



Backtesting Bootstrap Value-at-Risk and Expected Shortfall estimates in GARCH models

André David Vale Novais da Rocha Novo

davidrochanovo@gmail.com

Master Dissertation

Master in Finance

Supervisor

Maria Eduarda da Rocha Pinto Augusto da Silva, PhD

September 8, 2014

Biographical Note

André David Novo holds a Bachelor in Economics by the University of Porto School of Economics and Management and a Post-graduation in Finance and Taxation by the Porto Business School. Upon completion of his undergraduate studies he completed some Banking and Accounting internships in institutions, such as Banco de Portugal and Banco Espírito Santo. He is currently pursuing a Master in Finance by the University of Porto School of Economics and Management and the Certificate in Quantitative Finance by the CQF Institute.

Abstract

This work focuses on the, to our best knowledge, first application and backtesting of the bootstrap methods in GARCH models of Pascual et al. (2006) and Chen et al. (2011) to the estimation of value-at-risk and expected shortfall, using data from the FTSE 100 index, as well as the comparison of their performance with the ones of the Filtered Historical Simulation and Historical Simulation. The accurate estimation of these risk measures is significantly relevant to the risk management decisions of financial institutions, as well as to fulfill the regulatory requirements, such as the ones imposed by the Basel II Accords. Previous existing methods have some limitations, such not including the uncertainty due to parameter estimation. In addition to the computational costs of the method developed by Chen et al. (2011) being 100 times lower than the one of Pascual et al. (2006), the former assumes a symmetric conditional return distribution, which, according to the empirical application developed in this work, seems to have a positive impact on the accuracy of the risk measures generated. Moreover, both methods seem to outperform the Historical Simulation and the Filtered Historical Simulation, as the former imposes unnecessary capital requirements to financial institutions, while the latter fails to predict extreme losses.

Key-words: Value-at-risk, Expected Shortfall, Bootstrap, GARCH, Backtesting

JEL-Codes: C22; C53; G32

Contents

1	Introduction	1
2	Literature Review	3
2.1	Asset Returns	3
2.2	The GARCH (1,1) Model	4
2.3	Value-at-Risk and Expected Shortfall	5
2.4	Estimation of the Risk Measures	6
2.4.1	Value-at-Risk Estimation	6
2.4.2	Expected Shortfall Estimation	7
2.5	Bootstrap methods in GARCH (1,1)	7
2.5.1	Filtered Historical Simulation	7
2.5.2	Bootstrap in GARCH (1,1) under uncertainty	8
2.5.3	Sieve Bootstrap in GARCH (1,1) under uncertainty	10
2.5.4	Critical Analysis of the Methods	11
2.6	Value-at-Risk Backtesting	12
2.6.1	Testing the Accuracy of a Model	13
2.6.2	Choosing between Accurate Models	14
2.7	Expected Shortfall Backtesting	16
3	Empirical Application	19
3.1	Data and Sample	19
3.2	Model Implementation	23
3.3	Backtesting Results	26
3.3.1	Accuracy Backtesting	26
3.3.2	Model Comparison	27
3.3.3	Expected Shortfall Backtesting	29
4	Conclusions	32

List of Tables

3.1	Descriptive Statistics of the FTSE 100 Daily Returns	21
3.2	Jarque-Bera, D'Agostino and Anscombe-Glynn Tests of the FTSE 100 Daily Returns	21
3.3	Ljung-Box Test and Lagrange Multiplier Test of FTSE 100 Daily Returns Sample	22
3.4	Unconditional Coverage, Serial Independence and Conditional Coverage Tests of VaR Estimates	29
3.5	RLF Hypothesis Tests of VaR Estimates	29
3.6	QLF Hypothesis Tests of VaR Estimates	30
3.7	Bootstrap ES Test of Exceptions	30

List of Figures

3.1	FTSE 100 Daily Returns	20
3.2	FTSE 100 Daily Returns Empirical Distribution and Normal Quantile- Quantile Plot	20
3.3	p -values of Ljung-Box Test and Lagrange Multiplier Test of FTSE 100 Daily Returns Sub-Samples	23
3.4	p -values of Ljung-Box test and Lagrange Multiplier Test of GARCH (1,1) Standardized Residuals	24
3.5	FTSE 100 PRR and USB VaR and ES Estimates	25
3.6	FTSE 100 FHS and HS VaR and ES Estimates	26
3.7	FTSE 100 Daily Returns, PRR and USB VaR Estimates	27
3.8	FTSE 100 Daily Returns, FHS and HS VaR Estimates	28

1 Introduction

Since the financial crisis of 2007-2008, the importance of risk management has become increasingly recognized, as shown by the Dodd–Frank Wall Street Reform and Consumer Protection Act, and the Brown–Vitter bill, among others, with the latter currently under analysis in the U.S. Congress. In this context, the accurate estimation of the market risk incurred by financial institutions is essential, both for internal and regulatory purposes. As a result of the stock market crash that had occurred in the Black Monday of October 19, 1987, a new risk measure was developed known as Value-at-Risk, VaR. Since then, the adoption of this measure has significantly increased, firstly by option of some financial institutions in order to summarize different risks, such as the aggregated firm risk (e.g. JP Morgan’s Risk Metrics), and afterwards by imposition of the Basel II Accords. Later, Acerbi and Tasche (2002) introduced expected shortfall, ES, a risk measure that unlike the former is coherent. However, as the backtesting of the latter is harder than that of the former, value-at-risk remains as the benchmark for regulatory purposes.

There is a wide range of methodologies available in the literature to estimate the mentioned risk measures. However, these have some limitations, such as some methods based on conditional heteroscedastic models, which assume a specific distribution of conditional returns. Others, such as historical simulation and bootstrap methods directly applied to returns, do not account for volatility clustering. Lastly, Filtered Historical Simulation and Extreme Value Theory based methods do not consider the uncertainty due to parameter estimation. Pascual et al. (2006) proposed a new bootstrap method applied to GARCH models that overcomes the above mentioned limitations. Later, Chen et al. (2011) developed a similar methodology but reducing computational costs by a factor of 100. With the outputs of the former methods, the estimation of point forecasts of value-at-risk and expected shortfall is straightforward, as Pascual et al. (2006) indicated as a possible application of their method, but which, to our best knowledge, has never been done.

Given this background, the purpose of this work is three-fold: to study whether these methodologies are appropriate for VaR and ES estimation, according to the backtesting procedures defined by the Basel II Accords, as well as to those proposed in the literature; to perform a comparative analysis of both methods to assess the impact of reducing

computational costs in the accuracy of the risk estimation; and lastly, to compare the performance of the methods with the ones of the Filtered Historical Simulation and of the Historical Simulation. While the former comparison is important to conclude if considering the uncertainty due to parameter estimation, increases the accuracy of the methods, the latter is relevant to show the impact of the current trend of returning to the simpler Historical Simulation for regulatory purposes. As such, the accuracy of the VaR estimates are tested and compared, using the approaches of the Basel Committee on Banking Supervision (1996), Kupiec (1995), Christoffersen (1998) and Sarma et al. (2003), while the test proposed by Righi and Ceretta (2013) is applied to the ES estimates. These analysis are performed using daily closing prices of the FTSE 100 index.

On Section 2, a review of the relevant literature is presented. This includes the definition and analysis of asset returns and their stylized facts, of the risk measures, of the main methodologies to estimate them, of the bootstrap methods of Pascual et al. (2006) and Chen et al. (2011), and lastly of the main VaR and ES backtesting procedures. Section 3 contains the implementation procedures and backtesting results of the four mentioned methods, as well as an analysis of the data used. Lastly, Section 4 includes the main conclusions of this work.

2 Literature Review

In this section, the main studies present in the literature are examined. Firstly, the concept of asset returns is presented, as well as their main empirical stylized facts. Secondly, the relevant features of the GARCH (1,1) model are described. Then, value-at-risk and expected shortfall are defined, followed by a discussion of the main methods to estimate them. Afterwards, the bootstrap methods in GARCH models are analysed. Lastly, the relevant backtesting procedures defined by the Basel Committee on Banking Supervision (1996), as well as in the literature are described.

2.1 Asset Returns

The simple return of an asset on a day t , assuming no dividends, is given by

$$R'_t = \frac{P_t}{P_{t-1}} - 1, \quad (2.1)$$

where P_t is the the asset closing price on day t . Another definition of asset return is the continuously compounded or log return on a day t , which is given by

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right), \quad (2.2)$$

where \ln is the natural logarithmic function. According to Ruppert (2011), log and simple returns are approximately equal, as $\ln(1+x) \approx x$ for x close to 0, as are daily returns. In this context, loss can be defined as the negative of the log return. This second definition of asset return is the one used throughout the rest of this paper.

According to Tsay (2002) and Taylor (2007), volatility clustering is present in every long time series of daily asset returns. This phenomenon can be described by the remark of Mandelbrot (1963) on asset prices: "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes". Moreover, Taylor (2007) states that there are three statistical properties, known as stylized facts, that can be observed for almost all large datasets of daily returns, independently of the market or time period under analysis. Firstly, the distribution of daily returns is

not normal. Despite being approximately symmetric, the former has fat tails and a high peak, when compared with the latter. The other two stylized facts are that daily returns of different days are almost not correlated, while there is correlation between absolute or squared daily returns on nearby days.

2.2 The GARCH (1,1) Model

It is well known that financial returns exhibit volatility clusters or, in other words, their volatility appears to be higher in some periods and lower on others. Moreover, Tsay (2002) states that volatility is a continuous variable that oscillates between a fixed interval. Consequently, Bollerslev (1986) introduces the Generalized Conditional Heteroscedastic (GARCH) models, in which conditional variance depends not only on past returns, but also on past volatility. The most simple specification of these is the GARCH (1,1) model, which is widely used to model asset returns and can be defined as follows:

$$y_t = \sigma_t \varepsilon_t \quad (2.3)$$

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (2.4)$$

where y_t and σ_t are the conditional return and conditional volatility, respectively, ε_t is a sequence of independent and identically distributed random variables with mean 0 and variance 1, with the conditional distribution of y_t coinciding with the one of ε_t , and $\alpha + \beta < 1$. The latter condition ensures that the unconditional variance is finite, while the conditional variance changes over time. The conditional variance can also be written as

$$\sigma_t^2 = \frac{\omega}{1 - \alpha - \beta} + \alpha \sum_{j=0}^{\infty} \beta^j \left(y_{t-j-1}^2 - \frac{\omega}{1 - \alpha - \beta} \right) \quad (2.5)$$

Tsay (2002) shows that equations (2.3) and (2.4) can be rewritten as an ARMA (1,1) process, as follows:

$$y_t^2 = \omega + (\alpha + \beta)y_{t-1}^2 + v_t - \beta v_{t-1} \quad (2.6)$$

where $v_t = y_t^2 - \sigma_t^2$.

2.3 Value-at-Risk and Expected Shortfall

Value-at-risk (VaR) is defined as the maximum potential loss that a portfolio can suffer, during a specific time period, in the best $100(1 - \alpha)\%$ scenarios, with $0 \leq \alpha \leq 1$. According to Artzner et al. (1999) this can be expressed as

$$VaR_t^{(\alpha)} = -\sup_{r \in \mathfrak{R}} [r | P(R_t \leq r) \leq \alpha], \quad (2.7)$$

where R_t is the portfolio return at time t .

Artzner et al. (1999) state that a coherent risk measure should have four properties: monotonicity, translation invariance, homogeneity and subadditivity. Moreover, those authors shows that VaR does not have the fourth property, proposing another risk measure that has all of them: conditional value-at-risk (CVaR). This is equal to the expected loss in the $100\alpha\%$ worst scenarios and can be expressed as

$$CVaR_t^{(\alpha)} = -E[R_t | (R_t \leq -VaR_t^{(\alpha)})]. \quad (2.8)$$

Furthermore, Inui and Kijima (2005) mention that VaR ignores the statistical properties of the return distribution beyond the specified quantile, emphasizing the need for a different risk measure. Acerbi and Tasche (2002) show that CVaR only has the subadditivity property for continuous return distributions and develop another coherent risk measure, which the authors name expected shortfall (ES). ES is valid for both continuous and discrete distributions and is equal to the CVaR in the former case. This measure can be expressed as

$$ES_t^{(\alpha)} = CVaR_t^{(\alpha)} + (\lambda - 1)(CVaR_t^{(\alpha)} - VaR_t^{(\alpha)}), \quad (2.9)$$

where $\lambda \equiv \frac{P(R_t \leq -VaR_t^{(\alpha)})}{\alpha} \geq 1$.

2.4 Estimation of the Risk Measures

2.4.1 Value-at-Risk Estimation

The vast literature on the main methodologies to estimate VaR can be divided in three main groups, according to their underlying assumptions: non-parametric methods which assume that returns are independent and identically distributed (*iid*); parametric methods, which assume a distribution for the standardized returns; and lastly extreme value theory and other methods that do not take into account the uncertainty due to parameter estimation.

One example of a non-parametric method that relies on the *iid* assumption is the historical simulation approach, which is implemented by Hendricks (1996), Vlaar (2000) and Brandolini and Colucci (2012), among others. This method calculates the $VaR_t^{(\alpha)}$ as the α quantile of the empirical distribution of past daily losses. Barone-Adesi and Giannopoulos (2001) implement an alternative non-parametric method, in which VaR is estimated through a bootstrap procedure directly applied to returns. The authors state that this type of methods does not consider the existence of volatility clusters, as a consequence of the *iid* assumption.

Given the limitations of the non-parametric methods, conditional heteroscedastic models are introduced in VaR estimation, so that it is possible to take into account volatility clustering. The first set of these models assume a distribution for the conditional returns by assuming one for the standardized residuals of the model. An example of these can be found in Giot and Laurent (2003), who apply IGARCH (1,1) and APARCH (1,1) models, with normal, student-t and skewed-t distributed standardized residuals. On the other hand, Sadorsky (2005) and Chang et al. (2011) assume a normal distribution in a stochastic volatility and in an autoregressive jump intensity model, respectively.

McNeil and Frey (2000) use, in the context of conditional heteroscedastic models, an approach that falls into the Extreme Value Theory category. The authors estimate directly the quantiles of the tail of the distribution by assuming *iid* standardized residuals. According to Extreme Value Theory that is an acceptable assumption, on the condition that the conditional mean and variance are correctly specified. On the other hand, Barone-Adesi

and Giannopoulos (2001) and Brandolini and Colucci (2012) also implement a bootstrap method known as Filtered Historical Simulation, random drawing with replacement from the original sample of standardized residuals, with the parameters being constant in all bootstrap replicates.

Other approaches include the semi-parametric and non-parametric specifications of the conditional mean and variance, such as the ones of Fan and Gu (2003) and Cai and Wang (2008), respectively. Lastly, another group of authors, such as Engle and Manganelli (2004), Taylor (2008) and Gouriroux and Jasiak (2008), estimate VaR by defining the evolution of the quantile over time as the dependent variable of their models, with their methods having some of the previously mentioned limitations.

2.4.2 Expected Shortfall Estimation

The estimation of ES has not received as yet as much attention as VaR in the literature, as the former is more recent and regulators never gave it as much relevance as they did to the latter. Acerbi and Tasche (2002), who developed this measure, carry on its calculation based on the estimation of VaR through historical simulation. Other authors implement other methods, based on the previously described for VaR estimation, with the ES corresponding to the expected value of the losses that are higher than the VaR. Examples of this can be found in McNeil and Frey (2000) and Christoffersen and Gonçalves (2005), with the former adopting an extreme value theory approach and the latter assuming a specific conditional distribution for the returns. Giannopoulos and Tunaru (2005) implement a filtered historical simulation method and compute the ES as the bootstrap sample average of the forecasted losses that are higher than the VaR, while Taylor (2008) estimates ES through quantile regression.

2.5 Bootstrap methods in GARCH (1,1)

2.5.1 Filtered Historical Simulation

The Filtered Historical Simulation (FHS), implemented by Barone-Adesi and Giannopoulos (2001) and Brandolini and Colucci (2012), includes a bootstrap method in a GARCH

(1,1) model that does not take into consideration the uncertainty due to parameter estimation. This procedure includes the following steps:

- Step 1: Estimate the parameters of the GARCH (1,1) model, $\hat{\theta}_T = (\hat{\omega}, \hat{\alpha}, \hat{\beta})$ for a sequence of T returns, y_t , for $t = 1, \dots, T$;
- Step 2: Compute the standardized residuals $\hat{\epsilon}_t$, for $t = 1, \dots, T$, with

$$\hat{\epsilon}_t = \frac{y_t}{\hat{\sigma}_t} \quad (2.10)$$

where $\hat{\sigma}_t$ is the fitted conditional volatility of the estimated GARCH (1,1) model;

- Step 3: Estimate bootstrap forecasts of future returns and volatilities K periods using the following recursions

$$\hat{\sigma}_{T+k}^{*2} = \hat{\omega} + \hat{\alpha}y_{T+k-1}^{*2} + \hat{\beta}\hat{\sigma}_{T+k-1}^{*2} \quad (2.11)$$

$$y_{T+k}^* = \hat{\sigma}_{T+k}^* \hat{\epsilon}_{T+k}^*, k = 1, \dots, K, \quad (2.12)$$

where $y_t^* = y_t$, $\hat{\epsilon}_{T+k}^*$ are random draws with replacement from the original standardized residuals series and

$$\hat{\sigma}_T^{*2} = \frac{\hat{\omega}}{1 - \hat{\alpha} - \hat{\beta}} + \hat{\alpha} \sum_{j=0}^{T-2} \hat{\beta}^j \left(y_{T-j-1}^2 - \frac{\hat{\omega}}{1 - \hat{\alpha} - \hat{\beta}} \right); \quad (2.13)$$

- Step 4: Repeat step 3, B times, where B is the number of bootstrap replicates.

For each k period ahead, the bootstrap conditional distribution of future returns and volatilities is determined by the B forecasts obtained for each period, assuming equal probability of occurrence of each of the computed future scenarios.

2.5.2 Bootstrap in GARCH (1,1) under uncertainty

Pascual et al. (2006) propose a bootstrap method (PRR) to estimate densities of future conditional returns and volatilities in GARCH models that takes into account the uncertainty due to parameter estimation. This procedure is described the following steps:

- Step 1: Estimate the parameters of the GARCH (1,1) model, $\hat{\theta}_T = (\hat{\omega}, \hat{\alpha}, \hat{\beta})$ for a sequence of T returns, y_t , for $t = 1, \dots, T$;

- Step 2: Compute the standardized residuals $\hat{\varepsilon}_t$, for $t = 1, \dots, T$, with

$$\hat{\varepsilon}_t = \frac{y_t}{\hat{\sigma}_t} \quad (2.14)$$

where $\hat{\sigma}_t$ is the fitted conditional volatility of the GARCH (1,1) model estimated;

- Step 3: Obtain a y_t^* , for $t = 1, \dots, T$, replicate sample with size equal to the original sample from the following recursions

$$\hat{\sigma}_t^{*2} = \hat{\omega} + \hat{\alpha} y_{t-1}^{*2} + \hat{\beta} \hat{\sigma}_{t-1}^{*2} \quad (2.15)$$

$$y_t^* = \hat{\sigma}_t^* \hat{\varepsilon}_t^*, t = 1, \dots, T, \quad (2.16)$$

where $\hat{\sigma}_1^{*2} = \frac{\hat{\omega}}{1 - \hat{\alpha} - \hat{\beta}}$ and $\hat{\varepsilon}_t^*$ is a random draws with replacement from the original standardized residuals series;

- Step 4: Re-estimate the GARCH (1,1) model for the replicate sample, resulting in a new set of parameters, $\hat{\theta}_T^* = (\hat{\omega}^*, \hat{\alpha}^*, \hat{\beta}^*)$;
- Step 5: Estimate bootstrap forecasts of future returns and volatilities K periods ahead using the following recursions

$$\hat{\sigma}_{T+k}^{*2} = \hat{\omega}^* + \hat{\alpha}^* y_{T+k-1}^{*2} + \hat{\beta}^* \hat{\sigma}_{T+k-1}^{*2} \quad (2.17)$$

$$y_{T+k}^* = \hat{\sigma}_{T+k}^* \hat{\varepsilon}_{T+k}^*, k = 1, \dots, K, \quad (2.18)$$

where $y_t^* = y_t$, $\hat{\varepsilon}_{T+k}^*$ are random draws with replacement from the original standardized residuals series and

$$\hat{\sigma}_T^{*2} = \frac{\hat{\omega}^*}{1 - \hat{\alpha}^* - \hat{\beta}^*} + \hat{\alpha}^* \sum_{j=0}^{T-2} \hat{\beta}^{*j} \left(y_{T-j-1}^2 - \frac{\hat{\omega}^*}{1 - \hat{\alpha}^* - \hat{\beta}^*} \right); \quad (2.19)$$

- Step 6: Repeat steps 3 to 5, B times, where B is the number of bootstrap replicates.

For each k period ahead, the bootstrap conditional distribution of future returns and volatilities is determined by the B forecasts obtained for each period, assuming equal probability of occurrence of each of the computed future scenarios.

2.5.3 Sieve Bootstrap in GARCH (1,1) under uncertainty

The ARMA(1,1) representation in equation (2.6) allows to implement the unconditional sieve bootstrap procedure (USB) proposed by Chen et al. (2011). This procedure includes consists of the following steps:

- Step 1: Estimate the ARMA(1,1) parameters, $\hat{\theta}_T = (\hat{\omega}, \widehat{\alpha + \beta}, \hat{\beta})$ for a sequence of T squared returns, y_t^2 , for $t = 1, \dots, T$;
- Step 2: Estimate the residuals \hat{v}_t , for $t = 2, \dots, T$ with $\hat{v}_1=0$ and

$$\hat{v}_t = y_t^2 - \hat{\omega} - (\widehat{\alpha + \beta})y_{t-1}^2 + \hat{\beta}\hat{v}_{t-1}; \quad (2.20)$$

- Step 3: Compute the centered residuals $\tilde{v}_t, t = 2, \dots, T$, with

$$\tilde{v}_t = \hat{v}_t - \frac{1}{T-1} \sum_{t=2}^T \hat{v}_t \quad (2.21)$$

- Step 4: Obtain a y_t^{*2} , for $t = 1, \dots, T$ replicate with size equal to the original sample from the following recursion:

$$y_t^{*2} = \hat{\omega} + (\widehat{\alpha + \beta})y_{t-1}^{*2} + \tilde{v}_t^* - \hat{\beta}\tilde{v}_{t-1}^*, \quad (2.22)$$

where \tilde{v}_t^* are random draws with replacement from the original centred residuals series, $\tilde{v}_0^* = 0$ and $y_0^{*2} = \frac{\hat{\omega}}{1 - (\widehat{\alpha + \beta})}$;

- Step 5: Re-estimate the ARMA (1,1) model for the replicate, resulting in a new set of parameters, $\hat{\theta}_T^* = (\hat{\omega}^*, (\widehat{\alpha + \beta})^*, \hat{\beta}^*)$;
- Step 6: Obtain the bootstrap sample volatility, $\hat{\sigma}_t^{*2}$, for $t = 2, \dots, T$, with the following recursion:

$$\hat{\sigma}_t^{*2} = \hat{\omega}^* + \hat{\alpha}^*y_{t-1}^{*2} + \hat{\beta}^*\hat{\sigma}_{t-1}^{*2} \quad (2.23)$$

where $\hat{\sigma}_1^{*2} = \frac{\hat{\omega}^*}{1 - (\hat{\alpha}^* + \hat{\beta}^*)}$;

- Step 7: Estimate bootstrap forecasts of future returns and volatilities K periods ahead using the following recursions

$$\hat{y}_{T+k}^{*2} = \hat{\omega}^* + (\widehat{\alpha + \beta})^* y_{T+k-1}^{*2} + \tilde{v}_{T+k}^* - \hat{\beta}^* \tilde{v}_{T+k-1}^* \quad (2.24)$$

$$\hat{\sigma}_{T+k}^{*2} = \hat{\omega}^* + \hat{\alpha}^* y_{T+k-1}^{*2} + \hat{\beta}^* \hat{\sigma}_{T+k-1}^{*2}, k = 1, \dots, K, \quad (2.25)$$

where $\hat{y}_T^{*2} = y_T^2$ and \tilde{v}_{T+k}^* are random draws with replacement from the original centred residuals series;

- Step 8: Repeat steps 4 to 7, B times, where B is the number of bootstrap replicates.

For each k period ahead, the bootstrap conditional distribution of future squared returns and volatilities is determined by the B forecasts obtained for each period, assuming equal probability of occurrence of each of the computed future scenarios. Moreover, for each k period ahead, the $\alpha/2$ and $1 - \alpha/2$ quantiles of the bootstrap conditional distribution of future returns are equal to -1 and 1 times the square root of the $1 - \alpha$ quantile of the one of future squared returns, respectively.

2.5.4 Critical Analysis of the Methods

Despite considering the existence of volatility clusters and not assuming a particular distribution, the Extreme Value Theory and the Filtered Historical Simulation do not consider the uncertainty due to parameter estimation. In fact, the former assumes that the conditional mean and variance are correctly specified and the latter keeps the parameters fixed for all bootstrap replicates. On the other hand, the bootstrapping methods of Pascual et al. (2006) and Chen et al. (2011) to predict densities of future returns and volatilities in GARCH (1,1), not only take into account the same features as do the previously methodologies, but also consider the stated uncertainty. Consequently, it is relevant to study the performance of these methods to VaR estimation.

Inui and Kijima (2005) state that the ES estimator may suffer from the same limitations as does the underlying VaR estimator. As such, it is also relevant to study the applicability of the mentioned bootstrap methods of Pascual et al. (2006) and Chen et al. (2011) to ES estimation, as these do not suffer from the already stated limitations of the previous approaches.

According to Chen et al. (2011), the sieve bootstrap is one of the most efficient and used bootstrap techniques for linear time-series and is already extensively studied in the literature. Moreover, the authors state that this procedure decreases computational time by a factor of 100, relatively to other bootstrap techniques for GARCH models, such as

the one proposed by Pascual et al. (2006). This results from the ability to estimate the model linearly, using for example Least Squares, after representing the GARCH model, as a linear ARMA model.

Chen et al. (2011) also defend that this method does not rely on any underlying assumption on the distribution of the standardized residuals. However, from their computation of the quantiles of the conditional distribution of future returns, this seems to be at least questionable. The mentioned methodology divides evenly the $1 - \alpha$ quantile of the conditional distribution of future squared returns between the first and last $\alpha/2$ quantiles of the one of future returns. Consequently, this assumes that the values corresponding to the mentioned quantiles are symmetric. Given that this must be true for all values of α , this method relies on the assumption that the conditional distribution of future returns is symmetric.

Following this, it is relevant to study the absolute, as well as the relative, performance of PRR and USB in the estimation of VaR and ES, in order to observe if the theoretical limitations of the sieve bootstrap method in GARCH models proposed by Chen et al. (2011) constrains its applicability to risk management. Moreover, the performance of these methods should be compared with the one of FHS, to conclude if the inclusion of the uncertainty to parameter estimation increases the accuracy of the bootstrap method.

2.6 Value-at-Risk Backtesting

Given the wide range of methods that can be used to estimate VaR and ES, a procedure, known as backtesting, is defined to determine whether a specific model is appropriate for the mentioned goal. All backtesting procedures rely on comparing the risk measures estimated by the model under analysis with the actual trading results. The cases in which the actual loss exceeds the VaR estimate are called exceptions. According to Christoffersen (1998), the exception sequence, $I_t^{(\alpha)}$, can be defined mathematically as follows:

$$I_t^{(\alpha)} = \begin{cases} 1, & \text{if } R_t < -VaR_t^{(\alpha)} \\ 0, & \text{otherwise} \end{cases} \quad (2.26)$$

for $t = T + 1, \dots, T + n$, where T is the number of return observations used to estimate the VaR of the day $T + 1$ and n is the number of one-step ahead estimates of that risk measure included in the test. Consequently, VaR backtesting consists in determining whether the probability of occurrence of an exception is significantly different from the defined α , or testing the hypothesis:

$$H_0 : E[I_t^{(\alpha)} | I_{t-1}^{(\alpha)}] = \alpha \quad (2.27)$$

2.6.1 Testing the Accuracy of a Model

The Basel Committee on Banking Supervision (1996) outlines a framework for back-testing methods used by financial institutions to estimate VaR, known as traffic light approach. This methodology is based on the method that was used by a considerable number of banks to test their internal models. According to this framework, the backtesting should include approximately 250 daily return observations and the corresponding 250 one-step ahead estimates of the 99% ($\alpha = 1\%$) VaR.

If the number of exceptions generated is equal or lower than four, the model is considered accurate (green zone), while the opposite is concluded if the model generates ten or more exceptions (red zone). The latter scenario results in an automatic increase of the financial institution's capital requirements by 33.33% and in the need to develop a new model. Lastly, if the number of exceptions is between 5 and 9, the model falls in the yellow zone and the financial institution must demonstrate that the model is accurate. Failing to do so results in an increase of the capital requirements proportionally to the number of exceptions above 4 and, in the most problematic cases, in a mandatory redesign of the model.

Another approach is proposed by Kupiec (1995). This author develops an unconditional coverage test to determine whether the probability of an exception occurring under a specific VaR model is α , or in other words to test the null hypothesis $H_0 : E[I_t^{(\alpha)}] = \alpha$, with the m exceptions observed in the n trials, using the following likelihood ratio statistic:

$$LR_{uc} = 2 \ln \left[\left(1 - \frac{m}{n}\right)^{n-m} \left(\frac{m}{n}\right)^m \right] - 2 \ln [(1 - \alpha)^{n-m} \alpha^m] \stackrel{\text{asy}}{\sim} \chi_{(1)}^2, \quad (2.28)$$

where $m = \sum_{t=T+1}^{T+n} I_t^{(\alpha)}$. If the proportion of exceptions is significantly higher or lower than α , the likelihood ratio statistic will be high and so the H_0 hypothesis will be rejected.

As the previous test, checks whether the unconditional expectation of $I_t^{(\alpha)}$ is α , and not the conditional one as stated in equation (2.21), Christoffersen (1998) proposes another test, in which the latter case is analysed. The author combines Kupiec (1995)'s unconditional coverage test with another, in which the serial independence of the exception sequence is checked, obtaining a complete conditional coverage test, or in other words testing the null hypothesis $H_0 : E[I_t^{(\alpha)} | I_{t-1}^{(\alpha)}] = \alpha$. The likelihood ratio statistic of the serial independence test is the following:

$$LR_{ind} = 2 \ln [(1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}] - 2 \ln [(1 - \pi)^{n_{00}+n_{10}} \pi^{n_{01}+n_{11}}] \stackrel{\text{asy}}{\sim} \chi_{(1)}^2, \quad (2.29)$$

where n_{ij} is the number of observations of the exception sequence with value i that are followed by another with value j , $\pi_{01} = \frac{n_{01}}{n_{00}+n_{01}}$ and $\pi_{11} = \frac{n_{11}}{n_{10}+n_{11}}$. If the exceptions are concentrated in specific periods of the backtesting sample, the likelihood ratio statistic will be high and so the hypothesis of serial independence will be rejected. Lastly, the likelihood ratio statistic for the complete conditional coverage test is the following:

$$LR_{cc} = LR_{uc} + LR_{ind} \stackrel{\text{asy}}{\sim} \chi_{(2)}^2. \quad (2.30)$$

2.6.2 Choosing between Accurate Models

The tests described above allow to infer whether a specific method generates accurate VaR estimates. However, using just them, it is not possible to decide which of the two accurate models is better for the mentioned purpose. Lopez (1999) proposes a methodology to do this that the author denominates loss function evaluation method, that consists in choosing the model with lower C_m , given by:

$$C_m = \sum_{t=T+1}^{T+n} C_{m,t}, \quad (2.31)$$

where

$$C_{m,t} = \begin{cases} f(R_t, VaR_{m,t}^{(\alpha)}), & \text{if } R_t < -VaR_{m,t}^{(\alpha)} \\ g(R_t, VaR_{m,t}^{(\alpha)}), & \text{if } R_t \geq -VaR_{m,t}^{(\alpha)} \end{cases} \quad (2.32)$$

where the index m assumes a different identifier for each estimation method and $f(x,y)$ and $g(x,y)$ are functions such that $f(x,y) \geq g(x,y)$, for a given y .

Lopez (1999) only considers the point of view of the regulator by considering $g(x,y) = 0$, or in other words just penalizing the scenarios in which the actual returns were higher than the estimated VaR, proportionally to the difference. Consequently, Sarma et al. (2003) call the loss function defined by Lopez (1999) the Regulatory Loss function, and defines it as follows:

$$C_{m,t} = \begin{cases} (R_t - VaR_{m,t}^{(\alpha)})^2, & \text{if } R_t < -VaR_{m,t}^{(\alpha)} \\ 0, & \text{if } R_t \geq -VaR_{m,t}^{(\alpha)} \end{cases} \quad (2.33)$$

Moreover, according to Sarma et al. (2003), financial institutions that use VaR estimation models that overestimate the risk, will be imposed higher capital requirements than they should actually hold. As such, these authors propose an additional loss function, called the firm's loss function, that penalizes not only the scenarios in which exceptions occur, representing a failure of the bank's risk management policy, but also the ones in which the actual returns were lower than the estimated VaR:

$$C_{m,t} = \begin{cases} (R_t - (-VaR_{m,t}^{(\alpha)}))^2, & \text{if } R_t < -VaR_{m,t}^{(\alpha)} \\ -\gamma(-VaR_{m,t}^{(\alpha)}), & \text{if } R_t \geq -VaR_{m,t}^{(\alpha)} \end{cases} \quad (2.34)$$

where γ measures the opportunity cost of capital.

Angelidis et al. (2004) propose the Quantile Loss function, which also penalizes for higher capital requirements:

$$C_{m,t} = \begin{cases} (R_t - (-VaR_{m,t}^{(\alpha)}))^2, & \text{if } R_t < -VaR_{m,t}^{(\alpha)} \\ (q_{(\alpha)}[R_t]_{T+1}^{T+n} - (-VaR_{m,t}^{(\alpha)}))^2, & \text{if } R_t \geq -VaR_{m,t}^{(\alpha)} \end{cases} \quad (2.35)$$

where $q_{(\alpha)}[R_t]_{T+1}^{T+n}$ is the quantile α of the empirical distribution of the daily returns used for backtesting the model.

Sarma et al. (2003) also propose an one-sided sign hypothesis test to infer whether model i is significantly superior than another model j for VaR estimation, relatively to a

given loss function. The null hypothesis of this test is $H_0 : \theta = 0$, against the alternative hypothesis $H_0 : \theta < 0$, where θ is the median of the distribution of z_t , with

$$z_t = C_{i,t} - C_{j,t}. \quad (2.36)$$

As the variable $C_{m,t}$ is defined in such a way that higher values of C_m are obtained for worse models, negative values of z_t mean that model i is superior to model j .

Sarma et al. (2003) define another two variables, ψ_t and S_{ij} , with the former corresponding to the number of non-negative z_t 's and the latter to the sum of all the observations of the former:

$$\psi_t = \begin{cases} 1, & \text{if } z_t \geq 0 \\ 0, & \text{if } z_t < 0 \end{cases} \quad (2.37)$$

$$S_{ij} = \sum_{t=T+1}^{T+n} \psi_t \quad (2.38)$$

According to Diebold and Mariano (2002), the standardized statistic S_{ij} is asymptotically standard normal:

$$\frac{S_{ij} - 0.5n}{\sqrt{0.25n}} \stackrel{\text{asy}}{\sim} N(0, 1). \quad (2.39)$$

Consequently, if S_{ij} is too low, the H_0 hypothesis is rejected, or in other words the model i is significantly better than the model j for VaR estimation purposes. This is aligned with the idea that underlies the test, as a low value of S_{ij} is equivalent to model i outperforming model j more often than the opposite occurs.

2.7 Expected Shortfall Backtesting

According to Wong (2008), ES is not considered in Basel II Accords, more specifically by the Basel Committee on Banking Supervision (1996), because its backtesting is significantly harder than that of VaR. Nevertheless, different approaches can be found in the literature, such as the residual of McNeil and Frey (2000), the censored gaussian of Berkowitz (2001) and the functional delta of Kerkhof and Melenberg (2004).

Wong (2008) and Righi and Ceretta (2013) state that the previously mentioned methodologies can be inaccurate in samples, such as the ones used in practice and for regulation purposes, as these tend to be small, while the former depend on asymptotic test statistics or, in other words, on samples large enough so that the test statistic actual distribution converges to the one considered in the test. Moreover, they consider the full sample size to perform the test, and not just the observations in which exceptions occurred. This is a limitation as ES focus only on the return behaviour that represents a loss higher than the VaR.

As a result, Wong (2008) proposes another test that overcomes these limitations and so that is accurate even in small samples of exceptions. As the previously mentioned test assumes that the returns, which breach the computed VaR level, follow a normal distribution, Righi and Ceretta (2013) develop a different test that does not rely on any distribution. The authors propose to test whether an exception that occurs in a given day n is significantly worse than the one that is expected for the computed VaR, with the following test statistic:

$$BT_{T+n} = \frac{R_{T+n} - ES_{T+n}^{(\alpha)}}{SD_{T+n}^{(\alpha)}}, \quad (2.40)$$

where

$$SD_{T+n}^{(\alpha)} = (\text{Var}[R_{T+n} | (R_{T+n} \leq -\text{VaR}_{T+n}^{(\alpha)})])^{1/2} \quad (2.41)$$

As the authors defend that the probability distribution of R_{T+n} is not known, they use instead the following equivalent test statistic:

$$BT_{T+n} = \frac{\hat{\varepsilon}_{T+n} - E[\hat{\varepsilon}_{T+n} | \hat{\varepsilon}_{T+n} < F^{-1}(\alpha)]}{\text{Var}[\hat{\varepsilon}_{T+n} | \hat{\varepsilon}_{T+n} < F^{-1}(\alpha)]^{1/2}} \quad (2.42)$$

where $\hat{\varepsilon}_t$ is the standardized residual of a return series that follows a GARCH process as the one defined in equations (2.3) and (2.4), and $F^{-1}(\alpha)$ is the α quantile of the former's distribution, which is known.

In order to compute the p -value of the test without assuming a specific distribution for the test statistic, Righi and Ceretta (2013) propose the following Monte Carlo simulation:

- Step 1: Draw n observations u_i from the distribution F ;
- Step 2: Calculate $E[u_i|u_i < G^{-1}(\alpha)]$ and $Var[u_i|u_i < G^{-1}(\alpha)]$, where $G^{-1}(\alpha)$ is the α quantile of the u_i sample;
- Step 3: For every $u_i < G^{-1}(\alpha)$ compute BT_i :

$$BT_i = \frac{u_i - E[u_i|u_i < G^{-1}(\alpha)]}{Var[u_i|u_i < G^{-1}(\alpha)]^{1/2}} \quad (2.43)$$

- Step 4: Calculate $P(BT_i < BT_{T+n})$;
- Step 5: Repeat steps 1 to 5, N times;
- Step 6: Compute the p -value as the median of the probabilities calculated in the N steps 5, and the critical value as the median of the α quantile of each BT_i series, obtained in the N steps 3.

If the p -value leads to the rejection of the hypothesis, then it can be concluded that the loss is significantly higher than the ES. This methodology allows to test its exception individually, with Righi and Ceretta (2013) stating that this is necessary, as models that fail to predict extreme losses should not be used for risk estimation.

3 Empirical Application

In this section, one-step ahead VaR and ES estimates of the FTSE 100 index, for the 253 trading days of 2013 are obtained, with $\alpha = 1\%$, using PRR and USB, as well as Historical Simulation (HS) and FHS, for comparison purposes. In order to backtest these estimates, the tests proposed by the Basel Committee on Banking Supervision (1996), Kupiec (1995) and Christoffersen (1998) are performed to test the accuracy of the model, while the test implemented by Sarma et al. (2003), using both the Regulatory Loss function of Lopez (1999) and the Quantile Loss function of Angelidis et al. (2004), is used to choose the best model. Lastly, the method of Righi and Ceretta (2013) is used to backtest the obtained ES estimates.

In order to implement and backtest the methods presented in this work, code was written in the *R* programming language. The packages *fGarch* and *tseries* are used to estimate the GARCH (1,1) model in the representation used by PRR and FHS, and in the one used by USB, respectively. The package *moments* supplies the functions required to perform the Jarque-Bera test for normality, D'Agostino test of skewness and Anscombe-Glynn test of kurtosis. Lastly, the package *PerformanceAnalytics* is used for plotting.

3.1 Data and Sample

The data under analysis includes 1262 observations of the daily closing prices of the FTSE 100 index, from 31 December 2008 to 31 December 2013. This data is extracted from the Datastream database. Daily continuously compounded returns are calculated, using equation (2.2), and scaled by a factor of 100. The first four years of data (1009 observations) are used to obtain the one-step ahead VaR and ES estimates for the first backtesting day. The remaining year (253 observations) is used to backtest the PRR, USB, FHS and HS risk estimates. Rolling windows, with the same number of observations as the original sample, are used to estimate the risk measures for the second backtesting day onwards, with the backtesting sample providing each day the additional required daily observation. The plot of these returns, shown on Figure 3.1, suggests the existence of volatility clustering.

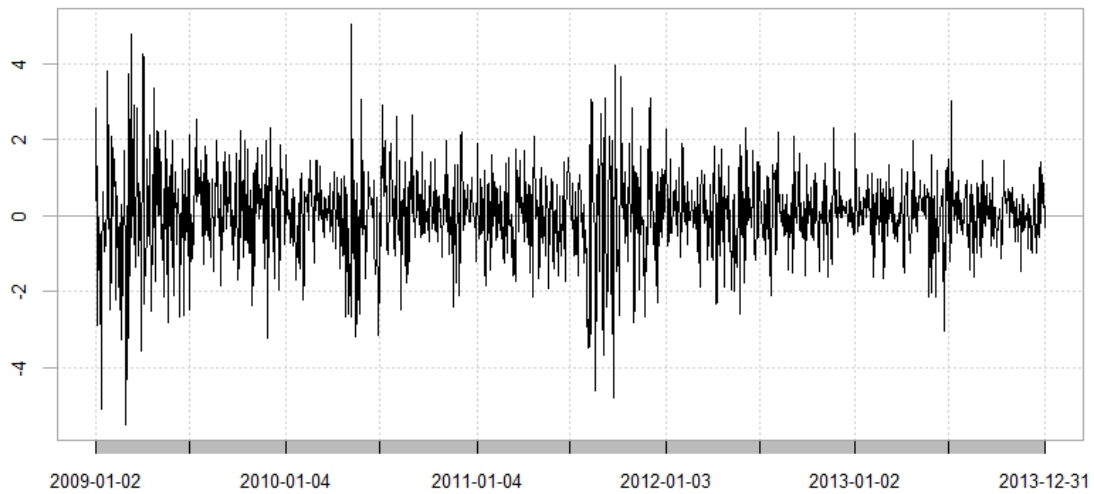


Figure 3.1: FTSE 100 daily continuously compounded returns from January 01, 2009 to December 31, 2013, totaling 1262 observations.

Plot (a) of Figure 3.2 shows the empirical density of the daily returns, as well as a normal distribution with the same mean and variance. Despite being approximately symmetric, the empirical density of the daily returns has a higher peak and seems to have fatter tails. The latter is confirmed in the normal quantile-quantile plot (b), in which can be seen that the first and last quantiles of the empirical distribution of the daily returns are respectively significantly lower and higher than the ones of the normal distribution.

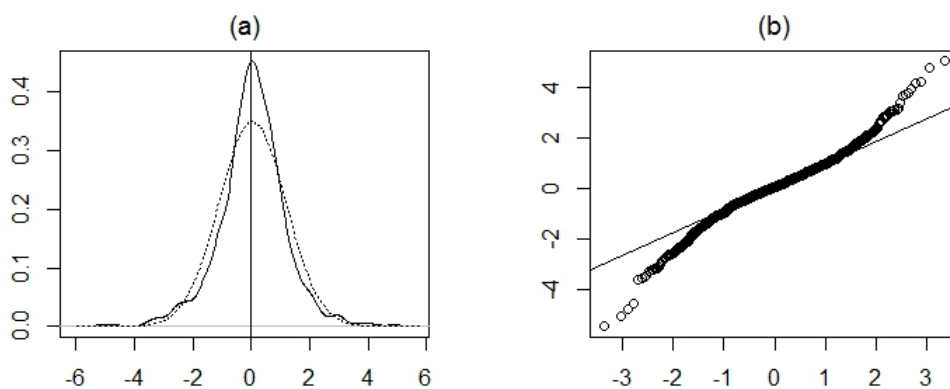


Figure 3.2: FTSE 100 daily returns empirical distribution (solid), normal distribution with the same mean and standard deviation (dotted) (a) and normal quantile-quantile plot (b).

Table 3.1 summarizes some descriptive statistics of the daily returns sample used both for estimation of the risk measures and for backtesting of the methods. The sample kurtosis above 3 suggests that the daily return distribution is *leptokurtic*, suggesting that this variable is not normally distributed. On the other hand, the skewness estimate is close to 0. In order to observe the statistical significance of the kurtosis and skewness estimates, the Jarque-Bera test for normality, D’Agostino test for skewness and Anscombe-Glynn test for kurtosis of the daily returns sample are performed. The results of these are shown in Table 3.2 and lead to the rejection of the hypothesis of normality, as well as the hypothesis of no excess kurtosis, for a 5% significance level. Nonetheless, the hypothesis of no skewness is not rejected, for the same significance level. As such, despite not being normal and showing evidence of excess kurtosis, the sample return distribution is approximately symmetric, confirming the conclusions previously derived from plotting.

Table 3.1: Descriptive statistics of the FTSE 100 daily continuously compounded returns from January 01, 2009 to December 31, 2013, totaling 1262 observations.

Mean	Median	Minimum	Maximum	Std. Deviation	Skewness	Kurtosis
0.0333	0.0592	-5.4805	5.0322	1.1437	-0.1722	5.2625

Table 3.2: Jarque-Bera test for normality, D’Agostino test of skewness and Anscombe-Glynn test of kurtosis of the FTSE 100 daily continuously compounded returns from January 01, 2009 to December 31, 2013, totaling 1262 observations.

Jarque-Bera test		D’Agostino test		Anscombe-Glynn test	
Statistic	<i>p</i> -value	Statistic	<i>p</i> -value	Statistic	<i>p</i> -value
277.19	0.0000	-1.6447	0.1000	8.4226	0.0000

Before applying the GARCH (1,1) model, described in section 2.2, some additional steps are required, as stated by Tsay (2002). According to the author, significant autocorrelation must be removed from the data, before applying a conditional heteroscedastic model, such as the GARCH (1,1). As such, significant autocorrelation was tested through a Ljung-Box test. This is equivalent to test $H_0 : \rho_1 = \dots = \rho_m = 0$, against the alternative

hypothesis $H_1 : \rho_l \neq 0, l \in \{1, \dots, m\}$, where ρ_l is the lag- l autocorrelation of the returns. If H_0 is rejected, the returns are significantly autocorrelated. Following the results of the simulation studies mentioned by Tsay (2002), $m \approx \ln(T)$, where T is the number of daily return observations included in the estimation of the risk measures. As the sample size is reduced by one when the daily returns are calculated, $m \approx \ln(1008) = 7$. This test is performed for the whole sample, as well as for every rolling window of data used, with their p -values being indicated in Table 3.3 and plotted in Figure 3.3, respectively. These do not indicate rejection of the hypothesis of no autocorrelation in either the sample or its sub-samples.

Table 3.3: Ljung-Box test and Lagrange Multiplier test of the FTSE 100 daily continuously compounded returns from January 01, 2009 to December 31, 2013, totaling 1262 observations.

Ljung-Box test		Lagrange Multiplier test	
Statistic	p -value	Statistic	p -value
4.4967	0.7211	124.213	0.0000

Following this, it is necessary to observe whether the squared residuals are serially correlated, as this is a requirement for applying a conditional heteroscedastic model, such as the GARCH (1,1). Therefore, the daily returns of the whole sample and of each rolling window are tested for autocorrelation in their squares, also known as ARCH effects, through a Lagrange Multiplier test. This is equivalent to perform the Wald test $H_0 : \alpha_1 = \dots = \alpha_m = 0$, against the alternative hypothesis $H_1 : \alpha_i \neq 0, i \in \{1, \dots, m\}$, in the following linear regression:

$$y_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \dots + \alpha_m y_{t-m}^2 + e_t, t = m + 1, \dots, T, \quad (3.1)$$

where e_t is the error term. The p -values for the whole sample and each rolling window are indicated in Table 3.3 and plotted in Figure 3.3, with the hypothesis of no autocorrelation in the squared returns being rejected for all the considered time frames. As such, given the presence of ARCH effects, the application of a conditional heteroscedastic model, such as

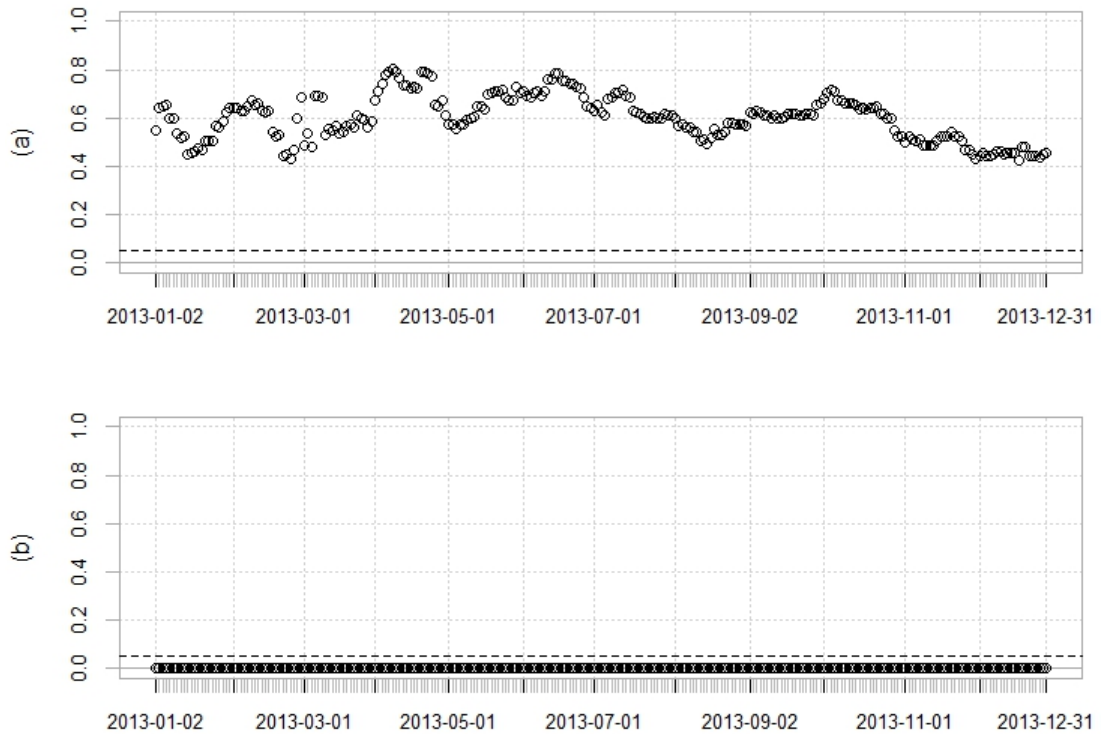


Figure 3.3: p -values of Ljung-Box test (a) and Lagrange Multiplier test (b) applied to the FTSE 100 daily continuously compounded returns sub-samples. The p -values of a given trading day result from performing the tests, on the data used to estimate the one-step ahead VaR and ES of that day.

the GARCH (1,1), is suitable. It is also worth noting that this daily returns sample shows all the stylized facts identified in the literature.

3.2 Model Implementation

After the previous tests, the parameters of a GARCH (1,1) model are estimated for each rolling window, and the corresponding standardized residuals computed. The latter are then tested for autocorrelation and ARCH effects, using the previously mentioned Ljung-Box test and Lagrange Multiplier test, respectively. The p -values of these can be found in Figure 3.4, indicating that the hypothesis of no autocorrelation in the standardized residuals and in their squares is not rejected. Consequently, the daily returns of each rolling window seem to be adequately described by a GARCH (1,1) process.

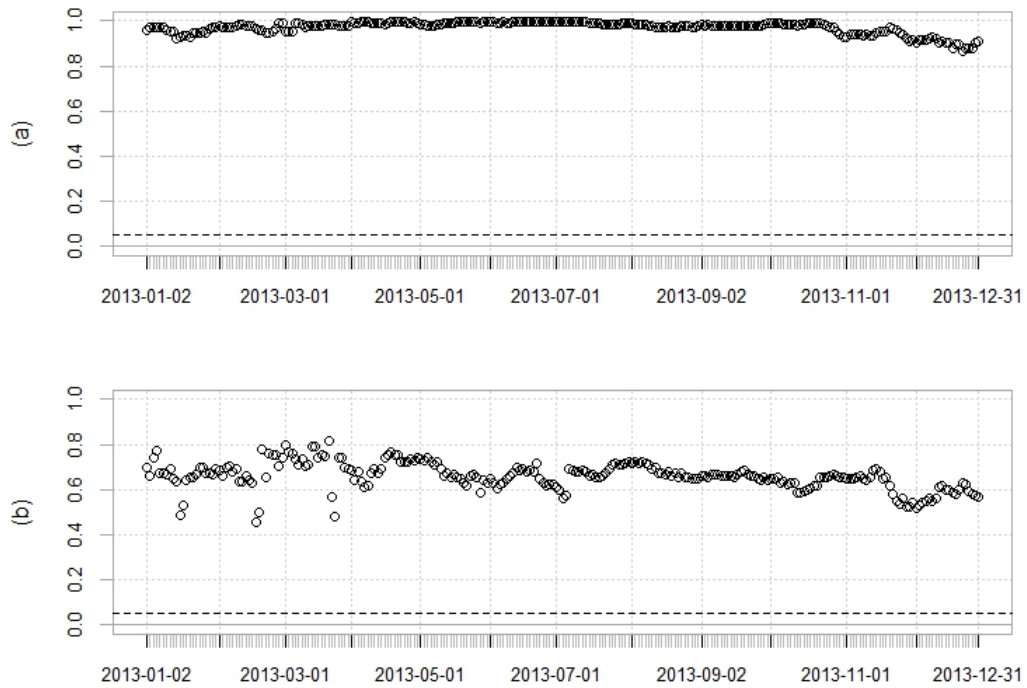


Figure 3.4: Ljung-Box test p -values (a) and Lagrange Multiplier test p -values (b) of GARCH (1,1) standardized residuals. The p -values of a given trading day result from performing the tests on the standardized residuals of the GARCH (1,1) model, estimated with the data used to estimate the one-step ahead VaR and ES of that day.

The FHS, PRR and USB bootstrap methods are then implemented. Regarding PRR and USB, instead of obtaining y_t^* or y_t^{*2} , for $t = 1, \dots, T$, in the respective step of section 2.5 that includes equation (2.16) or (2.22), they are generated for $t = 1, \dots, T + 150$ and the first 150 are discarded in the end of the process. This is done to reduce the effect of the initial values, which does not have a significant influence asymptotically, as stated by Chen et al. (2011). 1000 bootstrap replicates of y_{T+1}^* and y_{T+1}^{*2} are then generated or, in other words, B and K , mentioned in the description of the methods in section 2.5, were equal to 1000 and 1, respectively.

After sorting the bootstrap replicates in ascending order, the VaR estimates obtained with FHS and PRR are equal to the 1th quantile of the empirical distribution of the bootstrap replicates, while the ES ones are the average of the quantiles up to the first. All of these are multiplied by -1 to obtain the risk measure expressed as a loss and not as a

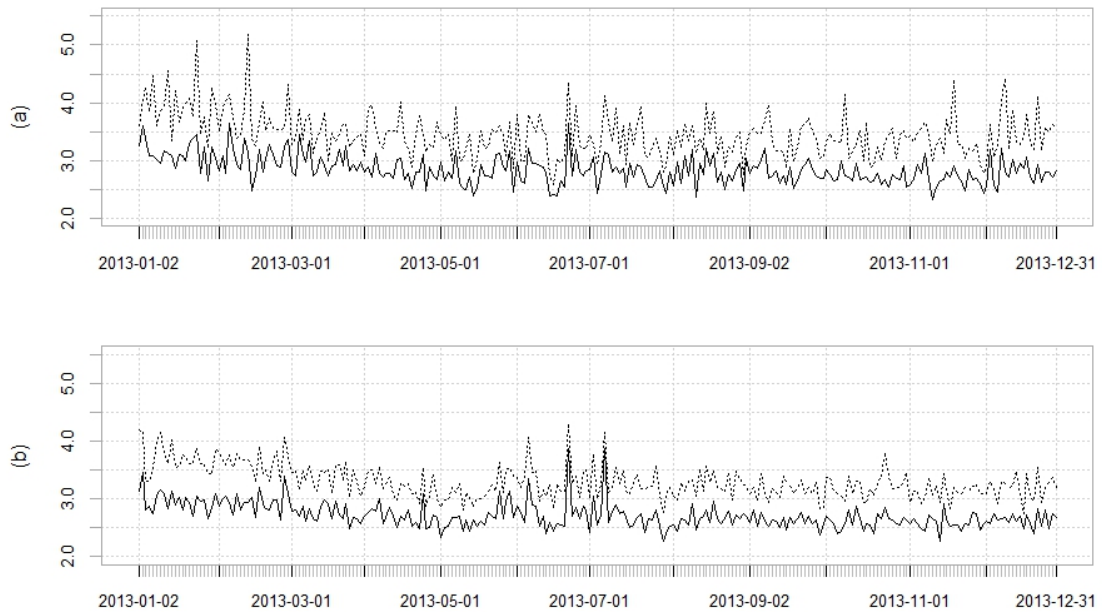


Figure 3.5: FTSE 100 PRR (a) and USB (b) VaR (solid) and ES (dotted) estimates, for the trading days from January 01, 2013 to December 31, 2013.

return. In the USB case, the VaR corresponds to the square root of the 98th quantile of the empirical distribution of the bootstrap replicates and the ES to the average of the square root of the last 20 replicates. Figure 3.5 represents the VaR and ES estimates obtained with PRR and USB, for the 253 trading days under analysis, for $\alpha = 1\%$.

The risk measures were also estimated using Historical Simulation (HS). On this case, the VaR corresponds to the 1th quantile of the empirical distribution of the past daily continuously compounded returns, while the ES one is the average of the quantiles up to the first. Figure 3.6 shows the VaR and ES estimates obtained with FHS and HS, for the 253 trading days under analysis, for $\alpha = 1\%$.

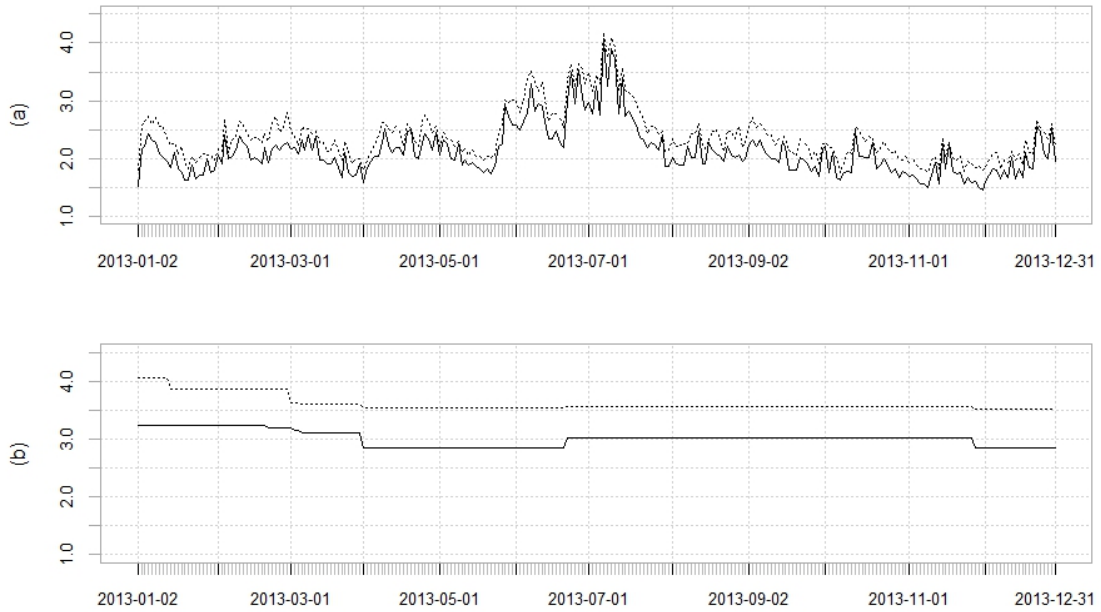


Figure 3.6: FTSE 100 FHS (a) and HS (b) VaR (solid) and ES (dotted) estimates, for the trading days from January 01, 2013 to December 31, 2013.

3.3 Backtesting Results

3.3.1 Accuracy Backtesting

After obtaining the VaR and ES estimates, their backtesting is required. As such, the backtest methodology proposed by the Basel Committee on Banking Supervision (1996) is firstly performed. After comparing the actual daily returns of the considered 253 trading days with the VaR estimates, it can be concluded that the number of exceptions generated by PRR and USB are 1 each. This can be observed on Figure 3.7. Regarding FHS and HS, the number of exceptions are 2 and 1, respectively, as shown in Figure 3.8. As the number of exceptions of the four methods is below 4, these are considered accurate for estimating VaR, according to the backtesting framework of the Basel Committee on Banking Supervision (1996).

Regarding the Unconditional Coverage Test of Kupiec (1995), the unconditional coverage hypothesis, $H_0 : E[I_t^{(\alpha)}] = \alpha$, is never rejected for any of the methods. The serial independence of the exceptions is then tested, through the Serial Independence Test of

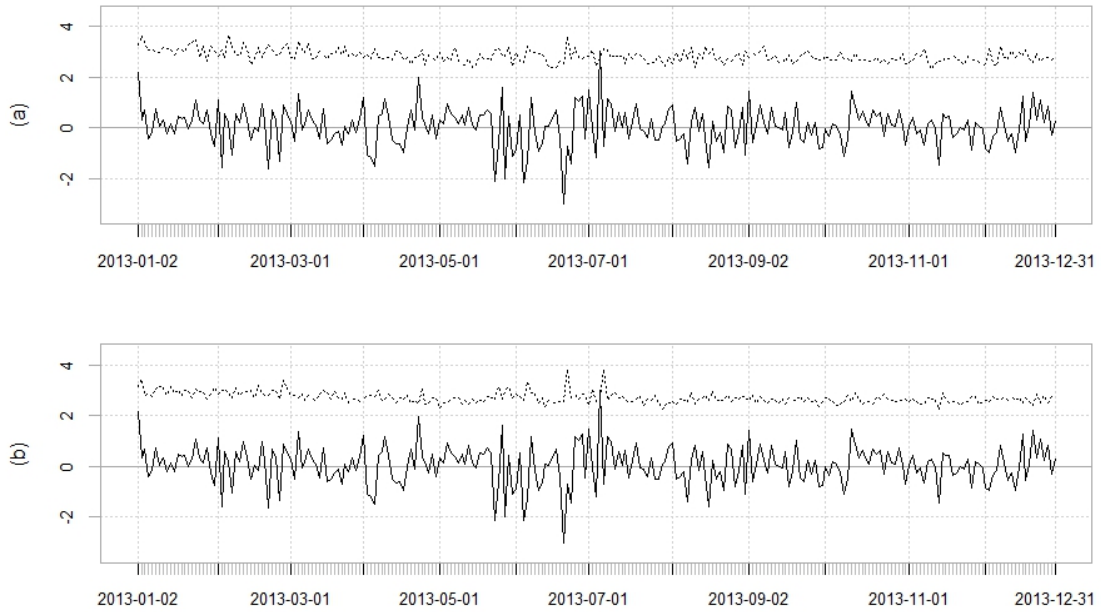


Figure 3.7: FTSE 100 daily continuously compounded returns from January 01, 2013 to December 31, 2013 (solid), PRR (a) and USB (b) VaR estimates (dotted).

Christoffersen (1998), with the hypothesis of serial independence not being rejected for any of the VaR estimates, given any significance level. Lastly, the Conditional Coverage Test of Christoffersen (1998), which combines the previous two, is performed. The conditional coverage hypothesis, $H_0 : E[I_t^{(\alpha)} | I_{t-1}^{(\alpha)}] = \alpha$, is not rejected for any of the mentioned methods. Consequently, the four methods are considered accurate for VaR estimation, according to the previous three tests. These results are shown in Table 3.4.

3.3.2 Model Comparison

As all four methods estimate VaR accurately, it is then necessary to determine which of them is better in doing so. Regarding the Regulatory Loss Function (RLF) of Lopez (1999), PRR and FHS have in 252 and 253 trading days, respectively, a similar or worse performance than the other three methods. USB has in 252 trading days a similar or worse performance than FHS and HS, and in 253 than PRR. Lastly, HS has in 253 trading days a similar or worse performance than PRR and USB, and in 252 than FHS. Nevertheless, these differences are not significant as, using Sarma et al. (2003) hypothesis test, the hypothesis of equal performance of both methods is not rejected, for any significance

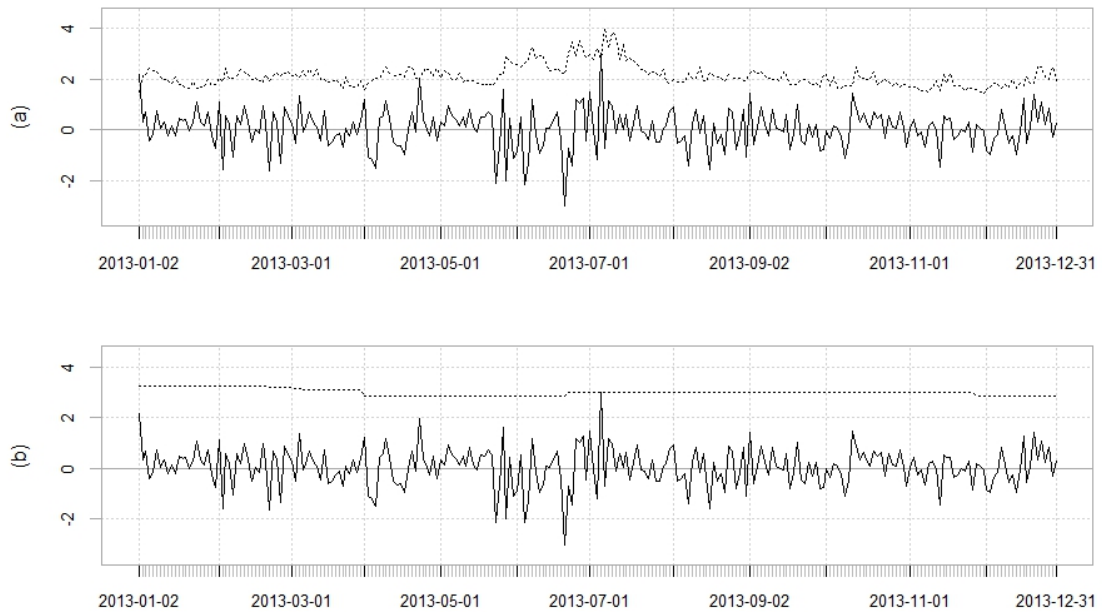


Figure 3.8: FTSE 100 daily continuously compounded returns from January 01, 2013 to December 31, 2013 (solid), FHS (a) and HS (b) VaR estimates (dotted).

level, independently of the alternative hypothesis. Consequently, regarding the regulatory side of risk or, in other words, just avoiding its underestimation, it is not significantly different to use one method or the other, for VaR estimation. These results are shown in Table 3.5.

On the other hand, regarding the Quantile Loss Function (QLF) of Angelidis et al. (2004), the difference is much wider. While, PRR has in 185, 24 and 56 trading days a similar or worse performance than USB, FHS and HS, USB has in 68, 230 and 14 days a similar or worse performance than PRR, FHS and HS, respectively. Regarding FHS and HS, these models have in 13, 23 and 13 trading days, and in 197, 239 and 240 trading days a similar or worse performance than PRR, USB, and HS and PRR, USB, and FHS, respectively. Using Sarma et al. (2003) hypothesis test, it is possible to reject some hypothesis of equal performance, with the alternative hypothesis being that PRR has a better performance than HS, USB has a better performance than PRR and HS, and lastly that FHS has a better performance than all the other methods. As such, the difference in performance of methods was deemed significant for any significance level, as the p -value

Table 3.4: Unconditional Coverage, Serial Independence and Conditional Coverage tests of VaR estimates

Method	LR_{uc}		LR_{ind}		LR_{cc}	
	Statistic	p -value	Statistic	p -value	Statistic	p -value
PRR	1.2129	0.2707	0.0080	0.9289	1.2209	0.5431
USB	1.2129	0.2707	0.0080	0.9289	1.2209	0.5431
FHS	0.1208	0.7281	0.0320	0.8580	0.1528	0.9264
HS	1.2129	0.2707	0.0080	0.9289	1.2209	0.5431

of the mentioned hypothesis tests was equal to 0. These results are shown in Table 3.6. The last mentioned fact leads to the conclusion that the worst performing methods significantly overestimate risk, and so their implementation over the better performers could result in higher unnecessary capital requirements for financial institutions.

Table 3.5: Regulatory Loss Function (RLF) hypothesis tests of VaR estimates. For each S_{ij} statistic and corresponding p -value, the models i and j are the one in the same column and row of the table, respectively.

	PRR		USB		FH		HS	
	S_{ij}	p -value	S_{ij}	p -value	S_{ij}	p -value	S_{ij}	p -value
PRR	253	1.0000	252	1.0000	252	1.0000	253	1.0000
USB	253	1.0000	253	1.0000	252	1.0000	253	1.0000
FHS	252	1.0000	252	1.0000	253	1.0000	252	1.0000
HS	252	1.0000	252	1.0000	252	1.0000	253	1.0000

3.3.3 Expected Shortfall Backtesting

Lastly, as in some of the trading days the actual daily return is lower than the VaR estimates, the hypothesis of no significant difference between the former (in loss form) and the ES estimate obtained for that day has to be tested. This is performed using the test

Table 3.6: Quantile Loss Function (QLF) hypothesis tests of VaR estimates. For each S_{ij} statistic and corresponding p -value, the models i and j are the one in the same column and row of the table, respectively.

	PRR		USB		FH		HS	
	S_{ij}	p -value	S_{ij}	p -value	S_{ij}	p -value	S_{ij}	p -value
PRR	253	1.0000	77	0.0000	20	0.0000	198	1.0000
USB	176	1.0000	253	1.0000	23	0.0000	239	1.0000
FHS	233	1.0000	230	1.0000	253	1.0000	240	1.0000
HS	55	0.0000	14	0.0000	13	0.0000	253	1.0000

proposed by Righi and Ceretta (2013), with the alternative hypothesis of the test being that the loss is higher than the ES estimate.

Table 3.7: Bootstrap ES test of exceptions

Method	Trading Day	Statistic	Critical Value			p -value
			1%	5%	10%	
PRR	118	0.1771	-2.1575	-2.1394	-1.8239	0.5400
USB	118	0.1827	-2.0429	-1.9841	-1.9128	0.3400
FHS	99	-0.4560	-0.9641	-0.9641	-0.9638	0.4300
FHS	118	-2.1403	-1.0820	-1.0820	-1.0808	0.0000
HS	118	0.8153	-2.0144	-2.0109	-1.8855	0.8600

In order to do so, a bootstrap method with 1000 simulations is performed, in which a sample with 10000 observations is randomly drawn from the estimated conditional distribution of daily returns obtained for that day. In other words, the statistic of the test is calculated with equation (2.40), with the corresponding changes on the "Monte Carlo simulation" proposed by the authors, which is in this case a bootstrap method, as in each iteration the sample is randomly drawn from the original one. This change is possible because PRR, USB, FHS and HS generate a conditional distribution of daily returns that allows the computation of equation (2.40), as well as the performance of the bootstrap.

Furthermore, in step 1 of this test, described in section 2.7, n is equal to 10000 and F corresponds to the estimated conditional distribution of the daily returns, while N is equal to 1000, in step 4.

The results of this test presented in Table 3.7 show that this hypothesis is not rejected for any of the exceptions observed, given any significance level, with the exception of the one generated on the 118th trading day of 2013 by FHS. As such, FHS should not be used for risk estimation, as it fails to forecast extreme losses. Regarding the other methods, it can be inferred that they are accurate for ES estimation.

4 Conclusions

In this work, the bootstrap methods in GARCH models of Pascual et al. (2006) and Chen et al. (2011) are analysed and extended to the estimation of VaR and ES for, to our best knowledge, the first time. Moreover, the previously implemented Filtered Historical Simulation and Historical Simulation are also considered in order to assess the relative performance of the two former methods. The applicability of these methods to risk estimation is then backtested, following the methodology required by the Basel Committee, as well as others present in the literature.

Regarding the assumptions underlying the different methods described in the literature, the methods of Pascual et al. (2006) and Chen et al. (2011) are not affected by the limitations of others, such as not taking into consideration the volatility clustering phenomenon, assuming a distribution for the conditional returns and not including the uncertainty due to parameter estimation. However, despite Chen et al. (2011) not specifying in their paper any assumption regarding the conditional return distribution, this method seems to require that the latter is symmetric.

From the empirical application presented in this work, it can be concluded that the methods proposed by Pascual et al. (2006) and Chen et al. (2011) are appropriate for risk estimation, with the confidence level required by Basel Committee on Banking Supervision (1996), as it was considered accurate by all VaR backtests and the exceptions that its estimates generated do not differ significantly from the ES estimated by the model. Nevertheless, despite their equal performance in avoiding risk underestimation, the method proposed by Chen et al. (2011) significantly outperforms the one of Pascual et al. (2006), when risk overestimation is also considered, as the latter overestimates risk more than the former, in the days in which exceptions do not occur. Consequently, it appears to be evident that the underlying assumption of the method of Chen et al. (2011) does not influence the risk estimation accuracy and leads to lower capital requirements for the financial institutions, which implement it over the one of Pascual et al. (2006), without having any significant disadvantage regarding avoiding risk underestimation, specially for the confidence level required by the Basel Committee on Banking Supervision (1996). As such, the method of Chen et al. (2011) seems to over perform the one of Pascual et al. (2006)

both in speed and accuracy of the risk estimates.

Moreover, the Filtered Historical Simulation and Historical Simulation are deemed accurate methods by all VaR backtests applied in this work. However, the Filtered Historical Simulation is rejected as an accurate model by the ES backtest of Righi and Ceretta (2013). As such, this method should not be used for risk estimation, as it fails to forecast extreme losses. This fact reinforces the importance of considering the uncertainty due to parameter estimation, on the bootstrap methods used to estimate risk. On the other hand, despite being completely accurate, the Historical Simulation imposes significantly higher capital requirements than the methods of Pascual et al. (2006) and Chen et al. (2011), showing that the current simplicity trend may in fact be a cost for financial institutions, without any apparent benefit.

Alternatively to the GARCH(1,1) process presented in this paper, different ones can be used to model financial returns, such as the GJR-GARCH model, which takes into account the fact that returns below their distribution mean have a higher impact on the volatility than do returns above it. As such, the analysis of the applicability to risk estimation of different models, such as the previous one, in the context of the described bootstrap techniques is one line of future research.

Bibliography

- Acerbi, C. and D. Tasche (2002). Expected Shortfall: A natural coherent alternative to Value at Risk. *Economic Notes* 31(2), 379–388.
- Angelidis, T., A. Benos, and S. Degiannakis (2004). The use of GARCH models in VaR estimation. *Statistical Methodology* 1(12), 105–128.
- Artzner, P., F. Delbaen, J. M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance* 9(3), 203–228.
- Barone-Adesi, G. and K. Giannopoulos (2001). Non-parametric VaR techniques: myths and realities. *Economic Notes* 30(2), 167–181.
- Basel Committee on Banking Supervision (1996). Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements.
- Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics* 19(4), 465–474.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327.
- Brandolini, D. and S. Colucci (2012). Backtesting value-at-risk: a comparison between filtered bootstrap and historical simulation. *Journal of Risk Model Validation* 6(4), 3–16.
- Cai, Z. and X. Wang (2008). Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics* 147(1), 120–130.
- Chang, T. H., H. M. Su, and C. L. Chiu (2011). Value-at-risk estimation with the optimal dynamic biofuel portfolio. *Energy Economics* 33(2), 264–272.
- Chen, B., Y. R. Gel, N. Balakrishna, and B. Abraham (2011). Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes. *Journal of Forecasting* 30(1), 51–71.
- Christoffersen, P. and S. Gonçalves (2005). Estimation risk in financial risk management. *Journal of Risk* 7, 1–28.

- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review* 39(4), 841–862.
- Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 20(1), 134–144.
- Engle, R. and S. Manganelli (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 22(4), 367–381.
- Fan, J. and J. Gu (2003). Semiparametric estimation of Value at Risk. *Econometrics Journal* 6(2), 261–290.
- Giannopoulos, K. and R. Tunaru (2005). Coherent risk measures under filtered historical simulation. *Journal of Banking & Finance* 29(4), 979–996.
- Giot, P. and S. Laurent (2003). Value-at-Risk for long and short trading positions. *Journal of Applied Econometrics* 18(6), 641–663.
- Gouriroux, C. and J. Jasiak (2008). Dynamic quantile models. *Journal of Econometrics* 147(1), 198–205.
- Hendricks, D. (1996). Evaluation of value-at-risk models using historical data. *Economic Policy Review* 2(1).
- Inui, K. and M. Kijima (2005). On the significance of expected shortfall as a coherent risk measure. *Journal of Banking & Finance* 29(4), 853–864.
- Kerkhof, J. and B. Melenberg (2004). Backtesting for Risk-based Regulatory Capital. *Journal of Banking & Finance* 28(8), 1845–1865. (Katholieke Universiteit Brabant).
- Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3(2), 73–84.
- Lopez, J. (1999). Methods for evaluating value-at-risk estimates. *Federal Reserve Bank of San Francisco Review* (2), 3–17.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business* 36(4), 394–419.
- McNeil, A. J. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7(34), 271–300.

- Pascual, L., J. Romo, and E. Ruiz (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics & Data Analysis* 50(9), 2293–2312.
- Righi, M. and P. Ceretta (2013). Individual and flexible expected shortfall backtesting. *Journal of Risk Model Validation* 7(3), 3–20.
- Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering*. New York: Springer.
- Sadorsky, P. (2005). Stochastic volatility forecasting and risk management. *Applied Financial Economics* 15(2), 121–135.
- Sarma, M., S. Thomas, and A. Shah (2003). Selection of Value-at-Risk models. *Journal of Forecasting* 22(4), 337–358.
- Taylor, J. W. (2008). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics* 6(3), 382–406.
- Taylor, S. J. (2007). *Asset Price Dynamics, Volatility, and Prediction*. New Jersey: Princeton University Press.
- Tsay, R. (2002). *Analysis of Financial Time Series*. New York: Wiley.
- Vlaar, P. J. G. (2000). Value at risk models for Dutch bond portfolios. *Journal of Banking & Finance* 24(7), 1131–1154.
- Wong, W. K. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking & Finance* 32(7), 1404–1415.