UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

Henrique Rodenbusch de Moura

On Range Value at Risk backtesting: A study from Expected Shortfall procedures

Porto Alegre, 2023.

Henrique Rodenbusch de Moura

On Range Value at Risk backtesting: A study from Expected Shortfall procedures

Dissertation presented as requirement for obtaining the degree of Master in Business Administration at the Business School of Federal University of Rio Grande do Sul.
Supervisor: Fernanda Maria Müller, PhD

Porto Alegre, 2023.

CIP - Catalogação na Publicação

Rodenbusch de Moura, Henrique
On Range Value at Risk backtesting: A study from Expected Shortfall procedures / Henrique Rodenbusch de Moura. -- 2023.
107 f.
Orientadora: Fernanda Maria Muller.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Escola de Administração, Programa de Pós-Graduação em Administração, Porto Alegre, BR-RS, 2023.

1. Risk Forecasting. 2. Risk Measures. 3. Range Value at Risk (RVaR). 4. Backtesting. 5. Monte Carlo simulation. I. Muller, Fernanda Maria, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Henrique Rodenbusch de Moura

On Range Value at Risk backtesting: A study from Expected Shortfall procedures

Dissertation presented as requirement for obtaining the degree of Master in Business Administration at the Business School of Federal University of Rio Grande do Sul.

Approved research. Porto Alegre, 2023:

Fernanda Maria Müller, PhD
Dissertation Advisor

Eduardo de Oliveira Horta, PhD
Committee Member

Marcelo Brutti Righi, PhD
Committee Member

Renata Rojas Guerra, PhD
Committee Member

Porto Alegre, 2023.

Ao meu pai.

Agradecimentos

Gostaria de agradecer, primeiramente, à minha mãe Gislaine Charão Rodenbusch, por sempre ter me dado o apoio e as condições necessárias para eu poder realizar os meus estudos até aqui, apesar de todas as adversidades da vida e, ao meu irmão Eduardo Rodenbusch de Moura que sempre ficou ao meu lado e acreditou em mim.

À minha companheira, Maria Victória Staggemeier Pasini, que esteve ao meu lado desde o início, por me encorajar a dar esse passo a favor do conhecimento e, mais ainda, pelo seu comprometimento e dedicação aos demais espectros das nossas vidas, proporcionando com que eu pudesse me dedicar a esta pesquisa, tornando possível a sua realização. Seu apoio foi fundamental nesta jornada e por isso eu sou grato.

À minha orientadora, Fernanda Maria Müller, quero agradecer, por ter me dado a oportunidade de ser seu orientando, acreditando no meu potencial e estando sempre presente para o esclarecimento de dúvidas e aperfeiçoamento do trabalho, em todas as suas etapas.

Por fim, quero agradecer ao Programa de Pós-Graduação em Administração e a todos os professores que de alguma forma contribuíram para a minha formação. Mais ainda, agradeço aos membros da banca examinadora pela disponibilidade e pelas contribuições pessoais acerca deste trabalho. Tais contribuições foram de grande valia e fundamentais para o aperfeiçoamento do mesmo.

"The difficult i'll do right now. The impossible will take a little while."

(Billie Holiday)

Abstract

Risk forecasting has played a major role in the risk management process, drawing the attention of financial organizations, regulators, and the academic community over the past decades to develop better tools to measure market risk. In this context, we slightly modified the definitions of Expected Shortfall (ES) backtesting procedures based on exceedance residuals with the main objective of making them effective for evaluating risk forecasts of Range Value at Risk (RVaR). To assess the performance of all procedures, we conducted Monte Carlo simulations to evaluate the size and power properties across different scenarios. In addition, we executed an empirical exercise with different asset classes, rolling window estimations, and significance levels. Jointly with each backtest, we applied the loss function of RVaR to assess the results and verify if they remained following both methods. We identified that none of the proposed backtesting procedures display significant superiority over the others in numerical and empirical analyses. Besides, the size and power of the tests deteriorate as the out-of-sample size increases. Also, with the increase of in-sample observations, we noted a degradation of sizes and improvement of powers for data generated with normal distribution at the significance levels of $\alpha = 1\%, \beta = 2.5\%$. For DGPs with third and fourth moments, the proposed procedures exhibit higher powers when the predictions are conducted using the normal distribution. Regarding the empirical application, we identified that the results closely follow both methods in the best-performing scenarios, exhibiting the opposite in the worst-performing ones.

Keywords: Risk Forecasting; Risk Measures; Range Value at Risk (RVaR); Backtesting; Monte Carlo simulation.

Resumo

Previsões de risco desempenham um papel importante no processo de gerenciamento de risco, atraindo a atenção de organizações financeiras, reguladores e da comunidade acadêmica nas últimas décadas para desenvolver ferramentas melhores para medir o risco de mercado. Nesse contexto, modificamos ligeiramente as definições dos procedimentos de backtesting do Expected Shortfall (ES) com base em resíduos de excedentes, com o principal objetivo de torná-los eficazes na avaliação das previsões de risco do Range Value at Risk (RVaR). Para avaliar o desempenho de todos os procedimentos, conduzimos simulações de Monte Carlo para examinar as propriedades de tamanho e poder em diferentes cenários. Além disso, realizamos um exercício empírico com diferentes classes de ativos, janelas rolantes e níveis de significância. Em conjunto com cada backtest, aplicamos a função de perda do RVaR para avaliar os resultados e verificar se eles se mantiveram seguindo ambos os métodos. Identificamos que nenhum dos procedimentos de backtesting propostos exibe superioridade significativa sobre os outros tanto nas análises numéricas quanto na análise empírica. Além disso, o tamanho e o poder dos testes se deterioram à medida que o tamanho da amostra fora do período de amostragem aumenta. Além disso, com o aumento das observações dentro da amostra, observamos uma degradação dos tamanhos e uma melhoria dos poderes para dados gerados com distribuição normal nos níveis de significância de $\alpha = 1\%, \beta = 2.5\%$. Para DGP's com terceiro e quarto momentos, os procedimentos propostos exibem maiores poderes quando as previsões são realizadas usando a distribuição normal. Em relação à aplicação empírica, identificamos que os resultados seguem de perto ambos os métodos nos cenários de melhor desempenho, exibindo o oposto naqueles de pior desempenho.

Palavras-Chave: Previsão de Risco; Medidas de Risco; Range Value at Risk (RVaR); Backtesting; Simulação de Monte Carlo.

Contents

1 INTRODUCTION **10**

2 BACKGROUND **13**
2.1 Notations and basic definitions **13**
2.2 Range Value at Risk **16**
2.3 Backtesting ES **16**
2.3.1 Acerbi and Szekely (2014) approach - Z_1 19

3 METHODOLOGY **21**
3.1 Proposed Approach **21**
3.1.1 Biswas and Sen (2023) approach - Z_1^A 23
3.2 Estimation Procedures **24**
3.2.1 Autoregressive model 24
3.2.2 GARCH model 25
3.2.3 Leverage effect 25
3.2.4 GAS model 26
3.2.5 Risk forecasting 27
3.3 Numerical Procedures **29**
3.4 Empirical Procedures **32**

4 RESULTS **34**
4.1 Numerical Analysis **34**
4.1.1 Size of the tests 34
4.1.2 Power of the tests 38
4.2 Empirical Analysis **46**

5 CONCLUSIONS **90**

BIBLIOGRAPHY **92**

APPENDIX A – TABLES **97**

1 Introduction

Risk management is a crucial competence of financial institutions such as banks, investment funds, and insurance industries and has drawn much attention in both the industry and academic community over the past decades. Techniques for measuring risk play a major role in the risk management process. Risk can be measured in terms of probability distributions, but it is useful to express it with a number that can be interpreted as a capital amount. Informally, quantitative tools that map random variables to capital amounts are called risk measures.

In 1996, the Basel Committee on Banking Supervision (BCBS) incorporated systematic risk as a supplement to credit risk and sanctioned the risk measure called Value at Risk (VaR) to be calculated for market risk capital requirement (TIAN et al., 2019). Since then, VaR has been one of the most popular risk management tools. In short, this measure describes, at a significance level, the maximum loss expected to occur by a financial position in a set time period. Despite its simplicity and ease of implementation, VaR has been criticized for not having important theoretical properties for measuring market risk. One of its deficiencies is the lack of subadditivity, which implies that it is not a coherent measure of risk (ARTZNER et al., 1999). Thus, the risk of a portfolio can be greater than the sum of individual asset risks, diverging from the diversification principle. Another weakness of VaR is that it ignores the potential losses beyond the quantile point of interest. In order to overcome these shortcomings, several alternative risk measures have been presented in the literature, the most accepted being the Expected Shortfall (ES), proposed by Acerbi and Tasche (2002). As stated by Daníelsson (2011), ES is the expected loss conditional on VaR being violated. By its definition, ES considers the magnitude of losses in addition to being a coherent measure of risk.

Although ES appears to be a more suitable risk measure than VaR, these are essentially theoretical properties. As stated by Müller and Righi (2018), there is a need for reliable estimates and forecasts under real applications. In this context, the literature has raised some statistical disadvantages of ES concerning VaR. As Gneiting (2011) showed, VaR is an elicitable measure of risk, while ES is not. A functional is called elicitable when it minimizes the expectation of some score function (ZIEGEL, 2016; ACERBI; SZEKELY, 2017). This property is quite important for risk management because it allows for assessing the quality of competing forecasting models by a scoring rule. Although ES is not directly elicitable, Fissler and Ziegel (2016) demonstrated that it could be joint elicitable with VaR. Also, Cont et al. (2010) point out the existence of a conflict between subadditivity and robustness (in a sense proposed by Hampel (1971)). In short, robustness can be understood as the stability of functional to small perturbations of the empirical distribution from the theoretical distribution of the data (KRÄTSCHMER et al., 2014). Hence, ES is not qualitatively robust, as it is a coherent risk measure.

To withdraw the weakness of ES not having robust estimators, Cont et al. (2010) slightly modified its definition and proposed the risk measure called Range Value at Risk (RVaR). This measure is defined as the average of all VaR over an interval of probability levels $\alpha, \beta \in [0,1]$ (FISSLER; ZIEGEL, 2021). RVaR is a conditional interquantile expectation since it lies in an interval generated by the VaR at the levels α and β (BARENDSE, 2017). Also, RVaR is a robust risk measure and comprises VaR and ES as limiting cases. Concerning elicitability, Wang and Wei (2020) demonstrated that RVaR does not have convex level sets and, therefore, is not directly elicitable. Fortunately, Fissler and Ziegel (2021) overcame this problem by showing that the triplet (VaR^{α}, VaR^{β}, $\mathrm{RVaR}^{\alpha,\beta}$) is elicitable, which implies that RVaR is jointly elicitable. Such features make RVaR a relevant measure of risk under real applications.

The key practical challenge for a given risk measure is to validate its estimates by checking whether realized losses observed ex-post align with the ex-ante estimates or forecasts. The statistical procedure by which realizations and forecasts can be compared is known as backtesting. This procedure plays a major role in financial management, as it can prevent underestimation of market risk and, hence, ensure that a financial agent carries sufficiently high capital and, at the same time, can reduce the likelihood of overestimating market risk, which can lead to excessive conservatism (DANÍELSSON, 2011). Despite the importance of this tool and the advantages presented by RVaR, there are still few studies on the subject, with particular attention given to Biswas and Sen (2023), who conducted a backtesting exercise modifying the Acerbi and Szekely (2014) approach for the RVaR measure and executed a simple simulation to evaluate their performance.

In this context, we propose modifications to established backtesting procedures in the literature to render them effective for evaluating Range Value at Risk forecasts. Over the last few years, theoretical progress in risk management has made backtesting ES possible with a sufficiently large variety of tests, each with its own particularity. Since ES is a special case of RVaR, we have addressed some of its best-known procedures and modified its definitions to our measure of interest. Against this background, we consider the violation residuals tests proposed by McNeil and Frey (2000), Righi and Ceretta (2015), and McNeil et al. (2015), which obtain the respective p-values from a nonparametric bootstrap test introduced by Efron and Tibshirani (1993). Such methods were selected based on their ease of implementation, coupled with their extensive utilization in the literature, with a particular emphasis on the former one (see, for instance, Singh et al. (2011), Siu (2021) and Nilsen (2022)).

The performance evaluation of each RVaR backtesting was done using both empirical and simulated data. We consider the Autoregressive (AR)-Generalized Autoregressive Score (GAS), the Autoregressive (AR)-Generalized Autoregressive Conditional Heteroskedasticity (GARCH), the Autoregressive (AR)-Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) and the Autoregressive (AR)-Glosten Jagannathan Runkle-Generalized Autoregressive Conditional Heteroskedasticity (GJR-GARCH) models with different probability distributions for the analyses. Garcia-Jorcano and Novales (2021) results indicate that the performance of risk models is associated with the probability distribution of innovations, and the choice of the models plays a secondary role. At first, we conducted Monte Carlo simulations to assess the size and power of the backtesting procedures. Monte Carlo simulation can be understood as a computational method that allows an experiment to be replicated enough times to the point that the empirical distribution of certain statistics of interest approximates the underlying probability measure (MCCRACKEN, 1955). So, with this finite samples method, we will investigate which backtestings perform best across different scenarios and sample sizes. In the second step, we performed an illustration using empirical financial data from distinct classes. Such a decision is due to the particularities of each asset class, providing an analysis of the developed backtesting behavior in diverse situations. Jointly with each backtest, we applied the loss function of RVaR proposed by Fissler and Ziegel (2021) to assess the results and verify if they remain following both methods. In accordance with what is commonly used in literature, we focus on the one-day-ahead estimation. Also, we consider different significance levels and rolling estimation windows for both numerical and empirical analysis.

The main contribution of our study lies in its pioneering exploration of different ES backtesting procedures based on violation residuals, modified to evaluate the quality of RVaR predictions, using Monte Carlo simulations to assess which proposed method presents the best size and power results. Previous studies, such as Du and Escanciano (2017) and Deng and Qiu (2021), considered the same approach but with the comparison of ES backtesting as the main objective. Furthermore, the work of Biswas and Sen (2023) proposes a backtesting procedure for RVaR employing only the Normal, Student-t, and Generalized Pareto distributions in conjunction with the standard GARCH model in its simulations. Furthermore, it lacks size and power tests, as well as an application with real financial data. On the other hand, our study aims to fill these gaps, contributing to the existing literature by conducting broader analyses with a greater number of parameters, models, and distributions, and thus, more rigorously evaluating the performance of the proposed approach. Finally, this study provides financial institutions, regulators, managers, and the academic community with a procedure that allows for evaluating risk forecasts from a measure different from those commonly used (VaR and ES) but with empirically relevant properties that make it an appropriate option in real situations.

The remainder of this project divides into the following contents: Chapter 2 addresses the notations and basic definitions used throughout the study, as well as the presentation of RVaR and descriptions of the ES backtesting; Chapter 3 addresses the proposed RVaR backtesting procedures, the description of the employed estimation methods and the methodological aspects of the numerical and empirical analysis; Chapter 4 presents the description of the size and power properties' results, as well as the empirical performance of the proposed backtesting procedures; and Chapter 5 summarizes and concludes the study. Finally, the Bibliography addresses the complete references cited throughout the text, and Appendix A gathers the additional tables developed for the empirical exercise.

2 Background

This chapter is divided into three sections. In section 2.1, we recall some of the most important axioms in the literature on risk measures and define the two most used risk measures, called Value at Risk (VaR) and Expected shortfall (ES). In section 2.2, we formally introduce the Range Value at Risk (RVaR). In the last section, we take a look at some of the most important backtesting procedures for the Expected Shortfall.

2.1 Notations and basic definitions

Consider the random result of any asset or portfolio X, where $X \geq 0$ is a profit and $X < 0$ a loss, defined for the space of feasible financial positions $\mathscr{X} := L^1(\Omega, \mathscr{F}, \mathbb{P})$, the space of essentially bounded random variables. We denote $\mathbb{E}[X] = \int_\Omega X d\mathbb{P}$, $F_X(x) = \mathbb{P}(X \leq x)$ and $F_X^{-1}(\alpha) = \inf\{x : F_X(x) \geq \alpha\}$, respectively, the expected value, the cumulative distribution function and the left quantile of $X \in \mathscr{X}$ for the significance level $\alpha \in (0,1)$. Also, we represent $\mathbb{1}(\cdot)$ as the indicator function that takes value 1 when the argument is true and zero otherwise. Furthermore, let $X^+ = \max\{X, 0\}$ and $X^- = \min\{X, 0\}$. A risk measure can be defined as the map $\rho : \mathscr{X} \to \mathbb{R}$ and is called monetary if, for any $X, Y \in \mathscr{X}$, it satisfies:

- **Monotonicity**: If $X \leq Y$, then $\rho(X) \geq \rho(Y)$;
- **Translation Invariance**: For any constant $m \in \mathbb{R}$, $\rho(X + m) = \rho(X) - m$.

If ρ fulfills both proprieties, it is monetary. A monetary risk measure is coherent in a sense proposed by Artzner et al. (1999) if it also satisfies:

- **Positive Homogeneity**: For any $\lambda \geq 0$, $\rho(\lambda X) = \lambda \rho(X)$;
- **Subadditivity**: $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

The wide majority of risk measures of practical interest in the academic literature also share the following property:

- **Law-Invariance**: If $F_X = F_Y$, then $\rho(X) = \rho(Y)$;

Briefly, Monotonicity is the requirement that the risk measure appraises as less risky those assets or portfolios with a lower loss. Translation Invariance means that adding an amount of cash m reduces the risk of the portfolio by the same amount m. Positive Homogeneity says that the risk of a position is directly proportional to its size. Subadditivity implies that the overall risk to which a portfolio is exposed cannot be worse than the sum of the individual risks of the positions that compose that portfolio – a manifestation of the diversification principle. Law-Invariance ensures that two positions with the same distribution function must have equal risks. For a detailed discussion of the axioms above, see Artzner et al. (1999), Frittelli and Gianin (2002) and Föllmer and Weber (2015).

Two of the most commonly used monetary risk measures (which fulfill law-invariance) are the Value at Risk (VaR) and Expected Shortfall (ES). Let $X \in \mathscr{X}$ arbitrary with the distribution function F_X. Given a significance level $\alpha \in (0,1)$, we can define the Value at Risk as:

$$VaR^{\alpha}(X) = -\inf\{x : F_X(x) \geq \alpha\} = -F_X^{-1}(\alpha). \tag{2.1}$$

Based on this formulation, we can observe that the VaR ignores the potential losses beyond the quantile point of interest. Furthermore, VaR has the drawback of not being coherent as it is not a subadditive measure of risk. The Expected Shortfall is defined as

$$ES^{\alpha}(X) = -\frac{1}{\alpha}\int_0^{\alpha} F_X^{-1}(s)ds, \tag{2.2}$$

was proposed by Acerbi and Tasche (2002) to overcome these shortcomings, as it considers the magnitude of losses, in addition to being a coherent risk measure.

For a number of competing forecasts or estimation procedures, we would like to compare them and decide which one performs best. This can be done if the risk measure has the statistical property called elicitability. To have a better understanding of this property, we must define a scoring function. In order to do so, we follow the definition proposed by Bellini and Bignozzi (2015), where, for any $x, y \in \mathbb{R}$, the function $S : \mathbb{R} \text{ x } \mathbb{R} \to [0, \infty)$ is called scoring function if it has the upcoming properties:

(i) $S(x,y) = 0$ if and only if $x = y$;

(ii) for any y, $x \mapsto S(x,y)$ is increasing for $x > y$ and decreasing for $x < y$;

(iii) for any y, $S(x,y)$ is continuous in x.

With this result, we can define elicitability. Thus, a law-invariant risk measure $\rho : \mathscr{X} \to \mathbb{R}$ is called elicitable if there exists a scoring function $S^{\rho} : \mathbb{R} \text{ x } \mathbb{R} \to [0, \infty)$ such that:

$$\rho(X) = -\underset{y \in \mathbb{R}}{\arg\min}\, \mathbb{E}[S^{\rho}(X,y)], \quad \forall X \in \mathscr{X}. \tag{2.3}$$

In this case, the scoring function S^{ρ} is strictly consistent with ρ.

As presented by Gneiting (2011), the VaR is elicitable and has the following consistent scoring function:

$$S^{VaR^{\alpha}}(x,y) = \alpha(x-y)^{+} + (1-\alpha)(x-y)^{-}. \tag{2.4}$$

On the other hand, Fissler and Ziegel (2016) showed that ES is not elicitable, but joint elicitable with Value-at-Risk and introduced a class of scoring functions that are strictly consistent with the pair $(\text{VaR}^{\alpha}(X), \text{ES}^{\alpha}(X))$. The scoring function of ES for real values, considering the adaption suggested by Gerlach et al. (2017), can be described by:

$$\begin{aligned} S^{ES^{\alpha}}(x,y,z) &= y(\mathbb{1}_{\{x<y\}} - \alpha) - x\mathbb{1}_{\{x<y\}} \\ &+ e^{z}\left(z - y + \frac{1}{\alpha}\mathbb{1}_{\{x<y\}}(y-x)\right) - e^{z} + 1 - \log(1-\alpha), \end{aligned} \tag{2.5}$$

where the function $S^{ES^{\alpha}}$ receives as input the values of the triplet $(X, \text{VaR}^{\alpha}(X), \text{ES}^{\alpha}(X))$. For more details regarding elicitability and scoring functions, see Gneiting (2011), Fissler and Ziegel (2016) and Ziegel (2016).

Another important statistical property when estimating risk measures is robustness. The notion of robustness adopted in this study is based on the concept of qualitative robustness, introduced by Hampel (1971). Informally, qualitative robustness refers to the stability of a statistical functional with respect to small perturbations of the empirical distribution from the 'true' theoretical distribution. According to Krätschmer et al. (2014), this concept is essentially equivalent to the weak continuity of the corresponding law-invariant risk measure at the true distribution with respect to the weak topology. Therefore, based on such equivalence, a law-invariant risk measure $\rho : \mathscr{X} \to \mathbb{R}$ is robust if:

$$\lim_{n\to\infty} d(F_n, F) = 0 \Longrightarrow \lim_{n\to\infty} |\rho(X_n) - \rho(X)| = 0, \tag{2.6}$$

where d is some metric, F_n, $n \geq 1$, and F are probability distributions, $X_n \sim F_n$, $n \geq 1$, and $X \sim F$. Usually, d is taken as the Prohorov or Lévy metrics (KRÄTSCHMER et al., 2014). For more details, we suggest Cont et al. (2010), Krätschmer et al. (2014) and Embrechts et al. (2015).

Cont et al. (2010) showed that the Value at Risk is a robust risk measure. More generally, they proved that if the quantile of the (true) loss distribution is uniquely determined, then the empirical quantile is a robust estimator. The authors also demonstrate the existence of a conflict between subadditivity and robustness. Thus, a law-invariant measure of risk ρ cannot be both coherent and robust. A direct implication of this result is that the ES is non-robust, as it is a coherent risk measure. To remove this drawback, Cont et al. (2010) slightly modified the definition of ES to be the average of VaR levels across a range of loss probabilities. In other words, they proposed a new risk measure called Range Value at Risk (RVaR). This measure is defined in the next subsection.

2.2 Range Value at Risk

Let α and β be two significance levels such that $0 < \alpha \leq \beta < 1$, the Range Value at Risk of $X \in \mathscr{X}$, $RVaR^{\alpha,\beta}(X)$, can be defined as:

$$RVaR^{\alpha,\beta}(X) = \begin{cases} \frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} VaR^{\gamma}(X)d\gamma, & \text{if } \alpha < \beta, \\ VaR^{\alpha}(X), & \text{if } \alpha = \beta. \end{cases} \tag{2.7}$$

The equation in (2.7) and the fact that the map $\gamma \mapsto VaR^{\gamma}(X)$ is decreasing imply that:

$$VaR^{\alpha}(X) \geq RVaR^{\alpha,\beta}(X) \geq VaR^{\beta}(X), \tag{2.8}$$

Thus, the RVaR lies in the closed interval generated by the VaR at the significance levels α, β.

The RVaR, as well as VaR, has the drawback of not being coherent, as it does not respect the axiom of subadditive. Moreover, Wang and Wei (2020) demonstrated that the RVaR, similarly to ES, is not elicitable. To overcome this issue, Fissler and Ziegel (2021) showed that RVaR is joint elicitable with Value-at-Risk and established a class of strictly consistent scoring functions $S : \mathbb{R} \text{ x } \mathbb{R}^3 \to \mathbb{R}$ for the triplet $(\text{VaR}^{\alpha}(X), \text{VaR}^{\beta}(X), \text{RVaR}^{\alpha,\beta}(X))$. A particular scoring function for RVaR, can be formulated as follows (FISSLER; ZIEGEL, 2021):

$$\begin{aligned} S^{RVaR^{\alpha,\beta}}(x,y,z,w) = {} & y(\mathbb{1}_{\{x<y\}} - \alpha) - x\mathbb{1}_{\{x<y\}} + z(\mathbb{1}_{\{x<z\}} - \beta) - x\mathbb{1}_{\{x<z\}} \\ & + (\beta-\alpha)\tanh((\beta-\alpha)w)\left[w + \frac{1}{\beta-\alpha}(S^{VaR^{\beta}}(x,z) - S^{VaR^{\alpha}}(x,y))\right] \\ & - \ln(\cosh((\alpha-\beta)w)) + 1 - \log(1-\alpha), \end{aligned} \tag{2.9}$$

where x, y, z, w denote the values of X, $\text{VaR}^{\alpha}(X)$, $\text{VaR}^{\beta}(X)$ and $\text{RVaR}^{\alpha,\beta}(X)$, respectively. The equation in (2.9) is also used by Müller et al. (2022).

Although RVaR has the drawbacks mentioned above, it has some advantages over VaR and ES, becoming a relevant risk measure in practice. Given its robustness, the RVaR has more stability than ES with respect to small perturbations of an empirical distribution $F \in \mathscr{F}$. Differently from VaR, it (partially) considers the magnitude of losses by averaging returns over an interval beyond a quantile point of interest. Also, as the RVaR provides the expectation over an interval, it is not subject to problems related to extreme tail values.

2.3 Backtesting ES

In the next subsections, we describe some of the best-known backtesting procedures for the Expected Shortfall. For a better understanding of the following procedures, it is necessary to make some initial definitions. Given an out-of-sample or future period $t = 1, \cdots, T$, we consider X_t the return of some asset at time t as well as μ_t and σ_t the conditional mean and standard deviation that are measurable with respect to the information available up to $t-1$ of the same asset. Besides, the measures VaR_t^{α} and ES_t^{α} denote the Value at Risk and Expected Shortfall forecasts for possible losses of some asset return at time t, for a significance level $\alpha \in (0,1)$, also measurable with respect to the information available up to $t-1$. For the definition of backtesting, we consider that the risk measures have their sign adjusted. We emphasize that this will be maintained in all situations where the backtesting is mentioned.

McNeil and Frey (2000) propose a backtest based on the series r, called exceedance residuals, which represents the size of the discrepancies between the observed returns and the estimated ES in the event of VaR violations[1]. For the out-of-sample period $t = 1, \cdots, T$ and a given significance level α, the residuals can be defined as:

$$r_t := \frac{X_t - ES_t^{\alpha}}{\sigma_t}. \tag{2.10}$$

Thus, (2.10) gives us information about how many dispersion measures (here, standard deviations) an observed return is apart from the estimated value of ES. In this way, the exceedance residuals series r can be denoted by:

$$r = \{r_t : \; X_t < VaR_t^{\alpha}\}. \tag{2.11}$$

The null and alternative hypotheses are given by:

$$\mathbb{H}_0 : \mathbb{E}[\, r\,] = 0$$
$$\mathbb{H}_1 : \mathbb{E}[\, r\,] < 0.$$

Under the null hypothesis, the residuals should behave like an i.i.d. sample with a mean of zero, against the alternative hypothesis that their mean is less than zero. Rejecting the null hypothesis implies that the risk is underestimated. To test the hypothesis of mean zero, the authors suggest using a standard non-parametric bootstrap test, introduced by Efron and Tibshirani (1993). Briefly, in the nonparametric approach, the underlying data distribution is not known and the resampling process is carried out based on the empirical distribution of the data. Thus, such process is obtained through resampling, with replacement, of the original sample. To assess forecasting results, this method is usually set with 1000 bootstrap replications; see Catania and Grassi (2022).

Righi and Ceretta (2015) propose an reconfiguration of McNeil and Frey (2000) procedure. Distinct from (2.10), the authors consider dispersion only in cases where violations occur. To be precise, the dispersion measure suggested in this approach is called Shortfall Deviation (SD), which is the truncated variance squared root for some quantile conditional to the probability α. The following formulation defines the SD:

$$SD^{\alpha}(X) = (\mathbb{V}[X|X < VaR^{\alpha}(X)])^{\frac{1}{2}} = \left(\frac{1}{\alpha}\int_0^{\alpha} (VaR^{\gamma}(X) - ES^{\alpha}(X))^2 \, d\gamma\right)^{\frac{1}{2}}. \tag{2.12}$$

[1] A violation occurs when an observed return is lower than the calculated VaR, for a certain period t.

Such dispersion is a better estimate than the standard deviation of the full sample since when extreme negative returns occur, the financial market agents are interested in the left tail risk (RIGHI; CERETTA, 2013). So, this backtest is based on the violation residuals series r', which differs from (2.10) by using the SD defined in (2.12) as the dispersion measure instead of the standard deviation of the full sample. Thus, for the out-of-sample period $t = 1, \cdots, T$ and a given significance level α, the residuals can be defined by:

$$r'_t = \frac{X_t - ES_t^{\alpha}}{SD_t^{\alpha}}. \tag{2.13}$$

Similarly to the previous procedures, the equation (2.13) gives us information about how many dispersion measures (here, shortfall deviations) an observed return is apart from the estimated value of ES. In this way, the series of violation residuals r' can be denoted by:

$$r' = \left\{ r'_t : \; X_t < VaR_t^{\alpha} \right\}. \tag{2.14}$$

The null and alternative hypotheses for long positions can be formulated as:

$$\begin{aligned} \mathbb{H}_0 &: \mathbb{E}[\, r'_1 \,] = 0 \\ \mathbb{H}_1 &: \mathbb{E}[\, r'_1 \,] < 0, \end{aligned}$$

and for short positions, they are given by:

$$\begin{aligned} \mathbb{H}_0 &: \mathbb{E}[\, r' \,] = 0 \\ \mathbb{H}_1 &: \mathbb{E}[\, r' \,] > 0. \end{aligned}$$

So, under the null hypothesis, the residuals should behave like an i.i.d. sample with a mean of zero, against the alternative hypothesis that the mean is negative for long positions and positive for short positions[2].

Another modification of McNeil and Frey (2000) procedure is suggested by McNeil et al. (2015). This backtest is based on the violation residuals series r'', which differs from (2.10) by using the difference between the estimated value of ES and the conditional mean of the full sample as the dispersion measure. Hence, for the out-of-sample period $t = 1, \cdots, T$ and a given significance level α, the residuals can be defined by:

$$r''_t = \frac{X_t - ES_t^{\alpha}}{ES_t^{\alpha} - \mu_t}. \tag{2.15}$$

So, this last procedure gives us information similar to the past tests, but now with a third dispersion measure, i.e., it provides how many differences between expected loss and conditional mean an observed return is apart from the same expected loss. In this way, the series of violation residuals r'' can be denoted by:

$$r'' = \left\{ r''_t : \; X_t < VaR_t^{\alpha} \right\}. \tag{2.16}$$

[2] Long position involves buying an asset with the hope of its value rising to make a profit, while a short position entails selling an asset not owned, betting on its price drop, aiming to repurchase it at a lower price for profit.

The null and alternative hypotheses are as follows:

$$\mathbb{H}_0 : \mathbb{E}[\, r'' \,] = 0$$
$$\mathbb{H}_1 : \mathbb{E}[\, r'' \,] < 0.$$

Under the null hypothesis, the residuals of r'' should behave like an i.i.d. sample with a mean of zero, against the alternative hypothesis that the mean is less than zero. Here, again, if we reject $\mathbb{H}_0$, the risk is underestimated. To test the zero mean hypothesis, the authors indicate two different approaches: a standard t-test, defined by:

$$M_{ES} = \frac{\sqrt{T}\,\bar{r}''}{\sigma}, \tag{2.17}$$

where $\bar{r}''$ is the sample mean of the violation residuals of size T, and σ denotes its standard deviation; and the same nonparametric bootstrap method already discussed.

2.3.1 Acerbi and Szekely (2014) approach - Z_1

For illustrative purposes, this subsection is dedicated to the definition of one of the proposed backtesting procedures of Acerbi and Szekely (2014), which was used and modified by Biswas and Sen (2023) for the case of the RVaR measure. More specifically, the former authors proposed three different backtestings for ES, where the approach of our interest is based on the conditional expectation of ES, which, given a significance level α and the out-of-sample period $t = 1, \cdots, T$, is equivalent to:

$$ES_t^{\alpha} = \mathbb{E}[X_t | X_t < VaR_t^{\alpha}]. \tag{2.18}$$

Thus, after some normalizations, we have the following equality:

$$\mathbb{E}\left[\frac{X_t}{ES_t^{\alpha}} - 1 \,\middle|\, X_t < VaR_t^{\alpha} \right] = 0. \tag{2.19}$$

If VaR_t^{α} was successfully backtested, the number of expected violations correspond to the theoretical level α. Thus, the authors suggest a test statistic based on (2.19) to evaluate only the magnitude of the realized exceptions against the model predictions. Define $\mathbb{1}_t = \mathbb{1}_{\{X_t < VaR_t^{\alpha}\}}$ and $\vec{X} = \{X_t\}_{t=1,\ldots,T}$. Thereby, the test statistic can be formulated as:

$$Z_1(\vec{X}) = \frac{1}{N} \sum_{t=1}^{T} \frac{X_t \mathbb{1}_t}{ES_t^{\alpha}} - 1, \tag{2.20}$$

if $N = \sum_{t=1}^{T} \mathbb{1}_t > 0$. The null hypothesis chosen for this test is given by:

$$\mathbb{H}_0 : F^{\alpha} = F_X^{\alpha}, \quad \text{for all period } t,$$

where F is the predictive distribution, F_X is the unknowable distribution of returns and F^{α} is the distribution tail conditioned to returns beyond α-quantile. The alternative hypothesis is:

$$\mathbb{H}_1 : ES_{\alpha,t}^{F_X} \leq ES_t^{\alpha}, \ \text{for all } t \text{ and} < \text{for some } t$$
$$VaR_{\alpha,t}^{F_X} = VaR_t^{\alpha}, \ \text{for all } t,$$

where $ES_{\alpha,t}^{F_X}$ and $VaR_{\alpha,t}^{F_X}$ are the values of the risk measures when $X \sim F_X$. The value of Z_1 is expected to be zero, i.e., the realized and the predicted left-tail are the same (ES is not rejected); or positive, i.e., the predicted values of ES underestimate the risk of the observed returns (reject the ES without rejecting VaR). As this test is subordinated to a preliminary VaR backtest, its predictions remain correct under the alternative hypothesis.

To test the significance of Z_1, independent simulations are performed using the distribution under the null hypothesis:

Simulate Independent	$X_t^i \sim F$	$\forall\, t, \forall\, i = 1, 2, \cdots, M$
Compute	$Z^i = Z(\vec{X}^i)$	
Estimate	$p = \sum_{i=1}^{M}(Z^i < Z(\vec{x}))/M$	

where M is the number of scenarios for independent simulations. Given a significance level ϕ, the test is accepted or rejected if $p \lesseqgtr \phi$.

3 Methodology

In this chapter, we define the procedures considered in our proposed objective. The chapter is divided into four sections. In section 3.1, we present the adaptions to RVaR of the backtesting procedures described in the last chapter. In section 3.2, we describe the estimation procedures, namely AR-GARCH-type and GAS models, that are used throughout the study. In section 3.3, we present the procedures necessary to assess the size and power properties of the backtesting procedures using Monte Carlo simulations. Lastly, in section 3.4, we describe the methodological procedures of the empirical analysis.

3.1 Proposed Approach

In this section, we present the adaptions to Range Value-at-Risk of the backtesting procedures for the Expected Shortfall described in Chapter 2. For a better understanding of the following procedures, it is necessary to make some initial definitions. Giving an out-of-sample or future period $t = 1, \cdots, T$, we consider X_t the return of some asset at time t as well as μ_t and σ_t the conditional mean and standard deviation that are measurable with respect to the information available up to $t-1$ of that asset. Besides, the measures VaR_t^{α} and $RVaR_t^{\alpha,\beta}$ denote the Value at Risk and Range Value at Risk forecasts for possible losses of some asset return at time t, for the significance levels $\alpha, \beta \in (0,1)$, also measurable with respect to the information available up to $t-1$. For the presentation of the backtesting, we consider that the risk forecasts have their sign adjusted.

Initially, we propose an modification of McNeil and Frey (2000) approach. As already described, this backtest is based on the size of the discrepancies between the observed returns and the estimated ES in the event of VaR violations. Using an analogous argument, we can derive the series k, which represents the size of the discrepancies between the observed returns and the estimated RVaR when the violation lies on an interval of two estimated VaRs. Thus, for the out-of-sample period $t = 1, \cdots, T$ and the significance levels α and β, the residuals can be defined by:

$$k_t = \frac{X_t - RVaR_t^{\alpha,\beta}}{\sigma_t} = Z_t - \mathbb{E}\left[Z_t | F^{-1}(\alpha) < Z_t < F^{-1}(\beta)\right]. \tag{3.1}$$

It is clear that the residuals are i.i.d.. Thus, (3.1) gives us information about how many standard deviations an observed return is apart from the estimated value of RVaR. In this way, the series of exceedance residuals k can be denoted by:

$$k = \left\{k_t : \ VaR_t^{\alpha} < X_t < VaR_t^{\beta}\right\}. \tag{3.2}$$

The appropriate null and alternative hypotheses are given by:

$$\mathbb{H}_0 : \mathbb{E}[\, k \,] = 0$$
$$\mathbb{H}_1 : \mathbb{E}[\, k \,] < 0.$$

So, the residuals should behave like an i.i.d. sample with a mean of zero under the null hypothesis, against the alternative hypothesis that the mean is less than zero. Following the original model, we used a non-parametric bootstrap test of Efron and Tibshirani (1993) to test the zero mean hypotheses. Also, 1000 bootstrap replications are considered to evaluate the results, following Catania and Grassi (2022).

Regarding the Righi and Ceretta (2015) procedure, additional adjustments are necessary. Different from (3.1), the authors consider dispersion only when violations occur and suggest the SD as a dispersion measure. In the case of RVaR, the violations that are taken into account occur in an interval of quantiles. Thus, we can modify the identity in (2.12) and use as a dispersion measure the SD defined by the following formulation:

$$SD^{\alpha,\beta}(X) = \left(\mathbb{V}\left[X|VaR^{\alpha} < X < VaR^{\beta}\right]\right)^{\frac{1}{2}} = \left(\frac{1}{\beta-\alpha}\int_{\alpha}^{\beta}\left(VaR^{\gamma}(X) - RVaR^{\alpha,\beta}(X)\right)^2 d\gamma\right)^{\frac{1}{2}}, \tag{3.3}$$

given the significance levels α and β. Thus, we followed the arguments of the authors and propose an equivalent backtest for RVaR based on the series of violation residuals k'. For the out-of-sample period $t = 1, \cdots, T$ and the significance levels α and β, the residuals can be defined by:

$$k'_t = \frac{X_t - RVaR_t^{\alpha,\beta}}{SD_t^{\alpha,\beta}} = \frac{Z_t - \mathbb{E}\left[Z_t|F^{-1}(\alpha) < Z_t < F^{-1}(\beta)\right]}{\left(\mathbb{V}\left[Z_t|F^{-1}(\alpha) < Z_t < F^{-1}(\beta)\right]\right)^{\frac{1}{2}}}. \tag{3.4}$$

It is clear that the residuals are i.i.d. and, similar to procedure (2.13), equation (3.4) gives us information about how many shortfall deviations an observed return is distant from the estimated value of RVaR. In this way, the violation residuals series k' can be denoted by:

$$k' = \left\{k'_t : \; VaR_t^{\alpha} < X_t < VaR_t^{\beta}\right\}. \tag{3.5}$$

The appropriate null and alternative hypotheses for long positions can be formulated as follows:

$$\mathbb{H}_0 : \mathbb{E}[\, k' \,] = 0$$
$$\mathbb{H}_1 : \mathbb{E}[\, k' \,] < 0,$$

and for short positions, they are given by:

$$\mathbb{H}_0 : \mathbb{E}[\, k'_1 \,] = 0$$
$$\mathbb{H}_1 : \mathbb{E}[\, k'_1 \,] > 0.$$

Under the null hypothesis, the residuals should behave like an i.i.d. sample with a mean of zero, against the alternative hypothesis that the mean is negative for long positions and positive for short positions. We used the same nonparametric bootstrap test of McNeil and Frey (2000) to test the zero mean hypotheses.

As the McNeil et al. (2015) backtest diverges from (2.10) and (2.13) by using as dispersion the difference between the estimated value of ES and the mean of the full sample, we suggest a series of violation residuals k'' exactly as (3.1), except by the fact that the dispersion measure considered is the difference between the estimated RVaR and the sample mean. Hence, for the out-of-sample period $t = 1, \cdots, T$ and the significance levels α and β, the residuals can be defined by:

$$k''_t = \frac{X_t - RVaR_t^{\alpha,\beta}}{RVaR_t^{\alpha,\beta} - \mu_t} = \frac{Z_t - \mathbb{E}\left[Z_t | F^{-1}(\alpha) < Z_t < F^{-1}(\beta)\right]}{\mathbb{E}\left[Z_t | F^{-1}(\alpha) < Z_t < F^{-1}(\beta)\right]}. \tag{3.6}$$

It is clear that the residuals are i.i.d., and also, that this last procedure provides how many differences between expected loss and conditional mean an observed return is apart from the same expected loss (here, RVaR). In this way, the violation residuals series k'' can be denoted by:

$$k'' = \left\{ k''_t : \; VaR_t^{\alpha} < X_t < VaR_t^{\beta} \right\}. \tag{3.7}$$

The null and alternative hypotheses are as follows:

$$\mathbb{H}_0 : \mathbb{E}[\, k'' \,] = 0$$
$$\mathbb{H}_1 : \mathbb{E}[\, k'' \,] < 0.$$

The residuals k'' are an i.i.d. sample with a mean of zero under the null hypothesis, against the alternative hypothesis that the mean is less than zero. Due to the absence of a suitable mathematical demonstration for a t-test to assess the null hypothesis of a mean of zero, we once again employed the non-parametric bootstrap test proposed by Efron and Tibshirani (1993) with 1000 replications. This method was selected because it is an alternative mentioned by the authors within the same study.

3.1.1 Biswas and Sen (2023) approach - Z_1^A

The modification proposed by Biswas and Sen (2023) is based on RVaR expressed as the conditional expectation, which, for the out-of-sample period $t = 1, \cdots, T$ and the significance levels α and β, is given by:

$$RVaR_t^{\alpha,\beta} = \mathbb{E}[X_t | VaR_t^{\alpha} < X_t < VaR_t^{\beta}]. \tag{3.8}$$

Thus, from (3.8), we can derive:

$$\mathbb{E}\left[\frac{X_t}{RVaR_t^{\alpha,\beta}} - 1 \middle| VaR_t^{\alpha} < X_t < VaR_t^{\beta} \right] = 0. \tag{3.9}$$

If VaR^{α} and VaR^{β} were successfully backtested, we could suggest a test statistic based on (3.9) to evaluate the magnitude of the realized exceptions against the model predictions. Define

$\mathbb{1}_t^{\alpha,\beta} = \mathbb{1}_{\{VaR_t^{\alpha} < X_t < VaR_t^{\beta}\}}$. The test statistic Z_1 of Acerbi and Szekely (2014), modified to RVaR, can be formulated by:

$$Z_1^A(\vec{X}) = \frac{1}{N}\sum_{t=1}^{T}\frac{X_t \mathbb{1}_t^{\alpha,\beta}}{RVaR_t^{\alpha,\beta}} - 1, \tag{3.10}$$

if $N = \sum_{t=1}^{T} \mathbb{1}_t^{\alpha,\beta} > 0$. The hypothesis test chosen for this test is given by:

$$\begin{aligned}
\mathbb{H}_0 : & F^{\alpha,\beta} = F_X^{\alpha,\beta}, \quad \text{for all } t, \\
\mathbb{H}_1 : & RVaR_{\alpha,\beta,t}^{F_X} \leq RVaR_t^{\alpha,\beta}, \quad \text{for all } t \text{ and} < \text{for some } t, \\
& VaR_{\alpha,t}^{F_X} = VaR_t^{\alpha} \text{ and } VaR_{\beta,t}^{F_X} = VaR_t^{\beta}, \quad \text{for all } t,
\end{aligned}$$

where $F_t^{\alpha,\beta}$ is the prediction distribution tail conditioned to returns between the α-quantile and β-quantile, and $RVaR_{\alpha,\beta,t}^{F_X}$ are the RVaR forecasts when $X \sim F_X$. The value of Z_1^A is expected to be zero, i.e., the realized and the predicted left-tail are the same (RVaR is not rejected), or positive, i.e., the RVaR forecasts underestimate the risk (reject RVaR without rejecting VaR). As a preliminary VaR backtest is performed, its forecasts remain correct even under $\mathbb{H}_1$. Finally, the same process of independent simulations is applied to test the significance of Z_1^A.

3.2 Estimation Procedures

In this section, we briefly describe the Autoregressive (AR) model, the standard GARCH model, and its two usual variations, namely Exponential Generalized Autoregressive Conditional Heteroskedastic (EGARCH) and Glosten Jagannanthan Runkle-Generalized Autoregressive Conditional Heteroskedastic (GJR-GARCH). Also, we describe an alternative model, which is driven by the score of the conditional density function, called the Generalized Autoregressive Score (GAS) model. Finally, we specify the parametric approach of risk forecasting and the distribution functions considered to perform such predictions.

3.2.1 Autoregressive model

While not of primary interest, the specification and estimation of the conditional mean are still important to construct consistent risk forecasts. A common model to estimate the conditional mean of financial returns is the Autoregressive (AR) model; see Li et al. (2002). Given a sequence of financial positions $\{X_t\}_{t=1,\ldots,T}$, the conditional mean can be estimated by the following equations:

$$\begin{aligned}
X_t &= \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t, \\
\varepsilon_t &= \sigma_t z_t, \qquad z_t \sim i.i.d.F(\theta),
\end{aligned} \tag{3.11}$$

where ε_t is the error term, σ_t is the conditional standard deviation, z_t is a white noise process, which can assume many distribution functions $F(\theta)$, where θ is a vector of parameters, including zero mean and unit variance in addition to additional parameters that vary as the distribution. Also, ϕ_i, for $i = 0, 1, ..., p$ being p term autoregressive order, are parameters of the AR model.

3.2.2 GARCH model

One of the main features of financial time series, in particular returns of assets, is volatility clustering. Informally, this means that returns exhibit highly volatile periods alternated with calm periods. Thus, by incorporating this characteristic into their structure, statistical models can more efficiently represent the stochastic behavior of financial returns. The majority of models that forecast conditional volatility and, therefore, capture the volatility clustering belong to the GARCH family of models. The first such model was the autoregressive conditional heteroskedasticity (ARCH) model proposed by Engle (1982). However, the generalized ARCH model (GARCH) of Bollerslev (1986) has become the most popular approach and the common denominator for most volatility models.

The GARCH(q, s) model can be defined as:

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{k=1}^{s} \beta_k \sigma_{t-k}^2, \tag{3.12}$$

where σ_t^2 is the conditional variance, ε_t is the error term in (3.11) and α_j, for $j = 1, ..., q$ as well as β_k, for $k = 1, ..., s$ are parameters of the GARCH model (with q and s being the GARCH model order). To ensure positive volatility forecasts, $\alpha_0 > 0$, $\alpha_j \geq 0$, $\beta_k \geq 0$ and to ensure covariance stationarity, $\sum_{j=1}^{q} \alpha_j + \sum_{k=1}^{s} \beta_k < 1$ (BOLLERSLEV, 1986).

3.2.3 Leverage effect

An asymmetric volatility response to positive and negative past returns has been empirically noted. In particular, increases in volatility are larger when previous returns are negative than when they have the same magnitude but are positive. Such behavior is known as the leverage effect. Although the standard GARCH model manages to capture the existence of volatility clustering, it cannot express the leverage effect since from (3.12) positive and negative error terms have the same effect on volatility.

A straightforward way to incorporate leverage effects in the GARCH model is to use the GJR-GARCH(q, s) model of Glosten et al. (1993). It can be described as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} [\alpha_i + \gamma_i \mathbb{1}_{\{\varepsilon_{t-i} < 0\}}] \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2, \tag{3.13}$$

where $\mathbb{1}(\cdot)$ is the indicator function, α_i, for $i = 1, ..., q$ as well as β_j, for $j = 1, ..., s$ are parameters of the GJR-GARCH model (with q and s being the GJR-GARCH model order). This model incorporates positive and negative shocks on the conditional variance asymmetrically via the use of the indicator function, where the volatility effect of a unit negative shock is $\alpha_i + \gamma_i$ while the effect of a unit positive shock is α_i. To ensure positivity, $\alpha_0 > 0$, $\alpha_i, \beta_j, \gamma_i \geq 0$ and to guarantee covariance stationarity, $\sum_{i=1}^{q} \gamma_i < 2(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{s} \beta_j)$ (RODRÍGUEZ; RUIZ, 2012).

Another widely used GARCH model allowing for leverage effects into its structure is the exponential GARCH (EGARCH(q,s)) proposed by Nelson (1991). It can be written as:

$$\ln \sigma_t^2 = \alpha_0 + \left(1 + \sum_{i=1}^{q} \alpha_i L^i\right)\left(1 - \sum_{j=1}^{s} \beta_j L^j\right)^{-1} g(\varepsilon_{t-1}), \tag{3.14}$$

where L^i denotes the back-shift operator, $g(\cdot)$ is a function on $\mathbb{R}$, α_i, for $i = 1,...,q$ as well as β_j, for $j = 1,...,s$ are parameters of the EGARCH model (with q and s being the EGARCH model order). To accommodate the asymmetric relation between returns and volatility, we need to model $g(\varepsilon_t)$ as a function of both the magnitude and the sign of ε_t. As suggested by the author, we can choose $g(\varepsilon_t)$ as a linear combination of ε_t and $|\varepsilon_t|$ of the form:

$$g(\varepsilon_t) = \theta\varepsilon_t + \gamma(|\varepsilon_t| - \mathbb{E}(|\varepsilon_t|)). \tag{3.15}$$

Thereby, the equation (3.15) gives rise to leverage effects. Furthermore, another advantage of this model in relation to the standard GARCH is that by modeling $\log \sigma_t^2$ rather than σ_t^2, there is no need for parameter constraints to ensure positive volatility predictions. Lastly, to guarantee weak stationarity, all the roots of $\left(1 - \sum_{j=1}^{s} \beta_j L^j\right)$ should be outside the unit circle (NELSON, 1991).

3.2.4 GAS model

Creal et al. (2013) and Harvey (2013) propose an alternative framework to estimate time-varying parameters, called GAS model. The main contribution of this approach is the use of the score of the conditional density function as the main driver of time-variation in the parameters of the time series process. Based on the score, the GAS model encompasses the traditional approaches and exploits the complete density structure rather than only means and higher moments (CREAL et al., 2013). A further advantage of using the conditional score as the driver is that the estimation by maximum likelihood is straightforward (ARDIA et al., 2019).

Let X_t denote the returns of some asset or portfolio, $\mathbf{y}_t$ a time-varying parameter vector, θ a vector of static parameters and $\mathscr{F}_{t-1}$ the available information set of X_t up to $t-1$. Consider $f(X_t|\, \mathbf{y}_t, \mathscr{F}_{t-1}; \theta)$ as the conditional density of X_t. The GAS model is specified as follows:

$$\mathbf{y}_t = \omega + \sum_{i=1}^{p} A_i\, s_{t-1} + \sum_{j=1}^{q} B_j\, \mathbf{y}_{t-1}, \tag{3.16}$$

where ω is a vector of constants, coefficient matrices A_i and B_j have appropriate dimensions for $i = 1,...,p$ and $j = 1,...,q$, while s_t is the scaled score vector:

$$s_t = S_t \cdot \nabla_t, \qquad \nabla_t = \frac{\partial \log f(X_t|\, \mathbf{y}_t, \mathscr{F}_{t-1}; \theta)}{\partial \mathbf{y}_t}. \tag{3.17}$$

The matrix S_t is a positive-definite scale matrix that adjusts the shape of the score ∇_t at the period t. It is natural to consider a form of scale that depends on the variance of the score, i.e., on:

$$S_t = \mathbb{E}_{t-1}[\nabla_t \nabla_t'], \tag{3.18}$$

where the expectation $\mathbb{E}_{t-1}$ is taken with respect to the conditional density $f(X_t | \mathbf{y}_t, \mathscr{F}_{t-1}; \theta)$.

3.2.5 Risk forecasting

Consider that the random returns of an asset or portfolio X_t have a fully parametric location-scale specification based on the expectation, dispersion and random component, conforming to $X_t = \mu_t + \sigma_t z_t$, where μ_t is the conditional mean (location), σ_t is the conditional standard deviation (scale) and z_t represents the i.i.d. innovations, which can assume different probability distribution functions. Also, consider a risk measure $\rho : \mathscr{X} \rightarrow \mathbb{R}$. Thus, for an out-of-sample day $t \in \{1, \cdots, T\}$, the risk forecasts can be obtained, in a parametric way, by: $\rho(X_t) = \rho(\mu_t + \sigma_t z_t) = -\mu_t + \sigma_t \rho(z_t)$[3]. Under this specification, the Range Value at Risk becomes:

$$RVaR_t^{\alpha,\beta} = -\left[\mu_t + \sigma_t \left((\beta-\alpha)^{-1} \int_{\alpha}^{\beta} F^{-1}(s)ds\right)\right], \tag{3.19}$$

where $0 < \alpha < \beta < 1$ and F denote some distribution function. As all backtesting procedures presented use VaR as an auxiliary measure and the adaptions of Righi and Ceretta (2013) and Righi and Ceretta (2015) backtests use the adapted SD as a dispersion measure, it is reasonable to illustrate them in their parametric form, as follows:

$$VaR_t^{\alpha} = -\left(\mu_t + \sigma_t F^{-1}(\alpha)\right), \tag{3.20}$$

$$SD_t^{\alpha,\beta} = \left[\sigma_t^2(\beta-\alpha)^{-1} \int_{\alpha}^{\beta} \left(F^{-1}(s) - \left((\beta-\alpha)^{-1} \int_{\alpha}^{\beta} F^{-1}(s)ds\right)\right)^2 ds\right]^{1/2}, \tag{3.21}$$

where, again, $0 < \alpha < \beta < 1$ and F represents some probability distribution. Thereby, in the parametric approach, we first filter μ_t and σ_t through some conditional location-scale time series model and then calculate the measures defined above for period t from the definition of function F (for details, see McNeil et al. (2015)). In our illustrations, the conditional mean and variance are predicted in two different ways. At first, we make use of AR models, described in (3.11), for the conditional mean ($\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}$) forecasts. In relation to conditional variance, we employed the GAS and GARCH-type models, specified in (3.12)-(3.16). The jointly application of these approaches are denoted as AR-GARCH, AR-EGARCH, AR-GJR-GARCH, and AR-GAS models. The first model is considered in our study because, in the risk forecasting literature, it is widely used to fit financial data and previous studies show good performance

[3] This result is obtained from the Translation Invariance and Positive Homogeneity properties.

compared to competitive models; see Ardia and Hoogerheide (2014) and Müller et al. (2022), for example. The second and third models were selected since they capture the existence of asymmetric volatility or the so-called leverage effect of financial data, and they are employed by other studies in risk forecasting, as in Alberg et al. (2008) and Umar et al. (2021), among others. Lastly, the fourth approach was selected because it shows competitive performances in recent risk forecasting studies compared to traditional approaches. See, for instance, Bartels and Ziegelmann (2016), Troster et al. (2019) and Xu and Lien (2021).

Table 1 – Parameters, support, and probability density function of the different used distributions

Probability distribution	Parameters*	Support	Probability density function
Normal	μ, σ	$\mathbb{R}$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Skewed normal	μ, σ, υ	$\mathbb{R}$	$\frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\upsilon\left(\frac{x-\mu}{\sigma}\right)\right)$
Student-*t*	η	$\mathbb{R}$	$\frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\eta\pi}\,\Gamma\left(\frac{\eta}{2}\right)}\left(1+\frac{x^2}{\eta}\right)^{-\left(\frac{\eta+1}{2}\right)}$
Skewed Student-*t*	η, υ	$\mathbb{R}$	$b\frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\pi(\eta-2)}\,\Gamma\left(\frac{\eta}{2}\right)}\left(1+\frac{\varepsilon^2}{\eta-2}\right)^{-\left(\frac{\eta+1}{2}\right)}$
Generalized error	μ, σ, β	$\mathbb{R}$	$\frac{\beta}{2^{1+\beta^{-1}}\sigma\,\Gamma\left(\frac{1}{\beta}\right)}\exp\left(-\frac{1}{2}\left\lvert\frac{x-\mu}{\sigma}\right\rvert^{\beta}\right)$
Skewed generalized error	$\mu, \sigma, \theta, \upsilon$	$\mathbb{R}$	$\frac{\theta}{\sigma}\left(2c\Gamma\left(\frac{1}{\theta}\right)\right)^{-1}\exp\left(-\frac{\lvert x-d\rvert^{\theta}}{[1-sign(x-d)\upsilon]^{\theta}c^{\theta}\sigma^{\theta}}\right)$
Normal-inverse Gaussian	$\mu, \sigma, \beta, \delta$	$\mathbb{R}$	$\frac{\sigma\delta}{\pi}\frac{e^{p(x)}}{q(x)}K_1[\delta\, q(x)]$
Johnson SU	$\mu, \sigma, \lambda, \delta$	$\mathbb{R}$	$\frac{\delta}{\sigma\sqrt{2\pi}\sqrt{z^2+1}}\exp\left(-\frac{1}{2}(\lambda+\delta\sinh^{-1}(z))\right)^2$

*μ, σ represent the location and scale parameters. υ is the skewness (asymmetry) parameter, ϕ and Φ denote the standard normal probability density function and cumulative distribution function, respectively. η denotes the number of degrees of freedom and $\Gamma(p) = \int_0^\infty x^{p-1}e^{-x}dx$ is the Gamma function. Regarding the skewed student-t distribution, $\varepsilon = (bx+a)/(1-\upsilon)$ if $x < -a/b$ or $\varepsilon = (bx+a)/(1+\upsilon)$ otherwise, and the constant terms are defined by: $a = 4\upsilon c(\eta-2)/(\eta-1)$, $b^2 = 1+3\upsilon^2-a^2$, with $c = \Gamma(\frac{\eta+1}{2})/\sqrt{\pi(\eta-2)}\,\Gamma\left(\frac{\eta}{2}\right)$. β represents the shape parameter. The parameter θ controls the height and fat-tail of the skewed generalized error distribution, $c = \Gamma(\frac{1}{\theta})^{1/2}\Gamma(\frac{3}{\theta})^{-1/2}h(\upsilon)^{-1}$, $d = \mu - \delta\sigma$, where $\delta = 2\upsilon A h(\upsilon)^{-1}$, with $h(\upsilon) = \sqrt{1+3\upsilon^2-4A^2\upsilon^2}$ and $A = \Gamma(\frac{2}{\theta})\Gamma(\frac{1}{\theta})^{-1/2}\Gamma(\frac{3}{\theta})^{-1/2}$. δ denotes the tail-heaviness parameter, K_1 is the Bessel function of the second kind with index 1, $p(x) = \sigma(\delta^2-\beta^2)^{1/2}+\beta(x-\mu)$, and $q(x) = (\sigma^2+(x-\mu)^2)^{1/2}$. For the Johnson SU distribution, λ and δ are the shape parameters and $z = \frac{x-\mu}{\sigma}$.

Regarding the distribution functions, we follow the results of Garcia-Jorcano and Novales (2021), which states that the important assumption for risk forecast performance is that of the probability distribution of the innovations, with the choice of volatility model playing a secondary role. For this reason, we consider different probability distribution functions for the AR-GARCH-type models, including normal, skewed normal (AZZALINI, 1985), Student-*t* (GOSSET, 1908), skewed Student-*t* (FERNANDEZ; STEEL, 1998), generalized error (MCDONALD; NEWEY, 1988), skewed generalized error (THEODOSSIOU, 1988), normal-inverse

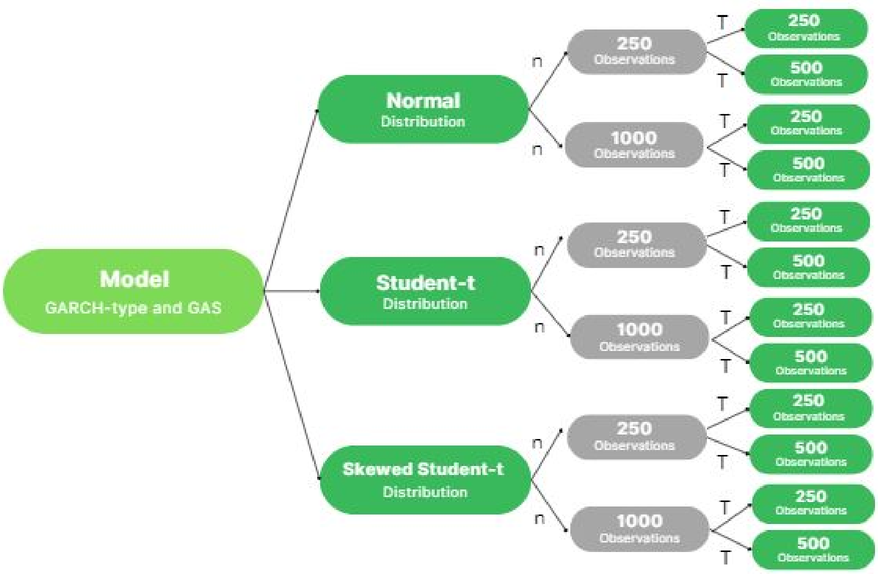
Gaussian (BARNDORFF-NIELSEN, 1977), or Johnson SU (JOHNSON, 1949) distributions. For the AR-GAS model, the assumed probability distributions are the normal, skewed normal, Student-t and skewed Student-t[4]. We used the quasi-maximum likelihood (QML) method to estimate the parameters of the models. In Table 1, we provide the underlying parameters, support, and probability density function for the different specifications of the distribution function F.

3.3 Numerical Procedures

In this section, we describe the Monte Carlo simulations exercise conducted to investigate the size and non-null rejection rates (power) of the three backtests defined in Equations (3.1) - (3.7) from Section 3.1. Briefly, the size of a test is the probability of incorrectly rejecting the null hypothesis if it is true (Type I error). In contrast, the power of a test is the probability of correctly rejecting the null hypothesis if it is false (Type II error). All computational implementations were performed using the `R` programming language (R Core Team, 2023), making use of the *rugarch* (GALANOS, 2019) and *GAS* (CATANIA et al., 2019) packages for data simulations as well as conditional mean and conditional variance predictions.

In our experiment, we decided to set the number of Monte Carlo replications at 1,000 since it can provide satisfactory results as observed, for example, by Escanciano and Velasco (2010) and Du and Escanciano (2017). For the remaining parameters, we set the in-sample sizes (n) at 250 and 1,000; the out-of-sample backtesting sample sizes (T) at 250 and 500; and the number of bootstrap or Monte Carlo samples at 1000. Such values are commonly employed in the literature; see Deng and Qiu (2021).

Figure 1 – Numerical Analysis Diagram



[4] This difference is due to the limitations of the *GAS* package.

We select three different models as data generating process (DGP) to generate the log-returns: (i) AR(1)-GARCH(1,1), (ii) AR(1)-EGARCH(1,1), and (iii) AR(1)-GAS(1,1). Following the results of Garcia-Jorcano and Novales (2021), we consider different probability distributions for prediction models and DGPs, assuming F to be normal (norm), student-t (std) and skewed student-t (sstd). Such distributions were considered because they are often applied in stock market analysis; see Diaz et al. (2017). As the last step before simulating the data, we need the parameters of the data generating processes to be well known. So, following Righi and Ceretta (2015) and Müller and Righi (2018), the chosen parameterization for (i) model is given by:

$$\begin{aligned} X_t &= 0.5X_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \qquad z_t \sim i.i.d.F(\theta), \\ \sigma_t^2 &= 4\text{x}10^{-6} + 0.10\varepsilon_{t-1}^2 + 0.85\sigma_{t-1}^2. \end{aligned} \tag{3.22}$$

With respect to the (ii) model, we choose the parameterization used by Du and Escanciano (2017), which is as follows:

$$\begin{aligned} X_t &= 0.5X_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \qquad z_t \sim i.i.d.F(\theta), \\ \ln\sigma_t^2 &= 0.01 + 0.9\ln\sigma_{t-1}^2 + 0.3(|\varepsilon_{t-1}| - \sqrt{2/\pi}) - 0.8\varepsilon_{t-1}. \end{aligned} \tag{3.23}$$

For both models, we also consider parameters for skewness (υ) and shape (η), equal to 0.92 and 6, respectively. The η parameter was taken from Müller and Righi (2018), once this parameterization represents an equity portfolio return distribution, while the υ parameter represents a similar value to the one we obtained, estimating the S&P 500 returns with a GARCH(1,1) model. To the best of our knowledge, no study still applies GAS models for evaluating the size and power of a test. Therefore, to use our (iii) model, we applied a similar procedure to that of Righi and Ceretta (2015) to determine the process that generates the data by fitting it to daily S&P500 Index returns.

To assess the size and power of the tests, we applied a technique similar to that of Escanciano and Velasco (2010), Du and Escanciano (2017) and Deng and Qiu (2021), considering nine different hypothesis tests where, in each of them, a single DGP is assumed as the null hypothesis. Consider, with some abuse of notation, $F_\Omega := \{\text{norm, std, sstd}\}$. So, for any $F_0 \in F_\Omega$, F_1 can be described as its complement, where $F_\Omega = F_0 \cup F_1$. To a better understanding of the method, suppose the hypothesis test of the first row of Table 2, with $F_0 = \text{norm}$. Under the specification of the null model, we simulate 1000 processes of length $n+T$. To evaluate the size property, for each simulated sample, we estimate the RVaR, VaR, and SD conforming to the explanation in subsection 3.2.5, employing the null forecasting model[5]. Thus, for each n, significance level, and simulated process, 1000 estimates of these measures are obtained. For all

[5] The null forecasting model is the same model and probability distribution $F_0 \in F_\Omega$ that generates the data in the null hypothesis. In the example, AR(1)-GARCH(1,1) model with normal distribution.

Table 2 – Hypothesis tests of size and size-corrected power of the backtesting procedures

Null hypothesis	Alternative hypothesis
$\mathbb{H}_0 : X_t = 0.5X_{t-1} + \varepsilon_t,$ $\varepsilon_t = \sigma_t z_t, \quad z_t \sim F_0,$ $\sigma_t^2 = 4\text{x}10^{-6} + 0.10\varepsilon_{t-1}^2 + 0.85\sigma_{t-1}^2$	$\mathbb{H}_1 : X_t = 0.5X_{t-1} + \varepsilon_t,$ $\varepsilon_t = \sigma_t z_t, \quad z_t \sim F_1,$ $\sigma_t^2 = 4\text{x}10^{-6} + 0.10\varepsilon_{t-1}^2 + 0.85\sigma_{t-1}^2$ $X_t = 0.5X_{t-1} + \varepsilon_t,$ $\varepsilon_t = \sigma_t z_t, \quad z_t \sim F_\Omega,$ $\ln \sigma_t^2 = 0.01 + 0.9 \ln \sigma_{t-1}^2 + 0.3(\lvert\varepsilon_{t-1}\rvert - \sqrt{2/\pi}) - 0.8\varepsilon_{t-1}$ AR(1)-GAS(1,1) $\sim F_\Omega$
$\mathbb{H}_0 : X_t = 0.5X_{t-1} + \varepsilon_t,$ $\varepsilon_t = \sigma_t z_t, \quad z_t \sim F_0,$ $\ln \sigma_t^2 = 0.01 + 0.9 \ln \sigma_{t-1}^2 + 0.3(\lvert\varepsilon_{t-1}\rvert - \sqrt{2/\pi}) - 0.8\varepsilon_{t-1}$	$\mathbb{H}_1 : X_t = 0.5X_{t-1} + \varepsilon_t,$ $\varepsilon_t = \sigma_t z_t, \quad z_t \sim F_1,$ $\ln \sigma_t^2 = 0.01 + 0.9 \ln \sigma_{t-1}^2 + 0.3(\lvert\varepsilon_{t-1}\rvert - \sqrt{2/\pi}) - 0.8\varepsilon_{t-1}$ $X_t = 0.5X_{t-1} + \varepsilon_t,$ $\varepsilon_t = \sigma_t z_t, \quad z_t \sim F_\Omega,$ $\sigma_t^2 = 4\text{x}10^{-6} + 0.10\varepsilon_{t-1}^2 + 0.85\sigma_{t-1}^2$ AR(1)-GAS(1,1) $\sim F_\Omega$
$\mathbb{H}_0$: AR(1)-GAS(1,1) $\sim F_0$	$\mathbb{H}_1$: AR(1)-GAS(1,1) $\sim F_1$ $X_t = 0.5X_{t-1} + \varepsilon_t,$ $\varepsilon_t = \sigma_t z_t, \quad z_t \sim F_\Omega,$ $\sigma_t^2 = 4\text{x}10^{-6} + 0.10\varepsilon_{t-1}^2 + 0.85\sigma_{t-1}^2$ $X_t = 0.5X_{t-1} + \varepsilon_t,$ $\varepsilon_t = \sigma_t z_t, \quad z_t \sim F_\Omega,$ $\ln \sigma_t^2 = 0.01 + 0.9 \ln \sigma_{t-1}^2 + 0.3(\lvert\varepsilon_{t-1}\rvert - \sqrt{2/\pi}) - 0.8\varepsilon_{t-1}$

of these sets, we perform all presented backtestings of RVaR, where the predictions are accepted or (incorrectly) rejected according to the results of each backtesting. Thereby, we denote the size of the backtestings as the null rejection rates of each procedure, considering all the different parameter specifications. Similarly, to evaluate the power of the tests, we repeat the scheme, but in this case, the null forecasting model is applied under the specification of each DGP and distribution function of $\mathbb{H}_1$, separately, resulting in eight distinct power property analyzes for each RVaR backtesting. In this way, the power of the backtestings is represented by the non-null rejection rates of the procedures across each scenario and different parameter specifications. The hypothesis tests are performed under all DGP models and all probability distribution functions $F_0 \in F_\Omega$.

The RVaR backtesting procedures are tested at 1%, 2.5%, 5% and 10% nominal levels. Regarding significance levels of the measures RVaR, VaR and SD, we consider 1%, 2.5%, and 5%. Such levels are employed since 1% and 5% are the most used values to forecast risk measures (see, for instance, Kuester et al. (2006), Escanciano and Pei (2012) and Müller and Righi (2018)), and 1% and 2.5% are the levels recommended by the Basel Committee on Banking Supervision (2013) for Value-at-Risk and Expected Shortfall predictions, respectively. In the case of RVaR, we always respect the condition $0 < \alpha < \beta < 1$ for the assumed levels, that is, $\beta = 2.5\%$ or $\beta = 5\%$ when $\alpha = 1\%$; or $\beta = 5\%$ when $\alpha = 1\%$ or $\alpha = 2.5\%$.

3.4 Empirical Procedures

To evaluate the performance of the proposed backtesting procedures for the RVaR measure, we also executed, as a secondary step, an empirical analysis. We conducted computational implementations using the R programming language (R Core Team, 2023), making use of the *quantmod* (RYAN et al., 2008) and *priceR* (CONDYLIOS et al., 2019) packages to download the databases, as well as the *rugarch* and *GAS* packages to estimate the model parameters.

We consider six factors of market risk: (i) equity, (ii) fixed income, (iii) exchange rates, (iv) commodities, (v) cryptocurrencies, and (vi) green finance. The first four factors were also considered by Righi and Ceretta (2015). We included cryptocurrencies because the studies that focus on risk measures are more restricted and are limited, for the most part, to VaR and ES measures; see, for instance, Trucíos (2019) and Trucíos et al. (2019). And also because they have been under increased scrutiny by policymakers, investors and researchers. Regarding green finance, we included this factor because of its rapid development and the ever-increasing concerns surrounding global environmental sustainability; see Gilchrist et al. (2021). For equity, we consider the data of the daily closing prices of the S&P500 and Ibovespa indices, representing the American and Brazilian markets. For fixed income, we consider US Treasury bonds rates 3-Years and 10-Years yield. For exchange rates, we consider the ratio to the US Dollar of the Euro and the Brazilian Real, contemplating both developed and emerging markets. For commodities, we consider WTI crude oil and Gold. For cryptocurrencies, we consider the ratio to the US Dollar of Bitcoin since it is the most popular and successful cryptocurrency to date, having a major role in studies on risk forecasting; see Subramoney et al. (2021). For green finance, we consider the STOXX Europe 600 and World Dow Jones Sustainability indexes due to its significant importance in the financial and business landscape, assessing performance of companies in terms of environmental sustainability, the increasing of temperature, and, at the same time, the scarcity of studies on the risk of this class of assets in the literature; see Bulai et al. (2022). As these six asset classes have different characteristics, our illustration provides an analysis of the developed backtesting procedure's behavior in distinct situations.

We computed one-day-ahead forecasts for RVaR and VaR using their parametric approaches (3.19) and (3.20), defined in subsection 3.2.5. Our financial positions X refers to log-returns of the considered assets, i.e., $X_t = (\ln P_t - \ln P_{t-1})$, where P refers to the closing price at time t and $t-1$, standard procedure in the literature. The analyzed period comprises data from January 2010 to December 2021. This span was chosen because it encompasses calm and turbulent periods caused by, for example, the Eurozone crisis and the COVID-19 pandemic.

The risk forecast of returns were quantified using the same models of the numerical illustration but with the addition of the AR(1)-GJR-GARCH(1,1) model. We consider all eight probability distribution functions described in Table 1 for the innovations z_t of the AR-GARCH-type models, while for the GAS model, the assumed probability distributions are the normal, skewed normal, student-t, and skewed student-t. These distributions were considered because they have the advantage of taking into account stylized facts of financial assets, such as heavy tails and/or negative skewness, which are common in financial data. The rolling estimation windows considered for the predictions have the length of 250, 500, and 1000 observations[6]. As significance levels, we consider the same values used in Section 3.3. Jointly with the backtesting, we made use of realized loss for the out-of-sample RVaR forecasts, which are calculated by a strictly consistent scoring function of RVaR. This method allows us to set up a contest of relative quality between different models and scenarios through the performance criterion that takes the following form:

$$\bar{S}_{RVaR} = \frac{1}{T}\sum_{t=1}^{T} S^{RVaR^{\alpha,\beta}}(x_t, y_t, z_t, w_t), \tag{3.24}$$

where T represents the number of out-of-sample terms and the scoring or loss function $S^{RVaR^{\alpha,\beta}}$ is defined in equation (2.9). The scoring functions are negatively oriented, that is, the smaller the realized loss $\bar{S}_{RVaR}$, the better the model. Thus, through the application of the backtestings and the loss function, we can assess the results and verify if they remain following both methods. In this way, we evaluated whether the worst models, according to the loss function, were rejected by the backtestings. To quantify realized loss, we consider the risk forecasts with signal adjusted.

[6] 250 is the minimum size recommended by the Basel Committee on Banking Supervision (2013), and 500 and 1,000 are the most common values used in the literature. See Kuester et al. (2006) and Righi and Ceretta (2015).

4 Results

In this chapter, we present the results of the size and power properties of the proposed backtesting procedures for the RVaR measure and those related to their empirical performance. For a better understanding of these results, the chapter is divided into two sections. In section 4.1, we present the numerical analysis results, where the size and power properties of the three proposed backtesting are described separately. In this way, we can more precisely highlight the behavior of each property, in addition to not mixing two concepts that can lead to a certain confusion. In section 4.2, we present the empirical performance of the proposed procedures.

4.1 Numerical Analysis

In this section we present a description of the numerical results.

4.1.1 Size of the tests

We organize the results from different perspectives to evaluate the size property of the proposed backtesting procedures for the RVaR. Tables 4 and 5 show, respectively, the rejection rate of the three tests under the null hypothesis for all previously defined models and distributions, using 250 and 1000 in-sample sizes (n). In each table, the results are separated by the considered nominal levels of 1%, 2.5%, 5% and 10%, where, in each of these levels, they are presented according to the out-of-sample sizes (T) of 250 and 500 obtained for each considered significance level. Tables 6, 7, and 8 show, respectively, the outcomes of the adapted approaches for RVaR regarding the evolution of the null rejection rate (size) with a look at the change in the number of in-sample observations. Figures 2 and 3 illustrate the behaviour of each null model over the nominal levels regarding the adapted McNeil and Frey (2000) approach[7]. Lastly, Table 9 highlights the model along with the distribution that most closely matches the nominal rejection levels for each backtesting and each scenario addressed. The McNeil and Frey (2000), Righi and Ceretta (2015) and McNeil et al. (2015) procedures adapted for RVaR are named in each table, respectively, by MFTest, RCTest, and MFETest. Given the computational burden, we were unable to get the results for the GAS model with skewed Student-t distribution for 1000 in-sample observations.

Initially, observing the behavior of the results in Tables 4 and 5, we verified that all backtesting procedures have similar test sizes for all scenarios under analysis. In other words, there is no major superiority in the size test performance of any approach considered, either being an option of equal or similar usefulness. By illustration, the models' results in Table 4 for

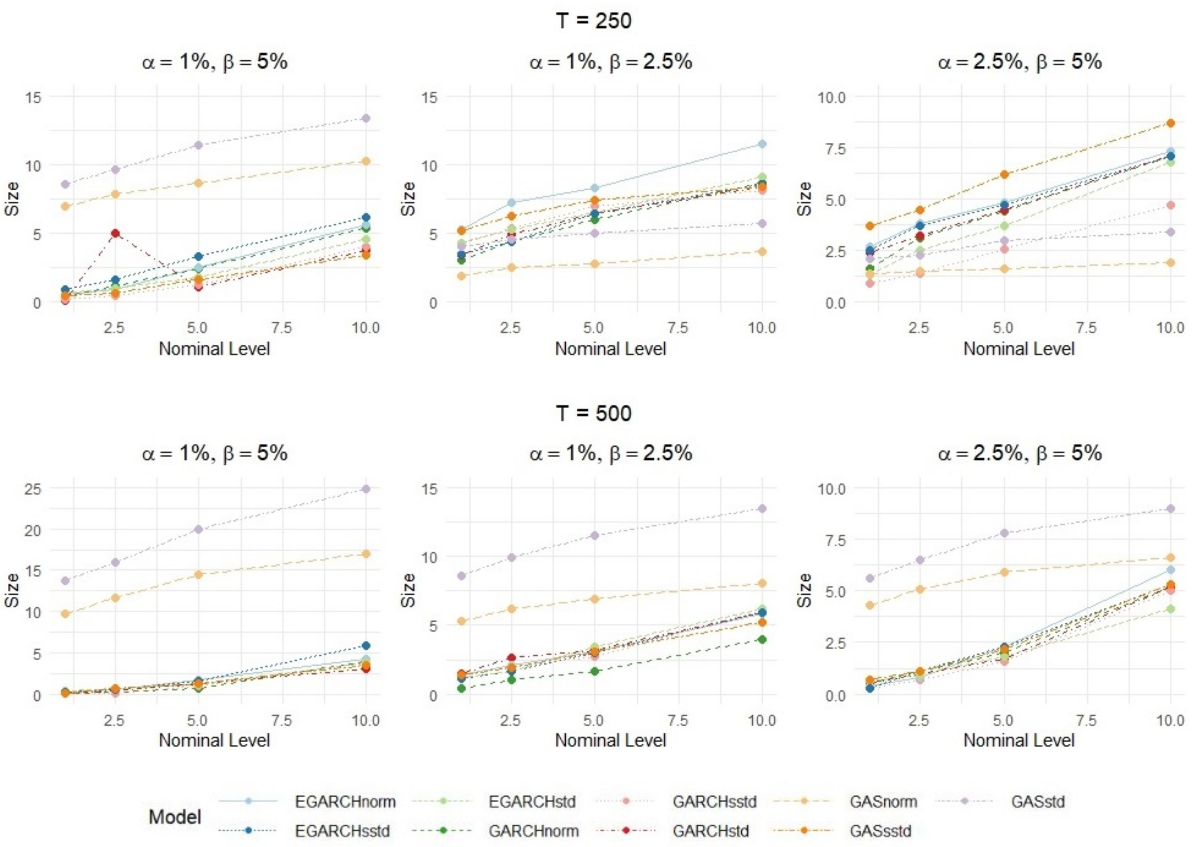
[7] The remaining models present very similar results and, for this reason, they were omitted from the analysis without loss of generality.

the nominal level of 1% and significance levels of $\alpha = 1\%$, $\beta = 5\%$ vary by a maximum of 0.1%. More specifically, for $T = 250$, the EGARCH$_{sstd}$, which exhibits the best performance in the three tests, has a size of 0.9% for the MFTest, 0.8% for the MFETest, and 0.8% for the RCTest. Similar results were observed in the other models and scenarios. Such a conclusion contrasts with what is seen in the literature, where a better choice among candidates is commonly observed; see, for instance, Deng and Qiu (2021). Another initial impression taken when analysing the same tables is that when looking at the GARCH-type models and the GAS$_{sstd}$, the highest null rejection rates are concentrated, with few exceptions, at $T = 250$. As the out-of-sample size moves toward 500, these rates tend to reduce. The opposite situation is seen for the GAS$_{norm}$ and GAS$_{std}$. Such behaviors of the proposed backtestings are commonly seen in the risk forecasting literature, where depending on the procedure and null model, the rejection rates can either escalate or diminish with the increase in the size of out-of-sample observations; see Du and Escanciano (2017). Furthermore, the proposed procedures exhibit generally the best sizes at the significance levels of $\alpha = 1\%, \beta = 2.5\%$, where for the nominal levels of 1% and 2.5%, these are found at $T = 500$; and for the nominal level of 10%, at $T = 250$. However, at the nominal level of 5%, the optimal size tests are observed at the significance levels of $\alpha = 2.5\%, \beta = 5\%$, and $T = 250$. Such behaviour stands for both in-sample sizes.

Upon conducting the further examination of the outcomes presented in Table 4, it is observed that the instances in which the null rejection rate is underestimated can be primarily attributed to the implementation of GARCH-type models and GAS$_{sstd}$ with $T = 500$, across all significance levels and nominal levels considered. Notably, this pattern deviates at the significance level of $\alpha = 1\%$, $\beta = 2.5\%$ for the nominal level of 1%. Furthermore, substantial underestimations can also be observed for the same models when $T = 250$ and $\alpha = 1\%$, $\beta = 5\%$. On the other hand, the backtest results underestimate the null rejection rate when the predictions are conducted by the GAS$_{norm}$ and GAS$_{std}$ models at the nominal levels of 5% and 10%, and significance levels of $\alpha = 1\%$, $\beta = 2.5\%$ and $\alpha = 2.5\%$, $\beta = 5\%$, for $T = 250$. Concerning the overestimation of the null rejection rate, the RVaR forecasts with GARCH-type models continue to display notable size distortions. Specifically, such distortions are observed at the significance level of $\alpha = 1\%$, $\beta = 2.5\%$, and nominal levels of 1%, 2.5%, and 5% for $T = 250$. Also, the GARCH$_{norm}$, GARCH$_{std}$, EGARCH$_{norm}$, and EGARCH$_{sstd}$ demonstrate overestimation at the significance level of $\alpha = 2.5\%$, $\beta = 5\%$, and nominal levels of 1% and 2.5%, again for 250 out-of-sample observations. When considering GAS$_{norm}$ and GAS$_{std}$, we observe the most pronounced distortions among all candidates at the significance level of $\alpha = 1\%$, $\beta = 5\%$, except for the nominal level of 10% and 250 out-of-sample observations, where GAS$_{norm}$ exhibits a size closer to the nominal level.

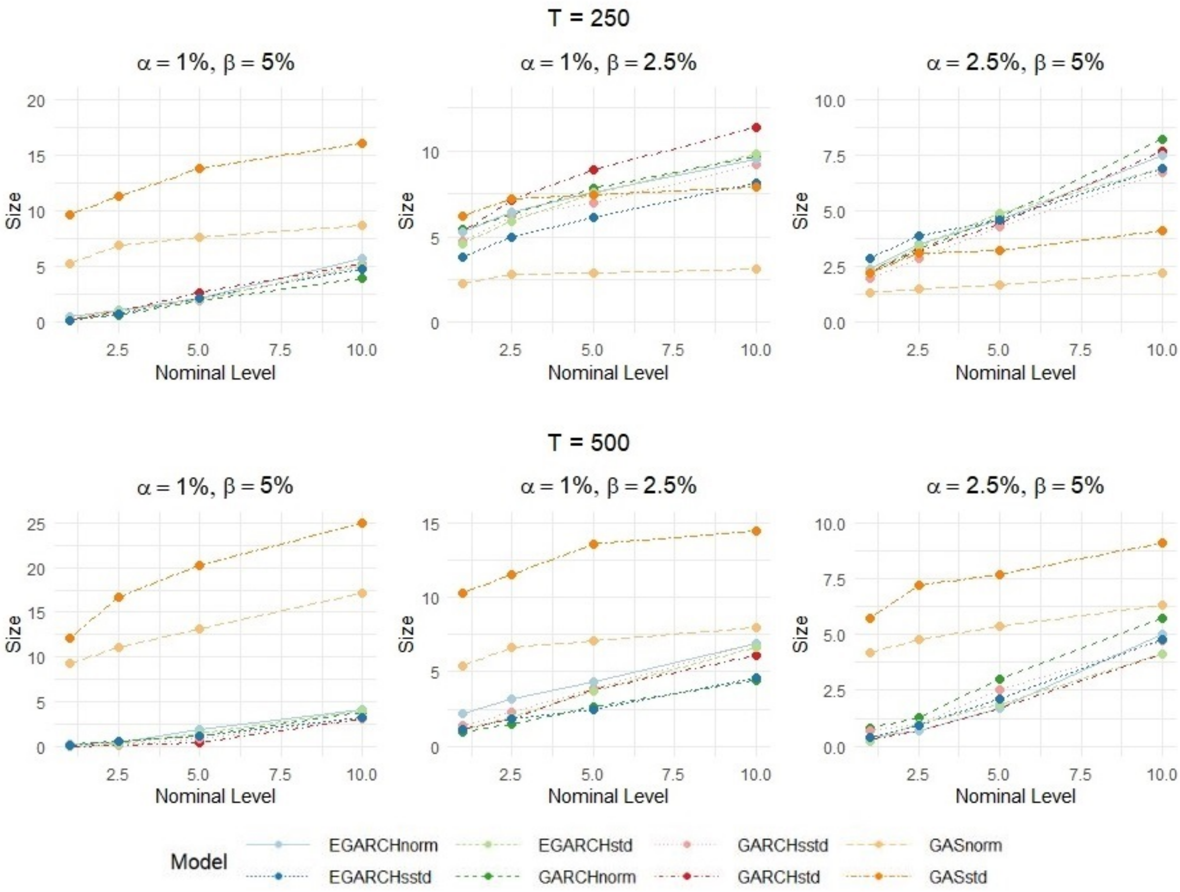
Observing the sizes for the three proposed backtesting procedures as the in-sample size increases from 250 to 1000 observations, Table 5 shows a similar behavior of the null models to the preceding scenario. More precisely, the proposed procedures underestimate and overestimate the null rejection rate for the same models at the same significance levels and out-of-sample observations described in the previous analysis, where n equals 250. Tables 6-8 provide a more comprehensive representation of the sizes of each model when the in-sample observations equal 250 and when they increase to 1000. Comparing each null model with its respective results in the two scenarios reveals a certain degree of similarity, except for the significance level of $\alpha = 1\%$, $\beta = 2.5\%$, and nominal levels of 1%, 2.5%, and 5%, coupled with 250 out-of-sample observations. In this cases, there is an escalation in distortions whereby, at $n = 1000$, the sizes of all null models become even further from the null rejection rate. By illustration, the size results of GARCH$_{norm}$ in Table 6 demonstrate an increase from 3.1% to 5.4% at the nominal level of 1%; from 4.4% to 6.3% at the nominal level of 2.5%; and from 6% to 7.8% at the nominal level of 5%. Such observation contradicts the literature, where what is actually seen is a considerable improvement of the size tests as there is an increase in the in-sample observations (see, for instance, Bayer and Dimitriadis (2020)).

Figure 2 – MF Test: Evolution of each null model for $n = 250$ and $T \in \{250, 500\}$, using the significance levels of 1%, 2.5% and 5%



Figures 2 and 3 provide a clearer visual representation of the information presented in Tables 6-8 for the adapted McNeil and Frey (2000) approach among all scenarios and parameters considered. By the graphs, it can be verified more evidently that, looking separately at each significance level and out-of-sample size, the null models maintain a similar behaviour for both in-sample observations. For both in-sample cases, when we analyse the GARCH-type models, the evolution of their sizes over nominal levels shows similar growth. Looking closer at the significance levels of $\alpha = 1\%$ and $\beta = 5\%$, the GARCH-type models show moderate increases and very low rejection rates for both in-sample observations. At the significance levels of $\alpha = 1\%$, $\beta = 2.5\%$ and $T = 250$, the aforementioned models have the highest rejection rates of the null hypothesis in addition to an average behavior about the other two when we look at $\alpha = 2.5\%$ and $\beta = 5\%$. Furthermore, when the out-of-sample observations tend to 500, the sizes of the GARCH-type models and GAS_{sstd} reduce; conversely, there is an increase in sizes for GAS_{norm} and GAS_{std}. Also, GAS_{norm} and GAS_{std} show the highest rejection rates of the null hypothesis at the significance levels $\alpha = 1\%$ and $\beta = 5\%$ and $T = 250$ as well as for all cases where $T = 500$. Lastly, although it was not possible to obtain results for the GAS_{sstd} model when $n = 1000$, we should point out that its behavior is similar to those mentioned for the GARCH-type models when looking at 250 in-sample observations (Figure 2).

Figure 3 – MF Test: Evolution of each null model for $n = 1000$ and $T \in \{250, 500\}$, using the significance levels of 1%, 2.5% and 5%



Continuing, Table 9 provides a clearer presentation of the models that have shown the best size test performances for the proposed backtesting procedures. Overall, the procedures tend to exhibit null rejection rates that align more closely with nominal levels when the data is normally distributed, with emphasis on 1000 out-of-sample observations. Nevertheless, the tests provide the best size results for the Skewed Student-t and Student-t distributions, primarily at the significance levels of: $\alpha = 1\%, \beta = 2.5\%$ and 250 out-of-sample observations for 5% and

10% nominal levels; $\alpha = 2.5\%, \beta = 5\%$, $T = 250$ for the nominal level of 2.5% and $T = 500$ for the nominal level of 10%; and $\alpha = 1\%, \beta = 5\%$ and 250 out-of-sample observations for 5% nominal level. Such relative superiority can be justified by the characteristics of the normal distribution, which, when compared to the others, does not have heavy tails (Student-t) and heavy tails with skewness (Skewed Student-t). Although, the characteristics of financial data resemble the Student-t and Skewed Student-t distributions more closely. In terms of the employed models, what we generally observe is that more satisfactory sizes are achieved through the null models EGARCH and GAS, with also excellent accomplishments for the GARCH model, mainly for nominal levels of 1% and 2.5%. This result may be related to the aforementioned nature of each model, where the GAS model encompasses the GARCH model and exploits the complete density of the returns rather than only means and higher moments. In contrast, the EGARCH model incorporates additional features such as asymmetry and leverage effect, allowing for a better representation of time-varying volatility. Besides, the GARCH-type models show superiority to the GAS model regarding the rejection rates at significance levels $\alpha = 1\%$ and $\beta = 5\%$, except at the nominal level of 10% and out-of-sample equal to 250, where the GAS$_{norm}$ presents the best results for all proposed procedures. Therefore, the introduced backtesting procedures seem to better encompass the risk predictions of RVaR which are based on the intrinsic characteristics of these models.

4.1.2 Power of the tests

We organize the results from different perspectives to evaluate the power property of the proposed backtesting procedures for the RVaR. Tables 10-21 show the alternative hypothesis rejection rate of the three tests, considering in the analysis the GARCH and EGARCH models, along with the previously defined distributions, and using in-sample sizes (n) of 250 and 1000 observations. In each table, the results are separated by the considered nominal levels of 1%, 2.5%, 5%, and 10%, where, in each of these levels, they are presented according to the out-of-sample sizes (T) of 250 and 500 obtained for each considered significance level. Tables 22-27 show, respectively, the outcomes of the proposed approaches for RVaR regarding the evolution of the non-null rejection rate with a look at the change in the number of in-sample observations. Lastly, Figures 4 and 5 illustrate the behaviour of each alternative model over the nominal levels within the context of the adapted McNeil and Frey (2000) approach[8]. The EGARCH$_{norm}$ model was considered the data generation process, given its robust performance in size tests, which positions it as a compelling candidate for graphical representation. Anew, the McNeil and Frey (2000), Righi and Ceretta (2015) and McNeil et al. (2015) procedures modified for RVaR are named in each table, respectively, by MFTest, RCTest, and MFETest. Given the computational burden, we could not get the results for the GAS model.

[8] The remaining models present very similar results and, for this reason, they were omitted from the analysis without loss of generality.

Observing the behavior of the results in Tables 10-21, we verified that, like the prior analysis, all backtesting procedures have similar test powers for all scenarios under analysis. In short, no major superiority exists in the power test performance of any approach considered. By illustration, the models' results in Table 12 for the nominal level of 5% and significance levels of $\alpha = 1\%$, $\beta = 5\%$ vary by a maximum of 0.2%. More specifically, for $T = 500$, the EGARCH$_{norm}$, which exhibits the best performance in the three tests, has a size of 4.1% for the MFTest, 4.0% for the MFETest, and 4.2% for the RCTest. Similar results were observed in the other models and scenarios. This conclusion differs from what is observed in the literature, where commonly one method stands out among candidates, as evidenced, for example, in Bayer and Dimitriadis (2020). Another initial impression when examining the same tables is that all non-null rejection rates underestimate the nominal levels for all backtest and parameters considered, not exceeding 15%. Values of this magnitude are observed at the significance level of $\alpha = 1\%, \beta = 2.5\%$ and $T = 250$. In the same way, with few exceptions, the highest non-null rejection rates are concentrated at 250 out-of-sample observations. As it increases to 500, these rates tend to reduce. The former case is usual in risk forecasting studies that use Monte Carlo simulations, as seen in Escanciano and Velasco (2010). Conversely, the latter case represents a divergence of what is often seen in literature, where the increase of the out-of-sample observations leads to a larger number of rejections regarding the alternative models; see, for instance, Deng and Qiu (2021).

From the results presented in Tables 10-15, we can observe which models exhibit the best power test performances across the nominal levels for the three backtesting procedures, considering 250 in-sample observations and distinct DGPs. To begin with, Table 10 shows the non-null rejection rates of the tests when employing GARCH$_{norm}$ as the data generating process. In this scenario, the proposed approaches demonstrate the best performances across the nominal levels when using the GARCH$_{std}$ as the alternative model, with the exception primarily at the significance levels of $\alpha = 1\%$, $\beta = 5\%$, where the EGARCH models stand out. When we generate the data using the GARCH$_{std}$ model (Table 11), the backtesting procedures show the best performances when the GARCH$_{norm}$ and EGARCH$_{norm}$ models are used for the predictions. More precisely, at nominal levels of 1%, 2.5%, and 5%, the proposed procedures yield better results for the former alternative model. Nonetheless, for significance levels of $\alpha = 2.5\%, \beta = 5\%$ and $T = 500$, and considering the nominal level of 1%, coupled with significance levels of $\alpha = 1\%, \beta = 2.5\%$ and $T = 250$, the tests are more efficacious when the latter model is used. As the nominal level tends to 10%, the highest non-null rejection rates are observed when applying the EGARCH$_{norm}$ model, especially for 500 out-of-sample observations. Using GARCH$_{sstd}$ as the DGP, we can observe from Table 12 that the highest non-null rejection rates are once again found when conducting backtests with forecasts generated by GARCH$_{norm}$ and EGARCH$_{norm}$ models. In the first case, the most satisfactory results are observed at significance levels of $\alpha = 1\%, \beta = 5\%$, and $T = 250$ for all nominal levels; at significance levels of $\alpha = 2.5\%, \beta = 5\%$, and $T = 250$ for all nominal levels except 10%; and finally, at significance levels of $\alpha = 1\%, \beta = 2.5\%$, and $T = 500$ for the 1% and 10%

nominal levels. Conversely, the second case prevails in most of the remaining scenarios, with greater prominence as n increases toward 10%. The exceptions are at significance levels of $\alpha = 1\%, \beta = 2.5\%$, and $T = 250$ for nominal levels of 1% and 2.5%. Applying the RVaR with the EGARCH$_{sstd}$ model achieves the best rejection rates on the simulated data.

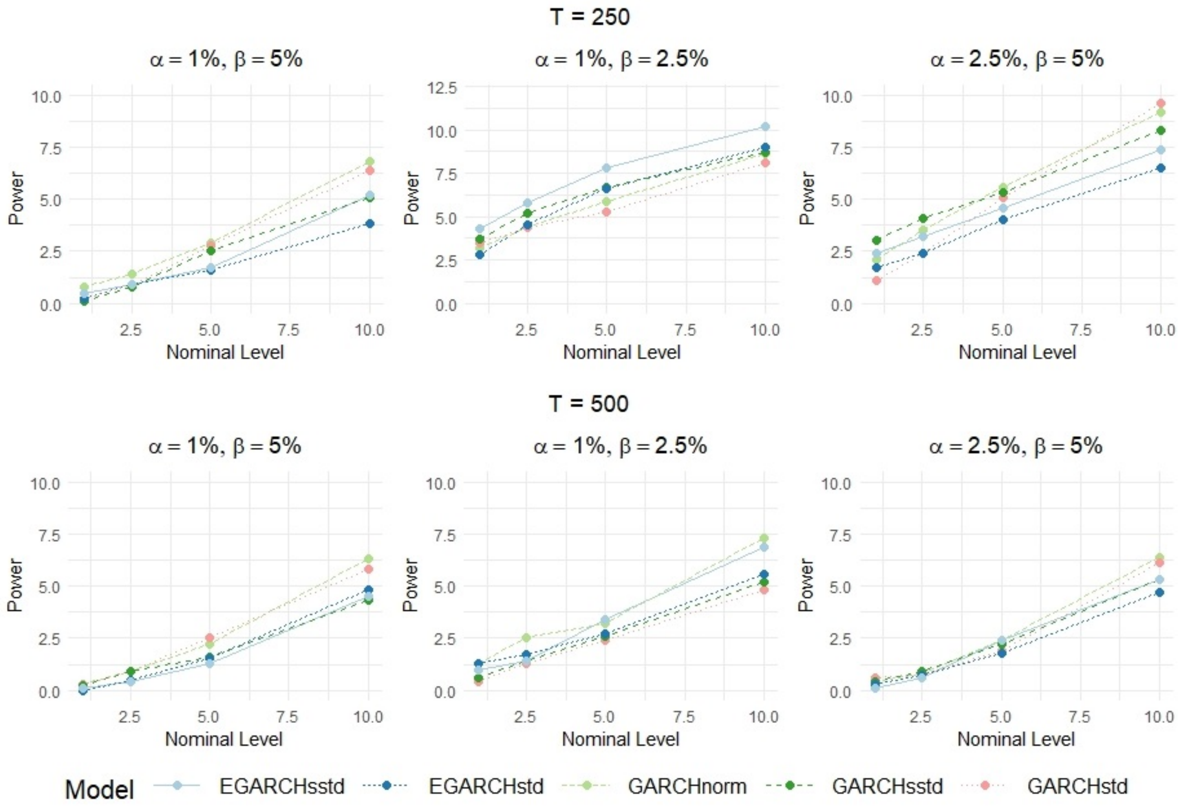
When the data is generated by the EGARCH$_{norm}$ model, we can see from Table 13 that the highest powers are predominantly evidenced when employing the GARCH$_{norm}$ model and, to a lesser extent, the EGARCH$_{sstd}$ model in the RVaR predictions. More specifically, the former case is observed primarily at the significance levels of $\alpha = 1\%, \beta = 5\%$ for all nominal levels and out-of-sample observations and at the significance levels of $\alpha = 1\%, \beta = 2.5\%$, for $T = 500$ and all nominal levels except 5%. The latter case is noted at the significance levels of $\alpha = 1\%, \beta = 2.5\%$, for all nominal levels and $T = 250$, and also for $T = 500$, when the nominal level is 5%. In Table 14, we can observe that when the data is generated by the EGARCH$_{std}$ model, the three backtesting procedures demonstrate their best performances when the EGARCH$_{norm}$ model is utilized as the alternative model, for all nominal levels and parameters considered, with exceptions only at the significance levels of $\alpha = 1\%, \beta = 5\%$, 250 out-of-sample observations and nominal levels of 5% and 10%; and at the significance levels of $\alpha = 2.5\%, \beta = 5\%$, 500 out-of-sample observations and nominal level of 1%. Table 15 shows the power tests of the three proposed procedures when using EGARCH$_{sstd}$ as the DGP. In this last scenario for 250 in-sample observations, the most satisfactory rejection rates are identified in the RVaR forecasts using the EGARCH$_{norm}$ model. In a few instances, the tests exhibit improved performance when the GARCH$_{norm}$ is employed as the alternative model. Such instances are found at $T = 250$, with significance levels of $\alpha = 2.5\%, \beta = 5\%$, across all nominal levels, as well as significance levels of $\alpha = 1\%, \beta = 5\%$, for the 1% and 2.5% nominal levels.

When looking at the power tests using 1000 in-sample observations, which are presented in Tables 16-21, we can observe differences in the performance of the three proposed backtesting procedures when compared to the previous framework. Initially, Table 16 shows the non-null rejection rates of the tests when employing GARCH$_{norm}$ as the data generating process. The proposed approaches demonstrate the best performances when utilizing the GARCH$_{std}$ as the alternative model, primarily at the significance levels of $\alpha = 1\%, \beta = 2.5\%$ across the nominal levels, and at the levels of $\alpha = 2.5\%, \beta = 5\%$ for nominal levels of 5% and 10%. For the remaining cases, the EGARCH model stands out. When we generate the data using the GARCH$_{std}$ model (Table 17), the backtesting procedures show the best performances primarily when the EGARCH$_{norm}$ and, to a lesser extent, GARCH$_{norm}$ models are used for the predictions. Specifically, the proposed procedures provide the most effective rejection rates when the first alternative model is utilized for calculating RVaR forecasts across all nominal levels. Although, for significance levels of $\alpha = 1\%, \beta = 5\%$, $T = 250$ and nominal levels of 1% and 10%; $\alpha = 2.5\%, \beta = 5\%$, $T = 250$ and nominal levels of 1% and 2.5%; and $\alpha = 1\%, \beta = 5\%$ jointly with the nominal level of 5%, the tests are more efficacious when the latter model is used. By simulating the data with the GARCH$_{sstd}$ model, we can observe from Table 18 that the highest non-null rejection rates are once again found strictly when conducting backtests with forecasts generated by GARCH$_{norm}$ and EGARCH$_{norm}$ models. In the former case, the most satisfactory results are concentrated at significance levels of $\alpha = 1\%, \beta = 5\%$ and $\alpha = 2.5\%, \beta = 5\%$, for all nominal levels and out-of-sample observations. On the other hand, the latter case prevails at significance levels of $\alpha = 1\%, \beta = 2.5\%$ across all nominal levels and out-of-sample sizes considered.

When the data is generated by the EGARCH$_{norm}$ model, we can observe from Table 19 that the highest power tests are predominantly evidenced when employing the EGARCH$_{std}$ and EGARCH$_{sstd}$ models, and to a lesser extent, the GARCH$_{norm}$ model, in the RVaR predictions. More specifically, the superior results of the backtesting procedures, primarily associated with applying GARCH models when $n = 250$, now exhibit improved rejection rates for EGARCH models. The prominence belongs to GARCH only for significance levels of $\alpha = 1\%, \beta = 2.5\%$, $T = 250$, and nominal levels of 2.5%, 5%, and 10%, as well as $\alpha = 1\%, \beta = 5\%$ for the nominal level of 10%. For EGARCH, it is worth mentioning that the best backtesting results for the $\alpha = 1\%, \beta = 5\%$, and $\alpha = 2.5\%, \beta = 5\%$ levels are associated with the use of EGARCH$_{sstd}$ model with T = 250, except for 10% nominal level. Also, for the significance levels of $\alpha = 2.5\%, \beta = 5\%$, these outcomes are achieved by applying the EGARCH$_{std}$ model for 500 out-of-sample observations. In Table 20, we can observe that when the data is generated by the EGARCH$_{std}$ model, the three backtesting procedures demonstrate their best power test performances at significance levels of 1% and 2.5% when the RVaR forecasts are computed using the EGARCH$_{norm}$ as the alternative model, across all considered nominal levels. Moreover, when the significance levels are $\alpha = 2.5\%, \beta = 5\%$, the rejection rates are superior at the 1% and 2.5% nominal levels when the same model is employed. However, at the 5% nominal level with $T = 500$ and the 10% nominal level with $T = 250$, the best rejection rates are given using the GARCH$_{norm}$ model on the simulated data. Also, at significance levels of $\alpha = 1\%$ and $\beta = 5\%$, the tests more efficiently reject forecasts produced by the GARCH$_{norm}$ model, particularly at the 1% and 10% nominal levels, as well as the 2.5% nominal level when $T = 500$. Lastly, Table 21 shows the power tests of the three proposed procedures when using EGARCH$_{sstd}$ as the DGP. The highest rejection rates for RVaR forecasts are achieved using the alternative models GARCH$_{norm}$ and EGARCH$_{norm}$. For the former case, these are observed at significance levels of $\alpha = 1\%, \beta = 5\%$, and $\alpha = 1\%, \beta = 2.5\%$;. In contrast, for the latter case, they are seen at the $\alpha = 2.5\%, \beta = 5\%$ levels, across all nominal levels considered.

Looking at the aforementioned results, it is reasonable to assert that they are connected to the employed distributions, wherein disparities in their respective characteristics contribute more expressively to the increase in power than the utilized models. Such a conclusion is consistent with similar studies that utilize Expected Shortfall, where non-null rejection rates are higher when the distributions employed in the data generating process differ from those used in the alternative model. See, for instance, Du and Escanciano (2017) and Bayer and Dimitriadis (2020).

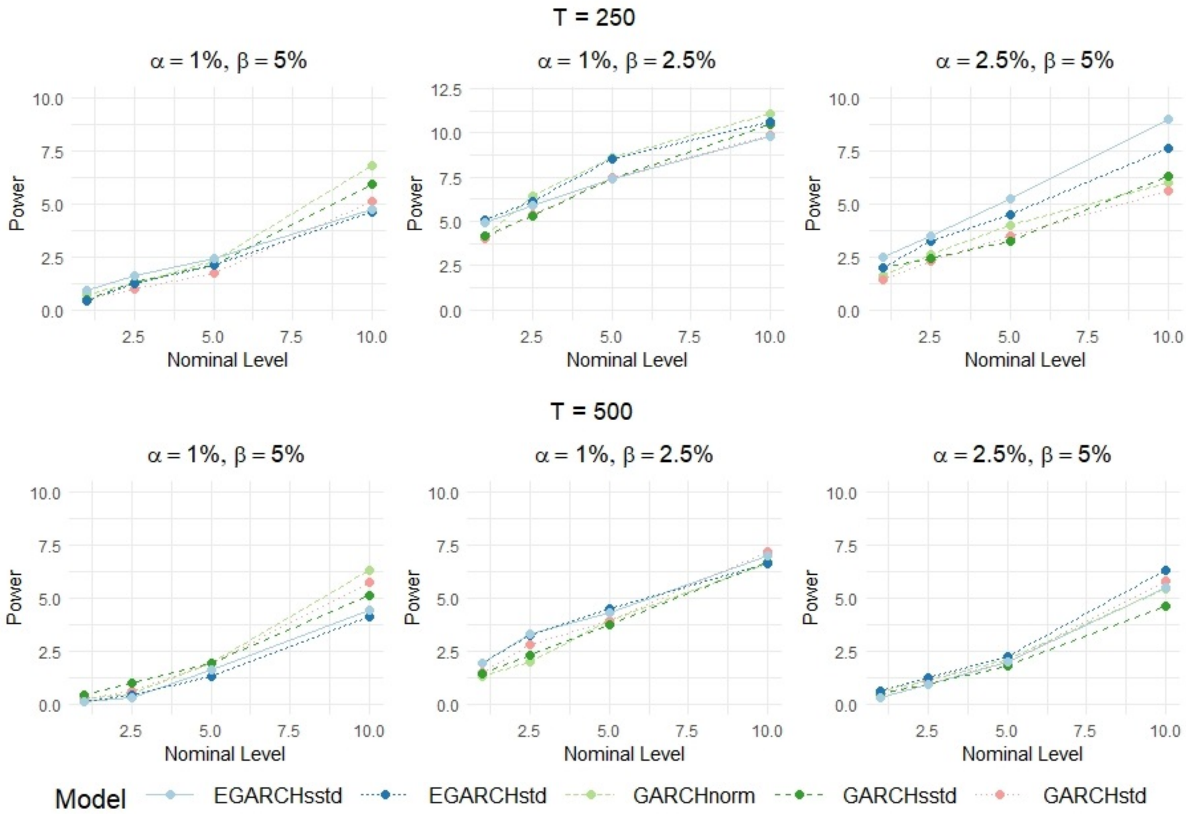
Figure 4 – MF Test: Evolution of each alternative model, using $EGARCH_{norm}$ as DGP, for $n = 250$, $T \in \{250, 500\}$ and significance levels of 1%, 2.5% and 5%.



Tables 22-27 provide a more comprehensive representation of the evolution of the non-null rejection rates, referring to each proposed backtesting procedure, individually, when the in-sample observations are equal to 250 and when they increase to 1000. Comparing each alternative model with its respective results in the two scenarios reveals substantial increases in test rejection rates as n tends to 1000 only in specific cases. Namely, these increases take place using as DGP the $GARCH_{norm}$ and $EGARCH_{norm}$ models at the significance level of $\alpha = 1\%$, $\beta = 2.5\%$, and 250 out-of-sample observations (Tables 22 and 25); the $GARCH_{std}$ model at the significance level of $\alpha = 2.5\%$, $\beta = 5\%$, along with 250 out-of-sample observations (Table 23); and the $EGARCH_{std}$ model at the significance level of $\alpha = 1\%$, $\beta = 2.5\%$, in conjunction with 250 out-of-sample observations, only for the GARCH alternative models (Table 26). In the remaining cases, it can be observed that the proposed procedures exhibit similarity or a reduction in power tests. Such observation presents a partial departure from the literature, where what is seen is a pronounced improvement in the efficacy of power tests in response to an increase of in-sample observations. See, for instance, Deng and Qiu (2021).

For illustrative purposes, Figures 4 and 5 provide a clearer visual representation of the power tests of the alternative McNeil and Frey (2000) procedure based on data generated by the $EGARCH_{norm}$ model. The choice of this scenario is merely due to its satisfactory performance in size tests, making it a reasonable candidate for an analysis of its behavior in terms of alternative models. When observing each graph individually, we can perceive that the rejection rates of

Figure 5 – MF Test: Evolution of each alternative model, using $EGARCH_{norm}$ as DGP, for $n = 1000$, $T \in \{250, 500\}$ and significance levels of 1%, 2.5% and 5%.



alternative hypotheses exhibit a certain degree of homogeneity across the nominal levels. Also, when comparing both in-sample sizes, it is possible to notice a similarity in the backtests results, emphasizing the above statement that there are no substantial improvements in the power test with increasing the size of n, with few exceptions. Looking closer at $T = 250$, we can more distinctly observe that the higher power tests are located at the significance level of $\alpha = 1\%$ and $\beta = 2.5\%$, while the lower ones are at $\alpha = 1\%$ and $\beta = 5\%$. In addition, an average behavior compared to the first two can be noted at the significance level of $\alpha = 2.5\%$ and $\beta = 5\%$. As the out-of-sample observations approach 500, the power of the GARCH-type models decreases. Also, the proposed backtesting procedures continue to show the highest rejection rates of alternative hypotheses at the significance level of $\alpha = 1\%$ and $\beta = 2.5\%$, but with reduced differences regarding the other levels. Conversely, the proposed approaches demonstrate comparable powers for the remaining significance levels, as indicated by the sharp decline in rejection rates at $\alpha = 2.5\%$ and $\beta = 5\%$. The same conclusions are extended to the RCTest and MFETest approaches.

In summary, when the data is generated by processes with Student-t or Skewed Student-t distribution, the proposed backtesting procedures exhibit higher non-null rejection rates when RVaR predictions are conducted using the normal distribution. Such behavior is observed for both 250 and 1000 in-sample observations. This can be justified by the capacity of the method to capture the properties of the normal distribution, which, when compared to the others, does not have heavy tails (Student-t) and heavy tails with skewness (Skewed Student-t). Moreover, the characteristics of financial data closely resemble the Student$-t$ and Skewed Student-t distributions. As a result, the obtained powers demonstrate consistency with empirical observations, as their rejection rates are higher when forecasts are made using a distribution that lacks inherent properties of financial asset returns. On the other hand, when DGPs with normal distribution are used, the proposed procedures exhibit some peculiarities. With data simulated by the GARCH$_{norm}$ model and 250 out-of-sample observations, the higher powers are linked to the Student-t distribution; however, as T tends to 1000 observations, there is an increase in the powers associated with the normal distribution. The opposite is observed when the data is generated by the EGARCH$_{norm}$ model, where for 250 out-of-sample observations, the higher powers are found in the normal distribution, and with an increase to 1000 out-of-sample observations, the higher rejection rates of the tests are related to the Student-t and Skewed Student-t distributions. Therefore, the proposed backtesting procedures exhibit greater consistency for normally distributed data for T equals 250 when generated by the former model and T equals 1000 when generated by the latter.

Our numerical analysis has focused on the aspects of size and power of the proposed backtesting procedures for the RVaR measure across different models and sample sizes. We can encapsulate the main results as follows:

- None of the proposed backtesting methods demonstrates significant superiority over the others: in both size and power tests, the variations between the backtestings hover around 0.2%. This result is of great significance, as it deviates from what is observed in the literature for other risk measures such as VaR and ES;
- With some exceptions, the size and power of the tests deteriorate as T increases: The highest powers are observed at $T = 250$, with particular emphasis on the significance levels of $\alpha = 1\%, \beta = 2.5\%$. The best sizes are also found at $T = 250$ but at the levels $\alpha = 2.5\%$, $\beta = 5\%$ for the nominal level of 5%, and at the levels $\alpha = 1\%$, $\beta = 2.5\%$ for the nominal level of 10%. However, best sizes are also observed at $T = 500$ for the nominal levels of 1% and 2.5%, together with the significance level of $\alpha = 1\%$, $\beta = 2.5\%$;
- As n increases to 1000, the proposed procedures demonstrate notable modifications in their power and size tests, particularly at the significance levels of $\alpha = 1\%, \beta = 2.5\%$, and $T = 250$. For such parameters, in general, the sizes deteriorate across the nominal levels, and the powers improve more prominently for data generated with the normal distribution;
- The proposed methodologies exhibit, in general, the best size tests primarily from null models with the normal distribution, with particular emphasis on 1000 in-sample observations;

- When the data is generated by processes that capture the properties of the financial assets (skewness or heavy tails), the proposed procedures exhibit higher powers when RVaR predictions are conducted using the normal distribution.

The best performing backtesting scenario depends on the interaction between the size and power properties. As aforementioned, coherence is evident in the size and power tests when the backtesting is executed using 250 out-of-sample observations. However, upon closer examination of the in-sample observations, a size worsening is noticeable, accompanied by an enhancement in power, particularly at the 1% and 2.5% significance levels. Notably, these power improvements apply solely to scenarios characterized by a normal distribution in the DGP, which does not hold for financial assets. Consequently, the optimal configuration can be identified at $n = 250$ and $T = 250$, where, although smaller than what is seen in the ES studies, it aligns with the minimum rolling estimation window size of 250 observations recommended by the Basel Committee on Banking Supervision (2013).

4.2 Empirical Analysis

In this section, we present the results of the three proposed backtesting procedures for RVaR, assessing their behaviors for the assets above in conjunction with their respective realized losses. To begin with, Table 3 displays the descriptive statistics of the daily returns of S&P 500, Ibovespa (IBOV), U.S. Treasury Bonds rates for 3 years (DGS3) and 10 years (DGS10), WTI Crude Oil, Gold, Real/U.S. Dollar and Euro/U.S. Dollar exchange rates, DJSI World, STOXX Europe 600, and Bitcoin spanning from January 2010 to December 2021[9].

Table 3 – Descriptive statistics of the daily returns of S&P 500, Ibovespa, 3-Years and 10-Years U.S. Treasury Bonds, WTI Crude Oil, Gold, Real/U.S. Dollar and Euro/U.S. Dollar exchange rates, DJSI World, STOXX Europe 600 and Bitcoin from January 2010 to December 2021.

Asset	Mean	Minimum	Maximum	Standard Deviation	Skewness	E.Kurtosis
S&P500	0.0005	−0.1276	0.0897	0.0108	−0.8561	16.2092
Ibovespa	0.0001	−0.1599	0.1302	0.0157	−0.8295	11.7415
Treasure 3-Years	−0.0003	−0.3101	0.3716	0.0503	0.0992	7.1422
Treasure 10-Years	−0.0002	−0.3151	0.3417	0.0291	0.0920	23.7933
WTI Crude Oil	0.0003	−0.2814	0.4258	0.0284	1.3551	45.9848
Gold	0.0002	−0.0982	0.0578	0.0103	−0.6438	6.5696
Real	−0.0004	−0.0725	0.0638	0.0111	−0.0523	3.5798
Euro	−0.0001	−0.0281	0.0313	0.0054	−0.0952	2.3674
DJSI World	−0.0003	−0.0834	0.1039	0.0094	1.0747	15.0287
STOXX Europe 600	0.0002	−0.1219	0.0807	0.0108	−0.8350	10.1350
Bitcoin	0.0023	−0.4647	0.2251	0.0393	−0.7999	11.2393

Note: E.Kurtosis refers to the excess kurtosis of asset returns.

We observe that the highest average returns in the period belong to Bitcoin (0.0023) and the lowest ratio to the U.S. Dollar of Real (-0.0004). In general, the mean of the assets is close to zero, as expected. Bitcoin also has the lowest minimum (-0.4647), while the highest maximum belongs to WTI Crude Oil (0.4258) returns. The highest volatility (0.0503) can be seen by the 3-year U.S. Treasury Bonds yield, as illustrated in the Standard Deviation column. Considering the assets' results, we expected that the 3 and 10-year U.S. Treasury Bonds yield, WTI Crude Oil, and Bitcoin would have the highest overall risk measure values. On the other hand, although Real had the lowest average returns, the lowest volatility belongs to the Euro (0.0054). The S&P 500, Ibovespa, and Dow Jones Sustainability indexes also exhibited lower volatility than Bitcoin (0.0393), likely due to the benefit of diversification. Notably, the series demonstrates higher kurtosis than expected for series that follow a normal distribution (kurtosis = 3), indicating that all of them are leptokurtic series, which is a standard behavior of return series. Nevertheless, certain series display positive skewness, which deviates from the stylized facts of financial asset returns. See Cont (2001) for details. Moreover, the high values of excess kurtosis among most assets can be attributed to the impact of the COVID-19 pandemic. Figure 6 presents the plots of the temporal evolution of each asset from January 2010 to December 2021 and elucidates the

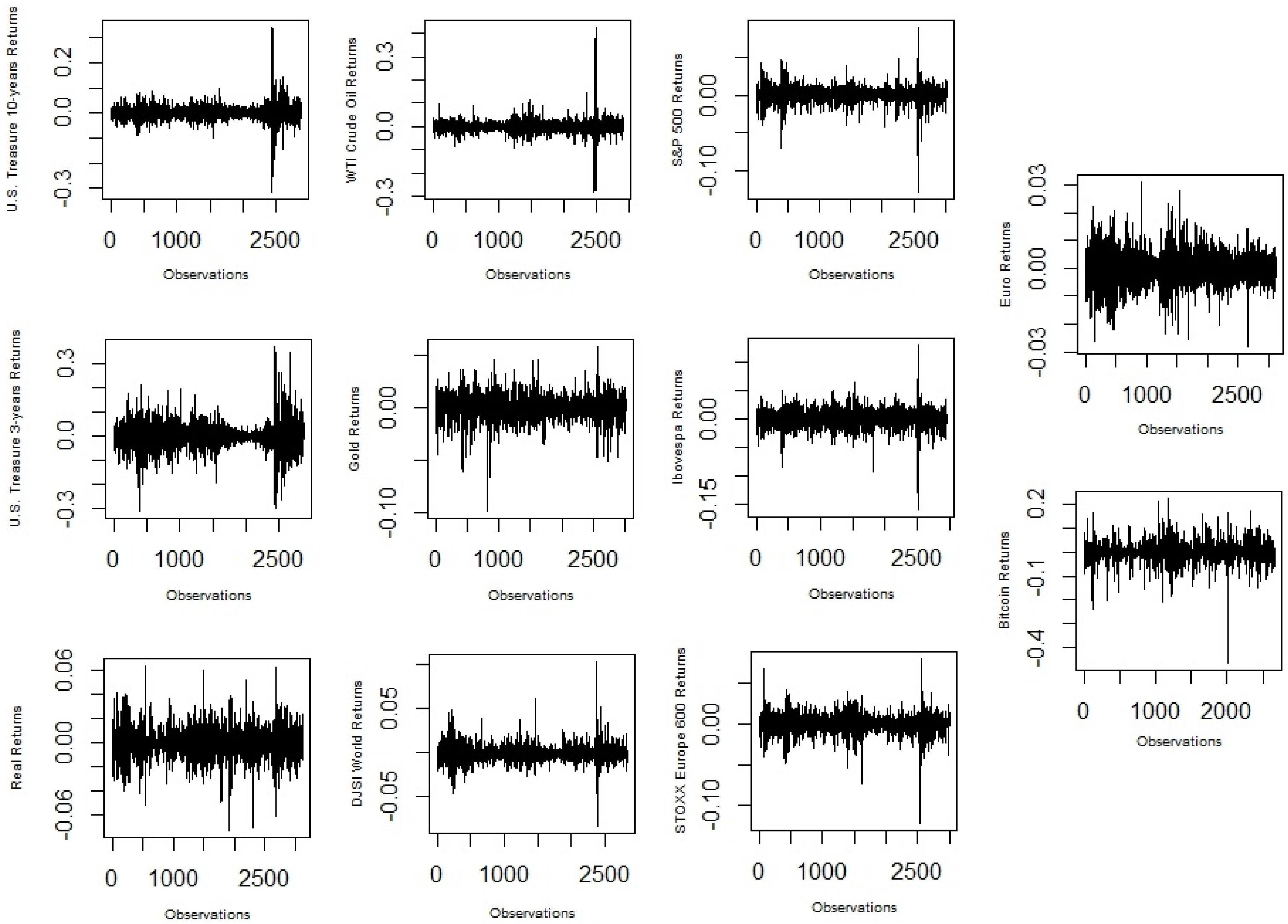
[9] Except for Bitcoin returns, which start in September 2014 due to the availability of data for this asset

presence of turbulent periods, especially in a more recent range of the data, corroborating with our previous COVID-19 statement. Also, it provides a clearer depiction of the high dispersion of returns for 3 and 10-year U.S. Treasury Bonds yield, WTI Crude Oil, and Bitcoin, as well as the low dispersion of returns for the ratio to the U.S. Dollar of Euro.

We organized the results from two different perspectives. At first, we commence by presenting the best results obtained by the proposed backtesting procedures, i.e., the scenarios where the p-values are closer to one, considering the different rolling estimation window and out-of-sample observation sizes, separated by significance level. In this way, we analyse these best-case scenarios and compare them with their respective outcomes obtained from the designated loss function $S^{RVaR^{\alpha,\beta}}$, as defined in equation (2.9). The same procedure was performed for the instances with the most unfavorable p-values, i.e., the tests that showed p-values closer to zero. In a second moment, we examine the finest and poorest outcomes through the lens of the same loss function $S^{RVaR^{\alpha,\beta}}$, contrasting them with the corresponding p-values yielded by the proposed approaches. As described earlier, this exercise allows us to evaluate the outcomes and ascertain if they remain true under both methods. The complete tables can be found in Appendix A.

Initially, Tables 28, 29, and 30 present the models with the best calculated p-values (closer to 1) of the proposed backtesting, along with the corresponding financial returns and loss function results, using rolling estimation windows of 250, 500, and 1000 observations. In each table, the results are separated by the significance levels of 1%, 2.5%, and 5%, wherein in each of these levels, the p-values are presented in descending order. Observing the results in such tables, we verified that there is a similar behavior of the backtesting procedures regarding the best scenarios under analysis. In other words, there is no relevant difference in the performance of the tests for any considered approach since they have similar best-performing scenarios. By illustration, the results in Table 28 for significance levels of $\alpha = 1\%$, $\beta = 5\%$ exhibit similarities in the MF Test, RC Test, and MFE Test, given that the GARCH_{ged}, GARCH_{sged} and GARCH_{std} performed better for Real, GAS_{sstd} for Euro and gjrGARCH_{std} for DGS10 in the three approaches. Similar results were observed in the other significance levels and rolling windows. Such a conclusion contrasts with what is commonly observed in the literature, where some backtesting procedure tends to outperform the other candidates; see, for example, Deng and Qiu (2021). Another initial impression taken when analysing the same tables is the prominence of GARCH-type models and, despite the aforementioned qualities, the absence of GAS models. Nevertheless, there is a greater emphasis on models with distributions that exhibit asymmetry or heavy tails, aligning with characteristics observed in financial assets. Furthermore, such results are consistent with studies of the same nature conducted with risk measures like VaR and ES, where models that perform better in empirical tests are associated with distributions that share characteristics related to stylized facts of stock returns. See, for instance, Mighri et al. (2010) and Nilsen (2022).

Figure 6 – Daily returns of S&P 500, Ibovespa, 3-Years and 10-Years U.S. Treasury Bonds, WTI Crude Oil, Gold, Real/U.S. Dollar and Euro/U.S. Dollar exchange rates, DJSI World, STOXX Europe 600 and Bitcoin from January 2010 to December 2021.



Upon further examination of the results presented in Table 28, it can be observed that at significance levels of $\alpha = 1\%, \beta = 5\%$, the majority of the best-performing models are related to exchange rate assets, with GARCH$_{ged}$, GARCH$_{std}$, and GARCH$_{sged}$ models presenting p-values greater than 0.97 when applied to the returns of Real/U.S. Dollar, as well as the GAS$_{sstd}$ model when employed to the returns of Euro/U.S. Dollar. When examining the significance levels of $\alpha = 1\%, \beta = 2.5\%$, most of the best-performing models are associated with commodity assets. For instance, in the RC Test, the gjrGARCH$_{nig}$, gjrGARCH$_{sstd}$, and EGARCH$_{ged}$ models demonstrate p-values close to one when applied to WTI Crude Oil, while the EGARCH$_{nig}$ model shows a high p-value for Gold returns. Lastly, no particular market risk factor is predominant at $\alpha = 2.5\%$ and $\beta = 5\%$. However, the GARCH-type models demonstrate the best results in the tests, with special attention to the GARCH models with distributions that capture the stylized facts (std, ged, and nig) of financial returns. Table 29 shows the differences in the test outcomes as the in-sample size increases to 500 observations compared to the preceding framework. With the increase of n, what is now observed for the significance levels of $\alpha = 1\%, \beta = 5\%$ is a higher number of models related to commodity returns and a decrease in those applied to the interest rate. Furthermore, the gjrGARCH and EGARCH models gain greater prominence, particularly in conjunction with the Skewed Generalized Error and Generalized Error distributions. For the significance levels of $\alpha = 1\%, \beta = 2.5\%$, the models with the best p-values are related to the Euro/U.S. Dollar returns. By illustration, in the MFE Test, the gjrGARCH$_{snorm}$, gjrGARCH$_{ged}$, gjrGARCH$_{std}$, GAS$_{snorm}$ and GAS$_{sstd}$ demonstrate notable results for the returns of such asset. When looking at $\alpha = 2.5\%, \beta = 5\%$, the best-performing models are associated with WTI Crude Oil and DGS10 returns. In these scenarios, we emphasize the performance of the GARCH-type models with Skewed Generalized Error and Generalized Error distributions. Finally, the behavior of the backtesting results as the rolling window increases to 1000 can be observed from Table 30. More specifically, at significance levels of $\alpha = 1\%, \beta = 5\%$, the dominance of GARCH models can be observed when applied to exchange rate assets (Real and Euro). There is no specific dominance of any distribution. At $\alpha = 1\%$ and $\beta = 2.5\%$, the best p-values are found in GARCH-type models related to the WTI Crude Oil returns. More precisely, the GARCH$_{jsu}$, gjrGARCH$_{nig}$, gjrGARCH$_{jsu}$, GARCH$_{nig}$, GARCH$_{sstd}$ show noteworthy p-values when applied to such asset across all proposed backtesting procedures. At last, when considering the significance levels of $\alpha = 2.5\%, \beta = 5\%$, the GARCH-type models showed test results closer to one when applied to Euro/USD Dollar returns, contrasting with what is seen when n equals 500 (Table 29). There is no predominance of any specific model or distribution, with no need to further comments.

Additional conclusions can be drawn with the addition of the RVaR loss function. When observing the corresponding results of $\bar{S}_{RVaR}$ for the models with superior performance regarding the backtesting procedures for the rolling estimation window of size 250 (Table 28), it is noticeable that these exhibit reasonably small realized losses for the significance level of $\alpha = 1\%$ and $\beta = 5\%$, where a substantial portion of the values remains below 1.05. There is a higher incidence of intermediate scores for the other levels, around 1.10. When the rolling window tends to 500 observations (Table 29), we observe that the best predictions, according to the proposed backtesting procedures, exhibit a combination of small and medium realized losses at the significance levels of $\alpha = 1\%, \beta = 5\%$; low values (around 1.026) across the majority of scores for $\alpha = 1\%, \beta = 2.5\%$; and moderate values (around 1.098) at the significance levels of $\alpha = 2.5\%, \beta = 5\%$. Finally, with a rolling window of 1000 observations (Table 30), it can be seen that the best-case scenarios show small $\bar{S}_{RVaR}$ results at the significance levels of $\alpha = 1\%, \beta = 5\%$ and $\alpha = 2.5\%, \beta = 5\%$, except for the EGARCH$_{sged}$ model for Bitcoin in the former case (1.191133), and the GAS$_{snorm}$ model for Bitcoin in the latter case (1.146886). Furthermore, the highest realized losses are observed at $\alpha = 1\%, \beta = 2.5\%$, reflected by GARCH-type models applied to WTI Crude Oil returns. Therefore, based on the forecasts with the best-calculated p-values from the proposed procedures, there exists a reasonable association with the RVaR loss function, as the realized losses are predominantly small, with the remaining ones falling within an intermediate range among those calculated in the study, with few exceptions.

Continuing, Tables 31, 32, and 33 present the models with the poorest calculated p-values of the proposed backtesting, along with the corresponding financial returns and loss function results, using rolling estimation windows of 250, 500, and 1000 observations. The structure of each table is the same as in the former case. Looking at the test results, we can observe that the backtesting procedures exhibited similar behavior compared to the previous analysis, i.e., concerning the worst scenarios, there was no meaningful distinction between any approach, as they have similar performance. By illustration, the results in Table 32 for significance levels of $\alpha = 1\%$, $\beta = 5\%$ are nearby for the MF Test, RC Test and MFE Test, given that the GARCH$_{std}$, GARCH$_{ged}$, GARCH$_{norm}$, EGARCH$_{std}$ and EGARCH$_{snorm}$ performed better for STOXX Euro 600 index, and gjrGARCH$_{norm}$ for S&P 500 index in the three approaches. Analogous findings are observed in the other significance levels and rolling windows. Concluding the initial impressions, we can verify anew the prominence of GARCH-type models in the tables under analysis, as well as distributions with properties of asymmetry or heavy tails. However, it is worth mentioning that there is a relative increase in scenarios where the normal distribution occurs in conjunction with GARCH and gjrGARCH models.

Upon conducting a more detailed examination of the results from Table 31, we can observe that at significance levels of $\alpha = 1\%, \beta = 5\%$, the scenarios exhibiting p-values close to zero are predominantly associated with the STOXX Euro 600 index in conjunction with the GARCH model, and to a lesser extent, the gjrGARCH model. Looking at the significance levels of $\alpha = 1\%, \beta = 2.5\%$, most of the worst-performing models are associated with fixed-income assets. For instance, in the MF Test, the gjrGARCH$_{snorm}$ and EGARCH$_{norm}$ models demonstrate p-values close to zero when applied to DGS3 returns, while the GARCH$_{sstd}$ and GARCH$_{norm}$ models show a low p-value for DGS10 returns. Finally, no particular market risk factor is predominant at $\alpha = 2.5\%$ and $\beta = 5\%$. Nevertheless, the gjrGARCH model shows most of the worst results in the tests, emphasizing the Normal and Skewed Student-t distributions. With the increase of the rolling window to 500 observations (Table 32), it can be noticed that for the significance levels of $\alpha = 1\%, \beta = 5\%$, the models with the poorest performances remain related to the STOXX Euro 600 index in conjunction with the GARCH model, presenting p-values smaller than 0.004 in the three approaches. For the significance levels of $\alpha = 1\%, \beta = 2.5\%$, the lower p-values are associated with the GARCH model in conjunction with the Skewed Normal distribution across all proposed approaches. When looking to $\alpha = 2.5\%, \beta = 5\%$, the gjrGARCH models with distributions that capture the stylized facts (*nig*, *jsu*, and *sged*) of financial returns, applied to S&P 500 index, must be mentioned, as they encompass half of the worst p-values for the reference significance levels. Finally, the behavior of the backtesting results as the rolling window increases to 1000 can be observed from Table 33. More specifically, at significance levels of $\alpha = 1\%, \beta = 5\%$, the dominance of gjrGARCH and EGARCH models can be observed when applied to DGS3 returns. There is no specific dominance of any distribution. When noticing $\alpha = 1\%$ and $\beta = 2.5\%$, the poorest p-values are found in GARCH-type model predictions related to STOXX Europe 600 and gold returns. More precisely, across all proposed backtesting procedures, the GARCH_{snorm}, GARCH_{norm} and GARCH_{std} show p-values close to zero when applied to the former asset, and the EGARCH_{sged} and EGARCH_{sstd} models exhibit the same behavior when applied to the lesser. At last, when considering the significance levels of $\alpha = 2.5\%$ and $\beta = 5\%$, we can observe the predominance of EGARCH models, unlike when n is equal to 500, where gjrGARCH models gain more prominence. There is no prominent return series to which the models demonstrate the worst results, except for the MFE Test, where DGS3 appears in half of the occurrences. Also, there is no predominance of any specific distribution.

Analysing the data from the perspective of the RVaR loss function, we can observe that a substantial part of the predictions do not remain consistent in both methods. Specifically, with a rolling window size of 250 (Table 31), it is noticeable that the worst performing models regarding each backtesting procedure exhibit small and medium realized losses at the significance levels of $\alpha = 1\%, \beta = 5\%$ and $\alpha = 2.5\%, \beta = 5\%$. More precisely, in the former level, the scores have values around 1.043, whereas for the latter one, these values range from 1.029186 to 1.088937. In contrast, for the significance levels $\alpha = 1\%$ and $\beta = 2.5\%$, some high values of $\bar{S}_{RVaR}$ were obtained, being related to the forecasts of the gjrGARCH_{snorm} and EGARCH_{norm} models applied to the returns of DGS3 (around 1.194) and the GARCH_{snorm} model applied to Bitcoin (1.156860). When the rolling window tends to 500 observations (Table 29), we verify that the worst predictions according to the proposed backtesting procedures exhibit small values for the realized losses at the three significance levels, with exceptions for the GAS_{snorm} and GARCH_{sged} models for Bitcoin (1.21689 and 1.18714, respectively) at the levels of $\alpha = 1\%, \beta = 5\%$; and the GARCH_{std} for Bitcoin at $\alpha = 2.5\%, \beta = 5\%$. Finally, with a rolling window of 1000 observations (Table 30), it can be observed a higher consistency of forecasts in both applied methods, especially at the significance levels of $\alpha = 1\%, \beta = 5\%$, and $\alpha = 2.5\%, \beta = 5\%$. In the first case, the scores are around 1.173, being reflected by the GARCH-type models applied to the returns of DGS3. In the second case, half of the scores show high values, identified by the GARCH-type models applied to the returns of Bitcoin and DGS3. At the significance levels of $\alpha = 1\%, \beta = 5\%$, the expected losses remain at low values, in contrast to when n is equal to 250, where the results were more satisfactory at the same levels. Therefore, the forecasts with the worst-calculated p-values by the proposed procedures exhibit a considerable proportion of small scores, with the greater agreement between the methods in specific situations, namely, for a rolling window of 250 observations and significance levels of $\alpha = 1\%, \beta = 2.5\%$; and a rolling window of 1000 observations and significance levels of $\alpha = 1\%, \beta = 5\%$ and $\alpha = 1\%, \beta = 2.5\%$, driven by GARCH-type models applied to DGS3 and, to a lesser extent, Bitcoin returns.

Moving to the second moment, Tables 34-36 present the models with the best predictions from the perspective of the realized loss $\bar{S}_{RVaR}$, along with the corresponding asset return series and the p-values calculated by the three proposed backtesting procedures, using 250, 500, and 1000 in-sample sizes. The results are arranged in each table based on the significance levels of 1%, 2.5%, and 5%, wherein the realized losses are presented in descending order for each level combination. Analysing such tables shows that the GARCH-type models with Normal, Skewed Normal, and Student-t distributions applied to the Euro/US Dollar returns demonstrate the lowest scores among the candidates for all considered significance levels and rolling windows. Also, the gjrGARCH model performs better, especially at the significance levels of $\alpha = 2.5\%$ and $\beta = 5\%$. By illustration, for a rolling window of 500 observations, gjrGARCH$_{norm}$, gjrGARCH$_{std}$, gjrGARCH$_{snorm}$ and gjrGARCH$_{ged}$ are the best performing models, along with GARCH$_{std}$ and GARCH$_{snorm}$, for the mentioned levels. Compared with the proposed backtesting procedures, the scenarios with the best performance according to the realized loss exhibit p-values greater than 0.05 in all backtesting results. They thus cannot be rejected for a significance level of 5%. Thereby, scenarios with lower scores exhibit satisfactory p-values from the proposed approaches, demonstrating consistency with both methods.

Continuing, Tables 37-39 present the models with the poorest predictions from the perspective of the realized loss $\bar{S}_{RVaR}$, along with the corresponding asset return series and the p-values calculated by the three proposed backtesting procedures, using 250, 500, and 1000 in-sample sizes. The structure of each table is the same as the previous ones. When observing such tables, we can identify the predominance of DGS3 returns across all significance levels in conjunction with EGARCH and GAS models for a rolling window of 250 observations. By illustration, at the three considered levels, the EGARCH$_{ged}$, EGARCH$_{sged}$, EGARCH$_{nig}$, and EGARCH$_{jsu}$ models exhibit the worst realized losses, alongside the GAS$_{snorm}$ and GAS$_{sstd}$. For the rolling window of 500 observations, DGS3 returns still remain prominent, but the EGARCH models are replaced by GARCH and gjrGARCH models in conjunction with Generalized error and Skeweed Generalized error distributions. It is worth noting that the poorest realized losses are similar for the significance levels of $\alpha = 1\%, \beta = 5\%$, as well as $\alpha = 1\%, \beta = 2.5\%$, i.e., the highest $\bar{S}_{RVaR}$ values are observed when applying the GAS$_{snorm}$ models to DGS3 and Bitcoin returns, and gjrGARCH$_{ged}$, gjrGARCH$_{sged}$, GARCH$_{ged}$, and GARCH$_{sged}$ models to DGS3 returns. The significance levels of $\alpha = 2.5\%, \beta = 5\%$ show slight differences among the others, with the worst outcomes observed when applying the GAS_{snorm}, GAS_{sstd}, $gjrGARCH_{sged}$, $gjrGARCH_{nig}$, and $GARCH_{nig}$ models to DGS3, and the GAS_{snorm} model to Bitcoin returns. As we transition to a rolling window of 1000 observations, the EGARCH models regain prominence alongside the GAS models, with greater prominence given to Bitcoin returns. With the addition of the proposed backtesting procedures, most scenarios with the poorest performance, according to the realized loss, exhibit p-values greater than 0.05 in all backtesting results, similar to what was identified in the previous framework. Exceptions are observed at the significance levels of $\alpha = 1\%, \beta = 2.5\%$ for the three backtesting when the $EGARCH_{jsu}$ model is applied to DGS3 returns; and for the RC Test, when the $EGARCH_{sged}$ model is employed to DGS3 returns ($p = 0.025$). Consequently, scenarios with higher scores present unsatisfactory p-values in the proposed approaches, indicating a disparity between both methods.

When analysing the results with regard only to the assets considered, it becomes evident that the proposed backtesting procedure demonstrates exceptional performance for the exchange rate returns, without any pronounced bias toward a specific distribution. However, the approaches overestimate tests carried out with commodity assets, especially WTI crude oil, mainly at the significance levels of $\alpha = 1\%$ and $\beta = 2.5\%$. On the other hand, the proposed methods fail to provide non-rejection forecasts related to green finance, especially those linked to STOXX Europe 600, with some amelioration when considering a sample size of $n = 1000$. Such behavior occurs in a more expressive way at the significance levels $\alpha = 1\%, \beta = 5\%$ and is related to the use of GARCH models, but without calling attention to any specific distribution.

In conclusion, when we assess both moments jointly, we can assert that there is a greater agreement between the methods when looking at the best results. Thereby, we observe predominantly small realized losses among those calculated when analysing the data from the perspective of the p-values closer to one obtained from each backtesting procedure. Similarly, from the perspective of realized losses, the best predictions exhibit high p-values at 5%. Such symmetry is absent when analysing the worst scenarios. In this latter case, we predominantly observe small scores from the perspective of the p-values obtained through each backtesting procedure. Similarly, from the perspective of realized losses, the worst predictions showed high p-values at 5%, with few exceptions. Hence, the methods exhibit inconsistency in evaluating models with poorer performance, whereas the opposite pattern is observed for those that performed better. Thus, the similarities between the processes are limited, with an inclination towards the best results. As with numerical analysis, none of the proposed backtesting methods stand out over the others.

Table 4 – Scenario 1 - Size of the three backtesting procedures in percentage using $n = 250$ and $T \in \{250, 500\}$

Backtesting	Model	1%						2.5%						5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH$_{norm}$	0.4%	0.1%	3.1%	0.4%	1.6%	0.5%	1.2%	0.2%	4.4%	1.0%	3.1%	**1.1%**	2.4%	0.7%	6.0%	1.7%	4.4%	2.0%	5.4%	4.0%	8.7%	4.0%	7.1%	5.2%
	GARCH$_{std}$	0.1%	0.1%	3.4%	1.5%	2.4%	0.5%	5.0%	0.4%	4.9%	**2.6%**	3.2%	1.0%	1.1%	1.3%	6.4%	3.2%	4.5%	1.7%	3.8%	3.1%	8.4%	5.9%	7.1%	5.2%
	GARCH$_{sstd}$	0.2%	0.1%	4.3%	1.3%	**0.9%**	0.3%	0.5%	0.1%	5.4%	2.0%	1.4%	0.7%	1.3%	1%	7.0%	2.7%	2.6%	1.6%	4.0%	3.8%	8.1%	6.1%	4.7%	5.0%
	EGARCH$_{norm}$	0.6%	0.2%	5.3%	1.4%	2.7%	0.4%	0.9%	0.5%	7.2%	2.1%	3.8%	0.8%	2.5%	**1.7%**	8.3%	3.1%	**4.8%**	2.3%	5.6%	4.2%	11.5%	5.8%	7.3%	6.0%
	EGARCH$_{std}$	0.7%	**0.4%**	4.3%	1.2%	1.3%	0.6%	1.0%	**0.7%**	5.3%	1.7%	**2.5%**	0.9%	1.8%	1.1%	6.5%	**3.4%**	3.7%	1.8%	4.6%	3.8%	**9.1%**	6.2%	6.8%	4.1%
	EGARCH$_{sstd}$	**0.9%**	0.2%	3.5%	**1.1%**	2.5%	0.3%	**1.6%**	0.6%	4.4%	1.7%	3.7%	**1.1%**	**3.3%**	1.6%	6.4%	3.0%	4.7%	2.3%	6.2%	**5.8%**	8.6%	5.9%	7.1%	5.3%
	GAS$_{norm}$	7.0%	9.7%	**1.9%**	5.3%	1.4%	4.3%	7.9%	11.6%	**2.5%**	6.2%	1.5%	5.1%	8.7%	14.4%	2.8%	6.9%	1.6%	**5.9%**	**10.3%**	17.0%	3.7%	**8.1%**	1.9%	6.6%
	GAS$_{std}$	8.6%	13.8%	4.0%	8.6%	2.1%	5.6%	9.6%	15.9%	4.6%	9.9%	2.3%	6.5%	11.4%	19.9%	**5.0%**	11.5%	3.0%	7.8%	13.4%	24.8%	5.7%	13.5%	3.4%	**9.0%**
	GAS$_{sstd}$	0.5%	0.1%	5.2%	1.4%	3.7%	**0.7%**	0.7%	**0.7%**	6.3%	1.9%	4.5%	**1.1%**	1.6%	1.3%	7.4%	3.1%	6.2%	2.2%	3.4%	3.5%	8.4%	5.2%	**8.7%**	5.3%
MFE Test	GARCH$_{norm}$	0.4%	0.1%	3.2%	0.5%	1.7%	0.6%	1.1%	0.2%	4.3%	1.1%	3.2%	**1.1%**	2.3%	1.1%	6.0%	1.5%	4.5%	2.0%	5.4%	4.4%	8.5%	3.9%	7.3%	5.2%
	GARCH$_{std}$	0.1%	0.1%	3.4%	1.6%	2.4%	0.4%	0.4%	0.4%	4.7%	**2.7%**	3.4%	0.9%	1.4%	1.0%	6.3%	3.4%	4.5%	1.6%	3.6%	3.1%	8.5%	5.5%	6.9%	5.3%
	GARCH$_{sstd}$	0.2%	0.1%	4.4%	1.3%	**1.0%**	0.2%	0.6%	0.2%	5.4%	1.8%	1.3%	0.6%	1.4%	1.2%	7.0%	2.6%	2.6%	1.8%	4.2%	3.7%	8.0%	6.2%	5.0%	5.3%
	EGARCH$_{norm}$	0.6%	0.2%	5.4%	1.4%	2.7%	0.4%	0.9%	0.5%	7.2%	2.1%	3.8%	0.8%	2.2%	**1.7%**	8.4%	3.2%	**4.9%**	2.0%	5.6%	4.1%	11.5%	6.0%	7.3%	6.1%
	EGARCH$_{std}$	0.7%	**0.3%**	4.4%	1.2%	1.4%	**0.7%**	1.0%	**0.7%**	5.2%	1.8%	**2.5%**	0.9%	1.7%	1.0%	6.5%	**3.6%**	3.7%	1.8%	4.3%	3.6%	**8.9%**	6.5%	6.8%	3.8%
	EGARCH$_{sstd}$	**0.8%**	0.2%	3.4%	**1.1%**	2.3%	0.3%	**1.8%**	0.6%	4.5%	1.6%	3.6%	**1.1%**	**3.0%**	1.6%	6.3%	2.8%	4.8%	2.0%	6.1%	**6.1%**	8.4%	5.6%	7.2%	5.1%
	GAS$_{norm}$	7.0%	9.7%	**2.0%**	5.5%	1.4%	4.3%	7.9%	11.5%	**2.5%**	6.1%	1.5%	5.1%	8.8%	14.4%	2.8%	6.9%	1.6%	**5.9%**	**10.2%**	17.0%	3.7%	**8.1%**	1.9%	6.6%
	GAS$_{std}$	8.6%	13.8%	4.1%	8.6%	2.1%	5.6%	9.6%	16.0%	4.6%	9.8%	2.3%	6.5%	11.3%	20.1%	**5.0%**	11.5%	3.0%	7.8%	13.4%	24.8%	5.6%	13.4%	3.4%	**9.1%**
	GAS$_{sstd}$	0.5%	0.1%	5.2%	1.4%	3.9%	**0.7%**	0.7%	**0.7%**	6.3%	1.9%	4.3%	**1.1%**	1.5%	1.3%	7.3%	2.9%	6.4%	2.5%	3.5%	3.7%	8.5%	5.1%	**8.7%**	5.5%
RC Test	GARCH$_{norm}$	0.4%	0.1%	3.2%	0.6%	1.7%	0.6%	1.2%	0.2%	4.3%	1.1%	3.1%	**1.2%**	2.4%	1.3%	5.9%	1.5%	4.5%	2.0%	5.2%	4.3%	8.3%	4.2%	6.9%	5.2%
	GARCH$_{std}$	0.1%	0.1%	3.4%	1.6%	2.4%	0.5%	0.5%	0.5%	4.7%	**2.4%**	3.3%	1.1%	1.3%	1.5%	6.2%	**3.6%**	4.4%	1.7%	3.9%	3.6%	8.5%	6.1%	6.9%	5.3%
	GARCH$_{sstd}$	0.3%	0.1%	4.2%	1.3%	**1.0%**	0.2%	0.6%	0.2%	5.3%	1.9%	1.6%	0.7%	1.4%	1.1%	7.0%	2.7%	2.5%	1.7%	4.0%	3.9%	8.3%	6.3%	4.6%	5.0%
	EGARCH$_{norm}$	0.6%	0.2%	5.3%	1.4%	2.7%	0.4%	0.9%	0.5%	7.3%	2.1%	3.9%	0.8%	2.4%	**1.7%**	8.4%	3.2%	4.6%	2.0%	5.4%	4.1%	11.5%	6.0%	7.4%	6.0%
	EGARCH$_{std}$	0.6%	**0.3%**	4.5%	1.2%	1.3%	**0.7%**	1.0%	**0.7%**	5.2%	1.8%	2.2%	0.9%	2.1%	1.0%	6.5%	**3.6%**	3.6%	1.8%	4.9%	3.6%	**8.9%**	6.5%	6.7%	3.8%
	EGARCH$_{sstd}$	**0.8%**	0.2%	3.5%	**1.1%**	2.6%	0.3%	**1.7%**	0.6%	4.7%	1.6%	3.4%	1.1%	**3.2%**	1.6%	6.5%	2.8%	**5.0%**	2.0%	6.7%	**6.1%**	8.3%	5.6%	7.1%	5.1%
	GAS$_{norm}$	7.1%	9.7%	**1.9%**	5.5%	1.4%	4.3%	7.9%	11.5%	**2.5%**	6.1%	1.5%	5.1%	8.8%	14.4%	2.9%	6.9%	1.6%	**5.9%**	**10.3%**	17.0%	3.7%	**8.1%**	1.9%	6.6%
	GAS$_{std}$	8.6%	13.8%	4.2%	8.6%	2.1%	5.6%	9.6%	16.0%	4.6%	9.8%	**2.3%**	6.5%	11.3%	20.1%	**5.0%**	11.5%	3.1%	7.8%	13.5%	24.8%	5.7%	13.4%	3.4%	**9.1%**
	GAS$_{sstd}$	0.5%	0.1%	5.2%	1.4%	3.6%	**0.7%**	0.7%	**0.7%**	6.3%	1.9%	4.4%	1.1%	1.7%	1.3%	7.4%	2.9%	6.3%	2.5%	3.4%	3.7%	8.3%	5.1%	**8.7%**	5.5%

Table 5 – Scenario 2 - Size of the three backtesting procedures in percentage using $n = 1000$ and $T \in \{250, 500\}$

Backtesting	**Model**	1%						2.5%						5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH$_{norm}$	0.3%	0.2%	5.4%	**0.9%**	2.3%	**0.8%**	0.6%	**0.6%**	6.3%	1.5%	3.3%	**1.3%**	2.0%	1.2%	7.8%	2.6%	4.7%	3.0%	4.0%	3.8%	9.7%	4.4%	**8.2%**	5.7%
	GARCH$_{std}$	0.2%	0.0%	5.3%	**1.1%**	2.2%	0.3%	1.0%	0.1%	7.1%	1.8%	3.2%	0.7%	**2.7%**	0.4%	8.9%	3.8%	4.4%	1.7%	5.3%	3.0%	11.4%	6.1%	7.7%	4.1%
	GARCH$_{sstd}$	**0.5%**	0.0%	4.7%	1.4%	2.0%	0.7%	**1.1%**	0.1%	6.2%	**2.3%**	**2.9%**	1.0%	2.0%	0.8%	7.0%	3.9%	4.3%	2.5%	5.3%	3.0%	9.2%	6.9%	6.7%	4.7%
	EGARCH$_{norm}$	**0.5%**	**0.3%**	5.3%	2.2%	2.4%	0.4%	**1.1%**	0.4%	6.4%	3.2%	3.5%	0.7%	2.2%	**1.9%**	7.6%	**4.3%**	4.6%	1.7%	5.7%	**4.1%**	9.5%	6.9%	7.5%	5.0%
	EGARCH$_{std}$	0.3%	0.0%	4.6%	**1.1%**	2.3%	0.2%	1.0%	0.2%	5.9%	1.9%	3.4%	1.0%	2.1%	1.4%	7.5%	3.7%	**4.9%**	1.9%	5.0%	4.0%	**9.8%**	6.6%	6.9%	4.1%
	EGARCH$_{sstd}$	0.2%	0.1%	3.8%	**1.1%**	2.9%	0.4%	0.8%	**0.6%**	5.0%	1.8%	3.9%	0.9%	2.2%	1.1%	**6.1%**	2.5%	4.6%	2.1%	4.8%	3.2%	8.1%	4.6%	6.9%	4.8%
	GAS$_{norm}$	5.3%	9.3%	**2.3%**	5.4%	**1.3%**	4.2%	6.9%	11.1%	**2.8%**	6.6%	1.5%	4.8%	7.6%	13.2%	2.9%	7.1%	1.7%	**5.4%**	**8.7%**	17.2%	3.1%	**8.0%**	2.2%	6.3%
	GAS$_{std}$	9.6%	12.1%	6.2%	10.3%	2.2%	5.7%	11.3%	16.7%	7.2%	11.5%	3.1%	7.2%	13.8%	20.3%	7.4%	13.6%	3.2%	7.7%	16.0%	25.0%	7.9%	14.5%	4.1%	**9.1%**
	GAS$_{sstd}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MFE Test	GARCH$_{norm}$	0.3%	0.2%	5.4%	**0.9%**	2.3%	**0.8%**	0.7%	**0.6%**	6.4%	1.7%	3.2%	**1.3%**	2.0%	1.3%	7.7%	2.6%	4.7%	2.7%	4.2%	3.9%	9.6%	4.5%	**7.9%**	5.6%
	GARCH$_{std}$	0.2%	0.0%	5.5%	**1.1%**	2.2%	0.3%	1.0%	0.1%	6.8%	1.7%	3.2%	0.8%	**2.7%**	0.5%	9.0%	3.6%	4.5%	1.8%	5.1%	2.9%	11.4%	6.3%	7.3%	4.1%
	GARCH$_{sstd}$	**0.5%**	0.0%	4.8%	1.4%	2.0%	0.7%	**1.2%**	0.1%	6.2%	**2.4%**	**2.9%**	1.0%	2.0%	0.8%	7.0%	3.7%	4.3%	2.5%	5.2%	3.1%	9.4%	6.8%	6.4%	4.6%
	EGARCH$_{norm}$	**0.5%**	**0.3%**	5.5%	2.2%	2.4%	0.4%	1.0%	0.4%	6.3%	3.2%	3.5%	0.8%	2.2%	**1.8%**	7.6%	**4.2%**	4.5%	1.6%	5.7%	4.1%	9.5%	6.9%	7.4%	4.6%
	EGARCH$_{std}$	0.3%	0.0%	4.6%	**1.1%**	2.3%	0.2%	0.8%	0.1%	5.8%	2.0%	3.4%	1.0%	2.3%	1.2%	7.6%	3.6%	**5.0%**	1.8%	4.7%	**4.2%**	**9.8%**	6.4%	7.2%	3.9%
	EGARCH$_{sstd}$	0.2%	0.1%	3.9%	1.2%	2.9%	0.4%	1.0%	**0.6%**	5.2%	1.8%	3.8%	0.9%	1.9%	1.1%	**6.3%**	2.6%	4.8%	2.2%	4.6%	3.0%	8.2%	4.8%	6.5%	4.7%
	GAS$_{norm}$	5.6%	9.2%	**2.3%**	5.4%	**1.3%**	4.2%	6.8%	11.1%	**2.8%**	6.7%	1.5%	4.8%	7.6%	13.1%	2.9%	7.1%	1.7%	**5.4%**	**8.7%**	17.2%	3.1%	**8.0%**	2.2%	6.3%
	GAS$_{std}$	9.6%	12.1%	6.3%	10.4%	2.2%	5.7%	11.2%	16.5%	7.2%	11.5%	3.1%	6.9%	13.8%	20.4%	7.4%	13.6%	3.2%	7.7%	15.9%	24.6%	7.9%	14.5%	4.1%	**9.3%**
	GAS$_{sstd}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RC Test	GARCH$_{norm}$	0.4%	0.2%	5.4%	0.9%	2.3%	**0.9%**	0.7%	**0.7%**	6.3%	1.5%	3.3%	**1.3%**	2.1%	1.2%	7.7%	2.6%	4.7%	3.1%	4.0%	3.9%	9.6%	4.5%	**7.9%**	5.7%
	GARCH$_{std}$	0.2%	0.0%	5.4%	**1.0%**	2.2%	0.3%	0.8%	0.1%	7.0%	1.7%	3.2%	0.7%	**2.7%**	0.5%	9.0%	3.6%	4.5%	1.7%	5.4%	2.9%	11.4%	6.3%	7.6%	3.9%
	GARCH$_{sstd}$	0.4%	0.0%	4.7%	1.4%	2.0%	0.7%	**1.1%**	0.2%	6.2%	**2.3%**	**2.9%**	1.0%	2.0%	0.8%	7.0%	3.8%	4.2%	2.5%	5.5%	3.0%	9.2%	7.0%	6.7%	4.7%
	EGARCH$_{norm}$	**0.5%**	**0.3%**	5.2%	2.1%	2.2%	0.4%	**1.1%**	0.5%	6.4%	3.2%	3.4%	0.7%	2.1%	**1.8%**	7.8%	**4.3%**	4.6%	1.8%	5.7%	**4.1%**	9.8%	6.9%	7.3%	4.7%
	EGARCH$_{std}$	0.3%	0.0%	4.7%	1.2%	2.3%	0.2%	1.0%	0.2%	5.8%	2.1%	3.4%	0.9%	2.2%	1.1%	7.7%	3.6%	**5.0%**	2.1%	5.0%	4.0%	**9.9%**	6.1%	7.2%	4.1%
	EGARCH$_{sstd}$	0.2%	0.1%	3.9%	1.1%	2.7%	0.4%	1.0%	0.6%	5.2%	1.8%	3.9%	1.0%	2.1%	1.1%	**6.2%**	2.7%	4.7%	2.0%	4.8%	3.1%	8.3%	5.0%	6.6%	4.5%
	GAS$_{norm}$	5.6%	9.5%	**2.3%**	5.4%	**1.3%**	4.2%	6.9%	11.2%	**2.8%**	6.6%	1.5%	4.8%	7.6%	12.9%	2.9%	7.1%	1.7%	**5.4%**	**8.8%**	17.0%	3.1%	**7.9%**	2.2%	6.3%
	GAS$_{std}$	9.6%	12.1%	6.2%	10.2%	2.2%	5.7%	11.4%	16.5%	7.1%	11.5%	3.1%	6.9%	13.8%	20.0%	7.4%	13.6%	3.2%	7.7%	15.8%	24.9%	8.0%	14.6%	4.1%	**9.3%**
	GAS$_{sstd}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 6 – Proposed McNeil and Frey (2000) backtesting size results, using a significance level of 1%, 2.5% and 5% with 250 and 1000 observations.

Size	Model	1%						2.5%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
250 Observations	GARCH$_{norm}$	0.4%	0.1%	3.1%	0.4%	1.6%	0.5%	1.2%	0.2%	4.4%	1.0%	3.1%	**1.1%**
	GARCH$_{std}$	0.1%	0.1%	3.4%	1.5%	2.4%	0.5%	5.0%	0.4%	4.9%	**2.6%**	3.2%	1.0%
	GARCH$_{sstd}$	0.2%	0.1%	4.3%	1.3%	**0.9%**	0.3%	0.5%	0.1%	5.4%	2.0%	1.4%	0.7%
	EGARCH$_{norm}$	0.6%	0.2%	5.3%	1.4%	2.7%	0.4%	0.9%	0.5%	7.2%	2.1%	3.8%	0.8%
	EGARCH$_{std}$	0.7%	**0.4%**	4.3%	1.2%	1.3%	0.6%	1.0%	**0.7%**	5.3%	1.7%	**2.5%**	0.9%
	EGARCH$_{sstd}$	**0.9%**	0.2%	3.5%	**1.1%**	2.5%	0.3%	**1.6%**	0.6%	4.4%	1.7%	3.7%	**1.1%**
	GAS$_{norm}$	7.0%	9.7%	**1.9%**	5.3%	1.4%	4.3%	7.9%	11.6%	**2.5%**	6.2%	1.5%	5.1%
	GAS$_{std}$	8.6%	13.8%	4.0%	8.6%	2.1%	5.6%	9.6%	15.9%	4.6%	9.9%	2.3%	6.5%
	GAS$_{sstd}$	0.5%	0.1%	5.2%	1.4%	3.7%	**0.7%**	0.7%	**0.7%**	6.3%	1.9%	4.5%	**1.1%**
1000 Observations	GARCH$_{norm}$	0.3%	0.2%	5.4%	**0.9%**	2.3%	**0.8%**	0.6%	**0.6%**	6.3%	1.5%	3.3%	**1.3%**
	GARCH$_{std}$	0.2%	0.0%	5.3%	**1.1%**	2.2%	0.3%	1.0%	0.1%	7.1%	1.8%	3.2%	0.7%
	GARCH$_{sstd}$	**0.5%**	0.0%	4.7%	1.4%	2.0%	0.7%	**1.1%**	0.1%	6.2%	**2.3%**	**2.9%**	1.0%
	EGARCH$_{norm}$	**0.5%**	**0.3%**	5.3%	2.2%	2.4%	0.4%	**1.1%**	0.4%	6.4%	3.2%	3.5%	0.7%
	EGARCH$_{std}$	0.3%	0.0%	4.6%	**1.1%**	2.3%	0.2%	1.0%	0.2%	5.9%	1.9%	3.4%	1.0%
	EGARCH$_{sstd}$	0.2%	0.1%	3.8%	**1.1%**	2.9%	0.4%	0.8%	**0.6%**	5.0%	1.8%	3.9%	0.9%
	GAS$_{norm}$	5.3%	9.3%	**2.3%**	5.4%	**1.3%**	4.2%	6.9%	11.1%	**2.8%**	6.6%	1.5%	4.8%
	GAS$_{std}$	9.6%	12.1%	6.2%	10.3%	2.2%	5.7%	11.3%	16.7%	7.2%	11.5%	3.1%	7.2%
	GAS$_{sstd}$	-	-	-	-	-	-	-	-	-	-	-	-

Size	Model	5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
250 Observations	GARCH$_{norm}$	2.4%	0.7%	6.0%	1.7%	4.4%	2.0%	5.4%	4.0%	8.7%	4.0%	7.1%	5.2%
	GARCH$_{std}$	1.1%	1.3%	6.4%	3.2%	4.5%	1.7%	3.8%	3.1%	8.4%	5.9%	7.1%	5.2%
	GARCH$_{sstd}$	1.3%	1%	7.0%	2.7%	2.6%	1.6%	4.0%	3.8%	8.1%	6.1%	4.7%	5.0%
	EGARCH$_{norm}$	2.5%	**1.7%**	8.3%	3.1%	**4.8%**	2.3%	5.6%	4.2%	11.5%	5.8%	7.3%	6.0%
	EGARCH$_{std}$	1.8%	1.1%	6.5%	**3.4%**	3.7%	1.8%	4.6%	3.8%	**9.1%**	6.2%	6.8%	4.1%
	EGARCH$_{sstd}$	**3.3%**	1.6%	6.4%	3.0%	4.7%	2.3%	6.2%	**5.8%**	8.6%	5.9%	7.1%	5.3%
	GAS$_{norm}$	8.7%	14.4%	2.8%	6.9%	1.6%	**5.9%**	**10.3%**	17.0%	3.7%	**8.1%**	1.9%	6.6%
	GAS$_{std}$	11.4%	19.9%	**5.0%**	11.5%	3.0%	7.8%	13.4%	24.8%	5.7%	13.5%	3.4%	**9.0%**
	GAS$_{sstd}$	1.6%	1.3%	7.4%	3.1%	6.2%	2.2%	3.4%	3.5%	8.4%	5.2%	**8.7%**	5.3%
1000 Observations	GARCH$_{norm}$	2.0%	1.2%	7.8%	2.6%	4.7%	3.0%	4.0%	3.8%	9.7%	4.4%	**8.2%**	5.7%
	GARCH$_{std}$	**2.7%**	0.4%	8.9%	3.8%	4.4%	1.7%	5.3%	3.0%	11.4%	6.1%	7.7%	4.1%
	GARCH$_{sstd}$	2.0%	0.8%	7.0%	3.9%	4.3%	2.5%	5.3%	3.0%	9.2%	6.9%	6.7%	4.7%
	EGARCH$_{norm}$	2.2%	**1.9%**	7.6%	**4.3%**	4.6%	1.7%	5.7%	**4.1%**	9.5%	6.9%	7.5%	5.0%
	EGARCH$_{std}$	2.1%	1.4%	7.5%	3.7%	**4.9%**	1.9%	5.0%	4.0%	**9.8%**	6.6%	6.9%	4.1%
	EGARCH$_{sstd}$	2.2%	1.1%	**6.1%**	2.5%	4.6%	2.1%	4.8%	3.2%	8.1%	4.6%	6.9%	4.8%
	GAS$_{norm}$	7.6%	13.2%	2.9%	7.1%	1.7%	**5.4%**	**8.7%**	17.2%	3.1%	**8.0%**	2.2%	6.3%
	GAS$_{std}$	13.8%	20.3%	7.4%	13.6%	3.2%	7.7%	16.0%	25.0%	7.9%	14.5%	4.1%	**9.1%**
	GAS$_{sstd}$	-	-	-	-	-	-	-	-	-	-	-	-

Table 7 – Proposed Righi and Ceretta (2015) backtesting size results, using a significance level of 1%, 2.5% and 5% with 250 and 1000 observations.

Size	Model	1%						2.5%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
250 Observations	$GARCH_{norm}$	0.4%	0.1%	3.2%	0.6%	1.7%	0.6%	1.2%	0.2%	4.3%	1.1%	3.1%	**1.2%**
	$GARCH_{std}$	0.1%	0.1%	3.4%	1.6%	2.4%	0.5%	0.5%	0.5%	4.7%	**2.4%**	3.3%	1.1%
	$GARCH_{sstd}$	0.3%	0.1%	4.2%	1.3%	**1.0%**	0.2%	0.6%	0.2%	5.3%	1.9%	1.6%	0.7%
	$EGARCH_{norm}$	0.6%	0.2%	5.3%	1.4%	2.7%	0.4%	0.9%	0.5%	7.3%	2.1%	3.9%	0.8%
	$EGARCH_{std}$	0.6%	**0.3%**	4.5%	1.2%	1.3%	**0.7%**	1.0%	**0.7%**	5.2%	1.8%	2.2%	0.9%
	$EGARCH_{sstd}$	**0.8%**	0.2%	3.5%	**1.1%**	2.6%	0.3%	**1.7%**	0.6%	4.7%	1.6%	3.4%	1.1%
	GAS_{norm}	7.1%	9.7%	**1.9%**	5.5%	1.4%	4.3%	7.9%	11.5%	**2.5%**	6.1%	1.5%	5.1%
	GAS_{std}	8.6%	13.8%	4.2%	8.6%	2.1%	5.6%	9.6%	16.0%	4.6%	9.8%	**2.3%**	6.5%
	GAS_{sstd}	0.5%	0.1%	5.2%	1.4%	3.6%	**0.7%**	0.7%	**0.7%**	6.3%	1.9%	4.4%	1.1%
1000 Observations	$GARCH_{norm}$	0.4%	0.2%	5.4%	0.9%	2.3%	**0.9%**	0.7%	**0.7%**	6.3%	1.5%	3.3%	**1.3%**
	$GARCH_{std}$	0.2%	0.0%	5.4%	**1.0%**	2.2%	0.3%	0.8%	0.1%	7.0%	1.7%	3.2%	0.7%
	$GARCH_{sstd}$	0.4%	0.0%	4.7%	1.4%	2.0%	0.7%	**1.1%**	0.2%	6.2%	**2.3%**	**2.9%**	1.0%
	$EGARCH_{norm}$	**0.5%**	**0.3%**	5.2%	2.1%	2.2%	0.4%	**1.1%**	0.5%	6.4%	3.2%	3.4%	0.7%
	$EGARCH_{std}$	0.3%	0.0%	4.7%	1.2%	2.3%	0.2%	1.0%	0.2%	5.8%	2.1%	3.4%	0.9%
	$EGARCH_{sstd}$	0.2%	0.1%	3.9%	1.1%	2.7%	0.4%	1.0%	0.6%	5.2%	1.8%	3.9%	1.0%
	GAS_{norm}	5.6%	9.5%	**2.3%**	5.4%	**1.3%**	4.2%	6.9%	11.2%	**2.8%**	6.6%	1.5%	4.8%
	GAS_{std}	9.6%	12.1%	6.2%	10.2%	2.2%	5.7%	11.4%	16.5%	7.1%	11.5%	3.1%	6.9%
	GAS_{sstd}	-	-	-	-	-	-	-	-	-	-	-	-

Size	Model	5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
250 Observations	$GARCH_{norm}$	2.4%	1.3%	5.9%	1.5%	4.5%	2.0%	5.2%	4.3%	8.3%	4.2%	6.9%	5.2%
	$GARCH_{std}$	1.3%	1.5%	6.2%	**3.6%**	4.4%	1.7%	3.9%	3.6%	8.5%	6.1%	6.9%	5.3%
	$GARCH_{sstd}$	1.4%	1.1%	7.0%	2.7%	2.5%	1.7%	4.0%	3.9%	8.3%	6.3%	4.6%	5.0%
	$EGARCH_{norm}$	2.4%	**1.7%**	8.4%	3.2%	4.6%	2.0%	5.4%	4.1%	11.5%	6.0%	7.4%	6.0%
	$EGARCH_{std}$	2.1%	1.0%	6.5%	**3.6%**	3.6%	1.8%	4.9%	3.6%	**8.9%**	6.5%	6.7%	3.8%
	$EGARCH_{sstd}$	**3.2%**	1.6%	6.5%	2.8%	**5.0%**	2.0%	6.7%	**6.1%**	8.3%	5.6%	7.1%	5.1%
	GAS_{norm}	8.8%	14.4%	2.9%	6.9%	1.6%	**5.9%**	**10.3%**	17.0%	3.7%	**8.1%**	1.9%	6.6%
	GAS_{std}	11.3%	20.1%	**5.0%**	11.5%	3.1%	7.8%	13.5%	24.8%	5.7%	13.4%	3.4%	**9.1%**
	GAS_{sstd}	1.7%	1.3%	7.4%	2.9%	6.3%	2.5%	3.4%	3.7%	8.3%	5.1%	**8.7%**	5.5%
1000 Observations	$GARCH_{norm}$	2.1%	1.2%	7.7%	2.6%	4.7%	3.1%	4.0%	3.9%	9.6%	4.5%	**7.9%**	5.7%
	$GARCH_{std}$	**2.7%**	0.5%	9.0%	3.6%	4.5%	1.7%	5.4%	2.9%	11.4%	6.3%	7.6%	3.9%
	$GARCH_{sstd}$	2.0%	0.8%	7.0%	3.8%	4.2%	2.5%	5.5%	3.0%	9.2%	7.0%	6.7%	4.7%
	$EGARCH_{norm}$	2.1%	**1.8%**	7.8%	**4.3%**	4.6%	1.8%	5.7%	**4.1%**	9.8%	6.9%	7.3%	4.7%
	$EGARCH_{std}$	2.2%	1.1%	7.7%	3.6%	**5.0%**	2.1%	5.0%	4.0%	**9.9%**	6.1%	7.2%	4.1%
	$EGARCH_{sstd}$	2.1%	1.1%	**6.2%**	2.7%	4.7%	2.0%	4.8%	3.1%	8.3%	5.0%	6.6%	4.5%
	GAS_{norm}	7.6%	12.9%	2.9%	7.1%	1.7%	**5.4%**	**8.8%**	17.0%	3.1%	**7.9%**	2.2%	6.3%
	GAS_{std}	13.8%	20.0%	7.4%	13.6%	3.2%	7.7%	15.8%	24.9%	8.0%	14.6%	4.1%	**9.3%**
	GAS_{sstd}	-	-	-	-	-	-	-	-	-	-	-	-

Table 8 – Proposed McNeil et al. (2015) backtesting size results, using a significance level of 1%, 2.5% and 5% with 250 and 1000 observations.

Size	Model	1%						2.5%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
250 Observations	$GARCH_{norm}$	0.4%	0.1%	3.2%	0.5%	1.7%	0.6%	1.1%	0.2%	4.3%	1.1%	3.2%	**1.1%**
	$GARCH_{std}$	0.1%	0.1%	3.4%	1.6%	2.4%	0.4%	0.4%	0.4%	4.7%	**2.7%**	3.4%	0.9%
	$GARCH_{sstd}$	0.2%	0.1%	4.4%	1.3%	**1.0%**	0.2%	0.6%	0.2%	5.4%	1.8%	1.3%	0.6%
	$EGARCH_{norm}$	0.6%	0.2%	5.4%	1.4%	2.7%	0.4%	0.9%	0.5%	7.2%	2.1%	3.8%	0.8%
	$EGARCH_{std}$	0.7%	**0.3%**	4.4%	1.2%	1.4%	**0.7%**	1.0%	**0.7%**	5.2%	1.8%	**2.5%**	0.9%
	$EGARCH_{sstd}$	**0.8%**	0.2%	3.4%	**1.1%**	2.3%	0.3%	**1.8%**	0.6%	4.5%	1.6%	3.6%	**1.1%**
	GAS_{norm}	7.0%	9.7%	**2.0%**	5.5%	1.4%	4.3%	7.9%	11.5%	**2.5%**	6.1%	1.5%	5.1%
	GAS_{std}	8.6%	13.8%	4.1%	8.6%	2.1%	5.6%	9.6%	16.0%	4.6%	9.8%	2.3%	6.5%
	GAS_{sstd}	0.5%	0.1%	5.2%	1.4%	3.9%	**0.7%**	0.7%	**0.7%**	6.3%	1.9%	4.3%	**1.1%**
1000 Observations	$GARCH_{norm}$	0.3%	0.2%	5.4%	**0.9%**	2.3%	**0.8%**	0.7%	**0.6%**	6.4%	1.7%	3.2%	**1.3%**
	$GARCH_{std}$	0.2%	0.0%	5.5%	**1.1%**	2.2%	0.3%	1.0%	0.1%	6.8%	1.7%	3.2%	0.8%
	$GARCH_{sstd}$	**0.5%**	0.0%	4.8%	1.4%	2.0%	0.7%	**1.2%**	0.1%	6.2%	**2.4%**	**2.9%**	1.0%
	$EGARCH_{norm}$	**0.5%**	**0.3%**	5.5%	2.2%	2.4%	0.4%	1.0%	0.4%	6.3%	3.2%	3.5%	0.8%
	$EGARCH_{std}$	0.3%	0.0%	4.6%	**1.1%**	2.3%	0.2%	0.8%	0.1%	5.8%	2.0%	3.4%	1.0%
	$EGARCH_{sstd}$	0.2%	0.1%	3.9%	1.2%	2.9%	0.4%	1.0%	**0.6%**	5.2%	1.8%	3.8%	0.9%
	GAS_{norm}	5.6%	9.2%	**2.3%**	5.4%	**1.3%**	4.2%	6.8%	11.1%	**2.8%**	6.7%	1.5%	4.8%
	GAS_{std}	9.6%	12.1%	6.3%	10.4%	2.2%	5.7%	11.2%	16.5%	7.2%	11.5%	3.1%	6.9%
	GAS_{sstd}	-	-	-	-	-	-	-	-	-	-	-	-

Size	Model	5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
250 Observations	$GARCH_{norm}$	2.3%	1.1%	6.0%	1.5%	4.5%	2.0%	5.4%	4.4%	8.5%	3.9%	7.3%	5.2%
	$GARCH_{std}$	1.4%	1.0%	6.3%	3.4%	4.5%	1.6%	3.6%	3.1%	8.5%	5.5%	6.9%	5.3%
	$GARCH_{sstd}$	1.4%	1.2%	7.0%	2.6%	2.6%	1.8%	4.2%	3.7%	8.0%	6.2%	5.0%	5.3%
	$EGARCH_{norm}$	2.2%	**1.7%**	8.4%	3.2%	**4.9%**	2.0%	5.6%	4.1%	11.5%	6.0%	7.3%	6.1%
	$EGARCH_{std}$	1.7%	1.0%	6.5%	**3.6%**	3.7%	1.8%	4.3%	3.6%	**8.9%**	6.5%	6.8%	3.8%
	$EGARCH_{sstd}$	**3.0%**	1.6%	6.3%	2.8%	4.8%	2.0%	6.1%	**6.1%**	8.4%	5.6%	7.2%	5.1%
	GAS_{norm}	8.8%	14.4%	2.8%	6.9%	1.6%	**5.9%**	**10.2%**	17.0%	3.7%	**8.1%**	1.9%	6.6%
	GAS_{std}	11.3%	20.1%	**5.0%**	11.5%	3.0%	7.8%	13.4%	24.8%	5.6%	13.4%	3.4%	**9.1%**
	GAS_{sstd}	1.5%	1.3%	7.3%	2.9%	6.4%	2.5%	3.5%	3.7%	8.5%	5.1%	**8.7%**	5.5%
1000 Observations	$GARCH_{norm}$	2.0%	1.3%	7.7%	2.6%	4.7%	2.7%	4.2%	3.9%	9.6%	4.5%	**7.9%**	5.6%
	$GARCH_{std}$	**2.7%**	0.5%	9.0%	3.6%	4.5%	1.8%	5.1%	2.9%	11.4%	6.3%	7.3%	4.1%
	$GARCH_{sstd}$	2.0%	0.8%	7.0%	3.7%	4.3%	2.5%	5.2%	3.1%	9.4%	6.8%	6.4%	4.6%
	$EGARCH_{norm}$	2.2%	**1.8%**	7.6%	**4.2%**	4.5%	1.6%	5.7%	4.1%	9.5%	6.9%	7.4%	4.6%
	$EGARCH_{std}$	2.3%	1.2%	7.6%	3.6%	**5.0%**	1.8%	4.7%	**4.2%**	**9.8%**	6.4%	7.2%	3.9%
	$EGARCH_{sstd}$	1.9%	1.1%	**6.3%**	2.6%	4.8%	2.2%	4.6%	3.0%	8.2%	4.8%	6.5%	4.7%
	GAS_{norm}	7.6%	13.1%	2.9%	7.1%	1.7%	**5.4%**	**8.7%**	17.2%	3.1%	**8.0%**	2.2%	6.3%
	GAS_{std}	13.8%	20.4%	7.4%	13.6%	3.2%	7.7%	15.9%	24.6%	7.9%	14.5%	4.1%	**9.3%**
	GAS_{sstd}	-	-	-	-	-	-	-	-	-	-	-	-

Table 9 – Summary of the best performing models for all approaches, $n \in \{250, 1000\}$ and $T \in \{250, 500\}$, using significance levels of 1%, 2.5%, 5%

Backtesting	**Size**	1%						2.5%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	**250 Observations**	EGARCH_{sstd}	EGARCH_{std}	GAS_{norm}	EGARCH_{sstd}	GARCH_{sstd}	GAS_{sstd}	EGARCH_{sstd}	EGARCH_{std} GAS_{sstd}	GAS_{norm}	GARCH_{std}	EGARCH_{std}	GARCH_{norm} EGARCH_{sstd} GAS_{sstd}
	1000 Observations	GARCH_{sstd} EGARCH_{norm}	EGARCH_{norm}	GAS_{norm}	GARCH_{norm} GARCH_{std} EGARCH_{std} EGARCH_{sstd}	GAS_{norm}	GARCH_{norm}	GARCH_{sstd} EGARCH_{norm}	GARCH_{norm} EGARCH_{sstd}	GAS_{norm}	GARCH_{sstd}	GARCH_{sstd}	GARCH_{norm}
MFE Test	**250 Observations**	EGARCH_{sstd}	EGARCH_{std}	GAS_{norm}	EGARCH_{sstd}	GARCH_{sstd}	EGARCH_{std} GAS_{sstd}	EGARCH_{sstd}	EGARCH_{std} GAS_{sstd}	GAS_{norm}	GARCH_{std}	EGARCH_{std}	GARCH_{norm} EGARCH_{sstd} GAS_{sstd}
	1000 Observations	GARCH_{sstd} EGARCH_{norm}	EGARCH_{norm}	GAS_{norm}	GARCH_{norm} GARCH_{std} EGARCH_{std}	GAS_{norm}	GARCH_{norm}	GARCH_{sstd}	GARCH_{norm} EGARCH_{sstd}	GAS_{norm}	GARCH_{sstd}	GARCH_{sstd}	GARCH_{norm}
RC Test	**250 Observations**	EGARCH_{sstd}	EGARCH_{std}	GAS_{norm}	EGARCH_{sstd}	GARCH_{sstd}	EGARCH_{std} GAS_{sstd}	EGARCH_{sstd}	EGARCH_{std} GAS_{sstd}	GAS_{norm}	GARCH_{std}	GAS_{std}	GARCH_{norm}
	1000 Observations	EGARCH_{norm}	EGARCH_{norm}	GAS_{norm}	GARCH_{std}	GAS_{norm}	GARCH_{norm}	GARCH_{sstd} $\text{EGARCH}norm$	GARCH_{norm}	GAS_{norm}	GARCH_{sstd}	GARCH_{sstd}	GARCH_{norm}

Backtesting	**Size**	5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	**250 Observations**	EGARCH_{sstd}	EGARCH_{norm}	GAS_{std}	EGARCH_{std}	EGARCH_{norm}	GAS_{norm}	GAS_{norm}	EGARCH_{sstd}	EGARCH_{std}	GAS_{norm}	GAS_{sstd}	GAS_{std}
	1000 Observations	GARCH_{std}	EGARCH_{norm}	EGARCH_{sstd}	EGARCH_{norm}	EGARCH_{std}	GAS_{norm}	GAS_{norm}	EGARCH_{norm}	EGARCH_{std}	GAS_{norm}	GARCH_{norm}	GAS_{std}
MFE Test	**250 Observations**	EGARCH_{sstd}	EGARCH_{norm}	GAS_{std}	EGARCH_{std}	EGARCH_{norm}	GAS_{norm}	GAS_{norm}	EGARCH_{sstd}	EGARCH_{std}	GAS_{norm}	GAS_{sstd}	GAS_{std}
	1000 Observations	GARCH_{std}	EGARCH_{norm}	EGARCH_{sstd}	EGARCH_{norm}	EGARCH_{std}	GAS_{norm}	GAS_{norm}	EGARCH_{std}	EGARCH_{std}	GAS_{norm}	GARCH_{norm}	GAS_{std}
RC Test	**250 Observations**	EGARCH_{sstd}	EGARCH_{norm}	GAS_{std}	GARCH_{std} EGARCH_{std}	EGARCH_{sstd}	GAS_{norm}	GAS_{norm}	EGARCH_{sstd}	EGARCH_{std}	GAS_{norm}	GAS_{sstd}	GAS_{std}
	1000 Observations	GARCH_{std}	EGARCH_{norm}	EGARCH_{sstd}	EGARCH_{norm}	EGARCH_{std}	GAS_{norm}	GAS_{norm}	EGARCH_{norm}	EGARCH_{std}	GAS_{norm}	GARCH_{norm}	GAS_{std}

Table 10 – Scenario 1 - Power of the three backtesting procedures in percentage, using $GARCH_{norm}$ as null hypothesis, $n = 250$ and $T \in \{250, 500\}$

Backtesting	Model	1%						2.5%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	$GARCH_{std}$	**0.6%**	0.0%	**4.7%**	**1.9%**	**3.0%**	**2.0%**	1.1%	0.3%	**5.8%**	**3.4%**	3.6%	**1.0%**
	$GARCH_{sstd}$	**0.6%**	0.0%	4.0%	1.2%	2.3%	0.3%	1.3%	0.5%	5.1%	2.0%	3.4%	0.6%
	$EGARCH_{norm}$	0.4%	**0.3%**	4.3%	1.7%	2.3%	0.3%	0.9%	**0.7%**	5.0%	2.4%	3.2%	0.8%
	$EGARCH_{std}$	0.5%	0.2%	3.8%	1.6%	**3.0%**	0.4%	**1.4%**	0.4%	5.1%	2.0%	**3.7%**	0.9%
	$EGARCH_{sstd}$	0.2%	0.0%	4.6%	1.6%	2.7%	0.4%	1.0%	0.4%	5.6%	2.2%	3.5%	0.8%
MFE Test	$GARCH_{std}$	**0.7%**	0.0%	**4.7%**	**1.9%**	**3.0%**	**2.0%**	1.1%	0.3%	**5.8%**	**3.4%**	3.5%	**0.9%**
	$GARCH_{sstd}$	0.5%	0.0%	3.9%	1.2%	2.3%	0.1%	1.3%	0.5%	5.0%	2.0%	3.4%	0.5%
	$EGARCH_{norm}$	0.5%	**0.3%**	4.3%	1.7%	2.3%	0.3%	0.9%	**0.7%**	5.0%	2.5%	3.3%	0.7%
	$EGARCH_{std}$	0.4%	0.2%	3.9%	1.5%	3.0%	0.4%	**1.5%**	0.4%	5.1%	2.0%	**3.7%**	**0.9%**
	$EGARCH_{sstd}$	0.2%	0.0%	4.6%	1.5%	2.7%	0.4%	0.8%	0.4%	5.6%	2.3%	3.5%	0.8%
RC Test	$GARCH_{std}$	**0.6%**	0.0%	**4.6%**	**1.8%**	2.9%	**2.0%**	1.1%	0.2%	**5.8%**	**3.3%**	3.5%	**1.0%**
	$GARCH_{sstd}$	**0.6%**	0.0%	3.9%	1.2%	2.3%	0.2%	1.3%	0.5%	5.2%	2.0%	3.3%	0.7%
	$EGARCH_{norm}$	0.5%	**0.3%**	4.3%	**1.8%**	2.2%	0.3%	0.8%	**0.7%**	5.0%	2.3%	3.1%	0.9%
	$EGARCH_{std}$	0.4%	**0.3%**	3.8%	1.5%	**3.0%**	0.4%	**1.5%**	0.4%	5.2%	2.0%	**3.8%**	0.9%
	$EGARCH_{sstd}$	0.2%	0.0%	4.5%	1.5%	2.7%	0.5%	1.0%	0.4%	5.6%	2.4%	3.5%	1.0%

Backtesting	Model	5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	$GARCH_{std}$	**2.8%**	1.5%	**7.7%**	**4.6%**	**5.1%**	**2.3%**	5.2%	3.7%	9.7%	**6.9%**	**8.3%**	**6.1%**
	$GARCH_{sstd}$	2.1%	**1.6%**	7.5%	3.0%	4.4%	1.8%	4.9%	4.1%	**10.2%**	5.8%	7.1%	5.2%
	$EGARCH_{norm}$	2.0%	1.5%	6.4%	3.9%	4.3%	1.9%	**5.3%**	4.3%	8.6%	6.7%	7.1%	4.6%
	$EGARCH_{std}$	2.1%	1.3%	6.6%	2.8%	4.7%	2.1%	**5.3%**	3.7%	8.5%	6.8%	7.8%	5.1%
	$EGARCH_{sstd}$	2.0%	1.4%	6.6%	3.4%	4.5%	1.5%	4.9%	**5.0%**	9.5%	6.1%	7.7%	4.7%
MFE Test	$GARCH_{std}$	**2.8%**	**1.6%**	**7.7%**	**4.5%**	**5.1%**	**2.4%**	5.4%	3.8%	9.8%	**7.1%**	**8.4%**	**6.0%**
	$GARCH_{sstd}$	2.1%	**1.6%**	7.4%	3.0%	4.5%	1.8%	4.7%	3.7%	**10.1%**	6.0%	7.2%	5.0%
	$EGARCH_{norm}$	2.1%	1.4%	6.7%	3.8%	4.3%	1.9%	**5.5%**	4.4%	8.4%	6.7%	6.9%	4.6%
	$EGARCH_{std}$	2.0%	1.3%	6.4%	3.0%	4.7%	2.0%	5.3%	3.5%	8.8%	6.6%	7.7%	5.0%
	$EGARCH_{sstd}$	2.1%	1.4%	6.7%	3.2%	4.6%	1.5%	5.2%	**4.8%**	9.6%	6.5%	7.6%	4.4%
RC Test	$GARCH_{std}$	**2.8%**	1.6%	**7.7%**	**4.6%**	**5.0%**	**2.5%**	5.4%	3.7%	9.6%	**7.1%**	**8.5%**	**6.2%**
	$GARCH_{sstd}$	2.2%	**1.7%**	7.3%	2.9%	4.5%	1.6%	4.6%	3.8%	**10.1%**	6.0%	7.3%	5.1%
	$EGARCH_{norm}$	1.9%	1.5%	6.5%	3.9%	4.4%	1.8%	**5.5%**	4.3%	8.7%	6.8%	7.2%	4.5%
	$EGARCH_{std}$	1.8%	1.2%	6.6%	2.9%	4.7%	2.0%	5.4%	3.7%	8.5%	6.9%	7.7%	4.8%
	$EGARCH_{sstd}$	2.2%	1.4%	6.7%	3.6%	4.4%	1.5%	5.0%	**4.9%**	9.7%	6.7%	7.7%	4.8%

Table 11 – Scenario 1 - Power of the three backtesting procedures in percentage, using GARCH$_{std}$ as null hypothesis, $n = 250$ and $T \in \{250, 500\}$

Backtesting	Model	1%						2.5%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH$_{norm}$	**2.2%**	**0.6%**	6.3%	**2.9%**	**3.9%**	0.5%	**3.1%**	**1.5%**	**8.1%**	**4.1%**	**5.4%**	1.0%
	GARCH$_{sstd}$	0.5%	0.1%	4.6%	1.2%	2.6%	0.0%	0.7%	0.2%	5.8%	1.8%	3.6%	0.5%
	EGARCH$_{norm}$	1.1%	0.2%	**7.0%**	2.2%	3.4%	**0.6%**	2.1%	1.0%	**8.1%**	3.6%	4.5%	**1.7%**
	EGARCH$_{std}$	0.6%	0.2%	4.3%	0.6%	2.4%	0.4%	1.1%	0.7%	5.7%	1.2%	3.4%	0.9%
	EGARCH$_{sstd}$	0.0%	0.2%	3.2%	0.7%	2.3%	0.5%	0.7%	0.4%	4.3%	1.7%	2.9%	1.0%
MFE Test	GARCH$_{norm}$	**2.2%**	**0.6%**	6.3%	**2.9%**	**4.1%**	0.4%	**3.2%**	**1.4%**	**8.1%**	**4.1%**	**5.4%**	1.0%
	GARCH$_{sstd}$	0.5%	0.1%	4.7%	1.3%	2.6%	0.0%	0.7%	0.5%	5.7%	1.9%	3.7%	0.4%
	EGARCH$_{norm}$	1.1%	0.2%	**6.8%**	2.2%	3.5%	**0.7%**	2.3%	1.2%	**8.1%**	3.5%	4.4%	**1.6%**
	EGARCH$_{std}$	0.6%	0.3%	4.4%	0.7%	2.4%	0.5%	1.1%	0.7%	5.7%	1.3%	3.4%	0.9%
	EGARCH$_{sstd}$	0.0%	0.2%	3.2%	0.7%	2.3%	0.5%	0.8%	0.5%	4.2%	1.8%	3.0%	1.1%
RC Test	GARCH$_{norm}$	**2.2%**	**0.6%**	6.3%	**2.9%**	**4.0%**	0.5%	**3.2%**	**1.3%**	**8.1%**	**4.4%**	**5.4%**	1.0%
	GARCH$_{sstd}$	0.5%	0.1%	5.0%	1.2%	2.7%	0.0%	0.7%	0.6%	5.8%	2.0%	3.8%	0.4%
	EGARCH$_{norm}$	1.1%	0.2%	**7.0%**	2.2%	3.4%	**0.7%**	2.2%	1.1%	8.0%	3.4%	4.6%	**1.6%**
	EGARCH$_{std}$	0.5%	0.2%	4.3%	0.8%	2.4%	0.4%	1.2%	0.8%	5.4%	1.2%	3.4%	0.9%
	EGARCH$_{sstd}$	0.1%	0.2%	3.0%	0.7%	2.3%	0.5%	0.7%	0.5%	4.0%	1.7%	3.0%	0.9%

Backtesting	Model	5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH$_{norm}$	**5.0%**	**3.5%**	**9.7%**	**5.7%**	**7.1%**	1.9%	**8.4%**	6.8%	**12.0%**	8.5%	9.3%	5.8%
	GARCH$_{sstd}$	1.6%	1.6%	7.1%	2.7%	4.5%	1.3%	4.8%	4.5%	9.7%	5.8%	7.3%	4.3%
	EGARCH$_{norm}$	3.7%	2.6%	9.6%	5.2%	**7.1%**	**3.1%**	7.2%	**7.7%**	11.7%	**9.3%**	**9.9%**	**6.5%**
	EGARCH$_{std}$	1.8%	2.0%	6.9%	1.9%	4.5%	2.1%	5.3%	5.8%	8.6%	4.9%	7.0%	5.8%
	EGARCH$_{sstd}$	1.3%	1.1%	5.2%	2.7%	4.4%	2.3%	4.2%	3.5%	8.3%	5.1%	8.0%	4.4%
MFE Test	GARCH$_{norm}$	**5.0%**	**3.5%**	**9.7%**	**5.8%**	**7.1%**	2.0%	**8.5%**	6.9%	**12.2%**	8.7%	9.3%	5.8%
	GARCH$_{sstd}$	1.7%	1.5%	6.9%	2.7%	4.6%	1.4%	5.3%	4.4%	9.7%	5.6%	7.5%	4.4%
	EGARCH$_{norm}$	3.6%	2.8%	9.7%	5.4%	7.0%	**3.1%**	7.2%	**7.9%**	11.7%	**9.1%**	**9.9%**	**6.6%**
	EGARCH$_{std}$	1.8%	1.9%	7.1%	2.0%	4.4%	2.0%	5.4%	5.9%	8.7%	5.2%	7.0%	6.2%
	EGARCH$_{sstd}$	1.1%	1.1%	5.1%	2.9%	4.3%	2.1%	3.8%	3.3%	8.2%	5.0%	8.0%	4.5%
RC Test	GARCH$_{norm}$	**5.0%**	**3.4%**	**9.8%**	**6.1%**	**7.1%**	2.0%	**8.3%**	6.7%	**12.0%**	8.5%	9.4%	5.8%
	GARCH$_{sstd}$	2.0%	1.4%	7.0%	2.9%	4.5%	1.7%	5.5%	4.5%	9.8%	5.9%	7.4%	4.1%
	EGARCH$_{norm}$	3.5%	2.5%	9.7%	5.3%	**7.1%**	**3.3%**	7.2%	**8.0%**	11.8%	**8.7%**	**10.0%**	**6.5%**
	EGARCH$_{std}$	1.8%	2.0%	6.9%	2.2%	4.5%	2.3%	5.4%	5.9%	8.8%	5.7%	7.1%	5.9%
	EGARCH$_{sstd}$	1.6%	1.1%	5.3%	2.8%	4.6%	2.2%	4.4%	4.0%	8.0%	5.1%	7.7%	4.2%

Table 12 – Scenario 1 - Power of the three backtesting procedures in percentage, using $GARCH_{sstd}$ as null hypothesis, $n = 250$ and $T \in \{250, 500\}$

Backtesting	Model	1%						2.5%						5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	$GARCH_{norm}$	**2.2%**	0.2%	4.5%	**2.0%**	**4.0%**	0.2%	**2.9%**	1.1%	5.8%	2.9%	**5.1%**	1.0%	**5.2%**	2.5%	7.6%	4.6%	**6.0%**	1.9%	**9.8%**	8.2%	10.1%	**8.5%**	9.6%	5.3%
	$GARCH_{std}$	0.5%	0.3%	4.5%	1.7%	2.1%	0.5%	1.1%	0.5%	5.7%	2.3%	3.7%	0.9%	2.1%	1.7%	7.1%	4.0%	4.7%	2.5%	6.0%	6.1%	8.7%	6.3%	7.4%	5.6%
	$EGARCH_{norm}$	0.8%	0.4%	5.2%	1.8%	3.8%	**1.0%**	1.7%	**1.5%**	6.9%	**3.2%**	4.6%	**1.5%**	4.4%	**4.1%**	**9.4%**	**5.2%**	**6.0%**	**2.8%**	9.4%	**8.9%**	**11.0%**	7.6%	**9.7%**	**6.5%**
	$EGARCH_{std}$	0.8%	**0.5%**	2.9%	1.2%	1.5%	0.7%	1.2%	0.9%	4.2%	2.0%	2.3%	1.1%	2.3%	2.0%	5.9%	3.1%	3.4%	2.3%	5.9%	5.8%	8.5%	6.5%	6.2%	4.7%
	$EGARCH_{sstd}$	0.3%	0.2%	**6.2%**	1.9%	2.1%	0.5%	1.0%	0.9%	**7.2%**	2.5%	3.1%	1.3%	2.2%	1.9%	9.2%	4.0%	4.7%	1.9%	5.4%	5.4%	10.5%	7.1%	7.1%	5.4%
MFE Test	$GARCH_{norm}$	**2.2%**	0.3%	4.5%	**2.0%**	**4.0%**	0.2%	**2.7%**	1.1%	5.8%	**3.0%**	**5.1%**	0.8%	**5.2%**	2.5%	7.6%	4.5%	**6.2%**	1.9%	**9.8%**	8.1%	10.3%	**8.8%**	9.6%	5.2%
	$GARCH_{std}$	0.5%	0.3%	4.6%	1.7%	2.1%	0.5%	1.2%	0.7%	5.7%	2.3%	3.7%	1.1%	2.2%	1.8%	7.1%	3.7%	4.7%	2.5%	6.2%	6.1%	8.5%	6.6%	7.7%	5.4%
	$EGARCH_{norm}$	0.8%	0.3%	5.2%	1.8%	3.8%	**1.0%**	1.7%	**1.5%**	6.9%	**3.0%**	4.7%	**1.5%**	4.6%	**4.0%**	**9.5%**	**5.1%**	6.1%	**2.8%**	9.7%	**9.2%**	**11.2%**	7.9%	**9.7%**	**6.4%**
	$EGARCH_{std}$	0.8%	**0.5%**	2.8%	1.2%	1.5%	0.8%	1.2%	1.1%	4.2%	2.1%	2.2%	1.1%	2.2%	2.0%	5.9%	3.0%	3.5%	2.4%	5.9%	5.7%	8.3%	6.4%	6.3%	4.8%
	$EGARCH_{sstd}$	0.4%	0.3%	**6.3%**	1.9%	2.0%	0.5%	1.0%	0.7%	**7.2%**	2.6%	3.1%	1.3%	2.2%	2.1%	9.1%	3.8%	4.6%	1.7%	5.5%	5.4%	10.6%	7.2%	6.9%	5.7%
RC Test	$GARCH_{norm}$	**2.1%**	0.3%	4.5%	**2.1%**	**4.0%**	0.2%	**2.9%**	1.0%	5.8%	**3.1%**	**5.1%**	0.9%	**5.2%**	2.5%	7.7%	4.8%	**6.1%**	1.8%	**9.7%**	8.5%	10.2%	**8.5%**	9.5%	5.4%
	$GARCH_{std}$	0.6%	0.2%	4.5%	1.6%	2.3%	0.6%	1.1%	0.8%	5.6%	2.2%	3.6%	1.1%	2.4%	2.1%	7.1%	3.9%	4.8%	2.4%	5.9%	6.1%	8.6%	6.7%	7.5%	5.3%
	$EGARCH_{norm}$	0.8%	**0.4%**	5.2%	1.8%	3.8%	**0.9%**	1.7%	**1.3%**	6.9%	3.1%	4.6%	**1.5%**	4.4%	**4.2%**	**9.5%**	**4.9%**	5.9%	**2.8%**	9.3%	**8.9%**	**11.0%**	7.8%	**9.6%**	**6.6%**
	$EGARCH_{std}$	0.8%	**0.4%**	3.2%	1.3%	1.6%	0.6%	1.2%	1.0%	4.2%	2.2%	2.2%	1.2%	2.1%	1.9%	5.8%	2.9%	3.6%	2.2%	5.5%	6.0%	8.3%	6.6%	6.4%	4.8%
	$EGARCH_{sstd}$	0.4%	0.3%	**6.2%**	1.8%	2.0%	0.5%	1.1%	0.8%	**7.3%**	2.7%	3.4%	1.3%	2.0%	2.1%	9.2%	4.0%	4.4%	1.9%	5.4%	5.4%	10.5%	7.0%	7.0%	5.3%

Table 13 – Scenario 1 - Power of the three backtesting procedures in percentage, using EGARCH$_{norm}$ as null hypothesis, $n = 250$ and $T \in \{250, 500\}$

Backtesting	**Model**	1%						2.5%						5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH$_{norm}$	**0.8%**	**0.3%**	3.2%	**1.3%**	2.1%	0.4%	**1.4%**	**0.9%**	4.4%	**2.5%**	3.5%	0.8%	**2.9%**	2.2%	5.9%	3.2%	**5.6%**	**2.4%**	**6.8%**	**6.3%**	8.6%	**7.3%**	9.2%	**6.4%**
	GARCH$_{std}$	0.2%	**0.3%**	3.5%	0.4%	1.1%	**0.6%**	0.9%	**0.9%**	4.4%	1.3%	2.4%	0.8%	2.8%	**2.5%**	5.3%	2.4%	5.1%	1.9%	6.4%	5.8%	8.1%	4.8%	**9.6%**	6.1%
	GARCH$_{sstd}$	0.1%	0.2%	3.7%	0.6%	**3.0%**	0.4%	0.8%	**0.9%**	5.2%	1.4%	**4.1%**	**0.9%**	2.5%	1.6%	6.7%	2.6%	5.3%	2.2%	5.1%	4.3%	8.7%	5.2%	8.3%	5.3%
	EGARCH$_{std}$	0.2%	0.0%	2.8%	**1.3%**	1.7%	0.3%	0.9%	0.5%	4.5%	1.7%	2.4%	0.7%	1.6%	1.5%	6.6%	2.7%	4.0%	1.8%	3.8%	4.8%	9.0%	5.6%	6.5%	4.7%
	EGARCH$_{sstd}$	0.5%	0.1%	**4.3%**	1.0%	2.4%	0.1%	0.9%	0.4%	**5.8%**	1.4%	3.2%	0.6%	1.7%	1.3%	**7.8%**	**3.4%**	4.6%	**2.4%**	5.2%	4.5%	**10.2%**	6.9%	7.4%	5.3%
MFE Test	GARCH$_{norm}$	**0.8%**	**0.3%**	3.1%	**1.3%**	2.1%	0.4%	**1.4%**	**0.9%**	4.4%	**2.4%**	3.5%	**1.0%**	**2.9%**	2.1%	5.9%	3.2%	**5.7%**	**2.4%**	**7.0%**	**6.2%**	8.7%	**7.2%**	9.3%	**6.4%**
	GARCH$_{std}$	0.2%	0.2%	3.4%	0.5%	1.1%	**0.6%**	1.0%	0.8%	4.4%	1.2%	2.5%	0.8%	2.8%	**2.6%**	5.3%	2.6%	5.2%	1.8%	6.6%	6.0%	8.1%	4.9%	**9.5%**	6.0%
	GARCH$_{sstd}$	0.1%	0.2%	3.7%	0.6%	**2.9%**	0.3%	1.0%	**0.9%**	5.1%	1.4%	**4.1%**	**1.0%**	2.7%	1.7%	6.8%	2.6%	5.4%	2.3%	5.2%	4.3%	8.7%	5.4%	8.3%	5.7%
	EGARCH$_{std}$	0.2%	0.0%	2.8%	1.2%	1.7%	0.3%	0.9%	0.6%	4.4%	1.7%	2.5%	0.7%	1.8%	1.5%	6.6%	2.8%	3.9%	1.9%	4.1%	4.7%	8.8%	5.8%	6.5%	4.7%
	EGARCH$_{sstd}$	0.5%	0.1%	**4.4%**	1.0%	2.4%	0.2%	0.9%	0.5%	**5.9%**	1.5%	3.3%	0.5%	1.7%	1.4%	**8.0%**	**3.4%**	4.4%	**2.4%**	4.9%	4.4%	**10.3%**	6.8%	7.5%	5.4%
RC Test	GARCH$_{norm}$	**0.8%**	**0.3%**	3.2%	**1.3%**	2.1%	0.4%	**1.3%**	**1.1%**	4.3%	**2.3%**	3.5%	1.0%	**3.1%**	1.9%	5.8%	3.2%	**5.7%**	**2.4%**	**7.1%**	**6.2%**	8.7%	**7.4%**	9.1%	**6.4%**
	GARCH$_{std}$	0.2%	0.2%	3.4%	0.5%	1.3%	**0.5%**	1.1%	1.0%	4.3%	1.2%	2.6%	0.8%	3.0%	**2.4%**	5.4%	2.4%	5.4%	2.1%	7.0%	6.1%	8.0%	5.0%	**9.6%**	5.8%
	GARCH$_{sstd}$	0.1%	0.2%	4.0%	0.6%	**2.9%**	0.4%	0.9%	1.0%	5.2%	1.3%	**4.1%**	**1.1%**	2.6%	1.7%	6.6%	2.7%	5.4%	2.1%	5.2%	4.8%	8.5%	5.4%	8.0%	5.2%
	EGARCH$_{std}$	0.2%	0.0%	2.9%	1.3%	1.7%	0.2%	0.9%	0.5%	4.4%	1.8%	2.5%	0.7%	1.8%	1.3%	6.5%	2.6%	4.0%	1.7%	3.8%	4.6%	8.8%	5.9%	6.7%	4.4%
	EGARCH$_{sstd}$	0.5%	0.0%	**4.5%**	1.0%	2.5%	0.2%	0.8%	0.6%	**6.0%**	1.6%	3.1%	0.6%	1.6%	1.3%	**8.0%**	**3.5%**	4.5%	2.3%	4.8%	4.3%	**10.3%**	6.8%	7.7%	5.1%

Table 14 – Scenario 1 - Power of the three backtesting procedures in percentage, using EGARCH_{std} as null hypothesis, $n = 250$ and $T \in \{250, 500\}$

Backtesting	**Model**	1%						2.5%						5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH_{norm}	1.0%	0.3%	4.8%	2.0%	3.0%	**0.7%**	1.7%	1.0%	6.0%	3.1%	**4.2%**	**1.4%**	**3.6%**	2.8%	6.8%	3.9%	**5.7%**	2.3%	**7.7%**	5.9%	9.0%	7.2%	7.6%	5.4%
	GARCH_{std}	0.2%	0.1%	3.3%	1.3%	2.3%	0.4%	0.5%	0.7%	4.4%	1.7%	3.4%	1.1%	1.6%	1.4%	6.1%	2.6%	4.8%	2.1%	4.7%	4.7%	8.9%	4.6%	7.3%	5.2%
	GARCH_{sstd}	0.3%	0.1%	2.8%	1.1%	2.6%	0.2%	1.0%	0.4%	4.5%	1.5%	3.6%	0.9%	2.1%	1.1%	6.1%	2.5%	5.1%	1.8%	5.3%	4.3%	8.2%	5.2%	8.5%	5.8%
	EGARCH_{norm}	**1.3%**	**0.4%**	**6.2%**	**3.0%**	**3.2%**	0.5%	**2.2%**	**1.1%**	**8.4%**	**4.0%**	**4.2%**	**1.4%**	3.3%	**3.0%**	**10.5%**	**5.9%**	**5.7%**	**2.7%**	7.2%	**7.2%**	**12.7%**	**8.3%**	**8.9%**	**6.0%**
	EGARCH_{sstd}	0.3%	0.3%	3.7%	1.3%	2.6%	0.5%	0.6%	0.8%	4.9%	2.9%	3.7%	1.1%	1.8%	1.5%	6.9%	4.3%	5.1%	2.0%	3.4%	4.2%	8.8%	7.4%	6.9%	5.2%
MFE Test	GARCH_{norm}	1.0%	0.3%	4.8%	1.9%	3.1%	**0.7%**	1.7%	**1.1%**	6.0%	3.0%	4.2%	**1.4%**	**3.6%**	2.6%	6.8%	3.9%	5.6%	2.3%	**7.6%**	5.8%	9.0%	7.4%	7.6%	5.6%
	GARCH_{std}	0.1%	0.1%	3.3%	1.2%	2.4%	0.4%	0.6%	0.7%	4.4%	1.7%	3.4%	1.1%	1.4%	1.6%	6.1%	2.6%	4.9%	2.1%	4.5%	4.6%	9.1%	4.7%	7.7%	5.5%
	GARCH_{sstd}	0.3%	0.1%	2.8%	1.2%	2.7%	0.1%	1.1%	0.4%	4.6%	1.5%	3.6%	0.9%	2.2%	1.1%	6.1%	2.7%	4.9%	1.8%	5.6%	4.5%	8.0%	5.5%	8.5%	5.9%
	EGARCH_{norm}	**1.2%**	**0.4%**	**6.0%**	**3.0%**	**3.3%**	0.5%	**2.2%**	1.0%	**8.4%**	**4.1%**	**4.3%**	1.2%	3.2%	**2.7%**	**10.8%**	**5.8%**	**5.7%**	**2.7%**	7.4%	**7.1%**	**12.7%**	**8.4%**	**8.6%**	**6.2%**
	EGARCH_{sstd}	0.3%	0.3%	3.8%	1.5%	2.5%	0.6%	0.5%	0.7%	4.6%	2.9%	3.7%	1.0%	1.6%	1.7%	6.9%	4.3%	5.0%	2.1%	4.0%	4.3%	8.9%	7.1%	6.7%	5.3%
RC Test	GARCH_{norm}	1.1%	0.3%	4.8%	2.0%	3.2%	**0.7%**	1.6%	1.1%	5.9%	3.2%	**4.3%**	**1.4%**	**3.7%**	2.5%	6.9%	3.8%	**5.6%**	2.3%	**7.4%**	5.8%	8.9%	7.1%	7.6%	5.2%
	GARCH_{std}	0.1%	0.2%	3.3%	1.2%	2.2%	0.5%	0.5%	0.7%	4.3%	1.7%	3.4%	1.0%	1.7%	1.6%	6.0%	2.5%	4.6%	2.1%	5.0%	4.8%	9.1%	5.1%	7.5%	5.4%
	GARCH_{sstd}	0.2%	0.1%	2.9%	1.2%	2.5%	0.2%	0.9%	0.4%	4.5%	1.5%	3.3%	0.9%	2.3%	1.1%	6.0%	3.0%	5.0%	1.4%	5.3%	4.3%	8.3%	5.6%	8.7%	5.9%
	EGARCH_{norm}	**1.3%**	**0.4%**	**6.2%**	**3.0%**	**3.3%**	0.5%	**2.2%**	**1.1%**	**8.4%**	**4.1%**	**4.3%**	1.3%	3.2%	**2.9%**	**10.5%**	**6.0%**	**5.6%**	**2.8%**	7.3%	**7.2%**	**13.0%**	**8.5%**	**8.8%**	**6.2%**
	EGARCH_{sstd}	0.3%	0.2%	3.7%	1.2%	2.7%	0.6%	0.6%	0.8%	4.6%	2.8%	3.7%	1.0%	1.7%	1.8%	6.7%	4.2%	5.1%	1.7%	4.3%	4.1%	8.9%	7.2%	6.9%	5.2%

Table 15 – Scenario 1 - Power of the three backtesting procedures in percentage, using EGARCH_{sstd} as null hypothesis, $n = 250$ and $T \in \{250, 500\}$

Backtesting	Model	1%						2.5%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH_{norm}	**1.0%**	0.2%	5.0%	1.8%	**3.8%**	0.8%	**2.1%**	0.9%	6.9%	2.8%	**4.9%**	1.1%
	GARCH_{std}	0.3%	0.1%	2.8%	0.9%	2.0%	0.1%	1.1%	0.4%	3.4%	1.3%	2.9%	0.7%
	GARCH_{sstd}	0.3%	0.1%	2.5%	0.9%	1.6%	0.2%	0.9%	0.2%	3.6%	1.5%	2.2%	1.0%
	EGARCH_{norm}	0.9%	**0.5%**	**7.7%**	**3.8%**	3.2%	**1.2%**	1.5%	**1.6%**	**10.1%**	**4.9%**	4.2%	**1.7%**
	EGARCH_{std}	0.7%	0.3%	3.6%	1.1%	2.9%	0.6%	1.1%	1.0%	5.2%	1.7%	3.6%	1.3%
MFE Test	GARCH_{norm}	**0.9%**	0.2%	5.0%	1.8%	**3.8%**	0.8%	**2.1%**	1.0%	6.9%	3.0%	**5.0%**	1.1%
	GARCH_{std}	0.3%	0.1%	2.8%	0.8%	2.0%	0.1%	1.0%	0.4%	3.6%	1.5%	2.8%	0.7%
	GARCH_{sstd}	0.3%	0.1%	2.5%	0.9%	1.6%	0.1%	0.8%	0.3%	3.5%	1.4%	2.4%	0.9%
	EGARCH_{norm}	**0.9%**	**0.5%**	**7.5%**	**3.7%**	3.2%	**1.2%**	1.4%	**1.7%**	**10.1%**	**5.0%**	4.2%	**1.8%**
	EGARCH_{std}	0.7%	0.3%	3.7%	1.0%	2.9%	0.6%	1.1%	1.1%	5.2%	1.6%	3.4%	1.2%
RC Test	GARCH_{norm}	0.9%	0.3%	5.1%	1.8%	**3.8%**	0.8%	**2.1%**	1.1%	6.9%	2.9%	**4.9%**	1.1%
	GARCH_{std}	0.3%	0.1%	2.9%	0.9%	1.9%	0.1%	1.1%	0.7%	3.7%	1.4%	2.6%	0.4%
	GARCH_{sstd}	0.4%	0.0%	2.5%	0.8%	1.6%	0.2%	0.8%	0.3%	3.4%	1.4%	2.3%	0.8%
	EGARCH_{norm}	**1.0%**	**0.5%**	**7.6%**	**3.7%**	3.1%	**1.1%**	1.6%	**1.6%**	**10.1%**	**4.9%**	4.2%	**1.8%**
	EGARCH_{std}	0.8%	0.3%	4.0%	1.1%	2.9%	0.5%	1.4%	1.2%	5.1%	1.5%	3.5%	1.1%

Backtesting	Model	5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH_{norm}	3.7%	3.0%	8.7%	4.2%	**6.9%**	2.3%	7.6%	8.4%	11.9%	8.6%	**9.6%**	4.8%
	GARCH_{std}	2.3%	2.4%	5.0%	2.6%	4.4%	2.0%	6.4%	5.0%	8.2%	5.8%	6.9%	4.7%
	GARCH_{sstd}	1.5%	1.5%	4.8%	2.5%	3.1%	1.6%	4.0%	3.8%	7.3%	5.0%	6.0%	4.0%
	EGARCH_{norm}	**4.3%**	**3.3%**	**12.7%**	**6.4%**	5.6%	**3.2%**	**8.4%**	**8.6%**	**14.7%**	**10.0%**	8.5%	**6.7%**
	EGARCH_{std}	2.9%	2.1%	6.1%	2.7%	4.8%	2.7%	6.3%	5.5%	8.4%	6.0%	7.4%	5.9%
MFE Test	GARCH_{norm}	3.6%	2.9%	8.7%	3.9%	**6.9%**	2.3%	7.6%	8.1%	11.5%	8.7%	**9.7%**	4.6%
	GARCH_{std}	2.3%	2.4%	5.2%	2.6%	4.4%	2.0%	6.2%	4.8%	8.0%	6.0%	6.7%	4.8%
	GARCH_{sstd}	1.5%	1.5%	5.0%	2.5%	3.0%	1.3%	4.1%	3.7%	7.3%	4.9%	6.5%	4.1%
	EGARCH_{norm}	**3.9%**	**3.4%**	**12.5%**	**6.4%**	5.6%	**3.3%**	**8.4%**	**8.6%**	**15.0%**	**10.1%**	8.7%	**7.0%**
	EGARCH_{std}	2.6%	2.1%	6.1%	2.6%	4.7%	2.6%	6.2%	5.7%	8.3%	5.9%	7.3%	5.6%
RC Test	GARCH_{norm}	3.5%	3.0%	8.6%	4.3%	**6.8%**	2.2%	7.3%	7.7%	11.9%	8.8%	**9.4%**	4.8%
	GARCH_{std}	2.2%	2.6%	4.9%	2.6%	4.6%	1.7%	6.0%	5.0%	7.9%	6.0%	6.6%	4.5%
	GARCH_{sstd}	1.6%	1.5%	5.1%	2.6%	3.1%	1.5%	4.6%	4.4%	7.7%	5.1%	6.7%	4.3%
	EGARCH_{norm}	**4.2%**	**3.4%**	**12.6%**	**6.4%**	5.5%	**3.3%**	**7.9%**	**8.4%**	**14.9%**	**9.9%**	8.7%	**6.9%**
	EGARCH_{std}	2.8%	2.3%	6.1%	2.7%	4.8%	2.7%	6.7%	5.7%	8.3%	5.8%	7.3%	5.7%

Table 16 – Scenario 2 - Power of the three backtesting procedures in percentage, using $GARCH_{norm}$ as null hypothesis, $n = 1000$ and $T \in \{250, 500\}$

Backtesting	**Model**	1%						2.5%						5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	$GARCH_{std}$	0.6%	0.0%	5.9%	**2.3%**	2.3%	0.2%	0.8%	0.1%	**8.2%**	**2.8%**	3.2%	1.0%	1.6%	1.0%	**10.1%**	3.6%	**5.4%**	**2.5%**	4.2%	4.4%	**11.7%**	6.6%	7.5%	**5.3%**
	$GARCH_{sstd}$	0.8%	0.1%	4.7%	1.3%	**2.6%**	0.3%	1.2%	0.1%	5.9%	2.4%	3.6%	0.5%	2.1%	1.2%	8.2%	3.7%	4.8%	1.2%	4.3%	3.9%	10.8%	6.0%	7.8%	3.9%
	$EGARCH_{norm}$	**1.0%**	0.1%	5.7%	1.1%	2.3%	**0.8%**	**1.7%**	0.3%	6.7%	2.4%	3.5%	1.4%	**2.5%**	**1.8%**	7.8%	3.9%	4.4%	2.2%	4.7%	**4.5%**	10.4%	**7.3%**	7.4%	4.7%
	$EGARCH_{std}$	0.8%	0.1%	**6.1%**	1.5%	2.4%	0.6%	1.6%	0.4%	7.4%	2.2%	**4.0%**	1.0%	2.3%	1.5%	8.4%	**4.1%**	5.1%	2.4%	**4.9%**	3.8%	10.3%	7.3%	7.6%	4.8%
	$EGARCH_{sstd}$	0.6%	**0.2%**	5.1%	1.1%	2.2%	0.7%	1.1%	**0.8%**	6.3%	1.9%	3.0%	**1.5%**	2.0%	1.7%	8.0%	3.5%	4.4%	2.3%	4.8%	4.1%	10.4%	6.2%	**8.0%**	5.2%
MFE Test	$GARCH_{std}$	0.6%	0.0%	6.0%	**2.3%**	2.4%	0.2%	0.8%	0.0%	**8.2%**	**2.8%**	3.3%	1.0%	1.6%	1.0%	**10.1%**	3.6%	**5.3%**	**2.6%**	4.1%	**4.6%**	**11.7%**	6.8%	7.6%	**5.5%**
	$GARCH_{sstd}$	0.7%	0.0%	4.7%	1.3%	**2.5%**	0.3%	1.1%	0.1%	6.0%	2.4%	3.6%	0.5%	2.1%	1.0%	8.4%	3.7%	4.5%	1.3%	4.6%	3.7%	11.0%	6.0%	7.9%	3.9%
	$EGARCH_{norm}$	**0.8%**	0.1%	5.8%	1.2%	2.3%	**0.8%**	**1.7%**	0.4%	6.7%	2.5%	3.7%	1.2%	**2.6%**	1.6%	7.8%	4.1%	4.3%	2.0%	4.6%	4.4%	10.5%	7.3%	7.2%	4.6%
	$EGARCH_{std}$	**0.8%**	0.1%	**6.1%**	1.4%	2.4%	0.5%	1.6%	0.3%	7.4%	2.2%	**4.1%**	1.0%	2.3%	1.6%	8.5%	**4.2%**	5.1%	2.4%	4.4%	3.7%	10.4%	**7.5%**	7.9%	4.9%
	$EGARCH_{sstd}$	0.7%	**0.2%**	5.0%	1.1%	2.2%	0.7%	1.1%	**0.8%**	6.3%	1.9%	3.1%	**1.4%**	1.9%	**1.7%**	8.0%	3.6%	4.5%	2.4%	**4.8%**	4.1%	10.5%	6.3%	**8.2%**	5.4%
RC Test	$GARCH_{std}$	0.6%	0.0%	6.0%	**2.3%**	2.3%	0.2%	0.8%	0.0%	**8.2%**	**2.7%**	3.3%	1.0%	1.6%	1.0%	**10.0%**	3.8%	**5.4%**	**2.5%**	4.0%	**4.3%**	**11.9%**	6.9%	7.3%	**5.5%**
	$GARCH_{sstd}$	0.7%	0.1%	4.6%	1.3%	2.4%	0.3%	1.2%	0.1%	5.9%	2.4%	3.6%	0.5%	2.1%	1.2%	8.3%	3.7%	4.5%	1.3%	4.8%	3.9%	11.0%	6.0%	7.5%	4.0%
	$EGARCH_{norm}$	**0.9%**	0.1%	5.7%	1.1%	2.1%	**0.8%**	**1.9%**	0.3%	6.7%	2.5%	3.5%	1.2%	**2.5%**	**1.6%**	7.8%	4.1%	4.4%	1.9%	4.8%	**4.3%**	10.3%	7.1%	7.3%	4.7%
	$EGARCH_{std}$	0.8%	0.1%	**6.1%**	1.2%	**2.5%**	0.6%	1.6%	0.3%	7.5%	2.0%	**4.1%**	1.0%	2.5%	1.4%	8.4%	**4.2%**	5.2%	2.4%	4.8%	3.9%	10.0%	**7.5%**	7.7%	4.8%
	$EGARCH_{sstd}$	0.6%	**0.2%**	5.0%	1.1%	2.1%	0.7%	1.1%	**0.7%**	6.5%	1.8%	3.1%	**1.4%**	1.9%	**1.6%**	8.0%	3.6%	4.5%	2.3%	**5.2%**	4.0%	10.4%	6.4%	**7.9%**	5.3%

Table 17 – Scenario 2 - Power of the three backtesting procedures in percentage, using $GARCH_{std}$ as null hypothesis, $n = 1000$ and $T \in \{250, 500\}$

Backtesting	Model	1%						2.5%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	$GARCH_{norm}$	**1.6%**	0.0%	5.0%	2.6%	**4.4%**	0.8%	2.3%	0.8%	6.3%	3.3%	**5.8%**	1.3%
	$GARCH_{sstd}$	0.5%	0.1%	3.6%	1.0%	3.6%	0.7%	0.9%	0.6%	4.8%	1.7%	4.6%	1.1%
	$EGARCH_{norm}$	1.2%	**0.2%**	**5.4%**	**2.9%**	4.2%	**1.2%**	**2.4%**	**1.1%**	**6.9%**	**3.6%**	5.7%	**2.0%**
	$EGARCH_{std}$	0.3%	0.0%	4.1%	0.7%	2.7%	0.5%	0.6%	0.3%	5.5%	1.4%	3.8%	1.4%
	$EGARCH_{sstd}$	0.2%	0.1%	4.8%	1.0%	2.8%	0.2%	0.9%	0.2%	5.6%	1.8%	4.2%	0.5%
MFE Test	$GARCH_{norm}$	**1.6%**	0.0%	5.2%	2.6%	**4.5%**	0.8%	**2.3%**	0.7%	6.2%	3.3%	**5.9%**	1.3%
	$GARCH_{sstd}$	0.5%	0.0%	3.6%	1.0%	3.6%	0.9%	1.0%	0.6%	4.9%	1.8%	4.6%	1.1%
	$EGARCH_{norm}$	1.2%	**0.2%**	**5.5%**	**2.9%**	4.2%	**1.1%**	**2.3%**	**1.0%**	**6.8%**	**3.5%**	**5.9%**	**2.0%**
	$EGARCH_{std}$	0.3%	0.0%	4.2%	0.7%	2.6%	0.5%	0.5%	0.3%	5.6%	1.6%	3.8%	1.3%
	$EGARCH_{sstd}$	0.2%	0.1%	4.9%	1.0%	2.9%	0.2%	0.9%	0.2%	5.7%	1.8%	4.3%	0.5%
RC Test	$GARCH_{norm}$	**1.4%**	0.0%	5.2%	**2.7%**	**4.4%**	0.8%	**2.3%**	0.6%	6.1%	3.3%	**5.7%**	1.3%
	$GARCH_{sstd}$	0.5%	0.0%	3.7%	1.0%	3.7%	0.8%	0.9%	0.6%	4.8%	1.8%	4.6%	1.1%
	$EGARCH_{norm}$	1.2%	**0.2%**	**5.5%**	**2.7%**	4.2%	**1.2%**	**2.3%**	**1.1%**	**6.9%**	**3.6%**	5.6%	**1.9%**
	$EGARCH_{std}$	0.3%	0.0%	4.1%	0.6%	2.5%	0.4%	0.6%	0.4%	5.5%	1.4%	3.7%	1.2%
	$EGARCH_{sstd}$	0.2%	0.1%	4.9%	1.0%	2.8%	0.2%	0.8%	0.2%	5.7%	1.8%	4.2%	0.5%

Backtesting	Model	5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	$GARCH_{norm}$	**3.9%**	**2.1%**	7.8%	3.5%	7.6%	3.2%	**7.9%**	5.1%	9.4%	6.7%	10.8%	5.9%
	$GARCH_{sstd}$	2.0%	1.2%	6.6%	2.4%	5.2%	1.9%	4.5%	3.5%	9.5%	4.1%	8.1%	5.3%
	$EGARCH_{norm}$	3.5%	2.0%	**8.5%**	**5.5%**	**8.3%**	**3.3%**	6.5%	**5.2%**	**11.3%**	**7.1%**	**10.9%**	**6.5%**
	$EGARCH_{std}$	1.8%	1.0%	7.1%	2.6%	5.5%	2.6%	4.5%	3.0%	9.5%	5.6%	8.5%	5.6%
	$EGARCH_{sstd}$	1.4%	0.9%	7.2%	3.1%	5.9%	1.2%	4.0%	3.5%	8.8%	6.0%	8.5%	4.8%
MFE Test	$GARCH_{norm}$	**4.0%**	**2.0%**	7.8%	3.6%	7.6%	3.2%	**7.7%**	5.2%	9.6%	6.7%	**10.8%**	5.9%
	$GARCH_{sstd}$	2.1%	1.2%	6.7%	2.5%	5.2%	1.8%	4.6%	3.5%	9.4%	4.2%	8.0%	5.1%
	$EGARCH_{norm}$	3.7%	**2.0%**	**8.5%**	**5.1%**	**8.5%**	**3.4%**	6.9%	**5.3%**	**11.0%**	**7.2%**	**10.8%**	**6.2%**
	$EGARCH_{std}$	1.8%	0.9%	7.0%	2.5%	5.3%	2.4%	4.3%	3.0%	9.5%	5.6%	8.7%	6.0%
	$EGARCH_{sstd}$	1.5%	0.9%	7.4%	3.3%	5.8%	1.3%	3.9%	3.6%	9.2%	6.1%	8.4%	5.0%
RC Test	$GARCH_{norm}$	**3.9%**	2.1%	7.8%	3.5%	7.4%	3.2%	**7.7%**	5.0%	9.6%	6.7%	**11.1%**	5.7%
	$GARCH_{sstd}$	2.2%	1.1%	6.6%	2.5%	5.1%	1.9%	4.5%	3.6%	9.1%	4.1%	8.1%	4.9%
	$EGARCH_{norm}$	3.6%	**2.3%**	**8.6%**	**5.2%**	**8.5%**	**3.4%**	6.8%	**5.3%**	**11.1%**	**7.2%**	10.8%	**6.2%**
	$EGARCH_{std}$	1.8%	0.9%	7.1%	2.4%	5.5%	2.6%	4.6%	3.0%	9.3%	5.7%	8.4%	5.9%
	$EGARCH_{sstd}$	1.4%	0.8%	7.3%	2.9%	5.8%	1.3%	4.2%	3.4%	9.2%	6.0%	8.2%	5.1%

Table 18 – Scenario 2 - Power of the three backtesting procedures in percentage, using $GARCH_{sstd}$ as null hypothesis, $n = 1000$ and $T \in \{250, 500\}$

Backtesting	Model	1%						2.5%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
	$GARCH_{norm}$	**1.2%**	**0.4%**	5.5%	2.2%	3.9%	**1.0%**	2.2%	**1.1%**	7.0%	3.7%	**5.3%**	**2.0%**
	$GARCH_{std}$	0.5%	0.0%	4.1%	1.4%	1.7%	0.2%	0.9%	0.5%	5.3%	2.5%	2.0%	0.7%
MF Test	$EGARCH_{norm}$	**1.2%**	0.0%	**6.4%**	**2.6%**	**4.1%**	0.5%	**2.3%**	0.4%	**7.6%**	**4.2%**	5.1%	1.5%
	$EGARCH_{std}$	0.6%	0.1%	4.4%	0.8%	1.9%	0.5%	1.0%	0.3%	5.7%	1.4%	3.1%	0.9%
	$EGARCH_{sstd}$	0.5%	0.2%	4.2%	1.4%	1.7%	0.1%	1.1%	0.4%	5.5%	2.0%	2.5%	0.5%
	$GARCH_{norm}$	**1.2%**	**0.4%**	5.5%	2.3%	4.0%	**1.1%**	**2.3%**	**1.1%**	7.0%	3.8%	**5.3%**	**2.0%**
	$GARCH_{std}$	0.5%	0.0%	4.1%	1.3%	1.7%	0.2%	0.9%	0.5%	5.4%	2.4%	2.1%	0.7%
MFE Test	$EGARCH_{norm}$	**1.2%**	0.0%	**6.4%**	**2.5%**	**4.1%**	0.5%	**2.3%**	0.3%	**7.6%**	**4.1%**	5.1%	1.5%
	$EGARCH_{std}$	0.6%	0.1%	4.6%	0.8%	1.8%	0.5%	1.2%	0.2%	5.7%	1.3%	2.9%	0.9%
	$EGARCH_{sstd}$	0.6%	0.2%	4.2%	1.4%	1.7%	0.1%	1.2%	0.5%	5.6%	2.0%	2.5%	0.5%
	$GARCH_{norm}$	**1.2%**	**0.4%**	5.5%	2.3%	4.0%	**1.1%**	**2.4%**	**1.0%**	7.1%	3.8%	**5.3%**	**2.1%**
	$GARCH_{std}$	0.5%	0.0%	3.8%	1.2%	1.7%	0.2%	0.8%	0.6%	5.3%	2.2%	2.1%	0.7%
RC Test	$EGARCH_{norm}$	1.1%	0.0%	**6.4%**	**2.5%**	**4.1%**	0.5%	2.3%	0.5%	**7.6%**	**4.1%**	5.0%	1.4%
	$EGARCH_{std}$	0.6%	0.1%	4.6%	0.8%	1.8%	0.5%	1.3%	0.3%	5.7%	1.4%	2.8%	1.0%
	$EGARCH_{sstd}$	0.6%	0.2%	4.0%	1.4%	1.7%	0.1%	1.2%	0.2%	5.6%	2.0%	2.6%	0.6%

Backtesting	Model	5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
	$GARCH_{norm}$	**3.3%**	**2.7%**	8.5%	**5.4%**	**7.2%**	**3.6%**	**7.1%**	**7.0%**	11.2%	8.3%	**10.1%**	**7.1%**
	$GARCH_{std}$	2.4%	1.7%	7.3%	3.9%	4.2%	1.8%	5.7%	5.4%	10.0%	6.3%	6.6%	5.7%
MF Test	$EGARCH_{norm}$	**3.3%**	2.0%	**9.3%**	5.3%	6.3%	2.5%	6.2%	6.3%	**11.9%**	**8.5%**	8.7%	5.8%
	$EGARCH_{std}$	2.7%	1.6%	7.6%	3.0%	4.8%	2.1%	5.7%	5.0%	10.6%	5.7%	7.7%	3.6%
	$EGARCH_{sstd}$	1.6%	1.0%	6.8%	3.2%	3.5%	1.4%	5.1%	3.3%	8.7%	6.3%	5.9%	3.9%
	$GARCH_{norm}$	3.1%	**2.9%**	8.5%	5.3%	**7.2%**	**3.6%**	**7.0%**	**6.6%**	11.1%	8.4%	**10.1%**	**7.3%**
	$GARCH_{std}$	2.4%	1.6%	7.3%	4.0%	4.1%	1.8%	5.6%	5.4%	10.2%	6.4%	6.8%	5.7%
MFE Test	$EGARCH_{norm}$	**3.3%**	2.0%	**9.3%**	**5.4%**	6.4%	2.6%	6.3%	6.1%	**11.8%**	**8.7%**	9.0%	5.7%
	$EGARCH_{std}$	2.7%	1.6%	7.7%	2.9%	4.8%	1.9%	5.8%	5.3%	10.7%	5.7%	7.7%	4.0%
	$EGARCH_{sstd}$	1.8%	1.1%	6.6%	3.3%	3.5%	1.3%	4.6%	3.4%	8.7%	6.5%	6.0%	3.7%
	$GARCH_{norm}$	**3.4%**	**2.8%**	8.5%	**5.5%**	**7.2%**	**3.7%**	**7.1%**	**6.9%**	11.3%	8.3%	**9.8%**	**7.2%**
	$GARCH_{std}$	2.3%	1.8%	7.3%	4.0%	4.3%	1.7%	5.8%	5.2%	10.3%	6.3%	6.6%	5.8%
RC Test	$EGARCH_{norm}$	3.2%	2.0%	**9.3%**	5.4%	6.4%	2.5%	6.2%	6.3%	**11.7%**	**8.7%**	8.8%	5.8%
	$EGARCH_{std}$	2.7%	1.7%	7.6%	3.0%	4.8%	2.2%	5.9%	5.4%	10.8%	5.7%	7.7%	3.9%
	$EGARCH_{sstd}$	1.8%	0.9%	6.7%	3.3%	3.3%	1.2%	4.7%	3.8%	8.6%	6.7%	6.1%	3.8%

Table 19 – Scenario 2 - Power of the three backtesting procedures in percentage, using EGARCH_{norm} as null hypothesis, $n = 1000$ and $T \in \{250, 500\}$

Backtesting	Model	1%						2.5%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH_{norm}	0.7%	0.2%	4.2%	1.3%	1.6%	**0.6%**	1.2%	0.5%	**6.4%**	2.0%	2.6%	1.1%
	GARCH_{std}	0.5%	0.2%	4.0%	1.5%	1.4%	0.4%	1.0%	0.6%	5.4%	2.8%	2.3%	1.1%
	GARCH_{sstd}	0.5%	**0.4%**	4.2%	1.4%	2.0%	0.5%	1.3%	**1.0%**	5.3%	2.3%	2.4%	0.9%
	EGARCH_{std}	0.4%	0.1%	**5.1%**	**1.9%**	2.0%	**0.6%**	1.2%	0.4%	6.1%	3.2%	3.2%	**1.2%**
	EGARCH_{sstd}	**0.9%**	0.1%	4.9%	**1.9%**	**2.5%**	0.3%	**1.6%**	0.3%	5.9%	**3.3%**	**3.5%**	0.9%
MFE Test	GARCH_{norm}	0.7%	0.3%	4.3%	1.2%	1.6%	**0.7%**	1.2%	0.5%	**6.5%**	1.9%	2.6%	1.2%
	GARCH_{std}	0.5%	0.2%	4.0%	1.4%	1.5%	0.4%	1.1%	0.5%	5.5%	2.7%	2.2%	**1.3%**
	GARCH_{sstd}	0.5%	**0.5%**	4.2%	1.3%	2.0%	0.5%	1.4%	**1.0%**	5.5%	2.4%	2.5%	0.9%
	EGARCH_{std}	0.4%	0.1%	**5.0%**	**2.1%**	2.0%	0.6%	1.2%	0.4%	6.3%	3.1%	3.2%	1.0%
	EGARCH_{sstd}	**0.8%**	0.1%	4.8%	1.9%	**2.6%**	0.3%	**1.5%**	0.1%	6.0%	**3.2%**	**3.6%**	1.0%
RC Test	GARCH_{norm}	0.7%	0.2%	4.4%	1.3%	1.6%	**0.7%**	1.2%	0.4%	**6.6%**	2.0%	2.5%	1.1%
	GARCH_{std}	0.5%	0.2%	4.0%	1.5%	1.5%	0.4%	1.1%	0.5%	5.5%	3.0%	2.2%	**1.2%**
	GARCH_{sstd}	0.5%	**0.4%**	4.1%	1.3%	2.0%	0.5%	1.3%	**1.0%**	5.5%	2.1%	2.4%	0.9%
	EGARCH_{std}	0.5%	0.1%	**5.1%**	**2.0%**	2.0%	**0.7%**	1.2%	0.2%	6.1%	3.0%	3.4%	1.0%
	EGARCH_{sstd}	**0.9%**	0.1%	5.0%	2.0%	**2.5%**	0.4%	**1.6%**	0.2%	6.0%	**3.2%**	**3.6%**	1.1%

Backtesting	Model	5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH_{norm}	2.3%	**1.9%**	**8.6%**	3.9%	4.0%	2.1%	**6.8%**	**6.3%**	**11.1%**	6.6%	6.0%	5.4%
	GARCH_{std}	1.7%	**1.9%**	7.5%	3.9%	3.5%	2.0%	5.1%	5.7%	9.9%	**7.2%**	5.6%	5.8%
	GARCH_{sstd}	2.1%	**1.9%**	7.4%	3.7%	3.2%	1.8%	5.9%	5.1%	10.5%	6.7%	6.3%	4.6%
	EGARCH_{std}	2.1%	1.3%	8.5%	**4.5%**	4.5%	**2.2%**	4.6%	4.1%	10.6%	6.6%	7.6%	**6.3%**
	EGARCH_{sstd}	**2.4%**	1.6%	7.4%	4.3%	**5.2%**	2.0%	4.7%	4.4%	9.8%	7.0%	**9.0%**	5.5%
MFE Test	GARCH_{norm}	**2.4%**	**2.0%**	8.5%	3.9%	4.2%	1.9%	**6.5%**	**6.6%**	**11.2%**	6.5%	6.1%	5.2%
	GARCH_{std}	1.7%	2.0%	7.6%	4.0%	3.9%	2.0%	5.2%	5.5%	9.6%	**7.5%**	5.6%	5.6%
	GARCH_{sstd}	2.1%	1.8%	7.4%	3.7%	3.3%	1.8%	5.8%	5.2%	10.5%	6.5%	6.1%	4.9%
	EGARCH_{std}	1.9%	1.4%	**8.6%**	**4.4%**	4.2%	**2.2%**	4.5%	4.1%	10.5%	6.8%	7.5%	**6.0%**
	EGARCH_{sstd}	2.3%	1.2%	7.4%	4.0%	**5.3%**	2.0%	4.8%	4.1%	9.7%	6.8%	**8.7%**	5.5%
RC Test	GARCH_{norm}	**2.3%**	**1.9%**	8.5%	3.7%	4.1%	2.0%	**6.4%**	**6.6%**	**11.2%**	6.2%	5.9%	5.1%
	GARCH_{std}	1.6%	**1.9%**	7.5%	3.9%	3.7%	2.1%	5.1%	5.6%	9.7%	**7.3%**	5.7%	5.8%
	GARCH_{sstd}	2.2%	1.8%	7.4%	3.6%	3.2%	1.9%	5.7%	5.2%	10.5%	6.5%	6.1%	4.7%
	EGARCH_{std}	2.1%	1.5%	**8.7%**	**4.3%**	4.4%	**2.2%**	4.6%	4.0%	10.5%	6.7%	7.6%	**6.1%**
	EGARCH_{sstd}	2.2%	1.3%	7.5%	4.1%	**5.2%**	2.0%	5.0%	4.4%	9.7%	6.9%	**8.8%**	5.6%

Table 20 – Scenario 2 - Power of the three backtesting procedures in percentage, using EGARCH_{std} as null hypothesis, $n = 1000$ and $T \in \{250, 500\}$

Backtesting	Model	1%						2.5%						5%						10%					
		$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$
		250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH_{norm}	**1.0%**	**0.3%**	4.6%	2.2%	2.9%	1.0%	1.6%	**0.7%**	6.3%	2.8%	4.3%	**1.7%**	**2.9%**	**1.8%**	7.6%	4.6%	5.5%	**2.9%**	**6.6%**	**6.1%**	10.2%	7.6%	**7.6%**	5.1%
	GARCH_{std}	0.4%	0.2%	4.7%	0.5%	1.5%	0.4%	1.1%	0.3%	6.5%	1.4%	2.8%	0.7%	1.9%	0.9%	7.8%	2.1%	4.2%	2.4%	4.8%	3.6%	10.4%	4.9%	**7.6%**	**5.4%**
	GARCH_{sstd}	0.5%	0.1%	4.6%	0.9%	1.6%	0.3%	1.1%	0.3%	6.0%	1.8%	2.6%	0.8%	2.8%	1.2%	7.5%	3.4%	4.3%	1.7%	5.9%	3.9%	10.1%	6.6%	7.5%	**5.4%**
	EGARCH_{norm}	0.8%	0.1%	**6.0%**	**3.4%**	**3.8%**	**1.2%**	**1.7%**	0.6%	**7.9%**	**4.2%**	**4.7%**	**1.7%**	2.6%	**1.8%**	**9.1%**	**6.5%**	**6.0%**	2.7%	5.7%	5.0%	**11.3%**	**10.3%**	7.4%	4.9%
	EGARCH_{sstd}	0.2%	0.1%	4.0%	1.3%	2.5%	0.8%	0.6%	0.5%	5.1%	2.1%	2.9%	1.0%	1.5%	1.7%	6.4%	3.3%	4.5%	2.3%	5.5%	3.6%	8.8%	6.2%	6.5%	4.8%
MFE Test	GARCH_{norm}	**1.0%**	**0.3%**	4.7%	2.2%	3.0%	1.1%	1.6%	**0.8%**	6.4%	2.8%	4.3%	**1.8%**	**2.9%**	**1.8%**	7.6%	4.6%	5.3%	**2.9%**	**6.4%**	**6.0%**	10.1%	7.2%	**7.6%**	5.2%
	GARCH_{std}	0.4%	0.2%	4.7%	0.5%	1.5%	0.4%	1.0%	0.3%	6.5%	1.3%	2.9%	0.8%	1.9%	0.8%	7.7%	2.1%	4.4%	2.4%	4.6%	3.8%	10.1%	5.4%	7.2%	5.3%
	GARCH_{sstd}	0.5%	0.1%	4.7%	0.9%	1.5%	0.3%	1.0%	0.2%	6.1%	1.8%	2.6%	0.8%	2.6%	1.2%	7.4%	3.4%	4.4%	1.7%	5.5%	3.9%	9.9%	6.7%	7.5%	**5.5%**
	EGARCH_{norm}	0.8%	0.1%	**6.1%**	**3.4%**	**3.8%**	**1.2%**	**1.7%**	0.6%	**7.8%**	**4.3%**	**4.7%**	1.7%	2.5%	1.7%	**9.1%**	**6.8%**	**6.1%**	2.5%	5.8%	5.0%	**11.3%**	**10.3%**	7.4%	4.7%
	EGARCH_{sstd}	0.2%	0.3%	4.2%	1.1%	2.4%	0.9%	0.5%	0.6%	5.1%	1.7%	3.1%	1.0%	1.3%	1.6%	6.5%	3.2%	4.6%	2.2%	5.2%	3.6%	9.3%	5.8%	6.5%	4.8%
RC Test	GARCH_{norm}	**1.1%**	**0.3%**	4.7%	2.2%	3.0%	1.1%	1.6%	**0.8%**	6.4%	2.8%	4.3%	**1.9%**	**2.8%**	**1.7%**	7.6%	4.5%	5.5%	**2.8%**	**6.2%**	**6.4%**	10.0%	7.1%	**7.6%**	5.4%
	GARCH_{std}	0.4%	0.2%	4.8%	0.5%	1.5%	0.5%	1.1%	0.3%	6.5%	1.3%	3.0%	0.8%	1.8%	0.7%	7.7%	2.1%	4.3%	2.3%	4.8%	3.7%	10.0%	5.0%	7.4%	5.2%
	GARCH_{sstd}	0.5%	0.1%	4.7%	0.9%	1.5%	0.3%	1.0%	0.2%	6.1%	1.8%	2.6%	0.7%	2.5%	1.1%	7.6%	3.2%	4.4%	1.9%	5.6%	4.0%	9.9%	7.1%	7.4%	**5.7%**
	EGARCH_{norm}	0.8%	0.2%	**6.0%**	**3.3%**	**3.8%**	**1.2%**	**1.7%**	0.6%	**7.8%**	**4.2%**	**4.7%**	1.7%	2.5%	**1.7%**	**9.0%**	**6.7%**	**6.1%**	2.6%	5.7%	4.9%	**11.3%**	**10.7%**	7.5%	4.9%
	EGARCH_{sstd}	0.2%	0.2%	4.0%	1.3%	2.4%	0.9%	0.4%	0.6%	5.0%	1.8%	2.9%	1.0%	1.8%	1.5%	6.6%	3.3%	4.6%	2.1%	5.4%	3.6%	9.3%	5.9%	6.6%	4.8%

Table 21 – Scenario 2 - Power of the three backtesting procedures in percentage, using EGARCH_{sstd} as null hypothesis, $n = 1000$ and $T \in \{250, 500\}$

Backtesting	Model	1%						2.5%						5%						10%					
		$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
		250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	GARCH_{norm}	**1.4%**	**0.6%**	6.0%	**2.8%**	4.6%	0.8%	**2.2%**	**1.7%**	7.2%	**3.9%**	5.5%	1.4%	**3.9%**	**3.5%**	**9.6%**	**5.7%**	6.7%	2.7%	**7.9%**	**8.2%**	**12.1%**	**9.1%**	8.7%	6.2%
	GARCH_{std}	0.3%	0.0%	2.5%	1.1%	1.8%	0.6%	1.2%	0.7%	3.7%	1.6%	2.2%	1.1%	2.0%	1.9%	4.9%	2.7%	3.5%	2.5%	5.1%	6.0%	7.6%	5.8%	6.2%	4.8%
	GARCH_{sstd}	0.4%	0.0%	3.4%	0.9%	2.3%	0.2%	1.0%	0.2%	4.2%	1.8%	3.1%	0.3%	2.3%	1.7%	5.3%	2.4%	4.6%	1.3%	5.9%	4.8%	7.4%	5.3%	7.0%	5.0%
	EGARCH_{norm}	0.6%	0.3%	**6.3%**	2.7%	**4.8%**	**1.1%**	1.3%	0.6%	**7.4%**	3.4%	**6.2%**	**2.2%**	2.6%	2.1%	9.0%	4.7%	**7.5%**	**3.1%**	5.6%	5.7%	11.0%	6.9%	**9.3%**	**6.6%**
	EGARCH_{std}	0.4%	0.1%	4.6%	1.0%	2.8%	0.3%	0.7%	0.6%	5.9%	1.9%	3.7%	0.9%	2.3%	1.5%	7.3%	3.0%	5.6%	2.2%	5.8%	5.9%	9.2%	5.7%	8.8%	4.9%
MFE Test	GARCH_{norm}	**1.4%**	**0.6%**	6.0%	**2.7%**	4.6%	0.8%	**2.3%**	**1.7%**	7.3%	**4.0%**	5.5%	1.5%	**3.8%**	**3.6%**	**9.6%**	**5.8%**	6.6%	2.9%	**8.4%**	**8.0%**	**12.1%**	**9.2%**	8.8%	6.1%
	GARCH_{std}	0.3%	0.0%	2.5%	1.1%	1.8%	0.6%	1.1%	0.5%	3.6%	1.8%	2.2%	1.1%	2.2%	2.0%	4.5%	2.8%	3.7%	2.1%	4.9%	5.9%	7.6%	5.8%	6.5%	4.8%
	GARCH_{sstd}	0.4%	0.0%	3.4%	0.9%	2.3%	0.2%	1.0%	0.2%	4.1%	1.8%	3.0%	0.4%	2.2%	1.9%	5.3%	2.2%	4.6%	1.6%	5.7%	4.9%	7.2%	5.3%	7.0%	4.9%
	EGARCH_{norm}	0.6%	0.3%	**6.3%**	2.5%	**4.7%**	**1.1%**	1.3%	0.5%	**7.4%**	3.4%	**6.1%**	**2.1%**	2.4%	2.3%	9.1%	4.7%	**7.4%**	**3.0%**	5.8%	5.7%	11.0%	7.0%	**9.1%**	**6.5%**
	EGARCH_{std}	0.4%	0.1%	4.6%	0.9%	2.8%	0.3%	0.8%	0.5%	6.0%	1.8%	3.7%	1.0%	2.5%	1.5%	7.1%	3.0%	5.7%	2.2%	6.0%	5.6%	9.1%	5.2%	8.9%	4.9%
RC Test	GARCH_{norm}	**1.4%**	**0.6%**	6.1%	**2.8%**	4.6%	0.8%	**2.2%**	**1.6%**	7.3%	**3.9%**	5.5%	1.4%	**3.8%**	**3.6%**	**9.6%**	**5.6%**	6.8%	3.0%	**7.7%**	**7.9%**	**12.4%**	**9.0%**	8.6%	6.0%
	GARCH_{std}	0.3%	0.0%	2.5%	1.0%	1.9%	0.5%	1.2%	0.6%	3.6%	1.6%	2.2%	1.2%	2.1%	2.0%	4.7%	2.7%	3.6%	2.4%	5.0%	6.2%	7.5%	5.9%	6.3%	4.7%
	GARCH_{sstd}	0.4%	0.0%	3.3%	0.9%	2.3%	0.2%	1.0%	0.3%	4.2%	1.5%	3.1%	0.4%	2.3%	1.8%	5.3%	2.4%	4.6%	1.4%	5.5%	4.9%	7.2%	5.4%	6.9%	5.2%
	EGARCH_{norm}	0.6%	0.3%	**6.3%**	2.6%	**4.7%**	**1.1%**	1.3%	0.6%	**7.4%**	3.4%	**6.2%**	**2.1%**	2.5%	1.9%	9.0%	4.5%	**7.5%**	**3.1%**	6.0%	5.5%	10.9%	7.0%	**9.2%**	**6.5%**
	EGARCH_{std}	0.4%	0.1%	4.6%	1.0%	2.7%	0.3%	0.8%	0.5%	5.9%	1.8%	3.7%	0.9%	2.5%	1.5%	6.9%	2.9%	5.5%	2.2%	5.5%	5.9%	9.0%	6.0%	8.7%	4.6%

Table 22 – Proposed McNeil and Frey (2000), Righi and Ceretta (2015) and McNeil et al. (2015) backtesting size results, using $GARCH_{norm}$ as null hypothesis with 250 and 1000 in-sample observations and significance level of 1%, 2.5% and 5% .

Backtesting	Size	Model	1%						2.5%						5%						10%					
			$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$
			250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	**250 Observations**	$GARCH_{std}$	**0.6%**	0.0%	**4.7%**	**1.9%**	**3.0%**	**2.0%**	1.1%	0.3%	**5.8%**	**3.4%**	3.6%	**1.0%**	**2.8%**	1.5%	**7.7%**	**4.6%**	**5.1%**	**2.3%**	5.2%	3.7%	9.7%	**6.9%**	**8.3%**	**6.1%**
		$GARCH_{sstd}$	**0.6%**	0.0%	4.0%	1.2%	2.3%	0.3%	1.3%	0.5%	5.1%	2.0%	3.4%	0.6%	2.1%	**1.6%**	7.5%	3.0%	4.4%	1.8%	4.9%	4.1%	**10.2%**	5.8%	7.1%	5.2%
		$EGARCH_{norm}$	0.4%	**0.3%**	4.3%	1.7%	2.3%	0.3%	0.9%	**0.7%**	5.0%	2.4%	3.2%	0.8%	2.0%	1.5%	6.4%	3.9%	4.3%	1.9%	**5.3%**	4.3%	8.6%	6.7%	7.1%	4.6%
		$EGARCH_{std}$	0.5%	0.2%	3.8%	1.6%	**3.0%**	0.4%	**1.4%**	0.4%	5.1%	2.0%	**3.7%**	0.9%	2.1%	1.3%	6.6%	2.8%	4.7%	2.1%	**5.3%**	3.7%	8.5%	6.8%	7.8%	5.1%
		$EGARCH_{sstd}$	0.2%	0.0%	4.6%	1.6%	2.7%	0.4%	1.0%	0.4%	5.6%	2.2%	3.5%	0.8%	2.0%	1.4%	6.6%	3.4%	4.5%	1.5%	4.9%	**5.0%**	9.5%	6.1%	7.7%	4.7%
	1000 Observations	$GARCH_{std}$	0.6%	0.0%	5.9%	**2.3%**	2.3%	0.2%	0.8%	0.1%	**8.2%**	**2.8%**	3.2%	1.0%	1.6%	1.0%	**10.1%**	3.6%	**5.4%**	**2.5%**	4.2%	4.4%	**11.7%**	6.6%	7.5%	**5.3%**
		$GARCH_{sstd}$	0.8%	0.1%	4.7%	1.3%	**2.6%**	0.3%	1.2%	0.1%	5.9%	2.4%	3.6%	0.5%	2.1%	1.2%	8.2%	3.7%	4.8%	1.2%	4.3%	3.9%	10.8%	6.0%	7.8%	3.9%
		$EGARCH_{norm}$	**1.0%**	0.1%	5.7%	1.1%	2.3%	**0.8%**	**1.7%**	0.3%	6.7%	2.4%	3.5%	1.4%	**2.5%**	**1.8%**	7.8%	3.9%	4.4%	2.2%	4.7%	**4.5%**	10.4%	**7.3%**	7.4%	4.7%
		$EGARCH_{std}$	0.8%	0.1%	**6.1%**	1.5%	2.4%	0.6%	1.6%	0.4%	7.4%	2.2%	**4.0%**	1.0%	2.3%	1.5%	8.4%	**4.1%**	5.1%	2.4%	**4.9%**	3.8%	10.3%	7.3%	7.6%	4.8%
		$EGARCH_{sstd}$	0.6%	**0.2%**	5.1%	1.1%	2.2%	0.7%	1.1%	**0.8%**	6.3%	1.9%	3.0%	**1.5%**	2.0%	1.7%	8.0%	3.5%	4.4%	2.3%	4.8%	4.1%	10.4%	6.2%	**8.0%**	5.2%
RC Test	**250 Observations**	$GARCH_{std}$	**0.6%**	0.0%	**4.6%**	**1.8%**	2.9%	**2.0%**	1.1%	0.2%	**5.8%**	**3.3%**	3.5%	**1.0%**	**2.8%**	1.6%	**7.7%**	**4.6%**	**5.0%**	**2.5%**	5.4%	3.7%	9.6%	**7.1%**	**8.5%**	**6.2%**
		$GARCH_{sstd}$	**0.6%**	0.0%	3.9%	1.2%	2.3%	0.2%	1.3%	0.5%	5.2%	2.0%	3.3%	0.7%	2.2%	**1.7%**	7.3%	2.9%	4.5%	1.6%	4.6%	3.8%	**10.1%**	6.0%	7.3%	5.1%
		$EGARCH_{norm}$	0.5%	**0.3%**	4.3%	**1.8%**	2.2%	0.3%	0.8%	**0.7%**	5.0%	2.3%	3.1%	0.9%	1.9%	1.5%	6.5%	3.9%	4.4%	1.8%	**5.5%**	4.3%	8.7%	6.8%	7.2%	4.5%
		$EGARCH_{std}$	0.4%	**0.3%**	3.8%	1.5%	**3.0%**	0.4%	**1.5%**	0.4%	5.2%	2.0%	**3.8%**	0.9%	1.8%	1.2%	6.6%	2.9%	4.7%	2.0%	5.4%	3.7%	8.5%	6.9%	7.7%	4.8%
		$EGARCH_{sstd}$	0.2%	0.0%	4.5%	1.5%	2.7%	0.5%	1.0%	0.4%	5.6%	2.4%	3.5%	1.0%	2.2%	1.4%	6.7%	3.6%	4.4%	1.5%	5.0%	**4.9%**	9.7%	6.7%	7.7%	4.8%
	1000 Observations	$GARCH_{std}$	0.6%	0.0%	6.0%	**2.3%**	2.3%	0.2%	0.8%	0.0%	**8.2%**	**2.7%**	3.3%	1.0%	1.6%	1.0%	**10.0%**	3.8%	**5.4%**	**2.5%**	4.0%	**4.3%**	**11.9%**	6.9%	7.3%	**5.5%**
		$GARCH_{sstd}$	0.7%	0.1%	4.6%	1.3%	2.4%	0.3%	1.2%	0.1%	5.9%	2.4%	3.6%	0.5%	2.1%	1.2%	8.3%	3.7%	4.5%	1.3%	4.8%	3.9%	11.0%	6.0%	7.5%	4.0%
		$EGARCH_{norm}$	**0.9%**	0.1%	5.7%	1.1%	2.1%	**0.8%**	**1.9%**	0.3%	6.7%	2.5%	3.5%	1.2%	**2.5%**	**1.6%**	7.8%	4.1%	4.4%	1.9%	4.8%	**4.3%**	10.3%	7.1%	7.3%	4.7%
		$EGARCH_{std}$	0.8%	0.1%	**6.1%**	1.2%	**2.5%**	0.6%	1.6%	0.3%	7.5%	2.0%	**4.1%**	1.0%	2.5%	1.4%	8.4%	**4.2%**	5.2%	2.4%	4.8%	3.9%	10.0%	**7.5%**	7.7%	4.8%
		$EGARCH_{sstd}$	0.6%	**0.2%**	5.0%	1.1%	2.1%	0.7%	1.1%	**0.7%**	6.5%	1.8%	3.1%	**1.4%**	1.9%	**1.6%**	8.0%	3.6%	4.5%	2.3%	**5.2%**	4.0%	10.4%	6.4%	**7.9%**	5.3%
MFE Test	**250 Observations**	$GARCH_{std}$	**0.7%**	0.0%	**4.7%**	**1.9%**	**3.0%**	**2.0%**	1.1%	0.3%	**5.8%**	**3.4%**	3.5%	**0.9%**	**2.8%**	**1.6%**	**7.7%**	**4.5%**	**5.1%**	**2.4%**	5.4%	3.8%	9.8%	**7.1%**	**8.4%**	**6.0%**
		$GARCH_{sstd}$	0.5%	0.0%	3.9%	1.2%	2.3%	0.1%	1.3%	0.5%	5.0%	2.0%	3.4%	0.5%	2.1%	**1.6%**	7.4%	3.0%	4.5%	1.8%	4.7%	3.7%	**10.1%**	6.0%	7.2%	5.0%
		$EGARCH_{norm}$	0.5%	**0.3%**	4.3%	1.7%	2.3%	0.3%	0.9%	**0.7%**	5.0%	2.5%	3.3%	0.7%	2.1%	1.4%	6.7%	3.8%	4.3%	1.9%	**5.5%**	4.4%	8.4%	6.7%	6.9%	4.6%
		$EGARCH_{std}$	0.4%	0.2%	3.9%	1.5%	**3.0%**	0.4%	**1.5%**	0.4%	5.1%	2.0%	**3.7%**	**0.9%**	2.0%	1.3%	6.4%	3.0%	4.7%	2.0%	5.3%	3.5%	8.8%	6.6%	7.7%	5.0%
		$EGARCH_{sstd}$	0.2%	0.0%	4.6%	1.5%	2.7%	0.4%	0.8%	0.4%	5.6%	2.3%	3.5%	0.8%	2.1%	1.4%	6.7%	3.2%	4.6%	1.5%	5.2%	**4.8%**	9.6%	6.5%	7.6%	4.4%
	1000 Observations	$GARCH_{std}$	0.6%	0.0%	6.0%	**2.3%**	2.4%	0.2%	0.8%	0.0%	**8.2%**	**2.8%**	3.3%	1.0%	1.6%	1.0%	**10.1%**	3.6%	**5.3%**	**2.6%**	4.1%	**4.6%**	**11.7%**	6.8%	7.6%	**5.5%**
		$GARCH_{sstd}$	0.7%	0.0%	4.7%	1.3%	**2.5%**	0.3%	1.1%	0.1%	6.0%	2.4%	3.6%	0.5%	2.1%	1.0%	8.4%	3.7%	4.5%	1.3%	4.6%	3.7%	11.0%	6.0%	7.9%	3.9%
		$EGARCH_{norm}$	**0.8%**	0.1%	5.8%	1.2%	2.3%	**0.8%**	**1.7%**	0.4%	6.7%	2.5%	3.7%	1.2%	**2.6%**	1.6%	7.8%	4.1%	4.3%	2.0%	4.6%	4.4%	10.5%	7.3%	7.2%	4.6%
		$EGARCH_{std}$	**0.8%**	0.1%	**6.1%**	1.4%	2.4%	0.5%	1.6%	0.3%	7.4%	2.2%	**4.1%**	1.0%	2.3%	1.6%	8.5%	**4.2%**	5.1%	2.4%	4.4%	3.7%	10.4%	**7.5%**	7.9%	4.9%
		$EGARCH_{sstd}$	0.7%	**0.2%**	5.0%	1.1%	2.2%	0.7%	1.1%	**0.8%**	6.3%	1.9%	3.1%	**1.4%**	1.9%	**1.7%**	8.0%	3.6%	4.5%	2.4%	**4.8%**	4.1%	10.5%	6.3%	**8.2%**	5.4%

Table 23 – Proposed McNeil and Frey (2000), Righi and Ceretta (2015) and McNeil et al. (2015) backtesting size results, using GARCH$_{std}$ as null hypothesis with 250 and 1000 in-sample observations and significance level of 1%, 2.5% and 5%.

Backtesting	**Size**	**Model**	1%						2.5%						5%						10%					
			$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$
			250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	**250 Observations**	GARCH$_{norm}$	**2.2%**	**0.6%**	6.3%	**2.9%**	**3.9%**	0.5%	**3.1%**	**1.5%**	**8.1%**	**4.1%**	**5.4%**	1.0%	**5.0%**	**3.5%**	**9.7%**	**5.7%**	**7.1%**	1.9%	**8.4%**	6.8%	**12.0%**	8.5%	9.3%	5.8%
		GARCH$_{sstd}$	0.5%	0.1%	4.6%	1.2%	2.6%	0.0%	0.7%	0.2%	5.8%	1.8%	3.6%	0.5%	1.6%	1.6%	7.1%	2.7%	4.5%	1.3%	4.8%	4.5%	9.7%	5.8%	7.3%	4.3%
		EGARCH$_{norm}$	1.1%	0.2%	**7.0%**	2.2%	3.4%	**0.6%**	2.1%	1.0%	**8.1%**	3.6%	4.5%	**1.7%**	3.7%	2.6%	9.6%	5.2%	**7.1%**	**3.1%**	7.2%	**7.7%**	11.7%	**9.3%**	**9.9%**	**6.5%**
		EGARCH$_{std}$	0.6%	0.2%	4.3%	0.6%	2.4%	0.4%	1.1%	0.7%	5.7%	1.2%	3.4%	0.9%	1.8%	2.0%	6.9%	1.9%	4.5%	2.1%	5.3%	5.8%	8.6%	4.9%	7.0%	5.8%
		EGARCH$_{sstd}$	0.0%	0.2%	3.2%	0.7%	2.3%	0.5%	0.7%	0.4%	4.3%	1.7%	2.9%	1.0%	1.3%	1.1%	5.2%	2.7%	4.4%	2.3%	4.2%	3.5%	8.3%	5.1%	8.0%	4.4%
	1000 Observations	GARCH$_{norm}$	**1.6%**	0.0%	5.0%	2.6%	**4.4%**	0.8%	2.3%	0.8%	6.3%	3.3%	**5.8%**	1.3%	**3.9%**	**2.1%**	7.8%	3.5%	7.6%	3.2%	**7.9%**	5.1%	9.4%	6.7%	10.8%	5.9%
		GARCH$_{sstd}$	0.5%	0.1%	3.6%	1.0%	3.6%	0.7%	0.9%	0.6%	4.8%	1.7%	4.6%	1.1%	2.0%	1.2%	6.6%	2.4%	5.2%	1.9%	4.5%	3.5%	9.5%	4.1%	8.1%	5.3%
		EGARCH$_{norm}$	1.2%	**0.2%**	**5.4%**	**2.9%**	4.2%	**1.2%**	**2.4%**	**1.1%**	**6.9%**	**3.6%**	5.7%	**2.0%**	3.5%	2.0%	**8.5%**	**5.5%**	**8.3%**	**3.3%**	6.5%	**5.2%**	**11.3%**	**7.1%**	**10.9%**	**6.5%**
		EGARCH$_{std}$	0.3%	0.0%	4.1%	0.7%	2.7%	0.5%	0.6%	0.3%	5.5%	1.4%	3.8%	1.4%	1.8%	1.0%	7.1%	2.6%	5.5%	2.6%	4.5%	3.0%	9.5%	5.6%	8.5%	5.6%
		EGARCH$_{sstd}$	0.2%	0.1%	4.8%	1.0%	2.8%	0.2%	0.9%	0.2%	5.6%	1.8%	4.2%	0.5%	1.4%	0.9%	7.2%	3.1%	5.9%	1.2%	4.0%	3.5%	8.8%	6.0%	8.5%	4.8%
RC Test	**250 Observations**	GARCH$_{norm}$	**2.2%**	**0.6%**	6.3%	**2.9%**	**4.0%**	0.5%	**3.2%**	**1.3%**	**8.1%**	**4.4%**	**5.4%**	1.0%	**5.0%**	**3.4%**	**9.8%**	**6.1%**	**7.1%**	2.0%	**8.3%**	6.7%	**12.0%**	8.5%	9.4%	5.8%
		GARCH$_{sstd}$	0.5%	0.1%	5.0%	1.2%	2.7%	0.0%	0.7%	0.6%	5.8%	2.0%	3.8%	0.4%	2.0%	1.4%	7.0%	2.9%	4.5%	1.7%	5.5%	4.5%	9.8%	5.9%	7.4%	4.1%
		EGARCH$_{norm}$	1.1%	0.2%	**7.0%**	2.2%	3.4%	**0.7%**	2.2%	1.1%	8.0%	3.4%	4.6%	**1.6%**	3.5%	2.5%	9.7%	5.3%	**7.1%**	**3.3%**	7.2%	**8.0%**	11.8%	**8.7%**	**10.0%**	**6.5%**
		EGARCH$_{std}$	0.5%	0.2%	4.3%	0.8%	2.4%	0.4%	1.2%	0.8%	5.4%	1.2%	3.4%	0.9%	1.8%	2.0%	6.9%	2.2%	4.5%	2.3%	5.4%	5.9%	8.8%	5.7%	7.1%	5.9%
		EGARCH$_{sstd}$	0.1%	0.2%	3.0%	0.7%	2.3%	0.5%	0.7%	0.5%	4.0%	1.7%	3.0%	0.9%	1.6%	1.1%	5.3%	2.8%	4.6%	2.2%	4.4%	4.0%	8.0%	5.1%	7.7%	4.2%
	1000 Observations	GARCH$_{norm}$	**1.4%**	0.0%	5.2%	**2.7%**	**4.4%**	0.8%	**2.3%**	0.6%	6.1%	3.3%	**5.7%**	1.3%	**3.9%**	2.1%	7.8%	3.5%	7.4%	3.2%	**7.7%**	5.0%	9.6%	6.7%	**11.1%**	5.7%
		GARCH$_{sstd}$	0.5%	0.0%	3.7%	1.0%	3.7%	0.8%	0.9%	0.6%	4.8%	1.8%	4.6%	1.1%	2.2%	1.1%	6.6%	2.5%	5.1%	1.9%	4.5%	3.6%	9.1%	4.1%	8.1%	4.9%
		EGARCH$_{norm}$	1.2%	**0.2%**	**5.5%**	**2.7%**	4.2%	**1.2%**	**2.3%**	**1.1%**	**6.9%**	**3.6%**	5.6%	**1.9%**	3.6%	**2.3%**	**8.6%**	**5.2%**	**8.5%**	**3.4%**	6.8%	**5.3%**	**11.1%**	**7.2%**	10.8%	**6.2%**
		EGARCH$_{std}$	0.3%	0.0%	4.1%	0.6%	2.5%	0.4%	0.6%	0.4%	5.5%	1.4%	3.7%	1.2%	1.8%	0.9%	7.1%	2.4%	5.5%	2.6%	4.6%	3.0%	9.3%	5.7%	8.4%	5.9%
		EGARCH$_{sstd}$	0.2%	0.1%	4.9%	1.0%	2.8%	0.2%	0.8%	0.2%	5.7%	1.8%	4.2%	0.5%	1.4%	0.8%	7.3%	2.9%	5.8%	1.3%	4.2%	3.4%	9.2%	6.0%	8.2%	5.1%
MFE Test	**250 Observations**	GARCH$_{norm}$	**2.2%**	**0.6%**	6.3%	**2.9%**	**4.1%**	0.4%	**3.2%**	**1.4%**	**8.1%**	**4.1%**	**5.4%**	1.0%	**5.0%**	**3.5%**	**9.7%**	**5.8%**	**7.1%**	2.0%	**8.5%**	6.9%	**12.2%**	8.7%	9.3%	5.8%
		GARCH$_{sstd}$	0.5%	0.1%	4.7%	1.3%	2.6%	0.0%	0.7%	0.5%	5.7%	1.9%	3.7%	0.4%	1.7%	1.5%	6.9%	2.7%	4.6%	1.4%	5.3%	4.4%	9.7%	5.6%	7.5%	4.4%
		EGARCH$_{norm}$	1.1%	0.2%	**6.8%**	2.2%	3.5%	**0.7%**	2.3%	1.2%	**8.1%**	3.5%	4.4%	**1.6%**	3.6%	2.8%	9.7%	5.4%	7.0%	**3.1%**	7.2%	**7.9%**	11.7%	**9.1%**	**9.9%**	**6.6%**
		EGARCH$_{std}$	0.6%	0.3%	4.4%	0.7%	2.4%	0.5%	1.1%	0.7%	5.7%	1.3%	3.4%	0.9%	1.8%	1.9%	7.1%	2.0%	4.4%	2.0%	5.4%	5.9%	8.7%	5.2%	7.0%	6.2%
		EGARCH$_{sstd}$	0.0%	0.2%	3.2%	0.7%	2.3%	0.5%	0.8%	0.5%	4.2%	1.8%	3.0%	1.1%	1.1%	1.1%	5.1%	2.9%	4.3%	2.1%	3.8%	3.3%	8.2%	5.0%	8.0%	4.5%
	1000 Observations	GARCH$_{norm}$	**1.6%**	0.0%	5.2%	2.6%	**4.5%**	0.8%	**2.3%**	0.7%	6.2%	3.3%	**5.9%**	1.3%	**4.0%**	**2.0%**	7.8%	3.6%	7.6%	3.2%	**7.7%**	5.2%	9.6%	6.7%	**10.8%**	5.9%
		GARCH$_{sstd}$	0.5%	0.0%	3.6%	1.0%	3.6%	0.9%	1.0%	0.6%	4.9%	1.8%	4.6%	1.1%	2.1%	1.2%	6.7%	2.5%	5.2%	1.8%	4.6%	3.5%	9.4%	4.2%	8.0%	5.1%
		EGARCH$_{norm}$	1.2%	**0.2%**	**5.5%**	**2.9%**	4.2%	**1.1%**	**2.3%**	**1.0%**	**6.8%**	**3.5%**	**5.9%**	**2.0%**	3.7%	**2.0%**	**8.5%**	**5.1%**	**8.5%**	**3.4%**	6.9%	**5.3%**	**11.0%**	**7.2%**	**10.8%**	**6.2%**
		EGARCH$_{std}$	0.3%	0.0%	4.2%	0.7%	2.6%	0.5%	0.5%	0.3%	5.6%	1.6%	3.8%	1.3%	1.8%	0.9%	7.0%	2.5%	5.3%	2.4%	4.3%	3.0%	9.5%	5.6%	8.7%	6.0%
		EGARCH$_{sstd}$	0.2%	0.1%	4.9%	1.0%	2.9%	0.2%	0.9%	0.2%	5.7%	1.8%	4.3%	0.5%	1.5%	0.9%	7.4%	3.3%	5.8%	1.3%	3.9%	3.6%	9.2%	6.1%	8.4%	5.0%

Table 24 – Proposed McNeil and Frey (2000), Righi and Ceretta (2015) and McNeil et al. (2015) backtesting size results, using GARCH$_{sstd}$ as null hypothesis with 250 and 1000 in-sample observations and significance level of 1%, 2.5% and 5%.

Backtesting	Size	Model	1%						2.5%						5%						10%					
			$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
			250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	**250 Observations**	GARCH$_{norm}$	**2.2%**	0.2%	4.5%	**2.0%**	**4.0%**	0.2%	**2.9%**	1.1%	5.8%	2.9%	**5.1%**	1.0%	**5.2%**	2.5%	7.6%	4.6%	**6.0%**	1.9%	**9.8%**	8.2%	10.1%	**8.5%**	9.6%	5.3%
		GARCH$_{std}$	0.5%	0.3%	4.5%	1.7%	2.1%	0.5%	1.1%	0.5%	5.7%	2.3%	3.7%	0.9%	2.1%	1.7%	7.1%	4.0%	4.7%	2.5%	6.0%	6.1%	8.7%	6.3%	7.4%	5.6%
		EGARCH$_{norm}$	0.8%	0.4%	5.2%	1.8%	3.8%	**1.0%**	1.7%	**1.5%**	6.9%	**3.2%**	4.6%	**1.5%**	4.4%	**4.1%**	**9.4%**	**5.2%**	**6.0%**	**2.8%**	9.4%	**8.9%**	**11.0%**	7.6%	**9.7%**	**6.5%**
		EGARCH$_{std}$	0.8%	**0.5%**	2.9%	1.2%	1.5%	0.7%	1.2%	0.9%	4.2%	2.0%	2.3%	1.1%	2.3%	2.0%	5.9%	3.1%	3.4%	2.3%	5.9%	5.8%	8.5%	6.5%	6.2%	4.7%
		EGARCH$_{sstd}$	0.3%	0.2%	**6.2%**	1.9%	2.1%	0.5%	1.0%	0.9%	**7.2%**	2.5%	3.1%	1.3%	2.2%	1.9%	9.2%	4.0%	4.7%	1.9%	5.4%	5.4%	10.5%	7.1%	7.1%	5.4%
	1000 Observations	GARCH$_{norm}$	**1.2%**	**0.4%**	5.5%	2.2%	3.9%	**1.0%**	2.2%	**1.1%**	7.0%	3.7%	**5.3%**	**2.0%**	**3.3%**	**2.7%**	8.5%	**5.4%**	**7.2%**	**3.6%**	**7.1%**	**7.0%**	11.2%	8.3%	**10.1%**	**7.1%**
		GARCH$_{std}$	0.5%	0.0%	4.1%	1.4%	1.7%	0.2%	0.9%	0.5%	5.3%	2.5%	2.0%	0.7%	2.4%	1.7%	7.3%	3.9%	4.2%	1.8%	5.7%	5.4%	10.0%	6.3%	6.6%	5.7%
		EGARCH$_{norm}$	**1.2%**	0.0%	**6.4%**	**2.6%**	**4.1%**	0.5%	**2.3%**	0.4%	**7.6%**	**4.2%**	5.1%	1.5%	**3.3%**	2.0%	**9.3%**	5.3%	6.3%	2.5%	6.2%	6.3%	**11.9%**	**8.5%**	8.7%	5.8%
		EGARCH$_{std}$	0.6%	0.1%	4.4%	0.8%	1.9%	0.5%	1.0%	0.3%	5.7%	1.4%	3.1%	0.9%	2.7%	1.6%	7.6%	3.0%	4.8%	2.1%	5.7%	5.0%	10.6%	5.7%	7.7%	3.6%
		EGARCH$_{sstd}$	0.5%	0.2%	4.2%	1.4%	1.7%	0.1%	1.1%	0.4%	5.5%	2.0%	2.5%	0.5%	1.6%	1.0%	6.8%	3.2%	3.5%	1.4%	5.1%	3.3%	8.7%	6.3%	5.9%	3.9%
RC Test	**250 Observations**	GARCH$_{norm}$	**2.1%**	0.3%	4.5%	**2.1%**	**4.0%**	0.2%	**2.9%**	1.0%	5.8%	**3.1%**	**5.1%**	0.9%	**5.2%**	2.5%	7.7%	4.8%	**6.1%**	1.8%	**9.7%**	8.5%	10.2%	**8.5%**	9.5%	5.4%
		GARCH$_{std}$	0.6%	0.2%	4.5%	1.6%	2.3%	0.6%	1.1%	0.8%	5.6%	2.2%	3.6%	1.1%	2.4%	2.1%	7.1%	3.9%	4.8%	2.4%	5.9%	6.1%	8.6%	6.7%	7.5%	5.3%
		EGARCH$_{norm}$	0.8%	**0.4%**	5.2%	1.8%	3.8%	**0.9%**	1.7%	**1.3%**	6.9%	3.1%	4.6%	**1.5%**	4.4%	**4.2%**	**9.5%**	**4.9%**	5.9%	**2.8%**	9.3%	**8.9%**	**11.0%**	7.8%	**9.6%**	**6.6%**
		EGARCH$_{std}$	0.8%	**0.4%**	3.2%	1.3%	1.6%	0.6%	1.2%	1.0%	4.2%	2.2%	2.2%	1.2%	2.1%	1.9%	5.8%	2.9%	3.6%	2.2%	5.5%	6.0%	8.3%	6.6%	6.4%	4.8%
		EGARCH$_{sstd}$	0.4%	0.3%	**6.2%**	1.8%	2.0%	0.5%	1.1%	0.8%	**7.3%**	2.7%	3.4%	1.3%	2.0%	2.1%	9.2%	4.0%	4.4%	1.9%	5.4%	5.4%	10.5%	7.0%	7.0%	5.3%
	1000 Observations	GARCH$_{norm}$	**1.2%**	**0.4%**	5.5%	2.3%	4.0%	**1.1%**	**2.4%**	**1.0%**	7.1%	3.8%	**5.3%**	**2.1%**	**3.4%**	**2.8%**	8.5%	**5.5%**	**7.2%**	**3.7%**	**7.1%**	**6.9%**	11.3%	8.3%	**9.8%**	**7.2%**
		GARCH$_{std}$	0.5%	0.0%	3.8%	1.2%	1.7%	0.2%	0.8%	0.6%	5.3%	2.2%	2.1%	0.7%	2.3%	1.8%	7.3%	4.0%	4.3%	1.7%	5.8%	5.2%	10.3%	6.3%	6.6%	5.8%
		EGARCH$_{norm}$	1.1%	0.0%	**6.4%**	**2.5%**	**4.1%**	0.5%	2.3%	0.5%	**7.6%**	**4.1%**	5.0%	1.4%	3.2%	2.0%	**9.3%**	5.4%	6.4%	2.5%	6.2%	6.3%	**11.7%**	**8.7%**	8.8%	5.8%
		EGARCH$_{std}$	0.6%	0.1%	4.6%	0.8%	1.8%	0.5%	1.3%	0.3%	5.7%	1.4%	2.8%	1.0%	2.7%	1.7%	7.6%	3.0%	4.8%	2.2%	5.9%	5.4%	10.8%	5.7%	7.7%	3.9%
		EGARCH$_{sstd}$	0.6%	0.2%	4.0%	1.4%	1.7%	0.1%	1.2%	0.2%	5.6%	2.0%	2.6%	0.6%	1.8%	0.9%	6.7%	3.3%	3.3%	1.2%	4.7%	3.8%	8.6%	6.7%	6.1%	3.8%
MFE Test	**250 Observations**	GARCH$_{norm}$	**2.2%**	0.3%	4.5%	**2.0%**	**4.0%**	0.2%	**2.7%**	1.1%	5.8%	**3.0%**	**5.1%**	0.8%	**5.2%**	2.5%	7.6%	4.5%	**6.2%**	1.9%	**9.8%**	8.1%	10.3%	**8.8%**	9.6%	5.2%
		GARCH$_{std}$	0.5%	0.3%	4.6%	1.7%	2.1%	0.5%	1.2%	0.7%	5.7%	2.3%	3.7%	1.1%	2.2%	1.8%	7.1%	3.7%	4.7%	2.5%	6.2%	6.1%	8.5%	6.6%	7.7%	5.4%
		EGARCH$_{norm}$	0.8%	0.3%	5.2%	1.8%	3.8%	**1.0%**	1.7%	**1.5%**	6.9%	**3.0%**	4.7%	**1.5%**	4.6%	**4.0%**	**9.5%**	**5.1%**	6.1%	**2.8%**	9.7%	**9.2%**	**11.2%**	7.9%	**9.7%**	**6.4%**
		EGARCH$_{std}$	0.8%	**0.5%**	2.8%	1.2%	1.5%	0.8%	1.2%	1.1%	4.2%	2.1%	2.2%	1.1%	2.2%	2.0%	5.9%	3.0%	3.5%	2.4%	5.9%	5.7%	8.3%	6.4%	6.3%	4.8%
		EGARCH$_{sstd}$	0.4%	0.3%	**6.3%**	1.9%	2.0%	0.5%	1.0%	0.7%	**7.2%**	2.6%	3.1%	1.3%	2.2%	2.1%	9.1%	3.8%	4.6%	1.7%	5.5%	5.4%	10.6%	7.2%	6.9%	5.7%
	1000 Observations	GARCH$_{norm}$	**1.2%**	**0.4%**	5.5%	2.3%	4.0%	**1.1%**	**2.3%**	**1.1%**	7.0%	3.8%	**5.3%**	**2.0%**	3.1%	**2.9%**	8.5%	5.3%	**7.2%**	**3.6%**	**7.0%**	**6.6%**	11.1%	8.4%	**10.1%**	**7.3%**
		GARCH$_{std}$	0.5%	0.0%	4.1%	1.3%	1.7%	0.2%	0.9%	0.5%	5.4%	2.4%	2.1%	0.7%	2.4%	1.6%	7.3%	4.0%	4.1%	1.8%	5.6%	5.4%	10.2%	6.4%	6.8%	5.7%
		EGARCH$_{norm}$	**1.2%**	0.0%	**6.4%**	**2.5%**	**4.1%**	0.5%	**2.3%**	0.3%	**7.6%**	**4.1%**	5.1%	1.5%	**3.3%**	2.0%	**9.3%**	**5.4%**	6.4%	2.6%	6.3%	6.1%	**11.8%**	**8.7%**	9.0%	5.7%
		EGARCH$_{std}$	0.6%	0.1%	4.6%	0.8%	1.8%	0.5%	1.2%	0.2%	5.7%	1.3%	2.9%	0.9%	2.7%	1.6%	7.7%	2.9%	4.8%	1.9%	5.8%	5.3%	10.7%	5.7%	7.7%	4.0%
		EGARCH$_{sstd}$	0.6%	0.2%	4.2%	1.4%	1.7%	0.1%	1.2%	0.5%	5.6%	2.0%	2.5%	0.5%	1.8%	1.1%	6.6%	3.3%	3.5%	1.3%	4.6%	3.4%	8.7%	6.5%	6.0%	3.7%

Table 25 – Proposed McNeil and Frey (2000), Righi and Ceretta (2015) and McNeil et al. (2015) backtesting size results, using EGARCH$_{norm}$ as null hypothesis with 250 and 1000 in-sample observations and significance level of 1%, 2.5% and 5%.

Backtesting	Size	Model	1%						2.5%						5%						10%					
			$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
			250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	**250 Observations**	GARCH$_{norm}$	**0.8%**	**0.3%**	3.2%	**1.3%**	2.1%	0.4%	**1.4%**	**0.9%**	4.4%	**2.5%**	3.5%	0.8%	**2.9%**	2.2%	5.9%	3.2%	**5.6%**	**2.4%**	**6.8%**	**6.3%**	8.6%	**7.3%**	9.2%	**6.4%**
		GARCH$_{std}$	0.2%	**0.3%**	3.5%	0.4%	1.1%	**0.6%**	0.9%	**0.9%**	4.4%	1.3%	2.4%	0.8%	2.8%	**2.5%**	5.3%	2.4%	5.1%	1.9%	6.4%	5.8%	8.1%	4.8%	**9.6%**	6.1%
		GARCH$_{sstd}$	0.1%	0.2%	3.7%	0.6%	**3.0%**	0.4%	0.8%	**0.9%**	5.2%	1.4%	**4.1%**	**0.9%**	2.5%	1.6%	6.7%	2.6%	5.3%	2.2%	5.1%	4.3%	8.7%	5.2%	8.3%	5.3%
		EGARCH$_{std}$	0.2%	0.0%	2.8%	**1.3%**	1.7%	0.3%	0.9%	0.5%	4.5%	1.7%	2.4%	0.7%	1.6%	1.5%	6.6%	2.7%	4.0%	1.8%	3.8%	4.8%	9.0%	5.6%	6.5%	4.7%
		EGARCH$_{sstd}$	0.5%	0.1%	**4.3%**	1.0%	2.4%	0.1%	0.9%	0.4%	**5.8%**	1.4%	3.2%	0.6%	1.7%	1.3%	**7.8%**	**3.4%**	4.6%	**2.4%**	5.2%	4.5%	**10.2%**	6.9%	7.4%	5.3%
	1000 Observations	GARCH$_{norm}$	0.7%	0.2%	4.2%	1.3%	1.6%	**0.6%**	1.2%	0.5%	**6.4%**	2.0%	2.6%	1.1%	2.3%	**1.9%**	**8.6%**	3.9%	4.0%	2.1%	**6.8%**	**6.3%**	**11.1%**	6.6%	6.0%	5.4%
		GARCH$_{std}$	0.5%	0.2%	4.0%	1.5%	1.4%	0.4%	1.0%	0.6%	5.4%	2.8%	2.3%	1.1%	1.7%	**1.9%**	7.5%	3.9%	3.5%	2.0%	5.1%	5.7%	9.9%	**7.2%**	5.6%	5.8%
		GARCH$_{sstd}$	0.5%	**0.4%**	4.2%	1.4%	2.0%	0.5%	1.3%	**1.0%**	5.3%	2.3%	2.4%	0.9%	2.1%	**1.9%**	7.4%	3.7%	3.2%	1.8%	5.9%	5.1%	10.5%	6.7%	6.3%	4.6%
		EGARCH$_{std}$	0.4%	0.1%	**5.1%**	**1.9%**	2.0%	**0.6%**	1.2%	0.4%	6.1%	3.2%	3.2%	**1.2%**	2.1%	1.3%	8.5%	**4.5%**	4.5%	**2.2%**	4.6%	4.1%	10.6%	6.6%	7.6%	**6.3%**
		EGARCH$_{sstd}$	**0.9%**	0.1%	4.9%	**1.9%**	**2.5%**	0.3%	**1.6%**	0.3%	5.9%	**3.3%**	**3.5%**	0.9%	**2.4%**	1.6%	7.4%	4.3%	**5.2%**	2.0%	4.7%	4.4%	9.8%	7.0%	**9.0%**	5.5%
RC Test	**250 Observations**	GARCH$_{norm}$	**0.8%**	**0.3%**	3.2%	**1.3%**	2.1%	0.4%	**1.3%**	**1.1%**	4.3%	**2.3%**	3.5%	1.0%	**3.1%**	1.9%	5.8%	3.2%	**5.7%**	**2.4%**	**7.1%**	**6.2%**	8.7%	**7.4%**	9.1%	**6.4%**
		GARCH$_{std}$	0.2%	0.2%	3.4%	0.5%	1.3%	**0.5%**	1.1%	1.0%	4.3%	1.2%	2.6%	0.8%	3.0%	**2.4%**	5.4%	2.4%	5.4%	2.1%	7.0%	6.1%	8.0%	5.0%	**9.6%**	5.8%
		GARCH$_{sstd}$	0.1%	0.2%	4.0%	0.6%	**2.9%**	0.4%	0.9%	1.0%	5.2%	1.3%	**4.1%**	**1.1%**	2.6%	1.7%	6.6%	2.7%	5.4%	2.1%	5.2%	4.8%	8.5%	5.4%	8.0%	5.2%
		EGARCH$_{std}$	0.2%	0.0%	2.9%	1.3%	1.7%	0.2%	0.9%	0.5%	4.4%	1.8%	2.5%	0.7%	1.8%	1.3%	6.5%	2.6%	4.0%	1.7%	3.8%	4.6%	8.8%	5.9%	6.7%	4.4%
		EGARCH$_{sstd}$	0.5%	0.0%	**4.5%**	1.0%	2.5%	0.2%	0.8%	0.6%	**6.0%**	1.6%	3.1%	0.6%	1.6%	1.3%	**8.0%**	**3.5%**	4.5%	2.3%	4.8%	4.3%	**10.3%**	6.8%	7.7%	5.1%
	1000 Observations	GARCH$_{norm}$	0.7%	0.2%	4.4%	1.3%	1.6%	**0.7%**	1.2%	0.4%	**6.6%**	2.0%	2.5%	1.1%	**2.3%**	**1.9%**	8.5%	3.7%	4.1%	2.0%	**6.4%**	**6.6%**	**11.2%**	6.2%	5.9%	5.1%
		GARCH$_{std}$	0.5%	0.2%	4.0%	1.5%	1.5%	0.4%	1.1%	0.5%	5.5%	3.0%	2.2%	**1.2%**	1.6%	**1.9%**	7.5%	3.9%	3.7%	2.1%	5.1%	5.6%	9.7%	**7.3%**	5.7%	5.8%
		GARCH$_{sstd}$	0.5%	**0.4%**	4.1%	1.3%	2.0%	0.5%	1.3%	**1.0%**	5.5%	2.1%	2.4%	0.9%	2.2%	1.8%	7.4%	3.6%	3.2%	1.9%	5.7%	5.2%	10.5%	6.5%	6.1%	4.7%
		EGARCH$_{std}$	0.5%	0.1%	**5.1%**	**2.0%**	2.0%	**0.7%**	1.2%	0.2%	6.1%	3.0%	3.4%	1.0%	2.1%	1.5%	**8.7%**	**4.3%**	4.4%	**2.2%**	4.6%	4.0%	10.5%	6.7%	7.6%	**6.1%**
		EGARCH$_{sstd}$	**0.9%**	0.1%	5.0%	2.0%	**2.5%**	0.4%	**1.6%**	0.2%	6.0%	**3.2%**	**3.6%**	1.1%	2.2%	1.3%	7.5%	4.1%	**5.2%**	2.0%	5.0%	4.4%	9.7%	6.9%	**8.8%**	5.6%
MFE Test	**250 Observations**	GARCH$_{norm}$	**0.8%**	**0.3%**	3.1%	**1.3%**	2.1%	0.4%	**1.4%**	**0.9%**	4.4%	**2.4%**	3.5%	**1.0%**	**2.9%**	2.1%	5.9%	3.2%	**5.7%**	**2.4%**	**7.0%**	**6.2%**	8.7%	**7.2%**	9.3%	**6.4%**
		GARCH$_{std}$	0.2%	0.2%	3.4%	0.5%	1.1%	**0.6%**	1.0%	0.8%	4.4%	1.2%	2.5%	0.8%	2.8%	**2.6%**	5.3%	2.6%	5.2%	1.8%	6.6%	6.0%	8.1%	4.9%	**9.5%**	6.0%
		GARCH$_{sstd}$	0.1%	0.2%	3.7%	0.6%	**2.9%**	0.3%	1.0%	**0.9%**	5.1%	1.4%	**4.1%**	**1.0%**	2.7%	1.7%	6.8%	2.6%	5.4%	2.3%	5.2%	4.3%	8.7%	5.4%	8.3%	5.7%
		EGARCH$_{std}$	0.2%	0.0%	2.8%	1.2%	1.7%	0.3%	0.9%	0.6%	4.4%	1.7%	2.5%	0.7%	1.8%	1.5%	6.6%	2.8%	3.9%	1.9%	4.1%	4.7%	8.8%	5.8%	6.5%	4.7%
		EGARCH$_{sstd}$	0.5%	0.1%	**4.4%**	1.0%	2.4%	0.2%	0.9%	0.5%	**5.9%**	1.5%	3.3%	0.5%	1.7%	1.4%	**8.0%**	**3.4%**	4.4%	**2.4%**	4.9%	4.4%	**10.3%**	6.8%	7.5%	5.4%
	1000 Observations	GARCH$_{norm}$	0.7%	0.3%	4.3%	1.2%	1.6%	**0.7%**	1.2%	0.5%	**6.5%**	1.9%	2.6%	1.2%	**2.4%**	**2.0%**	8.5%	3.9%	4.2%	1.9%	**6.5%**	**6.6%**	**11.2%**	6.5%	6.1%	5.2%
		GARCH$_{std}$	0.5%	0.2%	4.0%	1.4%	1.5%	0.4%	1.1%	0.5%	5.5%	2.7%	2.2%	**1.3%**	1.7%	2.0%	7.6%	4.0%	3.9%	2.0%	5.2%	5.5%	9.6%	**7.5%**	5.6%	5.6%
		GARCH$_{sstd}$	0.5%	**0.5%**	4.2%	1.3%	2.0%	0.5%	1.4%	**1.0%**	5.5%	2.4%	2.5%	0.9%	2.1%	1.8%	7.4%	3.7%	3.3%	1.8%	5.8%	5.2%	10.5%	6.5%	6.1%	4.9%
		EGARCH$_{std}$	0.4%	0.1%	**5.0%**	**2.1%**	2.0%	0.6%	1.2%	0.4%	6.3%	3.1%	3.2%	1.0%	1.9%	1.4%	**8.6%**	**4.4%**	4.2%	**2.2%**	4.5%	4.1%	10.5%	6.8%	7.5%	**6.0%**
		EGARCH$_{sstd}$	**0.8%**	0.1%	4.8%	1.9%	**2.6%**	0.3%	**1.5%**	0.1%	6.0%	**3.2%**	**3.6%**	1.0%	2.3%	1.2%	7.4%	4.0%	**5.3%**	2.0%	4.8%	4.1%	9.7%	6.8%	**8.7%**	5.5%

Table 26 – Proposed McNeil and Frey (2000), Righi and Ceretta (2015) and McNeil et al. (2015) backtesting size results, using EGARCH$_{std}$ as null hypothesis with 250 and 1000 in-sample observations and significance level of 1%, 2.5% and 5%.

Backtesting	Size	Model	1%						2.5%						5%						10%					
			$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=5\%$	$\alpha=1\%$	$\beta=2.5\%$	$\alpha=2.5\%$	$\beta=5\%$
			250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	**250 Observations**	GARCH$_{norm}$	1.0%	0.3%	4.8%	2.0%	3.0%	**0.7%**	1.7%	1.0%	6.0%	3.1%	**4.2%**	**1.4%**	**3.6%**	2.8%	6.8%	3.9%	**5.7%**	2.3%	**7.7%**	5.9%	9.0%	7.2%	7.6%	5.4%
		GARCH$_{std}$	0.2%	0.1%	3.3%	1.3%	2.3%	0.4%	0.5%	0.7%	4.4%	1.7%	3.4%	1.1%	1.6%	1.4%	6.1%	2.6%	4.8%	2.1%	4.7%	4.7%	8.9%	4.6%	7.3%	5.2%
		GARCH$_{sstd}$	0.3%	0.1%	2.8%	1.1%	2.6%	0.2%	1.0%	0.4%	4.5%	1.5%	3.6%	0.9%	2.1%	1.1%	6.1%	2.5%	5.1%	1.8%	5.3%	4.3%	8.2%	5.2%	8.5%	5.8%
		EGARCH$_{norm}$	**1.3%**	**0.4%**	**6.2%**	**3.0%**	**3.2%**	0.5%	**2.2%**	**1.1%**	**8.4%**	**4.0%**	**4.2%**	**1.4%**	3.3%	**3.0%**	**10.5%**	**5.9%**	**5.7%**	**2.7%**	7.2%	**7.2%**	**12.7%**	**8.3%**	**8.9%**	**6.0%**
		EGARCH$_{sstd}$	0.3%	0.3%	3.7%	1.3%	2.6%	0.5%	0.6%	0.8%	4.9%	2.9%	3.7%	1.1%	1.8%	1.5%	6.9%	4.3%	5.1%	2.0%	3.4%	4.2%	8.8%	7.4%	6.9%	5.2%
	1000 Observations	GARCH$_{norm}$	**1.0%**	**0.3%**	4.6%	2.2%	2.9%	1.0%	1.6%	**0.7%**	6.3%	2.8%	4.3%	**1.7%**	**2.9%**	**1.8%**	7.6%	4.6%	5.5%	**2.9%**	**6.6%**	**6.1%**	10.2%	7.6%	**7.6%**	5.1%
		GARCH$_{std}$	0.4%	0.2%	4.7%	0.5%	1.5%	0.4%	1.1%	0.3%	6.5%	1.4%	2.8%	0.7%	1.9%	0.9%	7.8%	2.1%	4.2%	2.4%	4.8%	3.6%	10.4%	4.9%	**7.6%**	**5.4%**
		GARCH$_{sstd}$	0.5%	0.1%	4.6%	0.9%	1.6%	0.3%	1.1%	0.3%	6.0%	1.8%	2.6%	0.8%	2.8%	1.2%	7.5%	3.4%	4.3%	1.7%	5.9%	3.9%	10.1%	6.6%	7.5%	**5.4%**
		EGARCH$_{norm}$	0.8%	0.1%	**6.0%**	**3.4%**	**3.8%**	**1.2%**	**1.7%**	0.6%	**7.9%**	**4.2%**	**4.7%**	**1.7%**	2.6%	**1.8%**	**9.1%**	**6.5%**	**6.0%**	2.7%	5.7%	5.0%	**11.3%**	**10.3%**	7.4%	4.9%
		EGARCH$_{sstd}$	0.2%	0.1%	4.0%	1.3%	2.5%	0.8%	0.6%	0.5%	5.1%	2.1%	2.9%	1.0%	1.5%	1.7%	6.4%	3.3%	4.5%	2.3%	5.5%	3.6%	8.8%	6.2%	6.5%	4.8%
RC Test	**250 Observations**	GARCH$_{norm}$	1.1%	0.3%	4.8%	2.0%	3.2%	**0.7%**	1.6%	1.1%	5.9%	3.2%	**4.3%**	**1.4%**	**3.7%**	2.5%	6.9%	3.8%	**5.6%**	2.3%	**7.4%**	5.8%	8.9%	7.1%	7.6%	5.2%
		GARCH$_{std}$	0.1%	0.2%	3.3%	1.2%	2.2%	0.5%	0.5%	0.7%	4.3%	1.7%	3.4%	1.0%	1.7%	1.6%	6.0%	2.5%	4.6%	2.1%	5.0%	4.8%	9.1%	5.1%	7.5%	5.4%
		GARCH$_{sstd}$	0.2%	0.1%	2.9%	1.2%	2.5%	0.2%	0.9%	0.4%	4.5%	1.5%	3.3%	0.9%	2.3%	1.1%	6.0%	3.0%	5.0%	1.4%	5.3%	4.3%	8.3%	5.6%	8.7%	5.9%
		EGARCH$_{norm}$	**1.3%**	**0.4%**	**6.2%**	**3.0%**	**3.3%**	0.5%	**2.2%**	**1.1%**	**8.4%**	**4.1%**	**4.3%**	1.3%	3.2%	**2.9%**	**10.5%**	**6.0%**	**5.6%**	**2.8%**	7.3%	**7.2%**	**13.0%**	**8.5%**	**8.8%**	**6.2%**
		EGARCH$_{sstd}$	0.3%	0.2%	3.7%	1.2%	2.7%	0.6%	0.6%	0.8%	4.6%	2.8%	3.7%	1.0%	1.7%	1.8%	6.7%	4.2%	5.1%	1.7%	4.3%	4.1%	8.9%	7.2%	6.9%	5.2%
	1000 Observations	GARCH$_{norm}$	**1.1%**	**0.3%**	4.7%	2.2%	3.0%	1.1%	1.6%	**0.8%**	6.4%	2.8%	4.3%	**1.9%**	**2.8%**	**1.7%**	7.6%	4.5%	5.5%	**2.8%**	**6.2%**	**6.4%**	10.0%	7.1%	**7.6%**	5.4%
		GARCH$_{std}$	0.4%	0.2%	4.8%	0.5%	1.5%	0.5%	1.1%	0.3%	6.5%	1.3%	3.0%	0.8%	1.8%	0.7%	7.7%	2.1%	4.3%	2.3%	4.8%	3.7%	10.0%	5.0%	7.4%	5.2%
		GARCH$_{sstd}$	0.5%	0.1%	4.7%	0.9%	1.5%	0.3%	1.0%	0.2%	6.1%	1.8%	2.6%	0.7%	2.5%	1.1%	7.6%	3.2%	4.4%	1.9%	5.6%	4.0%	9.9%	7.1%	7.4%	**5.7%**
		EGARCH$_{norm}$	0.8%	0.2%	**6.0%**	**3.3%**	**3.8%**	**1.2%**	**1.7%**	0.6%	**7.8%**	**4.2%**	**4.7%**	1.7%	2.5%	**1.7%**	**9.0%**	**6.7%**	**6.1%**	2.6%	5.7%	4.9%	**11.3%**	**10.7%**	7.5%	4.9%
		EGARCH$_{sstd}$	0.2%	0.2%	4.0%	1.3%	2.4%	0.9%	0.4%	0.6%	5.0%	1.8%	2.9%	1.0%	1.8%	1.5%	6.6%	3.3%	4.6%	2.1%	5.4%	3.6%	9.3%	5.9%	6.6%	4.8%
MFE Test	**250 Observations**	GARCH$_{norm}$	1.0%	0.3%	4.8%	1.9%	3.1%	**0.7%**	1.7%	**1.1%**	6.0%	3.0%	4.2%	**1.4%**	**3.6%**	2.6%	6.8%	3.9%	5.6%	2.3%	**7.6%**	5.8%	9.0%	7.4%	7.6%	5.6%
		GARCH$_{std}$	0.1%	0.1%	3.3%	1.2%	2.4%	0.4%	0.6%	0.7%	4.4%	1.7%	3.4%	1.1%	1.4%	1.6%	6.1%	2.6%	4.9%	2.1%	4.5%	4.6%	9.1%	4.7%	7.7%	5.5%
		GARCH$_{sstd}$	0.3%	0.1%	2.8%	1.2%	2.7%	0.1%	1.1%	0.4%	4.6%	1.5%	3.6%	0.9%	2.2%	1.1%	6.1%	2.7%	4.9%	1.8%	5.6%	4.5%	8.0%	5.5%	8.5%	5.9%
		EGARCH$_{norm}$	**1.2%**	**0.4%**	**6.0%**	**3.0%**	**3.3%**	0.5%	**2.2%**	1.0%	**8.4%**	**4.1%**	**4.3%**	1.2%	3.2%	**2.7%**	**10.8%**	**5.8%**	**5.7%**	**2.7%**	7.4%	**7.1%**	**12.7%**	**8.4%**	**8.6%**	**6.2%**
		EGARCH$_{sstd}$	0.3%	0.3%	3.8%	1.5%	2.5%	0.6%	0.5%	0.7%	4.6%	2.9%	3.7%	1.0%	1.6%	1.7%	6.9%	4.3%	5.0%	2.1%	4.0%	4.3%	8.9%	7.1%	6.7%	5.3%
	1000 Observations	GARCH$_{norm}$	**1.0%**	**0.3%**	4.7%	2.2%	3.0%	1.1%	1.6%	**0.8%**	6.4%	2.8%	4.3%	**1.8%**	**2.9%**	**1.8%**	7.6%	4.6%	5.3%	**2.9%**	**6.4%**	**6.0%**	10.1%	7.2%	**7.6%**	5.2%
		GARCH$_{std}$	0.4%	0.2%	4.7%	0.5%	1.5%	0.4%	1.0%	0.3%	6.5%	1.3%	2.9%	0.8%	1.9%	0.8%	7.7%	2.1%	4.4%	2.4%	4.6%	3.8%	10.1%	5.4%	7.2%	5.3%
		GARCH$_{sstd}$	0.5%	0.1%	4.7%	0.9%	1.5%	0.3%	1.0%	0.2%	6.1%	1.8%	2.6%	0.8%	2.6%	1.2%	7.4%	3.4%	4.4%	1.7%	5.5%	3.9%	9.9%	6.7%	7.5%	**5.5%**
		EGARCH$_{norm}$	0.8%	0.1%	**6.1%**	**3.4%**	**3.8%**	**1.2%**	**1.7%**	0.6%	**7.8%**	**4.3%**	**4.7%**	1.7%	2.5%	1.7%	**9.1%**	**6.8%**	**6.1%**	2.5%	5.8%	5.0%	**11.3%**	**10.3%**	7.4%	4.7%
		EGARCH$_{sstd}$	0.2%	0.3%	4.2%	1.1%	2.4%	0.9%	0.5%	0.6%	5.1%	1.7%	3.1%	1.0%	1.3%	1.6%	6.5%	3.2%	4.6%	2.2%	5.2%	3.6%	9.3%	5.8%	6.5%	4.8%

Table 27 – Proposed McNeil and Frey (2000), Righi and Ceretta (2015) and McNeil et al. (2015) backtesting size results, using $EGARCH_{sstd}$ as null hypothesis with 250 and 1000 in-sample observations and significance level of 1%, 2.5% and 5%.

Backtesting	Size	Model	1%						2.5%						5%						10%					
			$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 5\%$	$\alpha = 1\%$	$\beta = 2.5\%$	$\alpha = 2.5\%$	$\beta = 5\%$
			250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500	250	500
MF Test	**250 Observations**	$GARCH_{norm}$	**1.0%**	0.2%	5.0%	1.8%	**3.8%**	0.8%	**2.1%**	0.9%	6.9%	2.8%	**4.9%**	1.1%	3.7%	3.0%	8.7%	4.2%	**6.9%**	2.3%	7.6%	8.4%	11.9%	8.6%	**9.6%**	4.8%
		$GARCH_{std}$	0.3%	0.1%	2.8%	0.9%	2.0%	0.1%	1.1%	0.4%	3.4%	1.3%	2.9%	0.7%	2.3%	2.4%	5.0%	2.6%	4.4%	2.0%	6.4%	5.0%	8.2%	5.8%	6.9%	4.7%
		$GARCH_{sstd}$	0.3%	0.1%	2.5%	0.9%	1.6%	0.2%	0.9%	0.2%	3.6%	1.5%	2.2%	1.0%	1.5%	1.5%	4.8%	2.5%	3.1%	1.6%	4.0%	3.8%	7.3%	5.0%	6.0%	4.0%
		$EGARCH_{norm}$	0.9%	**0.5%**	**7.7%**	**3.8%**	3.2%	**1.2%**	1.5%	**1.6%**	**10.1%**	**4.9%**	4.2%	**1.7%**	**4.3%**	**3.3%**	**12.7%**	**6.4%**	5.6%	**3.2%**	**8.4%**	**8.6%**	**14.7%**	**10.0%**	8.5%	**6.7%**
		$EGARCH_{std}$	0.7%	0.3%	3.6%	1.1%	2.9%	0.6%	1.1%	1.0%	5.2%	1.7%	3.6%	1.3%	2.9%	2.1%	6.1%	2.7%	4.8%	2.7%	6.3%	5.5%	8.4%	6.0%	7.4%	5.9%
	1000 Observations	$GARCH_{norm}$	**1.4%**	**0.6%**	6.0%	**2.8%**	4.6%	0.8%	**2.2%**	**1.7%**	7.2%	**3.9%**	5.5%	1.4%	**3.9%**	**3.5%**	**9.6%**	**5.7%**	6.7%	2.7%	**7.9%**	**8.2%**	**12.1%**	**9.1%**	8.7%	6.2%
		$GARCH_{std}$	0.3%	0.0%	2.5%	1.1%	1.8%	0.6%	1.2%	0.7%	3.7%	1.6%	2.2%	1.1%	2.0%	1.9%	4.9%	2.7%	3.5%	2.5%	5.1%	6.0%	7.6%	5.8%	6.2%	4.8%
		$GARCH_{sstd}$	0.4%	0.0%	3.4%	0.9%	2.3%	0.2%	1.0%	0.2%	4.2%	1.8%	3.1%	0.3%	2.3%	1.7%	5.3%	2.4%	4.6%	1.3%	5.9%	4.8%	7.4%	5.3%	7.0%	5.0%
		$EGARCH_{norm}$	0.6%	0.3%	**6.3%**	2.7%	**4.8%**	**1.1%**	1.3%	0.6%	**7.4%**	3.4%	**6.2%**	**2.2%**	2.6%	2.1%	9.0%	4.7%	**7.5%**	**3.1%**	5.6%	5.7%	11.0%	6.9%	**9.3%**	**6.6%**
		$EGARCH_{std}$	0.4%	0.1%	4.6%	1.0%	2.8%	0.3%	0.7%	0.6%	5.9%	1.9%	3.7%	0.9%	2.3%	1.5%	7.3%	3.0%	5.6%	2.2%	5.8%	5.9%	9.2%	5.7%	8.8%	4.9%
RC Test	**250 Observations**	$GARCH_{norm}$	0.9%	0.3%	5.1%	1.8%	**3.8%**	0.8%	**2.1%**	1.1%	6.9%	2.9%	**4.9%**	1.1%	3.5%	3.0%	8.6%	4.3%	**6.8%**	2.2%	7.3%	7.7%	11.9%	8.8%	**9.4%**	4.8%
		$GARCH_{std}$	0.3%	0.1%	2.9%	0.9%	1.9%	0.1%	1.1%	0.7%	3.7%	1.4%	2.6%	0.4%	2.2%	2.6%	4.9%	2.6%	4.6%	1.7%	6.0%	5.0%	7.9%	6.0%	6.6%	4.5%
		$GARCH_{sstd}$	0.4%	0.0%	2.5%	0.8%	1.6%	0.2%	0.8%	0.3%	3.4%	1.4%	2.3%	0.8%	1.6%	1.5%	5.1%	2.6%	3.1%	1.5%	4.6%	4.4%	7.7%	5.1%	6.7%	4.3%
		$EGARCH_{norm}$	**1.0%**	**0.5%**	**7.6%**	**3.7%**	3.1%	**1.1%**	1.6%	**1.6%**	**10.1%**	**4.9%**	4.2%	**1.8%**	**4.2%**	**3.4%**	**12.6%**	**6.4%**	5.5%	**3.3%**	**7.9%**	**8.4%**	**14.9%**	**9.9%**	8.7%	**6.9%**
		$EGARCH_{std}$	0.8%	0.3%	4.0%	1.1%	2.9%	0.5%	1.4%	1.2%	5.1%	1.5%	3.5%	1.1%	2.8%	2.3%	6.1%	2.7%	4.8%	2.7%	6.7%	5.7%	8.3%	5.8%	7.3%	5.7%
	1000 Observations	$GARCH_{norm}$	**1.4%**	**0.6%**	6.1%	**2.8%**	4.6%	0.8%	**2.2%**	**1.6%**	7.3%	**3.9%**	5.5%	1.4%	**3.8%**	**3.6%**	**9.6%**	**5.6%**	6.8%	3.0%	**7.7%**	**7.9%**	**12.4%**	**9.0%**	8.6%	6.0%
		$GARCH_{std}$	0.3%	0.0%	2.5%	1.0%	1.9%	0.5%	1.2%	0.6%	3.6%	1.6%	2.2%	1.2%	2.1%	2.0%	4.7%	2.7%	3.6%	2.4%	5.0%	6.2%	7.5%	5.9%	6.3%	4.7%
		$GARCH_{sstd}$	0.4%	0.0%	3.3%	0.9%	2.3%	0.2%	1.0%	0.3%	4.2%	1.5%	3.1%	0.4%	2.3%	1.8%	5.3%	2.4%	4.6%	1.4%	5.5%	4.9%	7.2%	5.4%	6.9%	5.2%
		$EGARCH_{norm}$	0.6%	0.3%	**6.3%**	2.6%	**4.7%**	**1.1%**	1.3%	0.6%	**7.4%**	3.4%	**6.2%**	**2.1%**	2.5%	1.9%	9.0%	4.5%	**7.5%**	**3.1%**	6.0%	5.5%	10.9%	7.0%	**9.2%**	**6.5%**
		$EGARCH_{std}$	0.4%	0.1%	4.6%	1.0%	2.7%	0.3%	0.8%	0.5%	5.9%	1.8%	3.7%	0.9%	2.5%	1.5%	6.9%	2.9%	5.5%	2.2%	5.5%	5.9%	9.0%	6.0%	8.7%	4.6%
MFE Test	**250 Observations**	$GARCH_{norm}$	**0.9%**	0.2%	5.0%	1.8%	**3.8%**	0.8%	**2.1%**	1.0%	6.9%	3.0%	**5.0%**	1.1%	3.6%	2.9%	8.7%	3.9%	**6.9%**	2.3%	7.6%	8.1%	11.5%	8.7%	**9.7%**	4.6%
		$GARCH_{std}$	0.3%	0.1%	2.8%	0.8%	2.0%	0.1%	1.0%	0.4%	3.6%	1.5%	2.8%	0.7%	2.3%	2.4%	5.2%	2.6%	4.4%	2.0%	6.2%	4.8%	8.0%	6.0%	6.7%	4.8%
		$GARCH_{sstd}$	0.3%	0.1%	2.5%	0.9%	1.6%	0.1%	0.8%	0.3%	3.5%	1.4%	2.4%	0.9%	1.5%	1.5%	5.0%	2.5%	3.0%	1.3%	4.1%	3.7%	7.3%	4.9%	6.5%	4.1%
		$EGARCH_{norm}$	**0.9%**	**0.5%**	**7.5%**	**3.7%**	3.2%	**1.2%**	1.4%	**1.7%**	**10.1%**	**5.0%**	4.2%	**1.8%**	**3.9%**	**3.4%**	**12.5%**	**6.4%**	5.6%	**3.3%**	**8.4%**	**8.6%**	**15.0%**	**10.1%**	8.7%	**7.0%**
		$EGARCH_{std}$	0.7%	0.3%	3.7%	1.0%	2.9%	0.6%	1.1%	1.1%	5.2%	1.6%	3.4%	1.2%	2.6%	2.1%	6.1%	2.6%	4.7%	2.6%	6.2%	5.7%	8.3%	5.9%	7.3%	5.6%
	1000 Observations	$GARCH_{norm}$	**1.4%**	**0.6%**	6.0%	**2.7%**	4.6%	0.8%	**2.3%**	**1.7%**	7.3%	**4.0%**	5.5%	1.5%	**3.8%**	**3.6%**	**9.6%**	**5.8%**	6.6%	2.9%	**8.4%**	**8.0%**	**12.1%**	**9.2%**	8.8%	6.1%
		$GARCH_{std}$	0.3%	0.0%	2.5%	1.1%	1.8%	0.6%	1.1%	0.5%	3.6%	1.8%	2.2%	1.1%	2.2%	2.0%	4.5%	2.8%	3.7%	2.1%	4.9%	5.9%	7.6%	5.8%	6.5%	4.8%
		$GARCH_{sstd}$	0.4%	0.0%	3.4%	0.9%	2.3%	0.2%	1.0%	0.2%	4.1%	1.8%	3.0%	0.4%	2.2%	1.9%	5.3%	2.2%	4.6%	1.6%	5.7%	4.9%	7.2%	5.3%	7.0%	4.9%
		$EGARCH_{norm}$	0.6%	0.3%	**6.3%**	2.5%	**4.7%**	**1.1%**	1.3%	0.5%	**7.4%**	3.4%	**6.1%**	**2.1%**	2.4%	2.3%	9.1%	4.7%	**7.4%**	**3.0%**	5.8%	5.7%	11.0%	7.0%	**9.1%**	**6.5%**
		$EGARCH_{std}$	0.4%	0.1%	4.6%	0.9%	2.8%	0.3%	0.8%	0.5%	6.0%	1.8%	3.7%	1.0%	2.5%	1.5%	7.1%	3.0%	5.7%	2.2%	6.0%	5.6%	9.1%	5.2%	8.9%	4.9%

Table 28 – Best p-values of the proposed backtesting procedures covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 250 observations, along with their respective loss function results.

Backtesting	$\alpha = 1\%, \beta = 5\%$			$\alpha = 1\%, \beta = 2.5\%$			$\alpha = 2.5\%, \beta = 5\%$		
	Asset-Model	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$
MF Test	Real-GARCH$_{ged}$	0.994912	1.048683	STOXX-GAS$_{sstd}$	0.994358	1.049042	Euro-GARCH$_{std}$	0.996699	1.028896
	Real-GARCH$_{std}$	0.992822	1.048213	Gold-EGARCH$_{nig}$	0.991265	1.052917	BTC-GARCH$_{ged}$	0.994227	1.135862
	Real-GARCH$_{sged}$	0.989912	1.048607	WTI-gjrGARCH$_{nig}$	0.988251	1.109245	DGS10-GARCH$_{norm}$	0.991646	1.099078
	Euro-GAS$_{sstd}$	0.987852	1.026580	DGS3-GARCH$_{sged}$	0.983950	1.208165	S&P500-gjrGARCH$_{nig}$	0.987662	1.044073
	DGS10-gjrGARCH$_{std}$	0.976534	1.103012	WTI-gjrGARCH$_{sstd}$	0.981309	1.107457	Real-EGARCH$_{std}$	0.983552	1.048477
	WTI-gjrGARCH$_{sstd}$	0.965680	1.097319	WTI-GARCH$_{sged}$	0.974093	1.108865	DGS10-GARCH$_{ged}$	0.981585	1.100609
RC Test	Real-GARCH$_{std}$	0.992554	1.048213	STOXX-GAS$_{sstd}$	0.991441	1.049042	Euro-GARCH$_{std}$	0.995418	1.028896
	Real-GARCH$_{ged}$	0.992183	1.048683	WTI-gjrGARCH$_{nig}$	0.990377	1.109245	BTC-GARCH$_{ged}$	0.995341	1.135862
	Euro-GAS$_{sstd}$	0.984260	1.026580	Gold-EGARCH$_{nig}$	0.988590	1.052917	DGS10-GARCH$_{norm}$	0.990915	1.099078
	Real-GARCH$_{sged}$	0.973521	1.048607	WTI-gjrGARCH$_{sstd}$	0.986116	1.107457	S&P500gjrGARCH$_{nig}$	0.990540	1.044073
	IBOV-GARCH$_{sstd}$	0.961731	1.062213	IBOV-GARCH$_{nig}$	0.970609	1.069022	Real-EGARCH$_{std}$	0.985444	1.048477
	DGS10-gjrGARCH$_{std}$	0.959928	1.103012	WTI-EGARCH$_{ged}$	0.969665	1.100800	DGS10-GARCH$_{ged}$	0.979334	1.100609
MFE Test	Real-GARCH$_{ged}$	0.994178	1.048683	STOXX-GAS$_{sstd}$	0.994926	1.049042	Euro-GARCH$_{std}$	0.997209	1.028896
	Real-GARCH$_{std}$	0.993102	1.048213	Gold-EGARCH$_{nig}$	0.992109	1.052917	BTC-GARCH$_{ged}$	0.992142	1.135862
	Euro-GAS$_{sstd}$	0.986131	1.026580	WTI-gjrGARCH$_{nig}$	0.989505	1.109245	Real-EGARCH$_{std}$	0.991846	1.048477
	DGS3-GARCH$_{sged}$	0.985343	1.186516	DGS3-GARCH$_{sged}$	0.982139	1.208165	DSG10-GARCH$_{norm}$	0.991591	1.099078
	Real-GARCH$_{sged}$	0.983798	1.048607	WTI-gjrGARCH$_{sstd}$	0.981278	1.107457	IBOV-gjrGARCH$_{nig}$	0.989622	1.044073
	DGS10-gjrGARCH$_{std}$	0.973880	1.103012	S&P500-GARCH$_{jsu}$	0.969397	1.047101	S&P500-gjrGARCH$_{sged}$	0.983659	1.044011

Table 29 – Best p-values of the proposed backtesting procedures covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 500 observations, along with their respective loss function results.

Backtesting	$\alpha = 1\%, \beta = 5\%$			$\alpha = 1\%, \beta = 2.5\%$			$\alpha = 2.5\%, \beta = 5\%$		
	Asset-Model	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$
MF Test	WTI-gjrGARCH$_{nig}$	0.997197	1.102357	Euro-gjrGARCH$_{snorm}$	0.998385	1.025718	DGS10-gjrGARCH$_{ged}$	0.997912	1.098865
	DGS3-GARCH$_{sged}$	0.996858	1.199320	WTI-EGARCH$_{sged}$	0.996465	1.109912	WTI-EGARCH$_{sged}$	0.990889	1.093509
	WTI-gjrGARCH$_{sged}$	0.995940	1.100861	Euro-GAS$_{snorm}$	0.995137	1.026936	WTI-gjrGARCH$_{sged}$	0.988470	1.094535
	Gold-EGARCH$_{sged}$	0.995329	1.045076	Euro-gjrGARCH$_{ged}$	0.986423	1.026535	DGS10-GARCH$_{jsu}$	0.987254	1.098261
	Euro-GAS$_{sstd}$	0.994302	1.026649	Euro-GAS$_{sstd}$	0.985931	1.028950	DGS10-GARCH$_{sged}$	0.984253	1.098772
	Real-EGARCH$_{ged}$	0.994253	1.048529	WTI-gjrGARCH$_{sged}$	0.971124	1.111306	DGS10-EGARCH$_{std}$	0.983306	1.097478
RC Test	Gold-EGARCH$_{sged}$	0.996589	1.045076	Euro-gjrGARCH$_{snorm}$	0.997842	1.025718	DGS10-gjrGARCH$_{ged}$	0.997617	1.098865
	Euro-GAS$_{sstd}$	0.994696	1.026649	Euro-GAS$_{snorm}$	0.995469	1.026936	WTI-gjrGARCH$_{sged}$	0.992872	1.094535
	WTI-gjrGARCH$_{nig}$	0.992168	1.102357	WTI-EGARCH$_{sged}$	0.994979	1.109912	DGS10-EGARCH$_{std}$	0.983813	1.097478
	WTI-gjrGARCH$_{sged}$	0.990046	1.100861	Euro-GAS$_{sstd}$	0.994127	1.028950	WTI-EGARCH$_{sged}$	0.976183	1.093509
	Real-EGARCH$_{ged}$	0.986029	1.048529	Euro-gjrGARCH$_{ged}$	0.982122	1.026535	DGS10-GARCH$_{sged}$	0.975361	1.098772
	Real-GARCH$_{sged}$	0.984979	1.049354	DGS3-EGARCH$_{nig}$	0.961192	1.203656	DGS10-GARCH$_{jsu}$	0.974569	1.098261
MFE Test	DGS3-GARCH$_{ged}$	0.996681	1.218456	Euro-gjrGARCH$_{snorm}$	0.998375	1.025718	WTI-gjrGARCH$_{sged}$	0.988968	1.094535
	WTI-gjrGARCH$_{nig}$	0.996413	1.102357	WTI-EGARCH$_{sged}$	0.996799	1.109912	WTI-EGARCH$_{sged}$	0.988962	1.093509
	Gold-EGARCH$_{sged}$	0.995833	1.045076	Euro-GAS$_{snorm}$	0.995459	1.026936	DGS10-GARCH$_{jsu}$	0.987773	1.098261
	WTI-gjrGARCH$_{sged}$	0.994483	1.100861	Euro-GAS$_{sstd}$	0.990266	1.028950	DGS10-GARCH$_{sged}$	0.983804	1.098772
	Euro-GAS$_{sstd}$	0.993720	1.026649	Euro-gjrGARCH$_{ged}$	0.980605	1.026535	DGS10-EGARCH$_{std}$	0.982679	1.097478
	Real-EGARCH$_{ged}$	0.993346	1.048529	Euro-gjrGARCH$_{std}$	0.967705	1.026332	DGS3-gjrGARCH$_{ged}$	0.976260	1.165772

Table 30 – Best p-values of the proposed backtesting procedures covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 1000 observations, along with their respective loss function results.

Backtesting	$\alpha = 1\%, \beta = 5\%$			$\alpha = 1\%, \beta = 2.5\%$			$\alpha = 2.5\%, \beta = 5\%$		
	Asset-Model	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$
MF Test	Real-GARCH$_{nig}$	0.987727	1.051332	WTI-GARCH$_{jsu}$	0.998240	1.120787	Euro-EGARCH$_{ged}$	0.999560	1.028379
	Euro-GARCH$_{sged}$	0.986382	1.024533	WTI-gjrGARCH$_{nig}$	0.989027	1.119710	Euro-gjrGARCH$_{norm}$	0.997933	1.027822
	Euro-GARCH$_{sstd}$	0.982394	1.024291	WTI-gjrGARCH$_{jsu}$	0.987740	1.119243	Euro-GARCH$_{sged}$	0.996424	1.028644
	Real-EGARCH$_{nig}$	0.982321	1.051935	WTI-GARCH$_{nig}$	0.980899	1.121421	IBOV-gjrGARCH$_{nig}$	0.996327	1.062147
	Real-GARCH$_{std}$	0.979671	1.049853	WTI-GARCH$_{sstd}$	0.966973	1.119111	BTC-GAS$_{snorm}$	0.994905	1.146886
	BTC-EGARCH$_{sged}$	0.974195	1.191133	DGS3-GARCH$_{sged}$	0.961193	1.196993	S&P500-GARCH$_{sged}$	0.994626	1.042434
RC Test	Euro-GARCH$_{sged}$	0.989693	1.024533	WTI-GARCH$_{jsu}$	0.997037	1.120787	Euro-EGARCH$_{ged}$	0.999628	1.028379
	Euro-GARCH$_{sstd}$	0.986760	1.024291	WTI-gjrGARCH$_{nig}$	0.976977	1.119710	Euro-gjrGARCH$_{norm}$	0.998126	1.027822
	Real-GARCH$_{nig}$	0.984122	1.051332	WTI-GARCH$_{nig}$	0.969150	1.121421	Euro-GARCH$_{sged}$	0.996660	1.028644
	Real-GARCH$_{std}$	0.979806	1.049853	WTI-gjrGARCH$_{jsu}$	0.966831	1.119243	Euro-EGARCH$_{snorm}$	0.995474	1.028348
	Real-EGARCH$_{nig}$	0.978313	1.051935	S&P500-gjrGARCH$_{ged}$	0.965934	1.043700	BTC-GAS$_{snorm}$	0.995140	1.146886
	BTC-GARCH$_{sged}$	0.977901	1.182580	WTI-GARCH$_{sstd}$	0.950826	1.119111	Euro-EGARCH$_{nig}$	0.994139	1.028519
MFE Test	Euro-GARCH$_{sged}$	0.987668	1.024533	WTI-GARCH$_{jsu}$	0.997337	1.120787	Euro-EGARCH$_{ged}$	0.999572	1.028379
	Real-GARCH$_{nig}$	0.986329	1.051332	WTI-gjrGARCH$_{nig}$	0.987547	1.119710	Euro-GARCH$_{sged}$	0.996414	1.028644
	Euro-GARCH$_{sstd}$	0.984490	1.024291	WTI-gjrGARCH$_{jsu}$	0.986181	1.119243	Euro-gjrGARCH$_{norm}$	0.996404	1.027822
	Real-EGARCH$_{nig}$	0.982565	1.051935	WTI-GARCH$_{nig}$	0.977493	1.121421	IBOV-gjrGARCH$_{nig}$	0.994993	1.062147
	Real-GARCH$_{std}$	0.978176	1.049853	WTI-GARCH$_{sstd}$	0.968841	1.119111	BTC-GAS$_{snorm}$	0.994887	1.146886
	BTC-EGARCH$_{sged}$	0.976720	1.191133	IBOV-gjrGARCH$_{ged}$	0.959197	1.043700	Euro-EGARCH$_{snorm}$	0.994610	1.028348

Table 31 – Worst p-values of the proposed backtesting procedures covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 250 observations, along with their respective loss function results.

Backtesting	$\alpha = 1\%, \beta = 5\%$			$\alpha = 1\%, \beta = 2.5\%$			$\alpha = 2.5\%, \beta = 5\%$		
	Asset-Model	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$
MF Test	STOXX-GARCH$_{std}$	0.000314	1.042327	DGS3-gjrGARCH$_{snorm}$	0.000017	1.195756	STOXX-gjrGARCH$_{std}$	0.007389	1.043794
	STOXX-gjrGARCH$_{std}$	0.000400	1.042537	DGS3-EGARCH$_{norm}$	0.000322	1.193886	WTI-EGARCH$_{snorm}$	0.008059	1.088937
	STOXX-GARCH$_{sstd}$	0.000460	1.045074	DGS10-GARCH$_{sstd}$	0.000732	1.111982	WTI-gjrGARCH$_{norm}$	0.011475	1.086169
	STOXX-GARCH$_{jsu}$	0.000726	1.045836	Gold-gjrGARCH$_{snorm}$	0.002045	1.045423	Euro-gjrGARCH$_{sstd}$	0.018616	1.029186
	STOXX-GARCH$_{snorm}$	0.001698	1.042984	BTC-GARCH$_{snorm}$	0.002374	1.156860	S&P500-GAS$_{sstd}$	0.020563	1.042341
	STOXX-gjrGARCH$_{ged}$	0.004019	1.042722	DGS10-GARCH$_{norm}$	0.007205	1.110485	STOXX-gjrGARCH$_{norm}$	0.021224	1.044012
RC Test	STOXX-GARCH$_{sstd}$	0.000235	1.045074	DGS3-gjrGARCH$_{snorm}$	0.000033	1.195756	STOXX-gjrGARCH$_{std}$	0.007709	1.043794
	STOXX-GARCH$_{jsu}$	0.000239	1.045836	DGS3-EGARCH$_{norm}$	0.000206	1.193886	WTI-EGARCH$_{snorm}$	0.009385	1.088937
	STOXX-GARCH$_{sstd}$	0.000345	1.042327	DGS10-GARCH$_{sstd}$	0.000375	1.111982	WTI-gjrGARCH$_{norm}$	0.011009	1.086169
	STOXX-gjrGARCH$_{std}$	0.000504	1.042537	BTC-GARCH$_{snorm}$	0.001145	1.156860	Euro-gjrGARCH$_{sstd}$	0.020036	1.029186
	DGS3-gjrGARCH$_{ged}$	0.000959	1.190563	Gold-gjrGARCH$_{snorm}$	0.002523	1.045423	STOXX-gjrGARCH$_{norm}$	0.021498	1.044012
	STOXX-GARCH$_{snorm}$	0.001307	1.042984	DGS10-GARCH$_{nig}$	0.008614	1.113341	S&P500-GAS$_{sstd}$	0.022842	1.042341
MFE Test	DGS3-gjrGARCH$_{ged}$	0.000089	1.190563	DGS3-gjrGARCH$_{snorm}$	0.000022	1.195756	WTI-EGARCH$_{snorm}$	0.006389	1.088937
	STOXX-GARCH$_{std}$	0.000323	1.042327	DGS3-EGARCH$_{norm}$	0.000281	1.193886	STOXX-gjrGARCH$_{std}$	0.008131	1.043794
	STOXX-gjrGARCH$_{std}$	0.000328	1.042537	DGS10-GARCH$_{sstd}$	0.000663	1.111982	WTI-gjrGARCH$_{norm}$	0.011587	1.086169
	STOXX-GARCH$_{jsu}$	0.000466	1.045836	BTC-GARCH$_{snorm}$	0.001344	1.156860	Euro-gjrGARCH$_{sstd}$	0.015802	1.029186
	STOXX-GARCH$_{sstd}$	0.000495	1.045074	Gold-gjrGARCH$_{snorm}$	0.002115	1.045423	STOXX-GARCH$_{ged}$	0.023498	1.044228
	STOXX-GARCH$_{snorm}$	0.001325	1.042984	DGS10-GARCH$_{norm}$	0.007636	1.110485	STOXX-gjrGARCH$_{norm}$	0.023619	1.044012

Table 32 – Worst p-values of the proposed backtesting procedures covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 500 observations, along with their respective loss function results.

Backtesting	$\alpha = 1\%, \beta = 5\%$			$\alpha = 1\%, \beta = 2.5\%$			$\alpha = 2.5\%, \beta = 5\%$		
	Asset-Model	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$
MF Test	STOXX-GARCH$_{std}$	0.000288	1.041233	STOXX-GARCH$_{ged}$	0.001872	1.045576	STOXX-EGARCH$_{std}$	0.000386	1.041377
	S&P500-gjrGARCH$_{norm}$	0.000910	1.036180	STOXX-GARCH$_{snorm}$	0.005508	1.044743	BTC-GARCH$_{std}$	0.004796	1.125754
	STOXX-GARCH$_{ged}$	0.000948	1.041517	S&P500-GARCH$_{snorm}$	0.005546	1.041456	S&P500-gjrGARCH$_{nig}$	0.016072	1.042734
	STOXX-GARCH$_{norm}$	0.002037	1.039374	BTC-GARCH$_{sged}$	0.009416	1.187140	S&P500-gjrGARCH$_{jsu}$	0.018078	1.042389
	STOXX-EGARCH$_{std}$	0.002976	1.039797	BTC-GAS$_{snorm}$	0.010326	1.216890	S&P500-gjrGARCH$_{sged}$	0.024758	1.042689
	STOXX-GARCH$_{snorm}$	0.003733	1.040949	DGS10-GARCH$_{snorm}$	0.011210	1.108962	Gold-GAS$_{snorm}$	0.027538	1.046495
RC Test	STOXX-GARCH$_{std}$	0.000343	1.041233	STOXX-GARCH$_{ged}$	0.002357	1.045576	STOXX-EGARCH$_{std}$	0.000211	1.041377
	S&P500-gjrGARCH$_{norm}$	0.000897	1.036180	STOXX-GARCH$_{snorm}$	0.005882	1.044743	BTC-GARCH$_{std}$	0.004950	1.125754
	STOXX-GARCH$_{ged}$	0.001016	1.041517	S&P500-GARCH$_{snorm}$	0.005953	1.041456	S&P500-gjrGARCH$_{nig}$	0.016596	1.042734
	STOXX-GARCH$_{norm}$	0.002224	1.039374	BTC-GAS$_{snorm}$	0.006762	1.216890	S&P500-gjrGARCH$_{jsu}$	0.018241	1.042389
	STOXX-EGARCH$_{std}$	0.003183	1.039797	S&P500-GARCH$_{norm}$	0.015304	1.038382	Gold-GAS$_{snorm}$	0.019106	1.046495
	STOXX-GARCH$_{snorm}$	0.003695	1.041260	DGS10-EGARCH$_{sstd}$	0.015532	1.109013	S&P500-gjrGARCH$_{sged}$	0.024612	1.042689
MFE Test	STOXX-GARCH$_{std}$	0.000289	1.041233	STOXX-GARCH$_{ged}$	0.001966	1.045576	STOXX-EGARCH$_{std}$	0.000468	1.041377
	STOXX-GARCH$_{ged}$	0.000841	1.041517	S&P500-GARCH$_{snorm}$	0.005271	1.041456	BTC-GARCH$_{std}$	0.008124	1.125754
	S&P500-gjrGARCH$_{norm}$	0.001249	1.036180	STOXX-GARCH$_{snorm}$	0.005808	1.044743	S&P500-gjrGARCH$_{nig}$	0.018165	1.042734
	STOXX-GARCH$_{norm}$	0.001742	1.039374	BTC-GAS$_{snorm}$	0.009663	1.216890	Gold-GAS$_{snorm}$	0.020438	1.046495
	STOXX-EGARCH$_{std}$	0.002716	1.039797	BTC-GARCH$_{sged}$	0.010058	1.187140	S&P500-gjrGARCH$_{jsu}$	0.024158	1.042389
	STOXX-GARCH$_{snorm}$	0.003445	1.041260	DGS10-GARCH$_{snorm}$	0.015850	1.108962	S&P500-gjrGARCH$_{sged}$	0.026828	1.042689

Table 33 – Worst p-values of the proposed backtesting procedures covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 1000 observations, along with their respective loss function results.

Backtesting	$\alpha = 1\%, \beta = 5\%$			$\alpha = 1\%, \beta = 2.5\%$			$\alpha = 2.5\%, \beta = 5\%$		
	Asset-Model	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$	**Asset-Model**	p-value	$\bar{S}_{RVaR}$
MF Test	DGS3-EGARCH$_{jsu}$	0.000464	1.170530	STOXX-GARCH$_{snorm}$	0.000180	1.044632	STOXX-EGARCH$_{sged}$	0.010335	1.046058
	DGS3-gjrGARCH$_{sged}$	0.000490	1.175926	STOXX-GARCH$_{norm}$	0.001360	1.042774	DGS3-EGARCH$_{jsu}$	0.011882	1.154663
	STOXX-GARCH$_{std}$	0.000923	1.041703	STOXX-GARCH$_{std}$	0.002859	1.045829	BTC-GARCH$_{std}$	0.011969	1.134720
	DGS3-gjrGARCH$_{ged}$	0.001609	1.173311	Gold-EGARCH$_{sged}$	0.005524	1.046503	DGS3-EGARCH$_{sstd}$	0.028897	1.153976
	DGS3-gjrGARCH$_{sstd}$	0.002645	1.173192	Gold-EGARCH$_{sstd}$	0.006808	1.046625	Gold-gjrGARCH$_{norm}$	0.035604	1.042204
	DGS3-EGARCH$_{sstd}$	0.003337	1.169723	DGS3-gjrGARCH$_{norm}$	0.007600	1.177150	S&P500-EGARCH$_{ged}$	0.048357	1.040433
RC Test	DGS3-EGARCH$_{jsu}$	0.000369	1.170530	STOXX-GARCH$_{snorm}$	0.000435	1.044632	BTC-GARCH$_{std}$	0.009697	1.134720
	DGS3-gjrGARCH$_{sged}$	0.000671	1.175926	STOXX-GARCH$_{norm}$	0.002493	1.042774	STOXX-EGARCH$_{sged}$	0.010684	1.046058
	STOXX-GARCH$_{std}$	0.000917	1.041703	STOXX-GARCH$_{std}$	0.002829	1.045829	DGS3-EGARCH$_{jsu}$	0.013013	1.154663
	DGS3-gjrGARCH$_{ged}$	0.002279	1.173311	Gold-EGARCH$_{sstd}$	0.005553	1.046625	DGS3-EGARCH$_{sstd}$	0.032641	1.153976
	DGS3-EGARCH$_{sstd}$	0.002857	1.169723	DGS3-gjrGARCH$_{norm}$	0.005934	1.177150	Gold-gjrGARCH$_{norm}$	0.034479	1.042204
	DGS3-EGARCH$_{nig}$	0.002893	1.171012	Gold-EGARCH$_{sged}$	0.006137	1.046503	S&P500-EGARCH$_{ged}$	0.049389	1.040433
MFE Test	DGS3-EGARCH$_{jsu}$	0.000435	1.170530	STOXX-GARCH$_{snorm}$	0.000205	1.044632	STOXX-EGARCH$_{sged}$	0.009748	1.046058
	DGS3-gjrGARCH$_{sged}$	0.000550	1.175926	STOXX-GARCH$_{norm}$	0.001389	1.042774	DGS3-EGARCH$_{jsu}$	0.012081	1.154663
	STOXX-GARCH$_{std}$	0.000863	1.041703	STOXX-GARCH$_{std}$	0.002611	1.045829	BTC-GARCH$_{std}$	0.012658	1.134720
	DGS3-gjrGARCH$_{ged}$	0.001776	1.173311	DGS3-gjrGARCH$_{norm}$	0.005290	1.177150	DGS3-EGARCH$_{sstd}$	0.029601	1.153976
	DGS3-EGARCH$_{nig}$	0.002975	1.171012	Gold-EGARCH$_{sstd}$	0.005751	1.046625	Gold-gjrGARCH$_{norm}$	0.034621	1.042204
	DGS3-gjrGARCH$_{sstd}$	0.002993	1.173192	Gold-EGARCH$_{sged}$	0.007499	1.046503	DGS3-EGARCH$_{nig}$	0.047433	1.155403

Table 34 – Best Realized Loss Function results covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 250 observations, along with their respective p-value of each backtesting procedure.

Significance Levels	**Asset-Model**	$\bar{S}_{RVaR}$	**MF Test**	**RC Test**	**MFE Test**
$\alpha = 1\%$, $\beta = 5\%$	Euro-gjrGARCH$_{norm}$	1.024131	0.105323	0.115929	0.106181
	Euro-GARCH$_{norm}$	1.024370	0.422968	0.436443	0.412557
	Euro-gjrGARCH$_{snorm}$	1.024539	0.137396	0.142310	0.143310
	Euro-EGARCH$_{norm}$	1.024652	0.364193	0.368746	0.356006
	Euro-gjrGARCH$_{std}$	1.024653	0.596078	0.542636	0.588725
	Euro-GARCH$_{snorm}$	1.024666	0.360389	0.390833	0.363055
$\alpha = 1\%$, $\beta = 2.5\%$	Euro-gjrGARCH$_{norm}$	1.025826	0.162740	0.166042	0.168426
	Euro-GARCH$_{norm}$	1.026093	0.146990	0.139264	0.144885
	Euro-EGARCH$_{norm}$	1.026241	0.348310	0.367677	0.358519
	Euro-gjrGARCH$_{snorm}$	1.026297	0.370449	0.409324	0.400820
	Euro-GARCH$_{snorm}$	1.026437	0.320144	0.337450	0.331225
	Euro-gjrGARCH$_{std}$	1.026626	0.547332	0.500613	0.539028
$\alpha = 2.5\%$, $\beta = 5\%$	Euro-gjrGARCH$_{std}$	1.028744	0.631398	0.651058	0.657625
	Euro-gjrGARCH$_{norm}$	1.028834	0.815259	0.822597	0.795260
	Euro-GARCH$_{std}$	1.028896	0.996699	0.995418	0.997209
	Euro-GARCH$_{norm}$	1.029041	0.975553	0.971635	0.974135
	Euro-gjrGARCH$_{ged}$	1.029084	0.401631	0.393558	0.383353
	Euro-gjrGARCH$_{snorm}$	1.029164	0.395253	0.379948	0.377244

Table 35 – Best Realized Loss Function results covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 500 observations, along with their respective p-value of each backtesting procedure.

Significance Levels	**Asset-Model**	$\bar{S}_{RVaR}$	**MF Test**	**RC Test**	**MFE Test**
$\alpha = 1\%$, $\beta = 5\%$	Euro-gjrGARCH$_{norm}$	1.023634	0.789215	0.794361	0.801092
	Euro-GARCH$_{norm}$	1.023888	0.570622	0.588460	0.577438
	Euro-gjrGARCH$_{snorm}$	1.024006	0.822765	0.837727	0.821482
	Euro-GARCH$_{snorm}$	1.024144	0.870880	0.850353	0.860041
	Euro-EGARCH$_{norm}$	1.024265	0.805168	0.786452	0.805482
	Euro-gjrGARCH$_{std}$	1.024350	0.975598	0.974507	0.977615
$\alpha = 1\%$, $\beta = 2.5\%$	Euro-gjrGARCH$_{norm}$	1.025290	0.926155	0.929077	0.913350
	Euro-GARCH$_{norm}$	1.025571	0.730437	0.744358	0.733113
	Euro-gjrGARCH$_{snorm}$	1.025718	0.998385	0.997842	0.998375
	Euro-GARCH$_{snorm}$	1.025868	0.907404	0.911305	0.907216
	Euro-EGARCH$_{norm}$	1.025967	0.526315	0.545410	0.538132
	Euro-gjrGARCH$_{std}$	1.026332	0.968617	0.960620	0.967705
$\alpha = 2.5\%$, $\beta = 5\%$	Euro-gjrGARCH$_{norm}$	1.028384	0.932210	0.932441	0.932104
	Euro-gjrGARCH$_{std}$	1.028397	0.901131	0.886554	0.910248
	Euro-GARCH$_{std}$	1.028523	0.462666	0.447107	0.465827
	Euro-GARCH$_{norm}$	1.028609	0.304998	0.292563	0.296404
	Euro-gjrGARCH$_{snorm}$	1.028689	0.757926	0.777557	0.768007
	Euro-gjrGARCH$_{ged}$	1.028764	0.495054	0.514839	0.484067

Table 36 – Best Realized Loss Function results covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 1000 observations, along with their respective p-value of each backtesting procedure.

Significance Levels	**Asset-Model**	$\bar{S}_{RVaR}$	**MF Test**	**RC Test**	**MFE Test**
	Euro-gjrGARCH$_{norm}$	1.023006	0.894992	0.900674	0.899034
	Euro-GARCH$_{norm}$	1.023100	0.900609	0.899867	0.904369
$\alpha = 1\%$,	Euro-gjrGARCH$_{snorm}$	1.023200	0.799032	0.798814	0.800213
$\beta = 5\%$	Euro-GARCH$_{snorm}$	1.023278	0.776501	0.796130	0.788328
	Euro-EGARCH$_{norm}$	1.023366	0.715503	0.717262	0.717277
	Euro-EGARCH$_{snorm}$	1.023606	0.905177	0.916939	0.911763
	Euro-gjrGARCH$_{norm}$	1.024603	0.422164	0.434688	0.454335
	Euro-GARCH$_{norm}$	1.024711	0.206571	0.215677	0.193978
$\alpha = 1\%$,	Euro-gjrGARCH$_{snorm}$	1.024827	0.700478	0.711163	0.731492
$\beta = 2.5\%$	Euro-GARCH$_{snorm}$	1.024916	0.085339	0.095393	0.085889
	Euro-EGARCH$_{norm}$	1.024989	0.311603	0.313382	0.301997
	Euro-EGARCH$_{snorm}$	1.025265	0.416269	0.397115	0.417132
	Euro-gjrGARCH$_{std}$	1.027768	0.966830	0.974930	0.973489
	Euro-gjrGARCH$_{norm}$	1.027822	0.997933	0.998126	0.996404
$\alpha = 2.5\%$,	Euro-GARCH$_{std}$	1.027894	0.906204	0.910668	0.906476
$\beta = 5\%$	Euro-GARCH$_{norm}$	1.027902	0.905508	0.922058	0.915681
	Euro-gjrGARCH$_{snorm}$	1.027981	0.979597	0.980007	0.980305
	Euro-EGARCH$_{std}$	1.027984	0.984940	0.984656	0.982652

Table 37 – Worst Realized Loss Function results covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 250 observations, along with their respective p-value of each backtesting procedure.

Significance Levels	**Asset-Model**	$\bar{S}_{RVaR}$	**MF Test**	**RC Test**	**MFE Test**
	DGS3-EGARCH$_{ged}$	1.354467	0.545578	0.384094	0.769160
	DGS3-EGARCH$_{sged}$	1.283317	0.523261	0.346309	0.644720
$\alpha = 1\%$,	DGS3-GAS$_{snorm}$	1.203957	0.204287	0.241465	0.210415
$\beta = 5\%$	DGS3-GAS$_{sstd}$	1.197593	0.083869	0.092539	0.088637
	DGS3-EGARCH$_{nig}$	1.196467	0.348463	0.288606	0.312758
	DGS3-EGARCH$_{jsu}$	1.195017	0.362548	0.212281	0.340578
	DGS3-EGARCH$_{ged}$	1.410225	0.456770	0.177518	0.727437
	DGS3-EGARCH$_{sged}$	1.321809	0.789078	0.517742	0.914337
$\alpha = 1\%$,	DGS3-GAS$_{snorm}$	1.221539	0.632737	0.625489	0.616218
$\beta = 2.5\%$	DGS3-GAS$_{sstd}$	1.217467	0.356457	0.302720	0.323365
	DGS3-EGARCH$_{nig}$	1.216766	0.132185	0.094864	0.116333
	DGS3-EGARCH$_{jsu}$	1.215048	0.045318	0.021880	0.034684
	DGS3-EGARCH$_{sged}$	1.225902	0.317300	0.176904	0.082316
	DGS3-EGARCH$_{ged}$	1.225097	0.463228	0.526023	0.315251
$\alpha = 2.5\%$,	DGS3-GAS$_{snorm}$	1.190727	0.714482	0.739216	0.723941
$\beta = 5\%$	DGS3-GAS$_{sstd}$	1.178171	0.913312	0.917897	0.916855
	DGS3-EGARCH$_{nig}$	1.177035	0.637423	0.636802	0.630993
	DGS3-EGARCH$_{jsu}$	1.175405	0.766182	0.739760	0.761662

Table 38 – Worst Realized Loss Function results covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 500 observations, along with their respective p-value of each backtesting procedure.

Significance Levels	**Asset-Model**	$\bar{S}_{RVaR}$	**MF Test**	**RC Test**	**MFE Test**
$\alpha = 1\%$, $\beta = 5\%$	DGS3-GAS$_{snorm}$	1.336191	0.071252	0.071646	0.064070
	DGS3-gjrGARCH$_{ged}$	1.220345	0.872335	0.823333	0.971878
	DGS3-GARCH$_{ged}$	1.218456	0.979656	0.961419	0.996681
	DGS3-gjrGARCH$_{sged}$	1.204826	0.558940	0.270783	0.804726
	BTC-GAS$_{snorm}$	1.199337	0.174759	0.131810	0.134096
	DGS3-GARCH$_{sged}$	1.199320	0.996858	0.940017	0.993068
$\alpha = 1\%$, $\beta = 2.5\%$	DGS3-GAS$_{snorm}$	1.366414	0.321229	0.284377	0.300956
	DGS3-gjrGARCH$_{ged}$	1.250801	0.385836	0.353005	0.420337
	DGS3-GARCH$_{ged}$	1.248819	0.490798	0.378036	0.717896
	DGS3-gjrGARCH$_{sged}$	1.230787	0.629786	0.454263	0.815528
	DGS3-GARCH$_{sged}$	1.225228	0.070354	0.024984	0.087643
	BTC-GAS$_{snorm}$	1.216890	0.010326	0.006762	0.009663
$\alpha = 2.5\%$, $\beta = 5\%$	DGS3-GAS$_{snorm}$	1.308299	0.424804	0.443116	0.418772
	BTC-GAS$_{snorm}$	1.185938	0.424664	0.449693	0.457230
	DGS3-GAS$_{sstd}$	1.173423	0.282658	0.303105	0.299611
	DGS3-gjrGARCH$_{sged}$	1.168969	0.054298	0.090789	0.070095
	DGS3-gjrGARCH$_{nig}$	1.167734	0.107154	0.125922	0.127943
	DGS3-GARCH$_{nig}$	1.166746	0.917215	0.904864	0.909213

Table 39 – Worst Realized Loss Function results covering all considered scenarios, using significance levels of 1%, 2.5%, and 5%, with a rolling window of 1000 observations, along with their respective *p*-value of each backtesting procedure.

Significance Levels	**Asset-Model**	$\bar{S}_{RVaR}$	**MF Test**	**RC Test**	**MFE Test**
$\alpha = 1\%$, $\beta = 5\%$	DGS3-GAS$_{snorm}$	1.589034	0.509870	0.515721	0.526070
	BTC-EGARCH$_{nig}$	1.205403	0.938455	0.933415	0.927145
	BTC-EGARCH$_{sstd}$	1.201824	0.781007	0.786422	0.758332
	BTC-EGARCH$_{jsu}$	1.201473	0.935242	0.931138	0.933506
	DGS3-GAS$_{sstd}$	1.195233	0.591042	0.530939	0.590893
	DGS3-EGARCH$_{sged}$	1.194968	0.229714	0.140618	0.367403
$\alpha = 1\%$, $\beta = 2.5\%$	DGS3-GAS$_{snorm}$	1.644496	0.453476	0.468582	0.454358
	BTC-EGARCH$_{nig}$	1.234562	0.080551	0.072913	0.073077
	BTC-EGARCH$_{sstd}$	1.229478	0.625803	0.692924	0.667126
	BTC-EGARCH$_{jsu}$	1.229409	0.452150	0.442200	0.452997
	DGS3-EGARCH$_{sged}$	1.217861	0.559012	0.458387	0.543303
	BTC-EGARCH$_{std}$	1.216277	0.556259	0.588511	0.601626
$\alpha = 2.5\%$, $\beta = 5\%$	DGS3-GAS$_{snorm}$	1.532448	0.622948	0.648268	0.662384
	DGS3-GAS$_{sstd}$	1.174773	0.414432	0.401065	0.415722
	BTC-EGARCH$_{nig}$	1.167796	0.592546	0.585741	0.563358
	DGS3-EGARCH$_{sged}$	1.166402	0.381381	0.290747	0.423106
	BTC-EGARCH$_{sged}$	1.164232	0.904336	0.896048	0.900017
	BTC-EGARCH$_{jsu}$	1.162938	0.366324	0.337173	0.352220

5 Conclusions

Risk measures constitute pivotal quantitative tools for financial agents, such as portfolio managers, investors, banks, and insurance industries. In recent years, risk measurement techniques have gained significant prominence, driven by the increasing complexity of financial markets and the need to comprehend and manage associated risks. By quantifying and forecasting risk, companies reduce uncertainty, enabling, for instance, a more efficient allocation of resources and the development of risk mitigation strategies. After all, the goal of measurement is to aim for decision-making. In our study, we modified the definitions of Expected Shortfall backtesting procedures based on exceedance residuals to make them effective for RVaR, a relatively new risk measure with pleasing theoretical and statistical properties. To assess the extent of the adequacy of these modifications, we conducted both numerical and empirical analyses. For the first, we employed rolling window sizes of 250 and 1000 observations and out-of-sample sizes of 250 and 500 observations. In contrast, for the second, an extension of the rolling window to 500 observations was introduced. We used significance levels of 1%, 2.5%, and 5% for both.

Initially, we present the main conclusions about the numerical analysis. Firstly, we can observe from the size and power tests that none of the proposed backtesting procedures display significant superiority over the others, either being an option of equal or similar usefulness. Also, the size and power of the tests deteriorate as T increases, with some exceptions. The highest powers are observed at $T = 250$, emphasizing the significance levels of $\alpha = 1\%, \beta = 2.5\%$. Some of the best sizes are also found at $T = 250$ but at the levels $\alpha = 2.5\%$, $\beta = 5\%$ for the nominal level of 5%, and at the levels $\alpha = 1\%$, $\beta = 2.5\%$ for the nominal level of 10%. As n increases to 1000, the sizes deteriorate, and the powers improve across the nominal levels for data generated with the Normal distribution. Both behaviours are particularly evident at the significance levels of $\alpha = 1\%, \beta = 2.5\%$, and $T = 250$. Furthermore, the proposed methodologies generally exhibit the best size tests primarily from null models with the normal distribution, emphasizing 1000 in-sample observations. When the data is generated by processes that capture the properties of the financial assets (skewness or heavy tails), the proposed procedures exhibit higher powers when RVaR predictions are conducted using the normal distribution.

In the empirical study, we note that, as the numerical analysis, none of the proposed backtesting methods stand out over the others. Also, when we assess the resulting p-values of the backtesting procedures jointly with the respective values obtained via the RVaR loss function, we can assert that there is a greater agreement between the methods when looking at the best scenarios. More precisely, when analyzing the data from the perspective of the p-values closer to one, we observe predominantly small realized losses among those calculated. Similarly, from the perspective of realized losses, the smaller scores exhibit significant p-values at a 5% significance level. On the other hand, from the perspective of the p-values closer to zero, we predominantly observe small scores. From the perspective of realized losses, the worst predictions showed significant p-values at 5%, with few exceptions. Hence, it becomes evident that the forecasts attained exhibit a higher degree of alignment with the two methods when scrutinized within the context of best-case scenarios while manifesting more pronounced divergences for the worst cases, thus presenting limited relations between the processes.

Although we carried out comprehensive analyzes to evaluate the proposed methods, it is essential to address the inherent limitations within the scope of this research. Due to technical and time constraints, the GAS models could not be implemented in their entirety, serving as the primary limiting factor of the Monte Carlo simulations applied in the study. Furthermore, the construction of the p-value was approached numerically, not proposing an analytical form and, therefore, a more robust backtesting procedure. For future research, we suggest advances in the mentioned limitations, expanding the scope of the scenarios considered in the numerical analysis, as well as the development of a more elaborate backtesting procedure, making use of a more sophisticated statistical theory and, therefore, making it more efficient for the evaluation of the risk predictions of the RVaR measure.

Bibliography

ACERBI, C.; SZEKELY, B. Backtesting expected shortfall. **Risk**, Incisive Media Limited, p. 76, 2014.

ACERBI, C.; SZEKELY, B. General properties of backtestable statistics. **Working Paper**, 2017. Available in: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2905109>.

ACERBI, C.; TASCHE, D. On the coherence of expected shortfall. **Journal of Banking & Finance**, Elsevier, v. 26, n. 7, p. 1487–1503, 2002.

ALBERG, D.; SHALIT, H.; YOSEF, R. Estimating stock market volatility using asymmetric garch models. **Applied Financial Economics**, Taylor & Francis, v. 18, n. 15, p. 1201–1208, 2008.

ARDIA, D.; HOOGERHEIDE, L. F. Garch models for daily stock returns: impact of estimation frequency on value-at-risk and expected shortfall forecasts. **Economics Letters 123**, p. 187–190, 2014.

ARDIA, D.; KRIS, B.; CATANIA, L. Generalized autoregressive score models in r: the gas package. **Journal of Statistical Software**, v. 88, n. 6, p. 1–28, 2019.

ARTZNER, P.; DELBAEN, F.; EBER, J.-M.; HEATH, D. Coherent measures of risk. **Mathematical finance**, Wiley Online Library, v. 9, n. 3, p. 203–228, 1999.

AZZALINI, A. A class of distributions which includes the normal ones. **Scandinavian Journal of Statistics**, v. 12, p. 171–178, 1985.

BARENDSE, S. Interquantile expectation regression. **Tinbergen Institute Discussion Paper 2017-034/III**, Tinbergen Institute, 2017.

BARNDORFF-NIELSEN, O. Exponentially decreasing distributions for the logarithm of particle size. **Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences**, Royal Society, v. 353, n. 1674, p. 401–419, 1977.

BARTELS, M.; ZIEGELMANN, F. A. Market risk forecasting for high dimensional portfolios via factor copulas with gas dynamics. **Insurance: Mathematics and economics**, Elsevier, v. 70, p. 66–79, 2016.

Basel Committee on Banking Supervision. **Fundamental review of the trading book: A revised market risk framework.** [S.l.], 2013. Available at: <https://www.bis.org/publ/bcbs265.pdf>.

BAYER, S.; DIMITRIADIS, T. Regression-based expected shortfall backtesting. **Journal of Financial Econometrics**, Oxford University Press, 2020.

BELLINI, F.; BIGNOZZI, V. On elicitable risk measures. **Quantitative Finance**, Taylor & Francis, v. 15, n. 5, p. 725–733, 2015.

BISWAS, S.; SEN, R. Nonparametric estimation of range value at risk. **Computation**, MDPI, v. 11, n. 28, p. 1–13, 2023.

BOLLERSLEV, T. Generalized autogressive conditional heteroskedasticity. **Journal of Econometrics**, Elsevier, v. 31, n. 3, p. 307–327, 1986.

BULAI, V.-C.; HOROBET, A.; POPOVICI, O. C.; BELASCU, L.; DUMITRESCU, S. A. A var-based methodology for assessing carbon price risk across european union economic sectors. **Energies**, MDPI, v. 14, p. 8424, 2022.

CATANIA, L.; BOUDT, K.; ARDIA, D. **GAS: Generalized Autoregressive Score Models**. [S.l.], 2019. R package version 0.3.4. Available at: <https://cran.r-project.org/web/packages/GAS/index.html>.

CATANIA, L.; GRASSI, S. Forecasting cryptocurrency volatility. **International Journal of Forecasting**, Elsevier, v. 38, n. 3, p. 878–894, 2022.

CONDYLIOS, S.; MIOTO, B.; SHALLOWAY, B. **priceR: Economics and Pricing Tools**. [S.l.], 2019. R package version 0.1.67. Available at: <https://cran.r-project.org/web/packages/priceR/index.html>.

CONT, R. Empirical properties of asset returns: stylized facts and statistical issues. **Quantitative Finance**, v. 1, n. 2, p. 223–236, 2001.

CONT, R.; DEGUEST, R.; SCANDOLO, G. Robustness and sensitivity analysis of risk measurement procedures. **Quantitative finance**, Taylor & Francis, v. 10, n. 6, p. 593–606, 2010.

CREAL, D.; KOOPMAN, S. J.; LUCAS, A. Generalized autoregressive score models with applications. **Journal of Applied Econometrics**, Wiley Online Library, v. 28, n. 5, p. 777–795, 2013.

DANÍELSSON, J. **Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and MATLAB**. [S.l.]: John Wiley & Sons Ltd, 2011.

DENG, K.; QIU, J. Backtesting expected shortfall and beyond. **Quantitative Finance**, Taylor & Francis, v. 21, n. 7, p. 1109–1125, 2021.

DIAZ, A.; GARCIA-DONATO, G.; MORA-VALENCIA, A. Risk quantification in turmoil markets. **Risk Management**, Springer, v. 19, n. 3, p. 202–224, 2017.

DU, Z.; ESCANCIANO, J. C. Backtesting expected shortfall: Accounting for tail risk. **Management Science**, INFORMS, v. 63, n. 4, p. 940–958, 2017.

EFRON, B.; TIBSHIRANI, R. J. **An Introduction to the Bootstrap**. [S.l.]: New York: Chapman & Hall, 1993.

EMBRECHTS, P.; WANG, B.; WANG, R. Aggregation-robustness and model uncertainty of regulatory risk measures. **Finance and Stochastics**, Springer, v. 19, n. 4, p. 763–790, 2015.

ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. **Econometrica**, The Econometric Society, v. 50, n. 4, p. 987–1007, 1982.

ESCANCIANO, J. C.; PEI, P. Pitfalls in backtesting historical simulation var models. **Journal of Banking & Finance**, Elsevier, v. 36, n. 8, p. 2233–2244, 2012.

ESCANCIANO, J. C.; VELASCO, C. Specification tests of parametric dynamic conditional quantiles. **Journal of Econometrics**, Elsevier, v. 159, n. 1, p. 209–221, 2010.

FERNANDEZ, C.; STEEL, M. F. J. On bayesian modelling of fat tails and skewness. **Journal of the American Statistical Association**, Taylor and Francis, v. 93, n. 441, p. 359–371, 1998.

FISSLER, T.; ZIEGEL, J. Higher order elicitability and osband's principle. **The Annals of Statistics**, v. 44, p. 1680–1707, 08 2016.

FISSLER, T.; ZIEGEL, J. F. Evaluating range value at risk forecasts. **arXiv preprint arXiv:1902.04489**, 2021.

FÖLLMER, H.; WEBER, S. The axiomatic approach to risk measures for capital determination. **Annual Review of Financial Economics**, Annual Reviews, v. 7, p. 301–337, 2015.

FRITTELLI, M.; GIANIN, E. R. Putting order in risk measures. **Journal of Banking & Finance**, Elsevier, v. 26, n. 7, p. 1473–1486, 2002.

GALANOS, A. **rugarch: Univariate GARCH models**. [S.l.], 2019. R package version 1.4-8. Available at: <https://cran.r-project.org/web/packages/rugarch/index.html>.

GARCIA-JORCANO, L.; NOVALES, A. Volatility specifications versus probability distributions in var forecasting. **Journal of Forecasting**, Wiley Online Library, v. 40, n. 2, p. 189–212, 2021.

GERLACH, R.; WALPOLE, D.; WANG, C. Semi-parametric bayesian tail risk forecasting incorporating realized measures of volatility. **Quantitative Finance**, Taylor & Francis, v. 17, n. 2, p. 199–215, 2017.

GILCHRIST, D.; YU, J.; ZHONG, R. The limits of green finance: a survey of literature in the context of green bonds and green loans. **Sustainability**, MDPI, v. 13, n. 2, 478, 2021.

GLOSTEN, L. R.; JAGANNATHAN, R.; RUNKLE, D. E. On the relation between the expected value and the volatility of the nominal excess return on stocks. **The Journal of Finance**, Wiley Online Library, v. 48, n. 5, p. 1779–1801, 1993.

GNEITING, T. Making and evaluating point forecasts. **Journal of the American Statistical Association**, Taylor & Francis, v. 106, n. 494, p. 746–762, 2011.

GOSSET, W. S. The probable error of a mean. **Biometrika Trust**, Oxford University Press, v. 6, n. 1, p. 1–25, 1908.

HAMPEL, F. R. A general qualitative definition of robustness. **The Annals of Mathematical Statistics**, JSTOR, p. 1887–1896, 1971.

HARVEY, A. C. **Dynamic models for volatility and heavy tails: with applications to financial and economic time series**. [S.l.]: Cambridge University Press, 2013.

JOHNSON, N. L. Systems of frequency curves generated by methods of translation. **Biometrika Trust**, Oxford University Press, v. 36, n. 1/2, p. 149–176, 1949.

KRÄTSCHMER, V.; SCHIED, A.; ZÄHLE, H. Comparative and qualitative robustness for law-invariant risk measures. **Finance and Stochastics**, Springer, v. 18, n. 2, p. 271–295, 2014.

KUESTER, K.; MITTNIK, S.; PAOLELLA, M. S. Value-at-risk prediction: A comparison of alternative strategies. **Journal of Financial Econometrics**, Oxford University Press, v. 4, n. 1, p. 53–89, 2006.

LI, W. K.; LING, S.; MCALEER, M. Recent theoretical results for time series models with garch errors. **Journal of Economic Surveys**, Wiley Online Library, v. 16, n. 3, p. 245–269, 2002.

MCCRACKEN, D. D. The monte carlo method. **Scientific American**, JSTOR, v. 192, n. 5, p. 90–97, 1955.

MCDONALD, J. B.; NEWEY, W. K. Partially adaptive estimation of regression models via the generalized t distribution. **Econometric Theory**, JSTOR, v. 4, n. 3, p. 428–457, 1988.

MCNEIL, A. J.; FREY, R. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. **Journal of Empirical Finance**, 2000.

MCNEIL, A. J.; FREY, R.; EMBRECHTS, P. **Quantitative risk management: concepts, techniques and tools-revised edition**. [S.l.]: Princeton university press, 2015.

MIGHRI, Z.; MOKNI, K.; MANSOURI, F. Empirical analysis of asymmetric long memory volatility models in value-at-risk estimation. **The Journal of Risk**, Incisive Risk Information, v. 13, n. 1, p. 55–128, 2010.

MÜLLER, F. M.; RIGHI, M. B. Numerical comparison of multivariate models to forecasting risk measures. **Risk Management**, Springer, v. 20, n. 1, p. 29–50, 2018.

MÜLLER, F. M.; SANTOS, S. S.; GÖSSLING, T. W.; RIGHI, M. B. Comparison of risk forecasts for cryptocurrencies: A focus on range value at risk. **Finance Research Letters**, Elsevier, v. 48, p. 102916, 2022.

NELSON, D. B. Conditional heteroskedasticity in asset returns: a new approach. **Econometrica**, JSTOR, v. 59, n. 2, p. 347–370, 1991.

NILSEN, J. A. S. **Market Risk Management Under the Fundamental Review of the Trading Book: Volatility Modeling and Historical Simulation of Value at Risk and Expected Shortfall**. Thesis (Master of Economic Theory and Econometrics) - University of Oslo, Oslo, 2022.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2023. Available at: <http://www.R-project.org/>.

RIGHI, M.; CERETTA, P. S. Individual and flexible expected shortfall backtesting. **Journal of Risk Model Validation**, v. 7, n. 3, p. 3–20, 2013.

RIGHI, M. B.; CERETTA, P. S. A comparison of expected shortfall estimation models. **Journal of Economics and Business**, Elsevier, v. 78, p. 14–47, 2015.

RODRÍGUEZ, M. J.; RUIZ, E. Revisiting several popular garch models with leverage effect: differences and similarities. **Journal of Financial Econometrics**, Oxford University Press, v. 10, n. 4, p. 637–668, 2012.

RYAN, J. A.; ULRICH, J. M.; SMITH, E. B.; THIELEN, W.; TEETOR, P.; BRONDER, S. **quantmod: Quantitative Financial Modelling Framework**. [S.l.], 2008. R package version 0.4.22. Available at: <https://cran.r-project.org/web/packages/quantmod/index.html>.

SINGH, A. K.; ALLEN, D. E.; POWELL, R. J. Value at risk estimation using extreme value theory. **19th International Congress on Modelling and Simulation**, 2011.

SIU, T. K. The risks of cryptocurrencies with long memory in volatility, non-normality and behavioural insights. **Applied Economics**, Taylor & Francis, v. 53, n. 7, p. 1991–2014, 2021.

SUBRAMONEY, S. D.; CHINHAMU, K.; CHIFURIRA, R. Value at risk estimation using gas models with heavy tailed distributions for cryptocurrencies. **International Journal of Finance and Banking Studies**, SSBFNET, v. 10, n. 4, p. 40–54, 2021.

THEODOSSIOU, P. Financial data and the skewed generalized t distribution. **Management Science**, Informs, v. 44, n. 12, p. 1650–1661, 1988.

TIAN, D.-s.; CAI, Z.-w.; FANG, Y. Econometric modeling of risk measures: A selective review of the recent literature. **Applied Mathematics-A Journal of Chinese Universities**, Springer, v. 34, n. 2, p. 205–228, 2019.

TROSTER, V.; TIWARI, A. K.; SHAHBAZ, M.; MACEDO, D. N. Bitcoin returns and risk: A general garch and gas analysis. **Finance Research Letters**, Elsevier, v. 30, p. 187–193, 2019.

TRUCÍOS, C. Forecasting bitcoin risk measures: A robust approach. **International Journal of Forecasting**, Elsevier, v. 35, n. 3, p. 836–847, 2019.

TRUCÍOS, C.; TIWARI, A. K.; ALQAHTANI, F. Value-at-risk and expected shortfall in cryptocurrencies' portfolio: A vine copula-based approach. **Available at SSRN 3441892**, 2019.

UMAR, M.; MIRZA, N.; RIZVI, S. K. A.; FURQAN, M. Asymmetric volatility structure of equity returns: Evidence from an emerging market. **Working Paper**, 2021. Available in: <https://www.sciencedirect.com/science/article/pii/S1062976921000788>.

WANG, R.; WEI, Y. Risk functionals with convex level sets. **Mathematical Finance**, Wiley Online Library, 2020.

XU, Y.; LIEN, D. Forecasting volatilities of oil and gas assets: A comparison of gas, garch, and egarch models. **Journal of Forecasting**, Wiley Online Library, v. 41, n. 2, p. 1–20, 2021.

ZIEGEL, J. F. Coherence and elicitability. **Mathematical Finance**, Wiley Online Library, v. 26, n. 4, p. 901–918, 2016.

APPENDIX A – Tables

Table 40 – Loss function results of Real/US Dollar using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n=250$			$n=500$			$n=1000$		
	$\alpha=1\%,\beta=5\%$	$\alpha=1\%,\beta=2.5\%$	$\alpha=2.5\%,\beta=5\%$	$\alpha=1\%,\beta=5\%$	$\alpha=1\%,\beta=2.5\%$	$\alpha=2.5\%,\beta=5\%$	$\alpha=1\%,\beta=5\%$	$\alpha=1\%,\beta=2.5\%$	$\alpha=2.5\%,\beta=5\%$
GARCH_{norm}	1.045891	1.049365	1.048576	1.045998	1.049484	1.048672	1.047541	1.051173	1.050045
GARCH_{std}	1.048213	1.052663	1.048711	1.048319	1.052820	1.048722	1.049853	1.054534	1.050000
GARCH_{sstd}	1.048002	1.052398	1.048582	1.048857	1.053446	1.049145	1.050670	1.055489	1.050621
GARCH_{snorm}	1.045754	1.049207	1.048468	1.046360	1.049909	1.048960	1.048120	1.051848	1.050509
GARCH_{ged}	1.048683	1.053184	1.049716	1.048515	1.053029	1.049515	1.050499	1.055268	1.051167
GARCH_{sged}	1.048607	1.053066	1.049703	1.049354	1.053984	1.050209	1.051453	1.056358	1.051954
GARCH_{jsu}	1.047902	1.052273	1.048696	1.048969	1.053604	1.049315	1.050996	1.055912	1.050924
GARCH_{nig}	1.048147	1.052595	1.048969	1.049312	1.054023	1.049714	1.051332	1.056326	1.051322
EGARCH_{norm}	1.045359	1.048768	1.048114	1.045197	1.048609	1.047957	1.047005	1.050593	1.049563
EGARCH_{std}	1.047893	1.052266	1.048477	1.047833	1.052251	1.048350	1.050114	1.054828	1.050230
EGARCH_{sstd}	1.048214	1.052646	1.048690	1.048638	1.053186	1.048968	1.051137	1.056020	1.051002
EGARCH_{snorm}	1.045435	1.048854	1.048182	1.045554	1.049023	1.048240	1.047576	1.051257	1.050022
EGARCH_{ged}	1.047694	1.051986	1.049004	1.048529	1.053052	1.049389	1.050174	1.054887	1.050909
EGARCH_{sged}	1.047982	1.052291	1.049296	1.048821	1.054367	1.049681	1.051578	1.056491	1.052071
EGARCH_{jsu}	1.050445	1.054990	1.051010	1.048543	1.053104	1.049010	1.051530	1.056520	1.051375
EGARCH_{nig}	1.048195	1.052656	1.048983	1.048871	1.053515	1.049376	1.051935	1.057009	1.051832
gjrGARCH_{norm}	1.047217	1.051137	1.048237	1.045696	1.049144	1.048411	1.047460	1.051079	1.049980
gjrGARCH_{std}	1.047977	1.052297	1.048075	1.048017	1.052460	1.048528	1.049658	1.054303	1.049870
gjrGARCH_{sstd}	1.047586	1.051831	1.047800	1.048580	1.053113	1.048974	1.050527	1.055315	1.050532
gjrGARCH_{snorm}	1.047228	1.051143	1.048261	1.046147	1.049667	1.048778	1.048119	1.051845	1.050512
gjrGARCH_{ged}	1.048079	1.052395	1.048452	1.048154	1.052609	1.049238	1.050279	1.055012	1.050999
gjrGARCH_{sged}	1.047889	1.052154	1.048335	1.049088	1.053676	1.050008	1.051324	1.056204	1.051861
gjrGARCH_{jsu}	1.047744	1.052017	1.047998	1.048636	1.053205	1.049088	1.050915	1.055812	1.050876
gjrGARCH_{nig}	1.047756	1.052038	1.048051	1.048885	1.053524	1.049395	1.051200	1.056167	1.051229
GAS_{sstd}	1.049672	1.054171	1.050308	1.048938	1.053345	1.049695	1.051694	1.056428	1.052016
GAS_{snorm}	1.048482	1.052178	1.050911	1.048501	1.052218	1.050914	1.047756	1.051383	1.050263

Table 41 – Loss function results of Bitcoin using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n = 250$			$n = 500$			$n = 1000$		
	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$
$GARCH_{norm}$	1.139853	1.152148	1.132512	1.148147	1.161289	1.139878	1.158762	1.172726	1.149580
$GARCH_{std}$	1.145191	1.162839	1.123219	1.149141	1.167484	1.125754	1.159041	1.178265	1.134720
$GARCH_{sstd}$	1.146328	1.164118	1.124259	1.151452	1.170173	1.127564	1.162609	1.182449	1.137354
$GARCH_{snorm}$	1.144006	1.156860	1.135997	1.153204	1.167006	1.144142	1.164464	1.179217	1.154349
$GARCH_{ged}$	1.155464	1.174712	1.135862	1.162318	1.182850	1.140879	1.176099	1.198269	1.152543
$GARCH_{sged}$	1.157633	1.177208	1.137545	1.166040	1.187140	1.143812	1.182580	1.205728	1.157717
$GARCH_{jsu}$	1.154777	1.174606	1.130295	1.161164	1.182269	1.134425	1.173570	1.196086	1.144951
$GARCH_{nig}$	1.161950	1.183427	1.136412	1.168936	1.191871	1.141107	1.182608	1.207271	1.152482
$EGARCH_{norm}$	1.149889	1.163061	1.141597	1.147374	1.160367	1.139238	1.158133	1.171966	1.149070
$EGARCH_{std}$	1.157186	1.177067	1.130474	1.172447	1.195279	1.139285	1.190750	1.216277	1.152255
$EGARCH_{sstd}$	1.160069	1.180439	1.132534	1.180633	1.205025	1.144505	1.201824	1.229478	1.159364
$EGARCH_{snorm}$	1.151869	1.165341	1.143207	1.151959	1.165587	1.143073	1.163579	1.178169	1.153633
$EGARCH_{ged}$	1.156325	1.175591	1.136948	1.166115	1.187266	1.143799	1.180824	1.203735	1.156128
$EGARCH_{sged}$	1.161353	1.181490	1.140539	1.173356	1.195683	1.149383	1.191133	1.215686	1.164232
$EGARCH_{jsu}$	1.165485	1.187319	1.137420	1.182440	1.207613	1.148315	1.201473	1.229409	1.162938
$EGARCH_{nig}$	1.170094	1.193124	1.141917	1.186137	1.212464	1.152672	1.205403	1.234562	1.167796
$gjrGARCH_{norm}$	1.137505	1.149620	1.130387	1.146268	1.159177	1.138270	1.158554	1.172457	1.149466
$gjrGARCH_{std}$	1.148458	1.166576	1.125682	1.148591	1.166774	1.125724	1.158250	1.177304	1.134257
$gjrGARCH_{sstd}$	1.150188	1.168569	1.127054	1.150544	1.169040	1.127258	1.161675	1.181308	1.136789
$gjrGARCH_{snorm}$	1.141212	1.153789	1.133554	1.150487	1.163975	1.141782	1.163888	1.178546	1.153885
$gjrGARCH_{ged}$	1.155110	1.174229	1.135852	1.161260	1.181594	1.140107	1.175908	1.198011	1.152368
$gjrGARCH_{sged}$	1.156463	1.175753	1.136901	1.164788	1.185657	1.142915	1.182744	1.205878	1.157842
$gjrGARCH_{jsu}$	1.155645	1.175521	1.131297	1.159766	1.180550	1.133692	1.172545	1.194823	1.144290
$gjrGARCH_{nig}$	1.161602	1.182843	1.136682	1.167371	1.189964	1.140143	1.182016	1.206528	1.152049
GAS_{sstd}	1.149992	1.166244	1.133922	1.155052	1.172033	1.137764	1.169115	1.187538	1.149743
GAS_{snorm}	1.156511	1.170450	1.147281	1.199337	1.216890	1.185938	1.155732	1.169340	1.146886

Table 42 – Loss function results of DGS3 using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n = 250$			$n = 500$			$n = 1000$		
	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$
$GARCH_{norm}$	1.178484	1.193585	1.167983	1.171708	1.186282	1.161805	1.163817	1.177750	1.154599
$GARCH_{std}$	1.186166	1.204851	1.167700	1.178467	1.196463	1.160999	1.170974	1.188009	1.155055
$GARCH_{sstd}$	1.189424	1.208595	1.170373	1.182830	1.201601	1.164376	1.174718	1.192436	1.157923
$GARCH_{snorm}$	1.180329	1.195731	1.169450	1.175269	1.190504	1.164589	1.167558	1.182159	1.157536
$GARCH_{ged}$	1.183204	1.204621	1.152365	1.218456	1.248819	1.163904	1.174902	1.192715	1.159510
$GARCH_{sged}$	1.186516	1.208165	1.159596	1.199320	1.225228	1.162857	1.178609	1.196993	1.162604
$GARCH_{jsu}$	1.190079	1.209531	1.170981	1.184355	1.203569	1.165552	1.175675	1.193678	1.158686
$GARCH_{nig}$	1.190307	1.209898	1.171519	1.185358	1.204800	1.166746	1.176179	1.194286	1.159447
$EGARCH_{norm}$	1.178840	1.193886	1.168505	1.170431	1.184817	1.160785	1.162343	1.176129	1.153339
$EGARCH_{std}$	1.188671	1.207524	1.170216	1.179299	1.197296	1.161924	1.166847	1.183304	1.151833
$EGARCH_{sstd}$	1.194155	1.213994	1.174242	1.183793	1.202632	1.165272	1.169723	1.186754	1.153976
$EGARCH_{snorm}$	1.182421	1.198034	1.171401	1.172775	1.187655	1.162528	1.164749	1.178999	1.155170
$EGARCH_{ged}$	1.354467	1.410225	1.225097	1.344821	1.399846	1.216224	1.181164	1.200659	1.160181
$EGARCH_{sged}$	1.283317	1.321809	1.225902	1.273671	1.311429	1.217029	1.194968	1.217861	1.166402
$EGARCH_{jsu}$	1.195017	1.215048	1.175405	1.184126	1.203172	1.165644	1.170530	1.187787	1.154663
$EGARCH_{nig}$	1.196467	1.216766	1.177035	1.184513	1.203656	1.166406	1.171012	1.188348	1.155403
$gjrGARCH_{norm}$	1.177781	1.192761	1.167452	1.171425	1.185966	1.161556	1.163301	1.177150	1.154139
$gjrGARCH_{std}$	1.184953	1.203259	1.167279	1.178764	1.196575	1.161820	1.169688	1.186391	1.154410
$gjrGARCH_{sstd}$	1.188305	1.207197	1.169953	1.183710	1.202406	1.165600	1.173192	1.190573	1.157084
$gjrGARCH_{snorm}$	1.180345	1.195756	1.169521	1.175046	1.190241	1.164403	1.166672	1.181136	1.156797
$gjrGARCH_{ged}$	1.190563	1.213120	1.156505	1.220345	1.250801	1.165772	1.173311	1.190736	1.158503
$gjrGARCH_{sged}$	1.190378	1.211869	1.164566	1.204826	1.230787	1.168969	1.175926	1.193777	1.160699
$gjrGARCH_{jsu}$	1.189558	1.208834	1.171229	1.185087	1.204171	1.166690	1.173953	1.191574	1.157682
$gjrGARCH_{nig}$	1.189921	1.209306	1.171936	1.185945	1.205222	1.167734	1.174258	1.191942	1.158259
GAS_{sstd}	1.197593	1.217467	1.178171	1.192693	1.212318	1.173423	1.195233	1.215547	1.174773
GAS_{snorm}	1.203957	1.221539	1.190727	1.336191	1.366414	1.308299	1.589034	1.644496	1.532448

Table 43 – Loss function results of Dow Jones Sustainability Index World using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n = 250$			$n = 500$			$n = 1000$		
	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$
$GARCH_{norm}$	1.037366	1.040083	1.040899	1.035340	1.037888	1.039061	1.036137	1.038757	1.039766
$GARCH_{std}$	1.039341	1.042840	1.041077	1.037660	1.041029	1.039575	1.038071	1.041450	1.040025
$GARCH_{sstd}$	1.038704	1.042092	1.040627	1.037865	1.041270	1.039738	1.038531	1.041990	1.040371
$GARCH_{snorm}$	1.036760	1.039383	1.040408	1.035274	1.037813	1.039010	1.036275	1.038915	1.039881
$GARCH_{ged}$	1.039444	1.042928	1.041723	1.037743	1.041094	1.040175	1.038284	1.041678	1.040670
$GARCH_{sged}$	1.039232	1.042670	1.041574	1.038262	1.041698	1.040594	1.038956	1.042455	1.041209
$GARCH_{jsu}$	1.038629	1.042014	1.040678	1.037886	1.041320	1.039811	1.038639	1.042149	1.040477
$GARCH_{nig}$	1.038669	1.042082	1.040780	1.038019	1.041493	1.040010	1.038794	1.042351	1.040677
$EGARCH_{norm}$	1.036625	1.039209	1.040312	1.035514	1.038039	1.039278	1.036885	1.039551	1.040490
$EGARCH_{std}$	1.038334	1.041567	1.040577	1.036823	1.039957	1.039238	1.038019	1.041250	1.040399
$EGARCH_{sstd}$	1.040243	1.043803	1.041940	1.038263	1.041651	1.040287	1.039314	1.042763	1.041375
$EGARCH_{snorm}$	1.036717	1.039315	1.040388	1.035744	1.038305	1.039467	1.037225	1.039943	1.040770
$EGARCH_{ged}$	1.037901	1.041023	1.040704	1.036852	1.039968	1.039658	1.038268	1.041536	1.040911
$EGARCH_{sged}$	1.039241	1.042561	1.041790	1.038358	1.041706	1.040872	1.039739	1.043233	1.042095
$EGARCH_{jsu}$	1.039505	1.042945	1.041544	1.038227	1.041636	1.040297	1.039455	1.042968	1.041454
$EGARCH_{nig}$	1.039415	1.042853	1.041560	1.038224	1.041641	1.040389	1.039576	1.043125	1.041602
$gjrGARCH_{norm}$	1.036608	1.039223	1.040258	1.034769	1.037245	1.038583	1.035692	1.038259	1.039406
$gjrGARCH_{std}$	1.037877	1.041058	1.040261	1.036063	1.039119	1.038588	1.036952	1.040084	1.039440
$gjrGARCH_{sstd}$	1.038384	1.041649	1.040656	1.036875	1.040073	1.039199	1.037874	1.041163	1.040138
$gjrGARCH_{snorm}$	1.036519	1.039126	1.040183	1.034971	1.037478	1.038750	1.035994	1.038606	1.039654
$gjrGARCH_{ged}$	1.038135	1.041358	1.040795	1.036380	1.039482	1.039186	1.037335	1.040532	1.040043
$gjrGARCH_{sged}$	1.038926	1.042266	1.041440	1.037447	1.040717	1.040046	1.038484	1.041860	1.040964
$gjrGARCH_{jsu}$	1.038277	1.041552	1.040618	1.036996	1.040246	1.039295	1.038016	1.041366	1.040219
$gjrGARCH_{nig}$	1.038562	1.041899	1.040832	1.037132	1.040419	1.039463	1.038186	1.041583	1.040395
GAS_{sstd}	1.039252	1.042574	1.041490	1.038032	1.041272	1.040376	1.038691	1.042024	1.040897
GAS_{snorm}	1.037627	1.040299	1.041221	1.036441	1.039005	1.040179	1.036245	1.038779	1.040012

Table 44 – Loss function results of Euro/US Dollar using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n = 250$			$n = 500$			$n = 1000$		
	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$
$GARCH_{norm}$	1.024370	1.026093	1.029041	1.023888	1.025571	1.028609	1.023100	1.024711	1.027902
$GARCH_{std}$	1.024813	1.026802	1.028896	1.024520	1.026529	1.028523	1.023904	1.025888	1.027894
$GARCH_{sstd}$	1.025318	1.027387	1.029294	1.024953	1.027035	1.028856	1.024291	1.026340	1.028184
$GARCH_{snorm}$	1.024666	1.026437	1.029279	1.024144	1.025868	1.028816	1.023278	1.024916	1.028048
$GARCH_{ged}$	1.025055	1.027065	1.029273	1.024757	1.026787	1.028927	1.024107	1.026103	1.028299
$GARCH_{sged}$	1.025520	1.027599	1.029647	1.025250	1.027356	1.029320	1.024533	1.026593	1.028644
$GARCH_{jsu}$	1.025406	1.027506	1.029359	1.025044	1.027158	1.028922	1.024341	1.026415	1.028235
$GARCH_{nig}$	1.025489	1.027612	1.029440	1.025107	1.027239	1.029006	1.024443	1.026538	1.028357
$EGARCH_{norm}$	1.024652	1.026241	1.029210	1.024265	1.025967	1.028963	1.023366	1.024989	1.028151
$EGARCH_{std}$	1.025253	1.027062	1.029361	1.025112	1.027166	1.029072	1.023934	1.025893	1.027984
$EGARCH_{sstd}$	1.026092	1.028040	1.030001	1.025785	1.027947	1.029588	1.024447	1.026489	1.028375
$EGARCH_{snorm}$	1.025272	1.026949	1.029724	1.024706	1.026471	1.029328	1.023606	1.025265	1.028348
$EGARCH_{ged}$	1.025529	1.027374	1.029602	1.024989	1.027004	1.029189	1.024131	1.026094	1.028379
$EGARCH_{sged}$	1.025869	1.027743	1.030013	1.025754	1.027882	1.029811	1.024641	1.026679	1.028792
$EGARCH_{jsu}$	1.026045	1.028001	1.029962	1.025832	1.028017	1.029634	1.024506	1.026571	1.028428
$EGARCH_{nig}$	1.026088	1.028056	1.030013	1.025872	1.028067	1.029702	1.024562	1.026638	1.028519
$gjrGARCH_{norm}$	1.024131	1.025826	1.028834	1.023634	1.025290	1.028384	1.023006	1.024603	1.027822
$gjrGARCH_{std}$	1.024653	1.026626	1.028744	1.024350	1.026332	1.028397	1.023696	1.025636	1.027768
$gjrGARCH_{sstd}$	1.025209	1.027265	1.029186	1.024879	1.026945	1.028808	1.024099	1.026106	1.028073
$gjrGARCH_{snorm}$	1.024539	1.026297	1.029164	1.024006	1.025718	1.028689	1.023200	1.024827	1.027981
$gjrGARCH_{ged}$	1.024828	1.026803	1.029084	1.024544	1.026535	1.028764	1.023927	1.025881	1.028178
$gjrGARCH_{sged}$	1.025369	1.027425	1.029524	1.025127	1.027208	1.029233	1.024382	1.026404	1.028546
$gjrGARCH_{jsu}$	1.025325	1.027383	1.029367	1.024982	1.027067	1.028900	1.024163	1.026195	1.028127
$gjrGARCH_{nig}$	1.025692	1.027777	1.029731	1.025155	1.027248	1.029055	1.024250	1.026299	1.028236
GAS_{sstd}	1.026580	1.028823	1.030269	1.026649	1.028950	1.030185	1.026742	1.029057	1.030237
GAS_{snorm}	1.025356	1.027176	1.029917	1.025132	1.026936	1.029713	1.025284	1.027094	1.029862

Table 45 – Loss function results of DGS10 using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n = 250$			$n = 500$			$n = 1000$		
	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$
GARCH_{norm}	1.102041	1.110485	1.099078	1.101044	1.109390	1.098160	1.102639	1.111140	1.099582
GARCH_{std}	1.104860	1.114494	1.099365	1.103981	1.113518	1.098589	1.105732	1.115456	1.100139
GARCH_{sstd}	1.102726	1.111982	1.097678	1.102657	1.111943	1.097613	1.104249	1.113730	1.098961
GARCH_{snorm}	1.101594	1.109943	1.098752	1.100682	1.108962	1.097880	1.102036	1.110435	1.099105
GARCH_{ged}	1.105075	1.114468	1.100609	1.104196	1.113561	1.099720	1.106125	1.115729	1.101358
GARCH_{sged}	1.103517	1.112678	1.099327	1.103030	1.112210	1.098772	1.104438	1.113782	1.099981
GARCH_{jsu}	1.103549	1.112904	1.098577	1.103281	1.112649	1.098261	1.105047	1.114648	1.099716
GARCH_{nig}	1.103908	1.113341	1.098939	1.103462	1.112869	1.098497	1.105147	1.114753	1.099920
EGARCH_{norm}	1.101669	1.109991	1.098831	1.100020	1.108168	1.097340	1.100426	1.108658	1.097690
EGARCH_{std}	1.104333	1.113613	1.099479	1.102063	1.111110	1.097478	1.102204	1.111333	1.097562
EGARCH_{sstd}	1.102315	1.111262	1.097876	1.100252	1.109013	1.096017	1.100143	1.108931	1.095933
EGARCH_{snorm}	1.100589	1.108718	1.097979	1.098504	1.106423	1.096091	1.098747	1.106700	1.096353
EGARCH_{ged}	1.103384	1.112296	1.099610	1.101687	1.110520	1.097906	1.102558	1.111599	1.098556
EGARCH_{sged}	1.101934	1.110621	1.098425	1.101429	1.110106	1.097837	1.100677	1.109425	1.097026
EGARCH_{jsu}	1.102759	1.111718	1.098483	1.100545	1.109324	1.096413	1.100601	1.109436	1.096433
EGARCH_{nig}	1.104035	1.113260	1.099411	1.100780	1.109610	1.096645	1.100635	1.109474	1.096536
gjrGARCH_{norm}	1.100438	1.108689	1.097684	1.100557	1.108825	1.097758	1.102223	1.110643	1.099262
gjrGARCH_{std}	1.103012	1.112343	1.098154	1.102859	1.112101	1.098003	1.104591	1.113997	1.099576
gjrGARCH_{sstd}	1.101350	1.110311	1.096885	1.101367	1.110336	1.096872	1.102923	1.112052	1.098259
gjrGARCH_{snorm}	1.099801	1.107925	1.097198	1.099681	1.107807	1.097059	1.101099	1.109329	1.098372
gjrGARCH_{ged}	1.102798	1.111795	1.098871	1.102908	1.111966	1.098865	1.104947	1.114273	1.100585
gjrGARCH_{sged}	1.101545	1.110347	1.097851	1.101717	1.110605	1.097872	1.103198	1.112257	1.099150
gjrGARCH_{jsu}	1.102123	1.111157	1.097726	1.101728	1.110733	1.097304	1.103570	1.112788	1.098883
gjrGARCH_{nig}	1.102310	1.111389	1.097946	1.100762	1.109704	1.096296	1.103628	1.112848	1.099021
GAS_{sstd}	1.111588	1.122272	1.104350	1.114440	1.125603	1.106413	1.117037	1.128465	1.108612
GAS_{snorm}	1.107138	1.116022	1.103706	1.143048	1.155292	1.135625	1.125852	1.136385	1.120595

Table 46 – Loss function results of Gold using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n = 250$			$n = 500$			$n = 1000$		
	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$
GARCH_{norm}	1.041999	1.045294	1.044882	1.041501	1.044738	1.044448	1.039338	1.042395	1.042489
GARCH_{std}	1.044908	1.049647	1.044293	1.044696	1.049408	1.044132	1.041783	1.046111	1.041871
GARCH_{sstd}	1.046661	1.051669	1.045684	1.045782	1.050657	1.045001	1.042338	1.046758	1.042293
GARCH_{snorm}	1.042998	1.046471	1.045661	1.042369	1.045748	1.045137	1.040125	1.043302	1.043129
GARCH_{ged}	1.044683	1.049256	1.045400	1.044579	1.049128	1.045317	1.042058	1.046302	1.043209
GARCH_{sged}	1.046479	1.051299	1.046899	1.045271	1.049894	1.045939	1.042417	1.046708	1.043515
GARCH_{jsu}	1.047121	1.052270	1.046238	1.046297	1.051305	1.045610	1.042909	1.047467	1.042900
GARCH_{nig}	1.047390	1.052619	1.046701	1.046610	1.051700	1.046106	1.043274	1.047914	1.043411
EGARCH_{norm}	1.041823	1.045089	1.044721	1.041666	1.044920	1.044591	1.039302	1.042355	1.042455
EGARCH_{std}	1.045558	1.050378	1.044791	1.044698	1.049373	1.044237	1.041616	1.045922	1.041748
EGARCH_{sstd}	1.047112	1.052172	1.046022	1.045995	1.050865	1.045263	1.042220	1.046625	1.042205
EGARCH_{snorm}	1.042765	1.046179	1.045476	1.042486	1.045866	1.045256	1.040075	1.043244	1.043090
EGARCH_{ged}	1.044739	1.049263	1.045436	1.044235	1.048684	1.045138	1.041848	1.046059	1.043052
EGARCH_{sged}	1.045784	1.050413	1.046483	1.045076	1.049622	1.045874	1.042240	1.046503	1.043386
EGARCH_{jsu}	1.047692	1.052885	1.046723	1.046296	1.051262	1.045716	1.042762	1.047300	1.042793
EGARCH_{nig}	1.047700	1.052917	1.047010	1.046399	1.051407	1.046052	1.043093	1.047708	1.043278
gjrGARCH_{norm}	1.041170	1.044372	1.044145	1.041316	1.044532	1.044281	1.039016	1.042041	1.042204
gjrGARCH_{std}	1.044612	1.049297	1.044084	1.044500	1.049161	1.044024	1.041551	1.045848	1.041697
gjrGARCH_{sstd}	1.046569	1.051555	1.045630	1.045541	1.050359	1.044856	1.042038	1.046412	1.042071
gjrGARCH_{snorm}	1.042068	1.045423	1.044851	1.042055	1.045395	1.044868	1.039704	1.042833	1.042762
gjrGARCH_{ged}	1.043746	1.048159	1.044679	1.044095	1.048558	1.044947	1.041649	1.045834	1.042878
gjrGARCH_{sged}	1.045699	1.050386	1.046230	1.044782	1.049319	1.045562	1.041944	1.046165	1.043132
gjrGARCH_{jsu}	1.046840	1.051917	1.046023	1.045977	1.050944	1.045421	1.042576	1.047082	1.042648
gjrGARCH_{nig}	1.047277	1.052442	1.046567	1.046637	1.051788	1.046092	1.042902	1.047486	1.043122
GAS_{sstd}	1.045506	1.049913	1.046074	1.045840	1.050335	1.046184	1.044017	1.048379	1.044467
GAS_{snorm}	1.043087	1.046517	1.045808	1.043898	1.047438	1.046495	1.042450	1.045832	1.045233

Table 47 – Loss function results of Ibovespa using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n = 250$			$n = 500$			$n = 1000$		
	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$
$GARCH_{norm}$	1.060055	1.064984	1.061062	1.059841	1.064766	1.060863	1.060465	1.065481	1.061387
$GARCH_{std}$	1.060857	1.066427	1.060396	1.061010	1.066639	1.060454	1.061821	1.067577	1.061097
$GARCH_{sstd}$	1.062213	1.067989	1.061568	1.062129	1.067924	1.061389	1.062355	1.068196	1.061527
$GARCH_{snorm}$	1.061271	1.066434	1.061991	1.061275	1.066438	1.061997	1.061572	1.066777	1.062265
$GARCH_{ged}$	1.061400	1.066991	1.061324	1.061623	1.067310	1.061405	1.062673	1.068547	1.062220
$GARCH_{sged}$	1.062874	1.068667	1.062585	1.063261	1.069168	1.062792	1.063612	1.069620	1.063007
$GARCH_{jsu}$	1.062661	1.068534	1.061944	1.062934	1.068881	1.062034	1.062953	1.068918	1.062013
$GARCH_{nig}$	1.063072	1.069022	1.062318	1.063250	1.069258	1.062345	1.063271	1.069300	1.062326
$EGARCH_{norm}$	1.060167	1.065072	1.061220	1.059500	1.064336	1.060625	1.059677	1.064554	1.060741
$EGARCH_{std}$	1.061477	1.067016	1.061147	1.060225	1.065676	1.059959	1.060868	1.066445	1.060380
$EGARCH_{sstd}$	1.063238	1.069027	1.062616	1.061617	1.067275	1.061102	1.061475	1.067156	1.060858
$EGARCH_{snorm}$	1.061594	1.066740	1.062357	1.060635	1.065670	1.061521	1.060527	1.065563	1.061405
$EGARCH_{ged}$	1.061575	1.067023	1.061663	1.060747	1.066202	1.060873	1.061542	1.067191	1.061368
$EGARCH_{sged}$	1.062723	1.068303	1.062792	1.062340	1.068026	1.062200	1.062450	1.068240	1.062116
$EGARCH_{jsu}$	1.063640	1.069523	1.062954	1.062127	1.067899	1.061515	1.061880	1.067657	1.061186
$EGARCH_{nig}$	1.064072	1.070048	1.063294	1.062361	1.068181	1.061750	1.062136	1.067967	1.061448
$gjrGARCH_{norm}$	1.060116	1.065007	1.061172	1.059937	1.064825	1.060998	1.060454	1.065429	1.061424
$gjrGARCH_{std}$	1.060804	1.066259	1.060600	1.060926	1.066456	1.060584	1.061593	1.067271	1.061024
$gjrGARCH_{sstd}$	1.062641	1.068388	1.062123	1.062303	1.068070	1.061737	1.062111	1.067879	1.061432
$gjrGARCH_{snorm}$	1.061521	1.066707	1.062277	1.061344	1.066469	1.062128	1.061203	1.066321	1.062000
$gjrGARCH_{ged}$	1.061113	1.066536	1.061315	1.061477	1.067043	1.061458	1.062349	1.068118	1.062058
$gjrGARCH_{sged}$	1.062983	1.068702	1.062915	1.063050	1.068860	1.062788	1.063278	1.069183	1.062823
$gjrGARCH_{jsu}$	1.063253	1.069105	1.062699	1.062530	1.068370	1.061989	1.062630	1.068506	1.061851
$gjrGARCH_{nig}$	1.063475	1.069366	1.062911	1.062517	1.068384	1.062041	1.062933	1.068870	1.062147
GAS_{sstd}	1.065239	1.071491	1.063689	1.065384	1.071690	1.063719	1.066949	1.073525	1.064859
GAS_{snorm}	1.063122	1.068365	1.063765	1.063978	1.069319	1.064519	1.064050	1.069398	1.064591

Table 48 – Loss function results of S&P500 using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n = 250$			$n = 500$			$n = 1000$		
	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$
$GARCH_{norm}$	1.036811	1.039817	1.039982	1.035479	1.038382	1.038770	1.036329	1.039300	1.039541
$GARCH_{std}$	1.038924	1.042893	1.039892	1.037538	1.041380	1.038682	1.038778	1.042786	1.039646
$GARCH_{sstd}$	1.041615	1.045981	1.042194	1.040056	1.044284	1.040784	1.040990	1.045343	1.041463
$GARCH_{snorm}$	1.039591	1.043076	1.042182	1.038102	1.041456	1.040844	1.038549	1.041917	1.041279
$GARCH_{ged}$	1.039627	1.043654	1.041129	1.038259	1.042157	1.039928	1.039395	1.043435	1.040872
$GARCH_{sged}$	1.041999	1.046346	1.043165	1.040528	1.044734	1.041868	1.041212	1.045493	1.042434
$GARCH_{jsu}$	1.042514	1.047101	1.042860	1.040973	1.045418	1.041485	1.041879	1.046442	1.042180
$GARCH_{nig}$	1.042835	1.047492	1.043226	1.041369	1.045901	1.041900	1.042269	1.046927	1.042591
$EGARCH_{norm}$	1.036752	1.039708	1.040002	1.035431	1.038291	1.038791	1.036195	1.039107	1.039489
$EGARCH_{std}$	1.038375	1.042124	1.039805	1.037002	1.040631	1.038613	1.038209	1.041999	1.039552
$EGARCH_{sstd}$	1.041800	1.046092	1.042583	1.040432	1.044604	1.041385	1.041450	1.045758	1.042148
$EGARCH_{snorm}$	1.039306	1.042707	1.042011	1.038083	1.041390	1.040900	1.038606	1.041933	1.041395
$EGARCH_{ged}$	1.038477	1.042169	1.040499	1.037301	1.040931	1.039405	1.038580	1.042383	1.040433
$EGARCH_{sged}$	1.041472	1.045576	1.043064	1.040452	1.044517	1.042062	1.041354	1.045534	1.042791
$EGARCH_{jsu}$	1.042490	1.046973	1.043089	1.041249	1.045635	1.041964	1.042279	1.046807	1.042740
$EGARCH_{nig}$	1.042746	1.047282	1.043400	1.041491	1.045928	1.042262	1.042552	1.047142	1.043065
$gjrGARCH_{norm}$	1.037488	1.040537	1.040633	1.036180	1.039121	1.039442	1.037111	1.040130	1.040290
$gjrGARCH_{std}$	1.039283	1.043184	1.040508	1.037801	1.041579	1.039157	1.039128	1.043083	1.040208
$gjrGARCH_{sstd}$	1.042495	1.046892	1.043128	1.041008	1.045286	1.041784	1.042120	1.046548	1.042621
$gjrGARCH_{snorm}$	1.040325	1.043856	1.042878	1.038906	1.042315	1.041603	1.039567	1.043014	1.042219
$gjrGARCH_{ged}$	1.039754	1.043679	1.041480	1.038313	1.042117	1.040165	1.039708	1.043700	1.041308
$gjrGARCH_{sged}$	1.042729	1.047055	1.044011	1.041279	1.045489	1.042689	1.042326	1.046668	1.043546
$gjrGARCH_{jsu}$	1.043336	1.047960	1.043730	1.041857	1.046356	1.042389	1.043012	1.047666	1.043286
$gjrGARCH_{nig}$	1.043622	1.048296	1.044073	1.042152	1.046712	1.042734	1.043395	1.048132	1.043698
GAS_{sstd}	1.041152	1.045213	1.042341	1.039855	1.043798	1.041191	1.040580	1.044656	1.041687
GAS_{snorm}	1.043594	1.047299	1.045968	1.043722	1.047483	1.046033	1.044776	1.048596	1.047026

Table 49 – Loss function results of STOXX Europe 600 using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n = 250$			$n = 500$			$n = 1000$		
	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$
GARCH_{norm}	1.040715	1.043968	1.043620	1.039374	1.042540	1.042384	1.039603	1.042774	1.042599
GARCH_{std}	1.042327	1.046428	1.043291	1.041233	1.045317	1.042165	1.041703	1.045829	1.042582
GARCH_{sstd}	1.045074	1.049612	1.045482	1.043675	1.048151	1.044114	1.044028	1.048532	1.044415
GARCH_{snorm}	1.042984	1.046606	1.045443	1.041260	1.044743	1.043887	1.041188	1.044632	1.043850
GARCH_{ged}	1.042731	1.046842	1.044228	1.041517	1.045576	1.043078	1.041875	1.045959	1.043396
GARCH_{sged}	1.045627	1.050172	1.046610	1.044039	1.048463	1.045166	1.044342	1.048793	1.045414
GARCH_{jsu}	1.045836	1.050566	1.046061	1.044345	1.048999	1.044612	1.044586	1.049255	1.044807
GARCH_{nig}	1.046081	1.050885	1.046339	1.044612	1.049334	1.044928	1.044786	1.049510	1.045076
EGARCH_{norm}	1.040196	1.043346	1.043233	1.038493	1.041514	1.041685	1.039761	1.042890	1.042829
EGARCH_{std}	1.041600	1.045455	1.043028	1.039797	1.043543	1.041377	1.040950	1.044820	1.042381
EGARCH_{sstd}	1.046375	1.051029	1.046589	1.043374	1.047677	1.044227	1.044399	1.048817	1.045093
EGARCH_{snorm}	1.042577	1.046122	1.045132	1.040949	1.044362	1.043663	1.042104	1.045613	1.044703
EGARCH_{ged}	1.041101	1.044781	1.043261	1.039821	1.043513	1.041925	1.041252	1.045126	1.043117
EGARCH_{sged}	1.044960	1.049217	1.046416	1.043572	1.047795	1.045035	1.044830	1.049217	1.046058
EGARCH_{jsu}	1.045798	1.050374	1.046375	1.043972	1.048431	1.044668	1.045060	1.049659	1.045529
EGARCH_{nig}	1.046330	1.051018	1.046867	1.044222	1.048731	1.044963	1.045246	1.049886	1.045776
gjrGARCH_{norm}	1.041104	1.044361	1.044012	1.039518	1.042654	1.042569	1.040596	1.043832	1.043541
gjrGARCH_{std}	1.042537	1.046540	1.043794	1.041004	1.044921	1.042320	1.042090	1.046127	1.043270
gjrGARCH_{sstd}	1.045733	1.050262	1.046301	1.044141	1.048558	1.044803	1.045133	1.049665	1.045654
gjrGARCH_{snorm}	1.043441	1.047085	1.045886	1.041839	1.045352	1.044436	1.042732	1.046320	1.045250
gjrGARCH_{ged}	1.042722	1.046695	1.044475	1.041162	1.045045	1.043005	1.042373	1.046406	1.044036
gjrGARCH_{sged}	1.046082	1.050559	1.047225	1.044478	1.048843	1.045742	1.045644	1.050161	1.046719
gjrGARCH_{jsu}	1.046354	1.051038	1.046776	1.044776	1.049359	1.045272	1.045724	1.050423	1.046056
gjrGARCH_{nig}	1.046612	1.051348	1.047073	1.044989	1.049620	1.045539	1.045948	1.050698	1.046333
GAS_{sstd}	1.044724	1.049042	1.045544	1.042975	1.047151	1.043982	1.044124	1.048513	1.044781
GAS_{snorm}	1.044896	1.048569	1.047313	1.045261	1.049038	1.047572	1.054241	1.058920	1.055500

Table 50 – Loss function results of WTI Crude Oil using significance levels of 1%, 2.5%, and 5%; with 250, 500, and 1,000 rolling window observation sizes.

Model	$n = 250$			$n = 500$			$n = 1000$		
	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$	$\alpha = 1\%, \beta = 5\%$	$\alpha = 1\%, \beta = 2.5\%$	$\alpha = 2.5\%, \beta = 5\%$
$GARCH_{norm}$	1.090657	1.098283	1.088639	1.093644	1.101559	1.091299	1.099629	1.108056	1.096694
$GARCH_{std}$	1.093352	1.102958	1.086614	1.096713	1.106716	1.089404	1.102546	1.112947	1.095078
$GARCH_{sstd}$	1.097808	1.108122	1.090052	1.101613	1.112430	1.093135	1.107837	1.119111	1.099178
$GARCH_{snorm}$	1.092563	1.100493	1.090163	1.096094	1.104400	1.093276	1.103198	1.112187	1.099594
$GARCH_{ged}$	1.094970	1.104683	1.089500	1.097858	1.107914	1.091976	1.103968	1.114485	1.097634
$GARCH_{sged}$	1.098631	1.108865	1.092499	1.102264	1.112978	1.095563	1.109838	1.121248	1.102402
$GARCH_{jsu}$	1.098979	1.109592	1.091108	1.102853	1.114002	1.094237	1.109155	1.120787	1.100174
$GARCH_{nig}$	1.099322	1.110034	1.091712	1.103129	1.114371	1.094793	1.109670	1.121421	1.100868
$EGARCH_{norm}$	1.088125	1.095401	1.086489	1.089831	1.097295	1.087996	1.095298	1.103226	1.092905
$EGARCH_{std}$	1.095771	1.105776	1.087658	1.095520	1.105386	1.088111	1.098634	1.108405	1.092043
$EGARCH_{sstd}$	1.105133	1.116756	1.094429	1.102931	1.114059	1.093526	1.104592	1.115359	1.096664
$EGARCH_{snorm}$	1.091164	1.098915	1.088937	1.093106	1.101077	1.090661	1.099166	1.107714	1.096047
$EGARCH_{ged}$	1.091739	1.100800	1.087155	1.093901	1.103327	1.088829	1.099346	1.109148	1.093910
$EGARCH_{sged}$	1.096964	1.106783	1.091442	1.099639	1.109912	1.093509	1.105877	1.116684	1.099220
$EGARCH_{jsu}$	1.101714	1.112737	1.093212	1.102180	1.113257	1.093561	1.105873	1.117006	1.097588
$EGARCH_{nig}$	1.101508	1.112513	1.093487	1.101989	1.113078	1.093823	1.106204	1.117428	1.098115
$gjrGARCH_{norm}$	1.087805	1.095068	1.086169	1.091437	1.099067	1.089420	1.097985	1.106191	1.095305
$gjrGARCH_{std}$	1.091217	1.100369	1.085165	1.094629	1.104275	1.087889	1.100628	1.110657	1.093803
$gjrGARCH_{sstd}$	1.097319	1.107457	1.089834	1.100895	1.111593	1.092608	1.106567	1.117576	1.098401
$gjrGARCH_{snorm}$	1.090728	1.098432	1.088542	1.094420	1.102523	1.091840	1.101709	1.110510	1.098329
$gjrGARCH_{ged}$	1.091482	1.100560	1.086838	1.095067	1.104661	1.089802	1.101765	1.111882	1.095983
$gjrGARCH_{sged}$	1.096748	1.106589	1.091156	1.100861	1.111306	1.094535	1.108363	1.119484	1.101333
$gjrGARCH_{jsu}$	1.098689	1.109188	1.091038	1.102213	1.113265	1.093759	1.107870	1.119243	1.099328
$gjrGARCH_{nig}$	1.098698	1.109245	1.091277	1.102357	1.113471	1.094226	1.108245	1.119710	1.099887
GAS_{sstd}	1.097873	1.107705	1.091392	1.101457	1.111816	1.094109	1.111545	1.123126	1.102423
GAS_{snorm}	1.098416	1.106549	1.095849	1.105899	1.114702	1.102575	1.118494	1.128341	1.113967