroscedastic asset returns: the SU-normal distribution. *Journal of Empirical finance*, 15(1), 41-63.

43. Clark, T.E., & McCraken, M.W. (2001). Test of equal forecast accuracy and encompassing for nested models. Journal of Econometrics, 105(1), 85-110.
44. Clift, S.S., Costanzino, N., & Curran, M. (2016). Empirical Performance of Backtesting Methods for Expected Shortfall. http://dx.doi.org/10.2139/ssrn.2618345
45. Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Events. Springer: London. 5
46. Cont, R., Deguest, R., & Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 16(6), 593-291.
47. Corlu, C.G., Meterelliyoz, M., & Tinic, M. (2016). Empirical distributions of daily equity index returns: A comparison. *Expert System with Applications*, 54, 170-192.
48. Costanzino, N., & Curran, M. (2015). Backtesting general spectral risk measures with application to Expected Shortfall. http://dx.doi.org/10.2139/ssrn.2514403
49. Crnkovic, C., & Drachman, J. (1996). Quality control. *Risk*, 9, 139-143.
50. Daniels, H.E. (1987). Tail Probability Approximations. *International Statistic Review*, 55, 37-48.
51. Danielsson, J., & de Vries, C.G. (1997b). Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, 4, 241-257.
52. Danielsson, J., & De Vries, C.G. (2000). Value-at-Risk and extreme returns. *Annales d'Economie et de Statistique*, 239-270.
53. Degiannakis, S., Floros, C. & Dent, P. (2013). Forecasting value-at-risk and expected shortfall using fractionally integrated models of conditional volatility: International evidence. *International Review of Financial Analysis*, 27, 21-33.
54. Dendramis, Y., Spungin, G.E., & Tzavalis, E. (2014). Forecasting VaR models under different volatility processes and distributions of return innovations. *Journal of Forecasting*, 33, 515-531.
55. Diamandis, P.F., Drakos, A.A., Kouretas, G.P., & Zarangas, L. (2011). Value-at-Risk for long and short trading positions: Evidence from developed and emerging equity markets. *International Review of Financial Analysis*, 20, 165-176.

56. Diebold, F.X., Gunther, T.A., & Tay, A.S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4), 863-883.
57. Diebold, F.X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 253-263.
58. Diebold, F.X., Schuermann, T., & Stroughair, J.D. (2000). Pitfalls and opportunities in the use of extreme value theory in risk management. *The Journal of Risk Finance*, 50, 264-272.
59. Ding, Z., Granger, C.W.J., & Engle, R.F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1, 83-106.
60. Du, Z., & Escanciano, J.C. (2016). Backtesting Expected Shortfall: Accounting for Tail Risk. *Management Science*, 63(4), 940-958.
61. Duffie, D., & Pan, J. (1997). An overview of value at risk. *Journal of Derivatives*, 4, 7-49.
62. Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26
63. Efron, B. & Tibshirani, R.J. (1993). An introduction to the Bootstrap. Chapman & Hall: New York.
64. El Babsiri, M., & Zakoian, J.M. (2001). Contemporaneous asymmetry in GARCH processes. *Journal of Econometrics*, 101, 257-294.
65. Embrechts, P. & Hofert, M. (2014). Statistics and quantitative risk management for banking and insurance. *Annual Review of Statistics and Its Applications*, 1, 493-514.
66. Embrechts, P., Lambrigger, D., & Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 176223.
67. Embrechts, P, Lambrigger, D., & Wüthrich, M. (2009). Multivariate extremes and the aggregation of dependence risks: examples and counter-examples. *Extremes*, 12(2), 107-127.
68. 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).
69. Engle R.F., & Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22, 367-381.
70. Ergen, I. (2015). Two-step methods in VaR prediction and the importance of fat tails. *Quantitative Finance*, 15(6), 1013-1030.

71. Ergün, A., & Jun, J. (2010). Time-varying higher-order conditional moments and forecasting intraday VaR and expected shortfall. *Quarterly Review of Economics and Finance*, 50, 264-272.
72. Fernandez, C., & Steel, M. (1998). On bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441), 359-371.
73. Filliben, J.J. (1975). The probability plot correlation coefficient test for normality. *Technometrics*, 17(1), 111-117.
74. Fissler, T., Ziegel, J.F., & Gneiting, T. (2015). Expected Shortfall is jointly elicitable with value at risk-implications for backtesting. arXiv preprint arXiv:1507.00244.
75. Föllmer, H., & Knispel, T. (2013). Convex risk measures: Basic facts, law-invariance and beyond, asymptotic for large portfolios. Handbook of the fundamentals of Financial Decision Making, Part II, 507-554.
76. Föllmer, H., & Schied, A. (2002). Stochastic finance: An introduction in discrete time. *Gruyters Studies in Mathematics*, 27.
77. Francioni, I., & Herzog, F. (2012). Probability-unbiased Value-at-Risk estimators, *Quantitative Finance*, 12, 755-768.
78. Garcia-Jorcano, L., & Novales, A. (2017). Volatility specifications versus probability distributions in VaR estimation. Manuscript.
79. Gerlach, R., Chen, C.W.S., Lin, E.M.H., & Lee, W.C.W. (2011). Bayesian forecasting for financial risk management, pre and post the global financial crisis. *Journal of Forecasting*, 31(8), 661-687.
80. Geweke, J. (1986). Modeling the persistence of conditional variances: a comment. *Econometric Review*, 5, 57-61.
81. Giacomini, R., & Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business and Economic Statistics*, 23(4), 416-431.
82. Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545-1578.
83. Giot, P., & Laurent, S. (2003a). Value-at-Risk for long and short trading positions. *Journal of Applied Econometrics*, 18, 641-664.
84. Giot, P., & Laurent, S. (2003b). Market risk in commodity markets: a VaR approach. *Energy Economics*, 25, 435-457.
85. Glosten, L., Jagannathan R., & Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48, 1779-1801.

86. Gneiting T. (2011) Making and evaluation point forecasts. *Journal of the American Statistical Association*, 106, 746-762.
87. Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and its Applications*, 1, 125-151.
88. Goncalves, S., & White, H. (2004). Maximum likelihood and the bootstrap for non- linear dynamics models. *Journal of Econometrics*, 119, 199-219.
89. Goncalves, S., & White, H. (2005). Bootstrap standard error estimates for linear regressions. *Journal of the American Statistical Association*, 100(471), 970-979.
90. Gonzalo, J., & Olmo, J. (2004). Which extreme values are really extreme? *Journal of Financial Econometrics*, 2(3), 349-369.
91. Graham, A., & Pál, J. (2014). Backtesting value-at-risk tail losses on a dynamic portfolio. *Journal of Risk Model Validation*, 8(2), 59-96.
92. Hansen, B. (1994). Autorregressive conditional density estimation. *International Economic Review*, 35, 705-730.
93. Hansen, P.R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7), 873-889.
94. Hansen, P.R., Lunde, A., & Nason J.M. (2003). Choosing the best volatility models: The model confidence set approach. *Oxford Bulletin of Economics and Statistics*, 65(s1), 839-861.
95. Hansen P.R., Lunde A., & Nason, J.M. (2011). The model confidence set. *Econometrica*, 79(2), 453-497.
96. Hentschel, L. (1995). All in the family nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics*, 39, 71-104.
97. Higgins, M.L., & Bera, A.K. (1992). A class of Non Linear Arch Models. *International Economic Review*, 33(1), 137-158.
98. Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistic*, 3(5), 1163-1174.
99. Hu, W. (2017). Calibration of multivariate generalized hyperbolic distributions using the EM algorithm, with applications in risk management, portfolio optimization and portfolio credit risk. Dissertation in the Florida State University.
100. Hull, J. (2015). Risk management and financial institutions. Pearson, Prentice Hall.

101. Hull, J., & White, A. (1998). Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk*, 1(1), 5-19.
102. Jackel, P. (2003). Monte Carlo methods in finance. Wiley Finance.
103. 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, 868-877.
104. Johnson, N.L. (1949). Systems of frequency curves generated by methods of translations. *Biometrika*, 36, 149-176.
105. Joiner, B.L., & Rosenblatt, J.R. (1971). Some properties of the range in samples from Tukey's symmetric lambda distributions. *Journal of the American Statistical Association*, 66(334), 394-399.
106. Jondeau, E., Poon, S.H., & Rockinger, M. (2000). Financial Modeling Under Non-Gaussian Distributions. Springer: London.
107. Jorion, P. (2007). Value at Risk: The new benchmark for controlling market risk. McGraw-Hill.
108. Kang, S.H., & Yong S-M. (2009). Value-at-Risk analysis for Asian emerging markets: asymmetry and fat tails in returns innovation. *The Korean Economic Review*, 25, 387-411.
109. Kerkhof, J., & Melenberg, B. (1995). Backtesting for risk-based regulatory capital. *Journal of Banking and Finance*, 28, 1845-1865.
110. Kilian, L. (1999). Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions? *Journal of Applied Econometrics*, 14(5), 491-510.
111. Kolmogorov, A. (1933). Sulla Determinazione Empirica di una Leggi di Distribucione. *Giornalle dell' Istitute Italiano degli Attuari*, 4, 1-11.
112. Kourouma, L., Dupre, D., Sanfilippo, G., & Taramasco, O. (2011). Extreme value at risk and expected shortfall during financial crisis. http://dx.doi.org/10.2139/ssrn.1744091
113. Kuester, K., Mittnik, S., & Paolella, M.S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1), 53-89.
114. Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17, 1217-1241.
115. Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 2, 174-184.

116. Lambert, P., & Laurent, S. (2001). Modelling financial time series using GARCH-type models with a skewed student distribution for the innovations. Mimeo, Université de Liege.
117. Lambert, N.S., Pennock, D.M., & Shoham, Y. (2008). Eliciting properties of probability distributions. In proceedings of the 9th ACM Conference on Electronic Commerce, ACM, 129-138.
118. Leccadito, A., Boffelli, S., & Urga, G. (2014). Evaluating the accuracy of value-at-risk forecasts: New multilevel tests. *International Journal of Forecasting*, 30, 206-216.
119. Lee, C.F., & Su, J.B. (2015). Value-at-Risk estimation via a semi-parametric approach: Evidence from the stock markets. Handbook of Financial Econometrics and Statistics. Springer Science Business Media: New York.
120. Liu, R.Y., & Singh, K. (1992a). Moving blocks jackknife and bootstrap capture weak dependence. LePage and Billiard (Eds), Exploring the limits of the bootstrap. Wiley: New York.
121. Liu, R.Y., & Singh, K. (1992b). Efficiency and robustness in resampling. *Annals of Statistics*, 20, 370-384.
122. Longin, F. (2005). The choice of the distribution of asset returns: how extreme value theory can help? *Journal of Banking and Finance*, 29, 1017-1035.
123. Lopez, J.A. (1998). Testing your risk tests. *Financial Survey* (May-Jun), 18-20.
124. Lopez, J.A. (1999). Methods for evaluating Value-at-Risk estimates. *Federal Reserve Bank of San Francisco Economic Review*, 2, 3-17.
125. Lopez, J.A., & Walter, C.A. (2000). Evaluating covariance matrix forecasts in a Value-at-Risk framework. http://dx.doi.org/10.2139/ssrn.305279
126. Louzis, D.P., Xanthopoulos-Sisinis, S., & Refenes, A.P. (2014). Realized volatility models and alternative Value-at-Risk prediction strategies. *Economic Modelling*, 40, 101-116.
127. Lugannani, R., & Rice, S.O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Advanced Applied Probability*, 12, 475-490.
128. Mabrouk, S., & Saadi, S. (2012). Parametric Value-at-Risk analysis: Evidence from stock indexes. *The Quarterly Review of Economics and Finance*, 52, 305-321.

129. 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, 1043-1075.
130. Massey, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46, 68-78.
131. McDonald, J.B., & Newey, W.K. (1988). Partially adaptive estimation of regression models via the generalized t distribution. *Econometric theory*, 4(3), 428-457.
132. McDonald, J., & Xu, Y. (1995). A generalization of the Beta distribution with applications. *Journal of Econometrics*, 66, 133-152.
133. McMillan, D.G., & Kambouroudis, D. (2009) Are RiskMetrics forecasts good enough? Evidence from 31 stock markets. *International Review of Financial Analysis*, 18, 117-124.
134. McNeil, A., & 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.
135. McNeil, A., Frey, A., & Embrechts, P. (2005). Quantitative risk management: Concepts, techniques and tools. Princeton University Press.
136. Miller, R. (1974). The jackknife: a review. *Biometrika,* 61, 1-15.
137. Mina, J., & Ulmer, A. (1999). Delta-Gamma four ways. Technical report, RiskMetrics Group. 1-17.
138. Mittnik, S., & Paolella, M.S. (2000). Prediction of financial downside-risk with heavy-tailed Conditional Distributions. http://dx.doi.org/10.2139/ssrn.391261
139. Nakajima, J., & Omori, Y. (2012). Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution. *Computational Statistics & Data Analysis*, 56(11), 690-3704.
140. Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347-370.
141. Nieto, M.R., & Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, 32, 475-501.
142. Osband, K.H. (1985). Providing incentives for better cost forecasting. PhD Thesis. University of California, Berkeley.

143. Ozun, A., Cifter, A., & Yilmazer, S. (2010). Filtered extreme-value theory for value- at-risk estimation: evidence from Turkey. *The Journal of Risk Finance*, 11(2), 164-179.
144. Pantula, S. (1986). Modeling the persistence in conditional variances: a comment. *Econometric Review*, 5, 71-74.
145. Paolella, M. (1997). Tail estimation and conditional modeling of heteroskedstic time-series. Ph.D Thesis, Institute of Statistics and Econometrics, Christian Albrechts University of Kiel.
146. Paolella, M.S., & Polak, P. (2015). COMFORT: A common market factor non- Gaussian returns model. *Journal of Econometrics*, 187(2), 593-605.
147. Pascual, L., Romo, J., & Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in garch models. *Computational Statistics and Data Analysis*, 50(9), 2293-2312.
148. Pearson, N.D., & Smithson, C. (2002). VaR the state of play. *Journal of Financial Econometrics*, 11, 175-189.
149. Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3, 119-131.
150. Pitera, M., & Schmidt, T. (2016). Unbiased estimation of risk. arXiv preprint arXiv:1603.02615.
151. Polanski, A., & Stoja, E. (2010). Incorporating higher moments into value-at-risk forecasting. *Journal of Forecasting*, 29, 523-535.
152. Politis, D., & Romano, J. (1994). The stationary bootstrap. *Journal of American Statistics Association*, 89, 1303-1313.
153. Quenouille, M. H. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Series B*, 11, 68-84.
154. Ramberg, J.S., & Schmeiser, B.W. (1974). An approximate method for generating asymmetric random variables. *Communications of the ACM*, 17(2), 78-82.
155. Righi, M.B., & Ceretta, P.S. (2015). A comparison of Expected Shortfall estimation models. *Journal of Economics and Business*, 78, 14-47.
156. Righi, M.B., & Ceretta, P.S. (2013). Individual and flexible Expected Shortfall backtesting. *Journal of Risk Model Validation*, 7(3), 3-20.
157. Riskmetrics, T.M. (1996). JP Morgan Technical Document.
158. Rocco, M. (2014). Extreme value theory in finance: a survey. *Journal of Economic Surveys*, 28(1), 82-108.

159. Romano, J.P., & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4), 1237-1282.
160. Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23, 470-472.
161. Ruiz, E., & Pascual, L. (2002). Bootstrapping financial time series. *Journal of Economic Surveys*, 16(3), 271-300.
162. Sarma, M., Thomas, S., & Shah, A. (2003). Selection of value at risk models. *Journal of Forecasting*, 22, 337-358.
163. Schaller, P. (2002). Uncertainty of parameter estimates in VaR calculations. http://dx.doi.org/10.2139/ssrn.308082
164. Schwert, W. (1990). Stock volatility and the crash of '87. *Review of Financial Studies*, 3, 77-102.
165. Sener, E., Baronyan, S., & Mengütürk, L.A. (2012). Ranking the predictive performances of value-at-risk estimation models. *International Journal of Forecasting*, 28, 849-873.
166. Simon, J.L. (1969). Basic research methods in social science. Random House.
167. Simonato, J.G. (2011). The performance of Johnson distributions for computing value at risk and expected shortfall. *The Journal of Derivatives*, 19(1), 7-24.
168. Smirnov, H. (1939). On the deviation of the empirical distribution function. *Recueil Mathematique (Matematiceskii Sbornik)*, N. S., 6, 3-26.
169. Smith, R. (1987). Estimating tails of probability distributions. *The Annals of Statistics*, 15, 1174-1207.
170. So, M.K.P., & Yu, P.L.H. (2006). Empirical analysis of GARCH models in value at risk estimation. *Journal of International Financial Markets, Institutions and Money*, 16, 180-197.
171. Stahl, G., Zheng, R., Kiesel, R., & Rühlicke (2012). Conceptualizing robustness in risk management. http://dx.doi.org/10.2139/ssrn.2065723
172. Tang, T.L., & Shieh, S.J. (2006). Long Memory in stock index futures markets: A value-at-risk approach. *Physica A*, 366, 437-448.
173. Taylor, S.J. (1986). Modelling Financial Time Series. John Wiley and Sons, Inc.
174. Theodossiou, P. (2001). Skewness and kurtosis in financial data and the pricing of options. Working Paper. Rutgers University.

175. Theodossiou, P. (1998). Financial data and skewed generalized t distribution. *Management Science*, 44, 1650-1661.
176. Tolikas, K. (2014). Unexpected tails in risk measurement: Some international evidence. *Journal of Banking and Finance*, 40, 476-493.
177. Tu, A.H., Wong, W.K., & Chang, M.C. (2008). Value-at-Risk long and short positions of Asian stock markets. *International Research Journal of Finance and Economics*, 22, 135-143.
178. Tukey, J. (1977). Exploratory data analysis. Addison-Wesley series in behavioral science. Adison-Wesley Publishing Company.
179. Tukey, J.W. (1958). Bias and confidence in not-quite so large samples. *Annals of Mathematical Statistics*, 29, 614.
180. Vose, D. (1996). Quantitative risk analysis: a guide to MC simulation modelling. John Wiley & Sons.
181. White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5), 1097-1126.
182. Wong, W.K. (2010). Backtesting value-at-risk based on tail losses. *Journal of Empirical Finance*, 17(3), 526-538.
183. Wong, W.K. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking and Finance*, 3, 1404-1415.
184. Wong, W.K., Fan, G., & Zeng, Y. (2012). Capturing tail risks beyond VaR. *Review of Pacific Basin Financial Markets and Polices*, 15(03), 1250015.
185. Zakoian, J.M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18, 931-955.
186. Zangari, P. (1996). An improved methodology for measuring VaR. *RiskMetrics Monitor*, 2nd quarter, 7-25.
187. Zhou, J. (2012). Extreme risk measures for REITs: A comparison among alternative methods. *Applied Financial Economics*, 22, 113-126.
188. Ziegel, J.F. (2016). Coherence and elicitability. *Mathematical Finance*, 26(4), 901-918.

Diciembre 2019

SANFI
SANTANDER FINANCIAL INSTITUTE
Fundación UCEIF

Títulos publicados

13/2014 Ana Carmen Díaz Mendoza y Miguel Ángel Martínez Sedano
Estudio sobre las sociedades gestoras de la industria de los fondos de inversión

14/2015 Pablo Ruiz-Verdú *et al.*
Riesgo bancario, regulación y crédito a las pequeñas y medianas empresas

15/2015 María Rodríguez Moreno
Systemic Risk: Measures and Determinants

16/2015 Federico Daniel Platania
Valuation of Derivative Assets under Cyclical Mean-Reversion Processes for Spot Prices

17/2016 Elena Cubillas Martín
Liberalización financiera y disciplina de mercado en diferentes entornos legales e institucionales. Implicaciones sobre el riesgo bancario

18/2016 Elena Ferrer
Investor Sentiment Effect in European Stock Markets

19/2016 Carlos González Pedraz
Commodity Markets: Asset Allocation, Pricing and Risk Management

20/2016 Andrea Ugolini
Modelling Systemic Risk in Financial Markets

21/2017 María Cantero Sáiz
Riesgo soberano y política monetaria: efectos sobre los préstamos bancarios y el crédito comercial

22/2017 Isabel Abínzano, Ana González, Luis F. Muga y Santiago Sánchez
Riesgo de crédito: análisis comparativo de las alternativas de medición y efectos industria-país

23/2018 Purificación Parrado-Martínez, Pilar Gómez y Antonio Partal
Probabilidad de incumplimiento e indicadores de riesgo en la banca europea: un enfoque regulatorio

24/2018 Iván Blanco y David Wehrheim
Three Essays in Financial Markets. The Bright Side of Financial Derivatives: Options Trading and Firm Innovation

25/2019 Alfredo Martín-Oliver, Anna Toldrà Simats y Sergio Vicente
Cambio tecnológico, reestructuración bancaria y acceso a financiación de las PYME

The thesis analyzes the effect that the sample size, the asymmetry in the distribution of returns and the leverage in their volatility have on the estimation and forecasting of market risk in financial assets. The goal is to compare the performance of a variety of models for the estimation and forecasting of Value at Risk (VaR) and Expected Shortfall (ES) for a set of assets of different nature: market indexes, individual stocks, bonds, exchange rates, and commodities.

The three chapters of the thesis address issues of greatest interest for the measurement of risk in financial institutions and, therefore, for the supervision of risks in the financial system. They deal with technical issues related to the implementation of the Basel Committee's guidelines on some aspects of which very little is known in the academic world and in the specialized financial sector.

In the first chapter, a numerical correction is proposed on the values usually estimated when there is little statistical information, either because it is a financial asset (bond, investment fund...) recently created or issued, or because the nature or the structure of the asset or portfolio have recently changed. The second chapter analyzes the relevance of different aspects of risk modeling. The third and last chapter provides a characterization of the preferable methodology to comply with Basel requirements related to the backtesting of the Expected Shortfall.

978-84-8102-912-3 14 €
www.editorial.unican.es
THEMA: KFFX