NOVA IMS Information Management School

MEGI

Mestrado em Estatística e Gestão de Informação
Master Program in Statistics and Information Management

Models for Forecasting Value at Risk

A comparison of the predictive ability of different VaR models to capture market losses incurred during the 2020 pandemic recession

Gonçalo de Sousa Neves

Dissertation report presented as partial requirement for obtaining the Master’s degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA IMS
Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

Models for Forecasting Value at Risk

A comparison of the predictive ability of different VaR models to capture market losses incurred during the 2020 pandemic recession

Gonçalo de Sousa Neves

Dissertation report presented as partial requirement for obtaining the Master’s degree in Statistics and Information Management

2022

Title: Models for Forecasting Value at Risk

Subtitle: A comparison of the predictive ability of different VaR models.

Gonçalo de Sousa Neves

MEGI

2022

Title: Models for Forecasting Value at Risk

Subtitle: A comparison of the predictive ability of different VaR models.

Gonçalo de Sousa Neves

MGI

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

MODELS FOR FORECASTING VALUE AT RISK

by

Gonçalo de Sousa Neves

Dissertation report presented as partial requirement for obtaining the Master's degree in Information Management/ Master's degree in Statistics and Information Management, with a specialization in Risk Analysis and Management

Advisor / Co Advisor: Maria Helena Miranda Flores Batista

August 2022

ABSTRACT

The purpose of this study is two-fold. First, it aims at providing a theoretical overview of the most widely adopted methods for forecasting Value-at-Risk (VaR). Second, through a practical implementation, it proposes a methodology to compare and evaluate the predictive ability of different parametric, non-parametric and semi-parametric models to capture the market losses incurred during the COVID-19 pandemic recession of 2020. To evaluate these models, it is applied a two-staged backtesting procedure based on accuracy statistical tests and loss functions. VaR forecasts are evaluated during a volatile and a stable forecasting periods. The results of the study suggest that, for the volatile period, the Extreme Value Theory with a peaks over threshold (EVT-POT) approach produces the most accurate VaR forecasts across all different methodologies. The Filtered Historical Simulation (FHS), Volatility Weighted Historical Simulation (VWHS) and the Glosten, Jagannathan and Runkle (GJR) GARCH with skewed generalized error distribution (GJR GARCH–SGED) models also produce satisfactory forecasts. Moreover, other parametric approaches, namely the GARCH and EWMA, despite less accurate, also produce reliable results. Furthermore, the overall performance of all models improves significantly during the stable forecasting period. For instance, the Historical Simulation with exponentially decreasing weights (BRW HS), one of the worst performers during the volatile forecasting period, produces the most accurate VaR forecasts, with the lowest penalty scores, during the stable forecasting period. Lastly, it was also found that as the level of conservativeness of the model increases, the overestimation of the actual incurred risk seems to a be recurrent event.

KEYWORDS

COVID-19; Financial Loss; Pandemic Recession; Stock Indexes; Value at Risk

Index

List of Equations

List of Figures

List of Tables

List of Abbreviations and Acronyms

ADIs	Authorized Deposit Taking Institutions
APARCH	Asymmetric Power ARCH Volatility
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroskedasticity
AS	Asymmetric slope
BBM	Block Maxima Model
BCBS	Basel Committee on Banking Supervision
BRW HS	Boudoukh, Richardson and Whitelaw Historical Simulation
CAViaR	Conditional Autoregressive Value at Risk by Regression Quantiles
CLRM	Classical Linear Regression Model
DJI 30	Dow Jones Industrial Average Index
EGARCH	Exponential Generalized Autoregressive Conditional Heteroskedasticity
EQWMA	Equally-Weighted Moving Average
ER	Exceedance Rate
EVT	Extreme Value Theory
EWMA	Exponentially Weighted Moving Average
FHS	Filtered Historical Simulation
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GEV	Generalized Extreme Value
GJR GARCH	Glosten, Jagannathan and Runkle Generalized Autoregressive Conditional Heteroskedasticity
GPD	Generalized Pareto Distribution
GSPC	Standard and Poor’s (S&P 500) index
HS	Historical Simulation
I.I.D	Independent and identically distributed
IG	Indirect GARCH
IMA	Internal Model Approach

IXIC	Nasdaq index
LLF	Lopez loss function
LM	Lagrange Multiplier
LR_{cc}	Conditional Coverage test
LR_{uc}	Unconditional Coverage test
MCS	Monte Carlo Simulation
N	Normal distribution
POT	Peaks Over Threshold
QLF	Quantile loss function
QPS	Quadratic Probability Score
RV	Realized Volatilities
SA	Standardized Approach
SAV	Symmetric absolute value
SGED	Skewed generalized error distribution
SMA	Simple Moving Average
SSTD	Skewed student t-distribution
STD	Student t-distribution
SWARCH	Markov Switching ARCH
SV	Stochastic Volatility
VaR	Value at Risk
VWHS	Volatility Weighted Historical Simulation

1. Introduction

The global financial crisis of 2008-09 highlighted the importance of ensuring that the financial system has enough liquidity to withstand adverse circumstances. The funding pressures that began in 2007 emphasized the severe deficiencies in the liquidity and risk management practices of some banks and financial institutions. As such, this lead the Basel Committee on Banking Supervision (BCBS) to ”intensify its efforts to strengthen the principals and standards for capital…” within the financial system (Gomes & Khan, 2011).

From a market risk perspective, the latest version of the Basel Accords (the III) introduced new requirements for both the Standardized Approach (SA) and Internal Model Approach (IMA) with the objective of ensuring that firms are able to meet capital and liquidity needs under stressed market conditions.

With this new set of stress testing regulations, regulatory institutions pledged banks to calculate their capital and liquidity requirements under stressed market scenarios. In this manner, banks would be able to assess whether if the estimates of loss produced by their internal models were adequate for predicting losses in abnormal market scenarios (Moody’s Analytics, 2011).

These new capital requirements lead banks and financial institutions to further adopt and evaluate alternative Value at Risk (VaR) models that could capture the potential risk of loss in the markets under these unusual scenarios (Holton, 2002).

Given these concerns, this study aims to assess the ability of different VaR models to predict the market losses incurred during the COVID-19 pandemic recession of 2020 for three stock market indexes (DJI 30, S&P500, NASDAQ). To achieve this purpose, VaR is forecasted for the year of 2020 using different parametric, non-parametric and semi-parametric methods. After that, the performance of each model is evaluated and compared through a backtesting procedure based on accuracy tests and loss functions.

Subsequently, a comparable analysis is performed for a stable forecasting period. This complementary analysis seeks to assess whether the relative performance of each model remains unchanged throughout both the volatile and stable forecasting periods.

The relevancy of this study lies in the fact that previous literature findings often report conflicting results. Additionally, as highlighted by Abad et al. (2014), most of these studies compare a very narrow range of models whose conclusions heavily depend on the data used for analysis. Lastly, at the best of our knowledge, there is not a substantial amount of research focused on assessing the adequacy of these models to capture the market losses incurred during the year of 2020 (see Lazar et al. (2022) and Shaik & Padmakumari (2022) for some examples).

As such, this study proposes a theoretical overview and practical application of different VaR models. Like so, the study not only contributes with a granular overview of the most used VaR methodologies, but it also proposes a detailed assessment and comparison of the performance of these approaches under various sets of data and different time frames.

The research process of this study is threefold. First, it is identified the most common techniques used for forecasting VaR. Then, with the purpose of selecting the models which are often found to produce the most accurate VaR forecasts, it is performed an exploratory research where it is featured some of the statistical assumptions, advantages and limitations of each model. Lastly, it is resorted to statistical software to produce VaR forecasts and to compare the relative performance of each model under both scenarios.

This study is structured the following way. **Section 2.1** provides some important remarks regarding the VaR computation and the characteristics of financial returns. **Section 2.2** provides an overview of the VaR history and its regulatory framework. **Section 2.3** explains the parity between the process of forecasting volatility and forecasting VaR. In **Section 2.4**, it is presented a literature review of the different VaR models. Moreover, an overview the models’ statistical specifications, advantages and shortcomings is also presented within this section. **Section 3.1** presents the data used for analysis while **Section 3.2** provides an overview of the methodology used to apply each of the VaR models. **Section 3.3** presents the evaluation framework used to compare the performance of each of the models. In **Section 4** are detailed all the empirical findings of this study. **Section 5** unveils the study's major conclusion and, lastly, **Section 6** displays the limitations and recommendations for future work.

2. Literature review

2.1. CHARACTERISTICS OF FINANCIAL RETURNS

2.1.1. Forecasting VaR: percent returns or continuously compounded returns

A return reflects the change in the value of a financial instrument or portfolio over a pre-determined evaluation period, for example over a business day, n. This measure is expressed by a percentage and reflects the profit or loss realized on an investment. This percentage change can be measured using simple percentage returns, see **Equation 1**, or continuously compounded, or log, returns, **see Equation 2**.

$$U_n = \frac{P_n - P_{n-1}}{P_{n-1}}$$

Equation 1. Formulation of simple percentage returns.

$$u_n = \ln(1 + U_n) = \ln\left(\frac{P_n}{P_{n-1}}\right)$$

Equation 2. Formulation of continuously compounded returns.

With P_n and P_{n-1} representing the price of the financial instrument at day n and at day $n-1$, respectively. U_n represents the percentage return while u_n denotes the log-return for a particular day n.

For tasks that aim to study returns throughout time, with temporal aggregation, it is simpler to apply the formula of log-returns. This is possible because log-returns are time additive and time additivity is not verified in simple returns.

On the other hand, if one desires to study the cross-sectional aggregation of returns (i.e., multiple individual returns that are measured at any point in time), the simple returns formula becomes more adequate. **Table 1** provides an overview of the two different ways to perform return aggregation and presents the formulas used to compute both percent and log-returns for each of these alternatives.

Table 1.

Return aggregation: percent returns and continuously compounded returns

Aggregation	Temporal	Cross-section
Percent returns	$U_{in}(k) = \prod_{n=1}^{N}(1 + U_{in}) - 1$	$U_{pn} = \sum_{i=1}^{N} w_i U_{in}$
Continuously compounded returns	$u_{in}(k) = \sum_{i=n}^{N} u_{in}$	$u_{pn} = ln\left(\sum_{i=1}^{N} wie^{u_{in}}\right)$

$U_{in}(k)$ represents the product of k-days 1-day gross returns, where U_{in} denotes the percentage return for a financial instrument i for a particular day n. $u_{in}(k)$ defines the sum of the k continuously compounded

1-day returns, where u_{in} represents the continuously compounded return for a financial instrument i for a particular day n. U_{pn} and u_{pn} present the percentage portfolio returns and the continuously compounded returns for a portfolio p, for a particular day n, respectively. Lastly, w_i presents the portfolio weights, that is, the fraction of the total portfolio, p, value allocated to the *ith* instrument, under the assumption that there are no short positions.

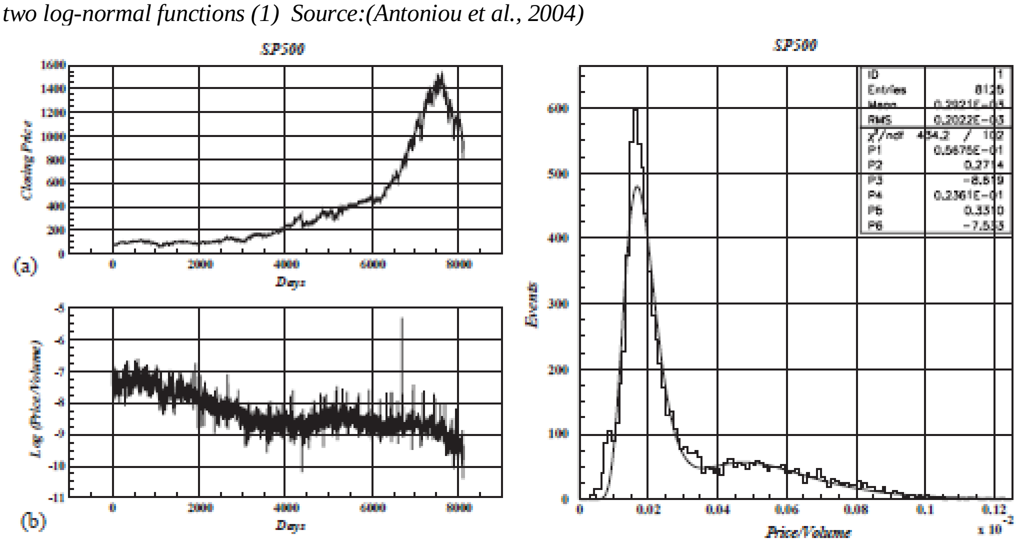
Given that the purpose of this study is to produce Value at Risk (VaR) forecasts which are dependent on the temporal aggregation of returns, log-returns ought to be considered to reflect the change in the value of a financial instrument or portfolio over a particular evaluation period.

Moreover the log-distribution also has proven to be able to effectively describe the distribution of a financial instruments' closing price, see Antoniou et al. (2004). The author demonstrates that the distribution of a financial instruments' daily closing price normalized by its traded volume fits well the log-normal distribution. See **Figure 1** for an example. As such, the distribution of the daily continuously compounded returns is described by these log-price changes, also called continuously compounded returns, or log-returns. Note that these three terms can be used interchangeably.

Additionally, the log-normal distribution also imposes a minimal bound of zero to a variable's value. This property is consistent with the behavior of stock prices which cannot go below the value of zero. Also, when considering very small returns, which is common for trades with short holding periods, the log-normal returns approximate the percent-returns (Morgan, 1996).

Figure 1

Left plots (from top to bottom): ***(a)*** *daily closing price data for S&P500 index;* ***(b)*** *the logarithm of daily closing prices normalized by traded volumes. Right plot: the distribution of daily closing prices normalized by traded volumes superimposed by the weighted sum of two log-normal functions (1) Source:(Antoniou et al., 2004)*



2.1.2. Modeling financial prices and returns

Most financial models used to describe the evolution of price returns are often based of the assumption that these prices follow a log-normal distribution assumption. However, this assumption presumes that log-prices have constant variance, which is not true (see **Section 2.3**). Also, standard distributions such as the normal and log-normal distribution fail to describe some features of the returns' distribution. Some of these features include: 1) volatility clustering: large changes in price tend to be followed by large changes, and small changes tend to followed by small changes (He et al., 2016); 2) heavy tails or extreme returns: the likelihood to encounter significant deviations from the mean is much greater than in the case of the normal distribution; iii) asymmetric distributions: the distribution of returns is skewed to the left with heavier tails than the distribution of gains (i.e., returns exhibit a leptokurtic distribution, see Bradley and Taqqu (2001)), iv) leverage effects: negative correlation of the lagged returns with the current continuous volatility component and/or current jump, which indicates that losses have a greater influence on future volatility than gains (Bollerslev et al., 2008), v) mean-reversion: the tendency of asset prices, and therefore, asset price changes, to return to a trend path. (Balvers et al., 2000)

As a result, in order to maximize the chances of producing accurate VaR forecasts, it is important that one applies VaR models that are capable of capturing these non-linear features of returns.

2.2. VAR HISTORY AND REGULATORY FRAMEWORK

Value at Risk (VaR) has been a recurring term commonly applied within the subject of Risk Management, and it is often linked with the terminology of risk. In fact, during the early 1990's, VaR was widely adopted for measuring market risk in trading portfolios. Thereby, it gained meaningful importance when, in 1998, it was set as the main risk measure reference for calculating a bank's capital requirements (Holton, 2002). Inevitably, VaR emerged as the financial service industry's premier risk management technique.

According to Philippe (2001), "VaR is defined as the worst expected loss over a given horizon under normal market conditions at a given confidence level". Hull (2018) simplifies this concept by referring that VaR allows a risk manager to state that he is X percent certain that he will not lose more than V dollars in time N.

VaR is computed from the inverse of the cumulative distribution function of returns r, for a set probability of q, see **Equation 3**. VaR is represented by the smallest value that exceeds q, or $1 - \dot{\alpha}$, with $\dot{\alpha}$ representing the significance level. It is a measure of loss for an investment, and it is typically used by financial entities to measure the number of assets required to cover eventual losses.

$$\overleftarrow{F(q)} = \inf \{r: F(r) \geq q\} = \sup \{r: F(r) < q\}$$

Equation 3. Empirical specification of Value at Risk.

The first reference of VaR as a risk measure in the scientific literature goes back to 1945. Although not explicitly, Leavens (1945) alluded multiple times to "the spread between losses and gains" to classify a portfolio's standard deviation.

In 1952, both Markowitz (1952) and Roy (1952) referred, and introduced a similar characterization for VaR. Both authors highlighted how probabilistic assumptions can be specified to measure VaR. However, given the limited availability of processing power at the time, VaR measures were mainly theoretical and were only published within the context of emerging portfolio theory.

Later in the 70's-80's, as the volume and complexity of market transactions increased, and as transactions became decentralized, trading firms had to find new ways to measure their risk exposure. Also, with the proliferation of leverage through the use of new derivative contracts, it became very difficult to evaluate portfolios. Resulting from this increased complexity, organizations sought to find a single risk measure that could be applied throughout multiple asset categories (Holton, 2002).

In 1988, the Basel Committee on Banking Supervision (BCBS) introduced the Basel I Accord, the first set of international banking regulations aimed at promoting banking stability. Nonetheless, despite defining new minimal emergency capital requirements for banks and, despite its efforts at promoting banking stability, the Basel I still failed to set specific requirements for market risk. For this reason, regulators soon recognized the inherit risks to the banking system "if insufficient capital were held to absorb the large sudden losses from huge exposures in capital markets" (Chang et al., 2011).

Subsequently, a new capital adequacy framework was implemented in 1998, through the Basel II Accord, which enhanced the requirements for market risk and aimed at rewarding institutions for their superior risk management systems. Thereby, with this new set of regulations, banks and other Authorized Deposit Taking Institutions (ADIs) became allowed to use internal models to forecast their daily VaR exposures. In that sense, a backtesting procedure was also implemented to compare actual returns with the corresponding VaR forecasts. This procedure allowed to assess the quality of the banks' and ADIs' internal models.

As a result, given these regulatory efforts, this decade had a significant growth of academic and professional literature focused on comparing alternative methodologies to measure VaR, particularly for large portfolios of financial instruments.

Once again, despite its efforts to standardize the banking regulatory framework and to assure banking stability, the Basel II failed to address and prevent the devastating consequences of the Global Financial Crisis of 2008-09. As such, in response to the crisis, the BCBS introduced a new regulatory regime for capital, liquidity and banking supervision (Moody's Analytics, 2011).

This new standard, the Basel III Accord, allowed banks to calculate a VaR forecast using different sets of weighting schemes, as long as the method results in a capital charge at least as conservative as the one defined for the Basel II Accord. This way, after recognizing the heavy dependence of VaR models on historical data, regulators defined that banks were no longer constraint to calculate VaR using a historical observation period with a minimum of one year (Sharma, 2012).

Given this increased flexibility to compute a bank's capital requirements and given the failure of previous VaR models to anticipate the losses incurred during the 2008 Financial Crisis, banks and scientific researchers have increased their efforts to explore alternative methodologies to compute their VaR forecasts.

2.3. FORECASTING VAR: MEAN AND VARIANCE PARAMETERS

As shown in **Equation 4**, the process of producing a forecast for the potential loss in the value of a portfolio generally depends upon two parameters, the mean μ and standard deviation σ.

$$VaR_{n+1} = \mu_{n+1} + F^{-1}\sigma_{n+1}$$

Equation 4. Specification of a VaR forecast.

Where μ_{n+1} and σ_{n+1} are the forecasts for the mean and standard deviation for the day $n+1$, given information of day n, and F^{-1} is the qth quantile of a particular distribution function F.

As the mean of the daily log-returns of financial instruments are usually not statistically different from zero, the mean parameter is often omitted from this equation, see Hull (2018). This simplication results in the following VaR specification,

$$VaR_{n+1} = F^{-1}\sigma_{n+1}$$

Therefore, the forecasting of VaR is closely linked with the forecasting of the variance of the financial instruments' log-returns. However, as briefly mentioned in **Section 2.1.2**, the variance of log-returns is not only non-constant throughout time but it also exhibits some dynamic features, such as volatility clustering.

Because the specification often used to compute the variance of a particular variable, see **Equation 5**, depends upon the assignment of an equal weight to each past observation, VaR cannot be forecasted using this volatility approach.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Equation 5. Specification of the equally weighted variance of a variable.

Where x_i represents the value of a particular random variable return of a financial instrument for a particular period i, $\bar{x}$ is the mean of that particular market variable for those N periods, N reflects the total number of periods.

Due to these reasons, there have been developed different VaR models whose formulations depend one way or another on the forecasting of a conditional variance which is able to accurately reflect these dynamic features of financial returns. These models will be explained in depth in the following sections. Also, note that the terms volatility and variance are used interchangeably throughout this study.

2.4. VALUE AT RISK METHODS

The literature review of this study is inspired by Abad et al. (2014)'s research paper. Given its comprehensive and extensive review on VaR models, this paper has been found to be, amongst researchers, one of the best papers at providing a detailed overview of the existing VaR methodologies. Thus, this study aims at not only extending, from a theoretical perspective, the overview of some of these methodologies but also at applying some of these methodologies in a practical scenario.

Due to the overwhelming amount of different VaR methodologies and model variations, and resulting from the impossibility of applying them all, this study address only a portion of the exiting approaches. The models were selected based on two distinct criteria. First, the relevance of the methodology for the research community, that is, whether if there is sufficient and relevant research focused on exploring the use of a specific approach. Second, the existence of enough resources to guarantee the practical implementation of the model, that is, whether if there is sufficient information available to assure that the application of the model is performed correctly.

To illustrate, in Abad et al. (2014)'s research, density estimation methods are found to be an existing alternative to other non-parametric methods, see **Section 2.4.2**. However, given the lack of scientific emphasis on the approach, it was not discussed in this study. On the other hand, there is extensive research focused on exploring the use of the Conditional Autoregressive Value At Risk By Regression Quantiles (CAViaR) approach for VaR forecasting purposes, see **Section 2.4.3.3**. Despite that, most studies disregard the practical process incurred to produce results, which makes it difficult to replicate. For that reason, the CAViaR is included only in the theorectical section of this study. Also, resulting from its poor performance and scarce application for forecasting VaR, the Stochastic Volatility (SV) model is also only discussed in the literature review section of this study, see **Section 2.4.1.4**. Lastly, the selection of the remaining volatility models was made with the purpose of highlighling the models which, in past literature, were proven to be the most capable of capturing the different non-linear features of financial returns.

Table 22 shown in the **Appendix** section of the study provides an overview of some research papers that compare the performace of different VaR methodologies.

2.4.1. Parametric methods

Parametric methods measure risk by fitting a probability distribution to the data. Then, VaR is inferred from the fitted curve. In essence, parametric methods assume that financial returns follow a pre-determined distribution.

The most common parametric approaches rely on the estimation of volatility. This is why parametric methods are often called volatility models. These volatility models include the Autoregressive Conditional Heteroskedasticity Model (ARCH), the Exponentially Weighted Moving Average (EWMA), the Generalized Autoregressive Conditional Heteroskedasticity (GARCH), the Glosten, Jagannathan and Runkle (GJR) Generalized Autoregressive Conditional Heteroskedasticity (GJR GARCH), the Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH), and Stochastic Volatility (SV) models.

2.4.1.1. Autoregressive Conditional Heteroskedasticity Model: ARCH (p)

Amongst all parametric approaches, the Autoregressive Conditional Heteroskedasticity (ARCH) model, introduced by Engle (1982), was the first process capable of capturing some of the non-linear features of financial returns. As volatilities of most financial instruments are proven to be heteroskedastic, the Classical Linear Regression Model (CLRM) is inappropriate to model their distributions. For this reason, the ARCH process introduces the concept of conditional volatilities. This way the method is capable of capturing the presence of volatility clustering of returns, and thus recognizing that returns' volatility is time dependent.

For a detailed analysis of the benefits brought by the ARCH process relative to the CLRM, see Morgan (1996).

According to this approach, the autocorrelation in volatility is modeled by allowing its conditional variance term to depend on previous *p* squared returns, see **Equation 6**. The older the return, the less weight given.

$$\sigma_n^2 = \zeta V_L + \sum_{i=1}^{p} \alpha_i u_{n-i}^2$$

Equation 6. Estimation of volatility using an ARCH (p) process.

The parameter *p* in the ARCH (*p*) specification refers to the number of lag returns to be included into the model. The estimate of the variance for a particular day n, σ_n^2, is based on a weighted long-run average variance term, ζV_L, often denoted by ω, plus the weighted sum of p past squared returns, where V_L represents the long-run variance rate and ζ is the weight assigned to V_L. α_i denotes the weight given to each past return i days ago and u_{n-i}^2 is the value of each squared return i days ago.

For a detailed discussion on the specification of the ARCH process see Brooks (2008), Engle (2001) and Kuester et al. (2006).

Despite capable of modeling some non-linear features of financial returns, such as volatility clustering and mean-reversion, the classic ARCH (*p*) process still is considered to be very unattractive not only to model return's distributions but also to forecast VaR.

First, while applying the model, it is difficult to determine the number of lags of past returns to be considered. Second, the number of lags required to capture all the dependence in the conditional variance can become very large. Third, there is a chance that non-negativity constraints are violated. These non-negativity constraints are defined so that the forecasted variance is always positive and that each of the model's parameter complies with its theoretical specifications, see Abad et al. (2014). Lastly, ARCH models often assume that the error term follows a Gaussian distribution, which is an inadequate assumption.

For these reasons, the classical ARCH process is found to be a very ineffective model to forecast VaR, often producing very inaccurate forecasts, see Angelidis et al. (2004) and Li and Lin (2004) as examples. As a result, scientific researchers have moved away from this classical formulation and have tried to explore different ARCH specifications whose empirical formulations seem likely to yield better VaR forecasts. These efforts include not only the application of the classical ARCH model under other distributional assumptions but also the adaption of this approach to different formulations. An example for this is provided by Giot and Laurent (2003) that compare the one-day-ahead forecasting accuracy of three different Asymmetric Power ARCH Volatility (APARCH) models, two based on the normal distribution and t-student distribution assumption, and other based on a skewed t-student distribution assumption. The authors find that models that rely on a skewed-distribution assumption considerably outperform models that rely on the normal distribution assumption. Also, Li and Lin (2004) suggest the implementation of a Markov Switching ARCH (SWARCH) model to capture the non-linear features of volatility. Its performance is compared with a linear ARCH and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) specification (see **Section 2.4.1.3**), which are shown to perform poorly in capturing non-linear features of financial returns. Finally, Angelidis et al. (2004) state that the ARCH process only yields acceptable results when the residuals are modeled under a student t-distribution (STD) or a generalized error distribution (GED).

2.4.1.2. Exponentially Weighted Moving Average Model: EWMA (Risk Metrics)

The Risk Metrics variance model, introduced by Morgan (1996), uses an Exponential Weighted Moving Average (EWMA) process to forecast variances. This was the first parametric model created specifically to forecast VaR and it relies on the assumption that financial returns are continuously compounded and that these follow a random walk process.

Since a random walk specification assumes that log-returns have constant variance, the EWMA model overcame this limitation by allowing variances to change over time, thus capturing some of the dynamic features of volatily, such as volatility clustering.

When compared with the Simple Moving Average (SMA) model the EWMA provides a meaningful advantage. While the SMA model assumes that volatiliy is estimated by giving the same weight to each observation, the EWMA allows recent observations to carry more weight. This way, the EWMA allows volatily to react faster to shocks in the markets. Also, following a market shock, the volatily estimate will decline exponentially as the weight given to the observation fades out with time. Contrastingly, when resorting to the SMA to forecast volatiliy, large shocks lead to abrupt changes in the variance once the shocks falls out of the measurement sample.

Equation 7 presents the formula used to compute the equally weighted volatily while **Equation 8** presents the formula for the exponentially decreasing weighted volatility. Volatily is denoted by σ^2.

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(u_n - \bar{u})^2$$

Equation 7. Estimation of equally weighted volatility using a SMA process.

$$\sigma^2 = (1-\lambda)\sum_{n=1}^{N}\lambda^{n-1}(u_n - \bar{u})^2$$

Equation 8. Estimation of volatility with exponentially decreasing weights using an EWMA process.

Where u_n represents the return of a financial instrument for a particular day n, $\bar{u}$ is the mean of returns, N reflects the number of days considered for the volatility computation and λ (or β) is called the decay factor.

Under the EWMA specification, the volatily estimate is a function of two factors: 1) the lagged variance and 2) the lagged squared return, where the weight parameter given to the latest lagged squared return is $(1-\lambda)$, or α. The estimate of variance for day n, σ_n^2 , is shown in **Equation 9,** where σ_{n-1}^2 represents the lagged estimated daily volatily and u_{n-1}^2 denotes the lagged squared return for day $n-1$. Accordingly, a (1,1) specification for the EWMA model states that the model accounts for only one lag variance and one lag return to compute a volatility estimate for day n.

$$\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2$$

$$where$$

$$(1-\lambda) = \alpha \; and \; \lambda = \beta$$

Equation 9. Estimation of volatility using an EWMA (1,1) process.

Despite contributing with meaninful improvements over the ARCH process, the EWMA model is still not able of capturing some dynamic features of volatilies, such as the leverage effects and the asymmetry of the return's distribution. Moreover, the model does not recognize the mean-reverting property of returns, thus failing to capture the persistence of volatilities throughout time (Abad et al., 2014).

Accordingly, given these limitations, research fidings show that the EWMA model, in general, produces innacurate VaR forecasts, being often outperformed by the GARCH (1,1) (see the following **Section 2.4.1.3** for a detailed overview of model) process. To illustrate, Abad and Benito (2013) produce VaR forecasts for eight different indexes and find that the EWMA model's forecasts are not accurate for any of the eight.

Similarly, Ñíguez (2008) compares the performance of various volatility models and finds that the EWMA produces the worst forecasts. Su and Knowles (2006) evaluate the out-of-sample VaR using a Quadratic Probability Score (QPS) function and find that the EWMA "tends to significantly understimate VaR at a 95 percent confidence level.". Moreover, Lu et al. (2009) compare VaR forecasts produced from two Equally-Weighted Moving Average (EQWMA) models, which rely a normal distribution and student t- distribution assumptions, and an EWMA model. The authors conclude that, at a 95% confidence level, the EQWMA based on the student t-distribution performs the worst and significantly overestimates VaR, whereas the EWMA performs "relatively well". On the other hand, at a 99% level, the EWMA methods performs the worst, deeply underestimating VaR.

2.4.1.3. Generalized Autoregressive Conditional Heteroskedasticity Model: GARCH (1,1)

Bollerslev (1986) extended the ARCH process by implementing a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. With this adaptation, the volatility estimate, σ_n^2 , became a weighted function based on three parameters 1) the weighted long-run average variance (ζV_L), often denoted by ω, 2) the weighted lagged squared return, (αu_{n-1}^2), and 3) the weighted volatility ($\beta\sigma_{n-1}^2$) , where α, β, and ζ denote the weights given to each of the parameters and V_L is the long-run variance rate, see **Equation 10**. This specification was later extended to allow for t-distributed errors, see Bollerslev (1987).

$$\sigma_n^2 = \zeta V_L + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$$

Equation 10. Estimation of volatility using a GARCH (1,1) process.

Fundamentally, a GARCH model can be viewed as an extension of an ARCH process that recognizes that more recent returns should be given more weight, or, as an EWMA model which recognizes that returns are mean-reverting. More specifically, the EWMA represents a particular case of a GARCH (1,1) model, where $\zeta = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$ (Hull, 2018).

Positivity constraints of the model are set so that $\alpha + \beta < 1$. Otherwise, the weight given to the long-run variance rate will be negative, which is an irrealistic outcome. Also, given that the sum of all parameters should be close to one, the long-run variance rate can be calculated the following way:

$$V_L = \frac{\omega}{\zeta}$$

Where $\zeta = 1 - \alpha - \beta$.

Regardless of its ability to model additional non-linear features, namely the mean-reversion property and volatility clustering of financial returns, literature fidings still suggest that the GARCH (1,1) process produces innacurate VaR forecasts. For instance, Ñíguez (2008) concludes that both EWMA (1,1) and GARCH (1,1) processes produce poor forecasts. Abad and Benito (2013) find that the GARCH model also provides imprecise forecasts. Gençay et al. (2003) observe that both the GARCH (1,1) based on normality and GARCH (1,1) based on a student-t distribution produce highly volatile forecasts across different stock indexes and confidence levels. Angelidis et al. (2004) also compare the performance of multiple GARCH models with different mean and volatility specifications, and observe that the models which are based on the assumption of normality often produce weak results.

2.4.1.4. Stochastic Volatility Model: SV

Another way of modeling conditional volatilities is through a stochastic volatility process. Oppositely to deterministic processes, such as the ARCH and GARCH models, which only have an error term in the mean equation, the Stochastic Volatility (SV) process adds a second error, ψ_n, term to forecast volatility.

The simplest autoregressive specification was introduced by Taylor (1982, 1986). The model is formulated as a log-normal autoregressive process with one lag, or AR (1). **Equation 11** and **12** present the model's mean and volatility specifications, respectively.

$$u_n = \mu_{n-1} + \sqrt{h_{n-1}}\delta_{n-1}, \ \ \sigma_n = \exp(h_n/2), \ \ \delta_{n-1} \sim N(0,1)$$

Equation 11. Estimation of the mean using Taylor's stochastic volatility specification.

$$h_n = \ \omega + \alpha h_{n-1} + \sigma_{n-1}\psi_{n-1}, \ \ \ \psi_{n-1} \sim N(0,1)$$

Equation 12. Estimation of volatility using Taylor's stochastic volatility specification.

Where u_n represents the estimate of the daily log-return for day n , μ_{n-1} denotes the conditional mean of the financial returns for day $n-1$ and δ_{n-1} refers to a zero-mean white noise process assumed to be normal. h_n represents the conditional variance, ω is the weighted long-run variance term and ψ_n is a stochastic white noise process with zero mean and variance one.

For this model, volatility is latent rather than observed, and as such, it is modeled indirectly. The logarithmic formulation ensures that the variance estimate is always positive. Also, the parameter α represents the persistence of the log-variance, and it is restricted to $|\alpha| < 1$ to assure stationarity (Shephard, 2007).

Despite being often used in options pricing literature, SV models are not very popular for discrete time series applications. This results from the complexity involved in estimating the model parameters (mean, variance and random white noise process) specially when accounting for different distributional assumptions and complex portfolios (Andersen et al., 2001).

According to scientific research, SV models often produce inaccurate VaR forecasts. Pederzoli, (2006) concludes that both the GARCH (1,1) and SV models underestimate real incurred losses. Nonetheless, the author also notes that, in general, the GARCH (1,1) performs slightly better than the SV model. Similar findings are reported by Chen et al. (2011).

2.4.1.5. Exponential Generalized Autoregressive Conditional Heteroskedasticity Model: EGARCH (1,1)

Despite the large popularity of the GARCH (1,1) model to forecast volatilities, evidence shows that the most simplistic form of a GARCH process still fails to provide reliable VaR forecasts. This often occurs because the model is not capable of capturing other prevalent asymmetric features of volatility.

Thus, Nelson (1991) recognized that GARCH models have two main limitations. First, these do not recognize the negative correlation between current returns and future returns volatility, also known as leverage effects (Black, 1976). Moreover, GARCH models often impose parameter restriction which often violate the theoretical assumptions of its coefficients and that restrict the dynamics of the conditional variance process.

Given the previous limitations, the author proposed the application of the Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) model, an extension of the GARCH process which is capable of modeling the leverage effects of financial returns. This way, the model recognizes that past negative shocks have a deeper impact on the current volatility than past positive shocks.

The leverage parameter γ presented in **Equation 13** measures the leverage effects and it is expected to be negative. Also, by measuring volatility as a logarithm, the model assures that the conditional variance forecasts are always positive.

$$log(\sigma_n^2) = \omega + \gamma \left(\frac{u_{n-1}}{\sigma_{n-1}}\right) + \alpha \left(\left|\frac{u_{n-1}}{\sigma_{n-1}}\right| - \sqrt{\frac{2}{\pi}} \right) + \beta \; log\,(\sigma_{n-1}^2) \quad with\, \alpha + \beta < 1$$

Equation 13. Estimation of volatility using an EGARCH (1,1) process.

In addition, the estimated volatility also depends on the size of the log-returns. If α is positive, the returns superior to the mean will have a deeper impact on the current volatility than those below the mean. (Abad et al., 2014) Past log-returns are represented by u_{n-1}.

Given its increased ability to model other dynamic features of financial returns, scientific findings suggest that the EGARCH model can be an effective parametric alternative to forecast losses. For instance, Angelidis et al. (2004) find that an AR(1)-EGARCH (0,1) specification produces the best VaR forecasts for different indexes. Pederzoli (2006) concludes that the EGARCH (1,1) clearly ourperforms both the SV and GARCH (1,1) models. Abad and Benito (2013) find that both the EGARCH based on the normal distribution assumption and the EGARCH based on student t- distribution produce accurate VaR forecasts. In addition, the authors also highlight that combining GARCH models with asymetric and heavy-tailed distributions can help to enhance their predictive ability. Similar conclusions are also reported by Ener et al. (2012).

2.4.1.6. GJR Generalized Autoregressive Conditional Heteroskedasticity Model: GJR GARCH (1, 1)

Another attempt to capture the asymmetry of financial returns was proposed by Glosten et al. (1993). The Glosten, Jagannathan and Runkle Generalized Autoregressive Conditional Heteroskedasticity (GJR-GARCH) model attempts to capture the asymmetry in volatility by introducing a dummy variable which takes the value of one if the mean of log-returns at day n is below zero ($\mu_n < 0$) and is zero if the mean of returns at day n is larger than zero ($\mu_n \geq 0$).

The conditional variance formulation for the GJR GARCH model is provided in **Equation 14**.

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 - \gamma u_{n-1}^2 S_{n-1}^- + \beta \sigma_{n-1}^2$$

$$S_{n-1}^- = 1 \; for \; u_{n-1} < 0 \; and \; S_{n-1}^- = 0 \; otherwise$$

Equation 14. Estimation of volatility using a GJR-GARCH (1,1) process.

If the parameter gamma, γ, is positive and statistically significant, there is enough statistical evidence to suggest the presence of a negative asymmetric volatility response. Like so, this process is also capable of modeling leverage effects by recognizing that negative shocks at time $n-1$ have a stronger impact in the variance forecast, σ_n^2, than positive shocks.

2.4.2. Non-parametric methods

Non-parametric models forecast VaR without making any assumptions about the distribution of financial returns, thus, allowing past data to describe their empirical distribution.

Non-parametric models rely on the assumption that, to forecast VaR, future return's performance will be similar to the most recent past. Therefore, these models assume that past losses will provide a satisfactory risk measure for the future.

Some of the most used non-parametric approaches are the Historical Simulation and the Boudoukh, Richardson and Whitelaw Historical Simulation (BRW HS). Although not examined in this study, there are additional non-parametric density estimation methods that can be used to forecast VaR. See Abad et al. (2014) for a detailed overview of these methods.

2.4.2.1. Historical Simulation: HS

Resulting from its simplicity, ease of application and lack of empirical and distributional assumptions, the Historical Simulation (HS) approach is one of the most widely implemented approaches used by banks and financial institutions to forecast VaR (Pérignon & Smith, 2009).

The HS uses past returns as a proxy for future losses and, accordingly, VaR is computed directly from the empirical distribution of past returns. For this reason, the HS has the benefit of allowing the accommodation of wide tails, skewness and other non-normal features. In addition, it is a cheap and simple methodology to implement from a computational perspective.

To estimate the $\dot{\alpha}\%$ VaR, the HS assigns an equal probability weight of $1/N$ to each day's return, (where N represents the total number of daily returns) therefore implying that returns are independent and identically distributed (i.i.d.). As a result, this assumption will neglect the immediate and real impact of a major crash in a portfolio's risk profile, given that volatility is already proven to be time-varying and clustered.

Furthermore, by using past returns as an empirical cumulative distribution function (CDF), losses need to exceed VaR for the risk estimate to rise. Hence, a VaR forecast is limited to the highest loss corresponding to the $\dot{\alpha}\%$ percentile of the selected historical period. Also, the empirical cumulative distribution function of the HS is discrete, which most times makes the VaR risk measure to be condensed between two major historical losses. By selecting one or the other as the real VaR measurement, one might be underestimating or overestimating the actual incurred risk, depending on the amplitude of these differences. Also, the model responds very passively to changes in conditional volatility and it responds asymmetrically to large price moves, in a sense that risk estimates increase after large losses and not large gains. For these reasons, VaR forecasts become highly dependent on the historical period used to perform the analysis.

Resulting from its limitations, literature findings agree that, despite very popular, the HS does not produce accurate VaR forecasts. Abad and Benito (2013) state that the VaR forecasts obtained with the HS are inaccurate for all of the indexes and sample sizes used to calculate the empirical distribution. Pérignon and Smith (2009) assert that computing VaR using the HS approach helps little to forecasting the future volatility of trading revenues of banking institutions. Similarly, Danielsson and Vries (2000) reach similar conclusions by stating that the HS is not capable of adressing losses outside the sample period.

2.4.2.2. Boudoukh, Richardson and Whitelaw (BRW) Historical Simulation: BRW HS

The BRW simulation proposed by Boudoukh et al. (1998) is a an extension of the HS approach which combines the benefits of the EWMA model with the HS. To illustrate, instead of applying equal weights to each individual past observations, the BRW applies exponentially declining weights to past returns. Then, the appropriate percentile is selected from the time-weighted empirical distribution.

By applying declining weights on past data, the BRW HS recognizes that recent losses are more relevant to forecast VaR than less recent ones. This way, the number of observations considered to compute the $\dot{\alpha}$% quantile will depend on whether extreme returns were observed recently or in the past. As a result, this method allows the model to react faster to changes in conditional volatilities.

For a detailed discussion on BRW models see Boudoukh et al. (1998) and Pritsker (2006).

Literature findings have provided some evidence which suggests that the BRW HS performs relatively better than the HS and other parametric models. For instance, Sašžiković and Aktan (2011) propose the implementation of the an optimised BRW historical simulation based on minimizing a size adjusted function. The authors find that this approach performs better than most volatility models and significantly better than the HS, which produces unsatisfatory results across all tested stock indexes. Boudoukh et al. (1998) compare the performance of a BRW HS, HS and a EWMA approach and conclude that while estimating the 1% VaR for the S&P500, the hybrid approach produced an absolute error which is 30% to 43% lower than the EWMA approach and 14% to 28% lower than the HS approach.

2.4.3. Semi-parametric methods

Semi-parametric methods incorporate both non-parametric and parametric approaches to produce a VaR forecast. The most relevant semi-parametric approaches include the Volatility Weighted Historical Simulation (VWHS), the Filtered Historical Simulation (FHS), the Conditional Autoregressive Value at Risk by Regression Quantiles (CAViaR), the Extreme Value Theory (EVT) and the Monte Carlo Simulation (MCS).

2.4.3.1. Volatility Weighted Historical Simulation: VWHS

One of the disadvantages of the classical HS approach is that it does not account for changes in volatilities. Hence, Hull and White (1998) proposed an approach which combines both the benefits of the

HS with the model-building approach. This new approach relaxes the multivariate normality condition implied by the model building approach by allowing daily returns to follow any probability distribution.

The Volatility Weighted Historical Simulation (VWHS), denoted by the authors as the Hull and White (HW) approach, can be viewed as extension of the HS which updates return information to account for the most recent changes in volatility. This way, historical changes are adjusted to reflect the ratio of the current daily volatility relative to the daily volatility at the observation's time.

Incorporating Volatility updating schemes into a Historical Simulation

One of the properties of the probability distribution of a market variable is that, when scaled by an estimate of its volatility, it becomes stationary, which suggests that the HS approach can be adapted to consider volatility changes (Hull & White, 1998). As such, this can be achieved by estimating the variance for a pre-determined historical period using an EWMA (1,1) or GARCH (1,1) model. After that, historical returns are standardized to account for changes in volatility throughout the respective period. It is also assumed that the probability distribution of u_n / σ_n is stationary.

$$u_n^* = \sigma_N \frac{u_n}{\sigma_n}$$

Where u_n represents the historical log-returns of a financial instrument or portfolio on day n, with $n < N$. σ_n corresponds to the historical GARCH/EWMA estimate of the daily standard deviation of log-returns made for at the end of day $n - 1$. σ_N denotes the most recent estimate of daily standard deviation computed at the end of the day $N - 1$.

$VaR(\dot{\alpha})$ can be computed as the $\dot{\alpha}$ quantile of the volatility adjusted empirical return distribution.

Despite the lack of research within the literature framework, some findings present promising results. For instance, Basu (2011) finds that, amongst multiple approaches, the VHWS is the best at capturing clustered volatility. However, the author recognizes that the amplitude of forecasted losses highly depends on the considered decay factor, or weight, assigned to each lag parameter of the EWMA or GARCH volatility specification (recall **Section 2.4.1.2** and **Section 2.4.1.3**, respectively) . Hull and White (1998) conclude that the VWHS provides better forecast than the BRW HS and HS.

2.4.3.2. Filtered Historical Simulation: FHS

The Filtered Historical Simulation (FHS) approach was introduced by Barone-Adesi et al. (1998) and combines some of the benefits provided by the HS with the flexibility of conditional volatily models.

Like the HS approach, the FHS method uses historical returns to define its empirical distribution. However, this approach recognizes that raw returns are unsuitable for historical simulation purposes since these do not fulfil the properties necessary for unbiased results. Raw returns cannot be used for the Historical Simulation process unless these are independent and identically distributed (i.i.d.). Hence, it is necessary to remove any serial correlation and volatility clusters from the data.

Having this concern in mind, the first step of this method involves fitting a conditional volatility model to the data, which is then used to forecast volatilities for each of the days in the sample period. The authors propose the use of an asymmetric GARCH process to model conditional volatilities.

Then, realized log-returns, denoted by u_n are standardized by the corresponding daily standard deviation estimate, σ_n, to produce standardized returns that are i.i.d. The standardized returns are given by,

$$z_n = u_n / \sigma_n$$

Lastly, it is performed a bootstrap procedure to the historical standardized returns. A segment of these return is drawn randomly with replacement, creating a sample of total dataset. Then, each of these drawn standardized returns is multiplied by the 1-day ahead standard deviation forecast, see **Equation 15**.

$$r_n^* = \mu_n + z^*$$

Equation 15. Simulated returns using the Filtered Historical Simulation approach.

Where $z^* = z_n * \sigma_{n+1}$ and μ_n being the mean of returns at time n and σ_{n+1} the standard deviation forecast for day $n + 1$.

By taking m drawings from the dataset, it is obtained a sample with M simulated returns. This way, the $VaR(\dot{\alpha})$ is computed as the $\dot{\alpha}$, quantile of the simulated return sample.

Scientific research has shown that the FHS is often found to be one of the best performing models to forecast VaR. When comparing with the other HS approaches, Barone-Adesi and Giannopoulos (2001) highlight the "impressive gains" which allow this approach to easily dominate over the HS. Sašžiković and Aktan (2011) find that the FHS model produces significant improvements over the BRW HS, the HS and VWHS. The same authors, in a different publication, see Žiković and Aktan (2009), reach the same consistent conclusions. Amongst different VaR models, the HHS (Hybrid HS-FHS), or FHS, was one of the best models at predicting losses during a period of extreme market volatility.

On the other hand, Nozari et al. (2010) provide contradictory results. These authous find that the FHS had "medium performance". It was accurate at a 95% confidence level but unprecise at a 99% and 99.5% confidence level.

2.4.3.3. Conditional Autoregressive Value at Risk by Regression Quantiles: CaViaR

Instead of modeling the whole distribution of financial returns, the Conditional Autoregressive Value at Risk by Regression Quantiles (CAViaR) approach, proposed by Engle and Manganelli (2004), seeks to only estimate the quantile of the distribution of returns. This proposal is based on a conditional autoregressive specification of VaR which recognizes that a VaR estimate and its standard deviation should be closely linked. According to the authors, given that returns tend to cluster over time while expressing some autocorrelation, a VaR forecast should reflect that same behavior. As a result, it was proposed and autoregressive specification of the CAViaR, see **Equation 16**.

Assume that U_n represents a vector of time n portfolio log-returns. Also, consider that $\dot{\alpha}$, denotes the probability, or significance level considered for the VaR computation (e.g., the estimation of a 5% VaR, where $\dot{\alpha} = 0.05$). $\beta_{\dot{\alpha}}$ represents a vector of unknown parameters. This vector defines the weights for each of the parameters also computed with a significance level of $\dot{\alpha}$.

$$VaR_n(\beta) = \beta_0 + \sum_{i=1}^{q} \beta_i VaR_{n-i}(\beta) + \sum_{j=1}^{r} \beta_j l(u_{n-j})$$

Equation 16. Estimation of VaR using the general CAViaR specification.

$VaR_n(\beta)$ denotes the time n, $\dot{\alpha}\%$ quantile of the distribution of financial returns formed at time $n - 1$. Also, $p = q + r + 1$ which represents the dimension of β (i.e., the number of underlying unknown factors, or weights, given to each parameter). l represents a function of a finite number of lagged financial returns. β_0 reflects a constant. The autoregressive term $\beta_i VaR_{n-i}(\beta)$ ensures that the estimated quantile forecast changes smoothly over time because it is always dependent, at some extent, on the estimate of the previous period. That happens because the $\dot{\alpha}\%$ VaR forecast for day n is always a function of the VaR estimated for the previous day, $n - 1$. Also, $\beta_j l(u_{n-j})$ links $VaR_n(\beta)$ to observable past returns. Lastly, u_{n-1} is defined a lagged return. VaR is also expected to increase as u_{n-1} becomes more negative.

In short, the VaR forecast for period n, under the CAViaR approach, is a function of the sum of three parameters: 1) a constant, β_0; 2) the weighted sum of the previously computed VaR estimates, $\sum_{i=1}^{q} \beta_i VaR_{n-i}(\beta)$; and 3) the weighted sum of the lagged past returns, $\sum_{i=1}^{r} \beta_j l(u_{n-j})$.

The authors also propose different autoregressive specifications for the CaViaR process. First, the adaptive approach, see **Equation 17**.

$$VaR_n(\beta_1) = VaR_{n-1}(\beta_1) + \beta_1\{[1 + \exp(G[u_{t-1} - VaR_{n-1}(\beta_1)])]^{-1} - \dot{\alpha}\}$$

Equation 17. Estimation of VaR using an adaptive CAViaR specification.

Where G denotes some positive finite number. As $G \rightarrow \infty$, the last term of the equation converges to $\beta_1[I(u_{t-1} \leq VaR_{n-1}(\beta_1)) - \dot{\alpha}]$, where I represents an indicator function.

Within this approach, whenever VaR is exceeded, its estimate should be imediately increased, however, when VaR is not exceeded, its estimate should be decreased slighly. This methodology reduces the probability of sequences of VaR hits and also makes it unlikely that there will nevel be hits. However, the adaptive approach learns very little from the actual size of the returns, specifically when these are close to VaR or extremely positive. As such, VaR is increased by the same amount regardless of whether the returns exceeded the VaR by a small margin or a large margin.

The symmetric absolute value (SAV) approach, see **Equation 18**, allows β_1 to assume other value rather than one. Also, it assumes that the effect of extreme returns on the VaR forecast is identical.

$$VaR_n(\beta) = \beta_0 + \beta_1 VaR_{n-1}(\beta) + \beta_2|u_{n-1}|$$

Equation 18. Estimation of VaR using a symmetric absolute value CAViaR specification.

In contrast, the asymmetric slope (AS) autoregressive process, see **Equation 19**, allows VaR forecasts to respond asymmetrically to positive and negative returns, thereby, accommodating the leverage effect often observed in financial returns.

$$VaR_n(\beta) = \beta_0 + \beta_1 VaR_{n-1}(\beta) + \beta_2 \max[u_{n-1}, 0] + \beta_3 \max[-u_{n-1}, 0]$$

Equation 19. Estimation of VaR using an asymmetric slope CAViaR specification.

In last, the indirect GARCH (1,1) (IG) process, see **Equation 20**, is often appropriate if the data was generated by a location-scale model with i.i.d errors.

$$VaR_n(\beta) = (\beta_0 + \beta_1 VaR_{n-1}^2(\beta) + \beta_2 u_{n-1}^2)^{1/2}$$

Equation 20. Estimation of VaR using an indirect GARCH (1,1) CAViaR specification.

All three extension of the CAViaR process present mean-reverting properties given that the coefficient on the lagged VaR is not constrained to one (Engle & Manganelli, 2004).

2.4.3.4. Extreme Value Theory: EVT

Extreme Value Theory (EVT) is often used to describe the science of estimating the tails of a distribution. The theory itself was first introduced by Gnedenko's research in 1943 which found that the wide tails of different probability distributions share some common properties (Gnedenko, 1943).

As the CAViaR approach, the EVT aims to estimate the quantile of a distribution. However, it achieves this by limiting the distribution of extreme returns observed over a long time period (Gilli & Këllezi, 2006).

There are two ways of limiting or identifying extreme values in real financial data. The Block Maxima Model (BMM) identifies a maximum negative return, taken in successive periods, named the block maxima. The Peaks Over Threshold (POT) model is focused on selecting observations that exceed a given (high) threshold.

Block Maxima Model: BMM

The BBM approach starts by splitting the selected dataset into *m* non-overlapping chunks of blocks, or sections, with approximately n days in each block. Then, it is selected the maximum value for each block, that is, the maximum loss registered within a group of *n* days. These selected observations form extreme events, also called a maximum block denoted by M_n. Given this grouping of observations, the BBM seeks to describe the distribution of the maxima.

$$M_n = \max(u_1, u_2, u_3, \dots, u_n)$$

Where u_n denotes the log-return for day n within a particular block *m.*

The Distribution of the Maxima

Fisher and Tippett (1928) and Gnedenko (1943) developed the theorem of the limit law for the block maxima, where n is defined as the size of the subsample, block or section. This theorem is commonly referred to as Fisher-Tipper-Gnekenko theorem and it states the following:

"Let (X_n) be a sequence of i.i.d. random variables. If there exist constants $c_n > 0, d_n \in R$ and some non-degenerate distribution function H such that:

$$\frac{M_n - d_n}{c_n} \to H$$

then H, belongs to one of three standard extreme distributions".

In essence, the theorem states that when one standardizes the maxima of multiple i.i.d. variables, these standardized variables can only converge to one of three standard extreme distributions, a Frétcher, a Weibull or a Gumbel distribution. See **Equation 21** for each specification.

$$Fr\acute{e}chet: \quad \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ e^{-x-\alpha}, & x > 0 \end{cases}$$

$$Weibull: \quad \Psi_\alpha(x) = \begin{cases} e^{-(-x)^{-\alpha}}, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

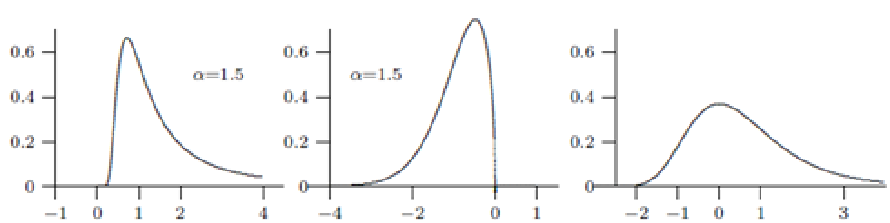
$$Gumbel: \quad \Lambda(x) = e^{-e^{-x}}$$

Equation 21. Specification of the Frétcher, Weibull and Gumbel extreme distributions.

The Fréchet cumulative distribution function, is characterized by a polynomically decaying tail and, therefore, suits well heavy tailed distributions. Oppositely, the Gumbel cumulative distribution function is described by exponentially decaying and thin tails. Lastly, the Weibull cumulative distribution function is the asymptotic distribution of finite endpoint distributions.

Figure 2

Graphical representation of the Frétcher (left), Weibull (middle) and Gumbel (right) extreme distributions. Source: (Gilli & Këllezi, 2006)



Jenkinson (1955) and von Mises (1954) proposed a representation of these three standard distributions using only one parameter, ξ, which reflets the heaviness of the tail, or shape, of a distribution. This representation is commonly referred to as the Generalized Extreme Value (GEV) distribution, see **Equation 22**.

$$H_\xi(x) = \begin{cases} e^{-(1+\xi x)^{-1/\xi}}, & \xi \neq 0 \\ e^{-e^{-x}}, & \xi = 0 \end{cases}$$

Equation 22. Specification of the Generalized Extreme Value (GEV) distribution.

With x, such that $1+\xi x>0$.

The GEV distribution is obtained by setting $\xi=\alpha^{-1}$, for the Fréchet distribution; $\xi=-\alpha^{-1}$ for the Weibull distribution and $\xi=0$ for the Gumbel distribution. Also, as it is not possible to known in advance the type of limiting distribution for the sample of the maxima, the distribution parameters are estimated using maximum likelihood methods.

Moreover, since the GEV function is defined by the limiting distribution of the normalized extrema and given that no one knowns the actual distributions of returns, it becomes impossible to identify the norming constants c_n and d_n.

As a result, the GEV formulation is simplified by including μ as the location and σ the scale parameter in its specification, see **Equation 23**.

$$H_{\xi,\sigma,\mu}(x)=H_{\xi}\left(\frac{x-\mu}{\sigma}\right)\quad x\in D\,,D\;=\begin{cases}]-\infty,\mu\;-\dfrac{\sigma}{\xi}\;[& \xi<0\\ ]-\infty,\infty[& \xi=0\\ ]\mu\;-\dfrac{\sigma}{\xi},\;\infty[& \xi>0\end{cases}$$

Equation 23. Specification of the Generalized Extreme Value (GEV) distribution using μ as the location and σ the scale parameter.

Ultimately, VaR is estimated by substituting the respective parameters of the GEV distribution by their estimates, see **Equation 24**.

$$\hat{R}^k\;=\begin{cases}\hat{\mu}-\dfrac{\hat{\sigma}}{\hat{\xi}}\;(1-(-\log\,(1-\dfrac{1}{k}))^{-\hat{\xi}}\;, & \hat{\xi}\neq 0\\ \hat{\mu}-\hat{\sigma}log\left(-log\left(1-\dfrac{1}{k}\right)\right), & \hat{\xi}=0\end{cases}$$

Equation 24. Estimation of VaR using the simplified specification of the GEV distribution.

To illustrate, the value of $\hat{R}^{10}$ of 9 means that the maximum loss observed during a period of one year will exceed 9% once in ten years on average.

Peaks Over Threshold: POT

Alternatively, the POT extreme value theory approach seeks to describe the distribution of exceedances over a certain threshold, defined as u^*.

Distribution of Exceedances

Considering an unknown distribution F for a random variable X. The distribution function of exceedances placed above a threshold of u^* will be given by F_{u^*}. This is called the conditional excess distribution function and is defined as

$$F_{u^*}(y)=P(X-u^*\leq y\,|X>u^*),\quad 0\leq y\leq x_F-u^*$$

Where X is a random variable, u^* is a given threshold, $y = x - u^*$ are the excesses and $x_F \leq \infty$ is the right endpoint of F.

The conditional distribution function, F_{u^*} can also be written as a function of F,

$$F_{u^*}(y) = \frac{F(u^* + y) - F(u^*)}{1 - F(u^*)} = \frac{F(x) - F(u^*)}{1 - F(u^*)}$$

The process of estimating F is theoretically simple. The reason for this is that the set of realizations that the random variable X takes mainly lie between the interval of $[0, u^*]$. By contrast, the estimation of F_{u^*} is far more complicated because there are not a lot of observations remaining to apply to its estimation.

The EVT becomes a very helpful approach because it provides a lot of powerful properties for the conditional excess distribution function. For instance, Pickands (1975) and Balkema and de Haan (1974) developed a theorem that states the following:

"*For a large class of underlying distribution functions F, the conditional access distribution function $F_{u^*}(y)$, for u^* large, is well approximated by*:

$$F_{u^*}(y) \approx G_{\xi.\sigma}(y), \quad u^* \to \infty$$

$$where$$

$$G_{\xi.\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-\frac{1}{\xi}} & if\ \xi \neq 0 \\ 1 - e^{-y/\sigma} & if\ \xi = 0 \end{cases}$$

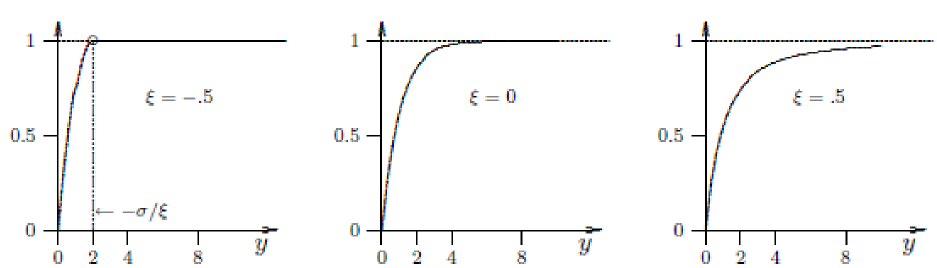
$$For\ y \in [0, (x_F - u^*)\ if\ \xi \geq 0\ and\ y \in \left[0, \frac{\sigma}{\xi}\right]\ if\ \xi < 0.$$

$G_{\xi.\sigma}$ represents the generatized Pareto distribution (GPD).

ξ refers to the shape parameter or tail index and expresses the heaviness of the tail. The larger its value the heavier the tail of the distribution. ξ can assume positive, negative or zero value, see **Figure 3**.

Figure 3

Graphical representation of the different shapes of the Generalized Pareto Distribution (GPD) depending on the value of the shape parameter ξ. Source: (Gilli & Këllezi, 2006)



Because it is impossible to assign an upper bound, or maximum loss, to the distribution of financial returns, only distributions with $\xi > 0$ are qualified to describe the returns' distribution.

From the previous GPD specification, if one isolates $F(x)$, it is possible to derive $VaR(\dot{\alpha})$, see **Equation 25**.

If F_{u^*} is replaced by the GPD distribution function and $F(u^*)$ by the estimate $(n - N_{u^*})/n$, where n represents the total number of observations and N_{u^*} the number of observations above a threshold u^*, it is obtained the following expression

$$\hat{F}(x) = \frac{N_{u^*}}{n}(1-(1+\frac{\hat{\xi}}{\sigma}(x-u^*))^{-1/\hat{\xi}} + (1-\frac{N_{u^*}}{n})$$

which simplifies to

$$\hat{F}(x) = 1-\frac{N_{u^*}}{n}(1+\frac{\hat{\xi}}{\sigma}(x-u^*))^{-1/\hat{\xi}}$$

By inverting the previous equation, and given a certain probability $\dot{\alpha}$, VaR can be estimated using the following formulation:

$$\widehat{VaR}_{\dot{\alpha}} = u^* + \frac{\hat{\xi}}{\sigma}((\frac{n}{N_{u^*}}\dot{\alpha})^{-\hat{\xi}} - 1)$$

Equation 25. Estimation of VaR using a GPD distribution specification.

By comparing the performance of multiple VaR models, literature suggest, in almost unanimous faction, that the EVT is the best approach for forecasting VaR. To exemplify, Bao et al. (2006) compares the performance of multiple parametric, non-parametric and semi-parametric approaches and find that EVT models perform the best at predicting VaR during a crisis period. Danielsson and Vries (2000) compare the performance of conditional volatility models with the HS and EVT approaches and reach similar conclusions.

Additional literature Bekiros and Georgoutsos, (2005); Marimoutou et al. (2009) and Reeves et al. (2005) also suggest that the EVT approach is the best performing model for forecasting VaR, especially when considering very high confidence levels.

2.4.3.5. Monte Carlo Simulation: MCS

The computational process of the Monte Carlo Simulation (MCS) is somewhat similar to the HS. However, while the HS approach carries out a simulation using the observed changes in returns over the last N periods to generate a hypothetical portfolio's performance, the MCS approach relies on the assumption that a specific statistical distribution is adequate to capture or approximate the changes in the market factors. Then, the MCS uses that distributional assumption to generate thousands of hypotheticals changes in the market factors. Thus, these hypothetical simulated changes create thousands of hypothetical portfolios $P\&L$ outcomes.

VaR is estimated empirically as the quantile of the simulated distribution of returns. To illustrate, to forecast a 1-day 99% VaR on date n, one can simulate N draws from the distribution of returns for date

$n + 1$. Then, the 99% VaR is forecasted by reading off the $\frac{N}{100}th$ element of the distribution. This selection if performed after sorting the different draws of the 1-day log-returns.

By combining parametric distributions with a simulation process, the MCS can capture a significant number of dynamic features in the data. Also, because it does not require any distribution assumptions about the underlying joint distribution, the MCS became a technique often used to forecast VaR for non-linear portfolios (Estrella et al., 1994).

Despite all these benefits, the MCS can become a computationally complex and demanding approach, especially when considered the increasing complexity of portfolios, which are often composed by elaborate hedging instruments such as financial derivatives.

Tolikas et al. (2007) perform a comparison of the performance of parametric, non-parametric and semi-parametric VaR forecasting methods and find that a MCS based on a normal distribution assumption, provides in general, disappointing forecasts (with exception of VaR at 95% confidence interval). Bao et al. (2006) also reach a similar conclusion. However, the authors find that, for pre-crises periods with lower volatility, the MCS provides a good prediction of the quantile loss. Also, Huang (2009) finds that the MCS based on the normal distribution performs better than the classical HS.

3. Methodology

3.1. DATA

The methodology used to implement this study includes the estimation and forecasting of the VaR risk measure for three of the most well-known American indexes. The data used for this purpose are the adjusted daily closing prices for the US Dow Jones Industrial Average (^DJI), the Standard & Poor 500 (^GSPC) and the NASDAQ (^IXIC) indexes. The data was extracted from Yahoo Finance for the period between January 3, 2000, until December 30, 2020. All statistical models and computations were performed using the R Statistical Software (v4.0.5; R Core Team 2021).

The adjusted closing prices are considered to compute the log-returns (u_n) using the formula $u_n = ln\ (P_n/P_{n-1})$, where P_n and P_{n-1} are the adjusted closing prices of the index for the periods n and $n-1$, respectively. **Figure 4** presents a graphical representation of the daily-adjusted prices while **Figure 5** shows the evolution of daily log-returns for each of the three stock indexes during the historical period. In **Table 2** it are presented some preliminary and descriptive statistics of log-returns for the underlying period.

Figure 4

Stock indexes daily-adjusted closing prices. The evolution of daily-adjusted closing prices of three indexes (DJI 30, S&P 500 and Nasdaq) from January 3, 2000, to December 30, 2020

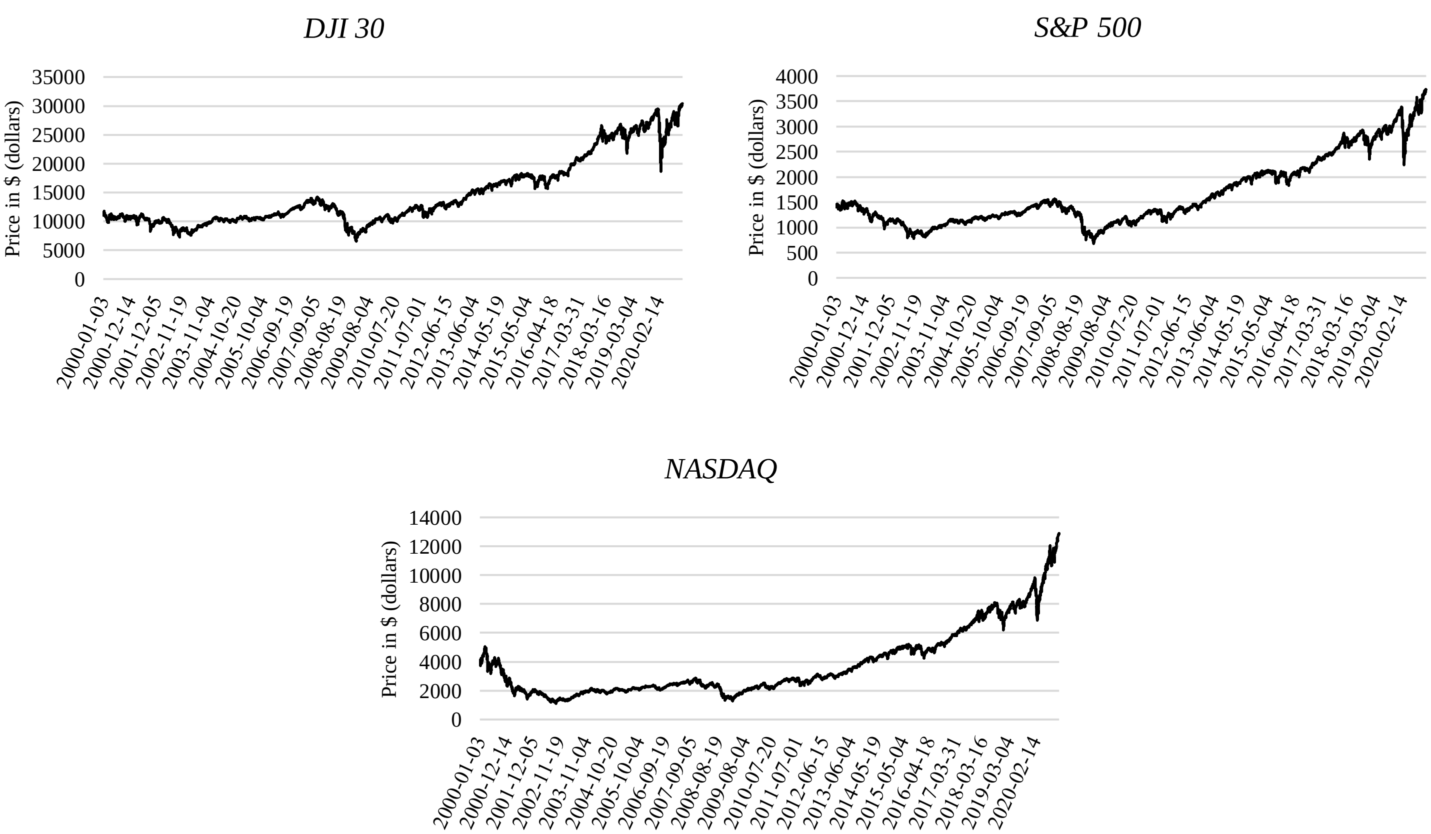


Figure 5

Stock indexes daily log-returns. The evolution of daily log-returns of three indexes (DJI 30, S&P 500 and Nasdaq) from January 3, 2000, to December 30, 2020

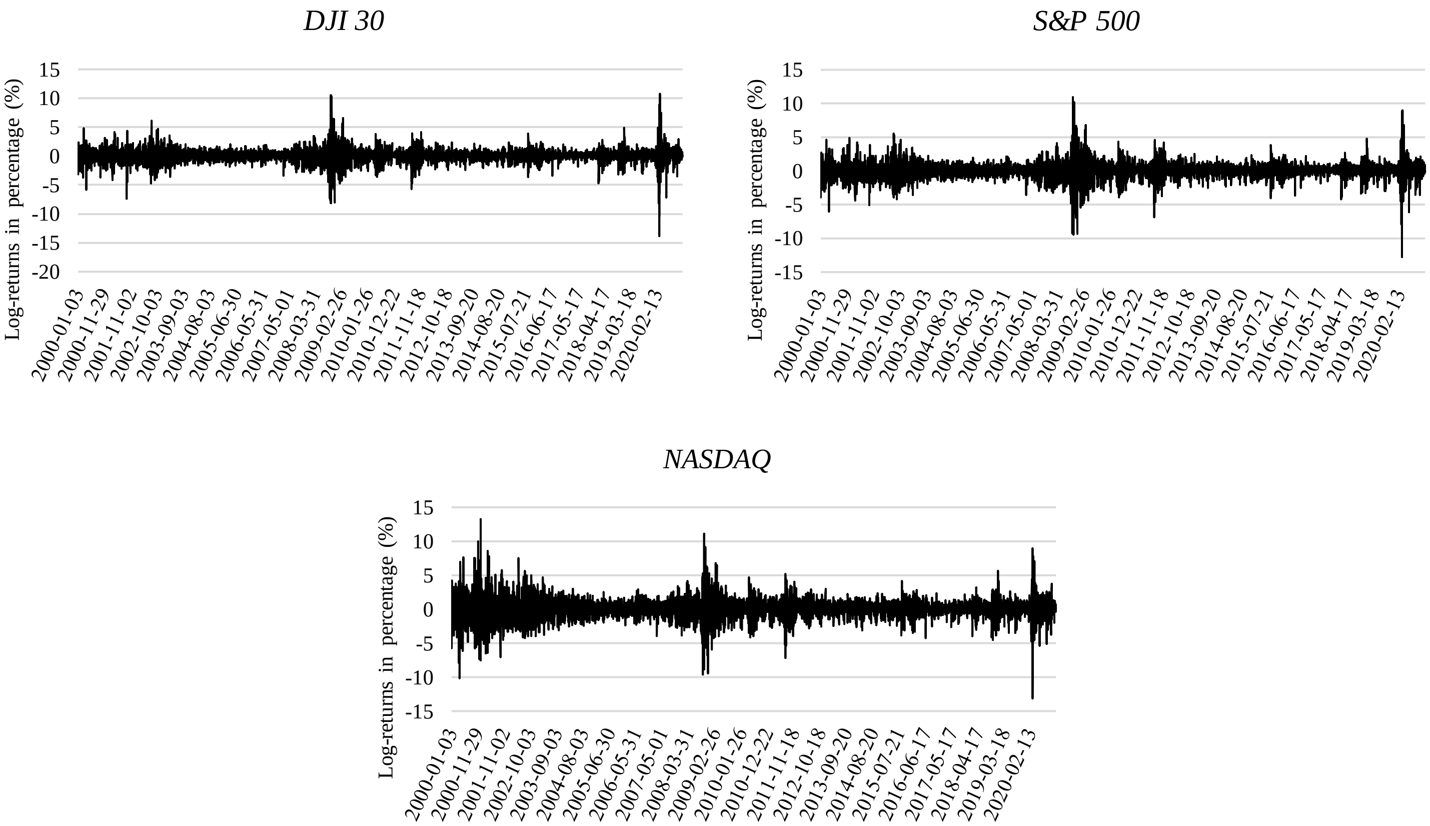


Table 2.

Summary statistics, normality, autocorrelation and statistical tests of the daily log- returns distribution for the period between January 3, 2000, and Dec 30, 2020, for the three indexes

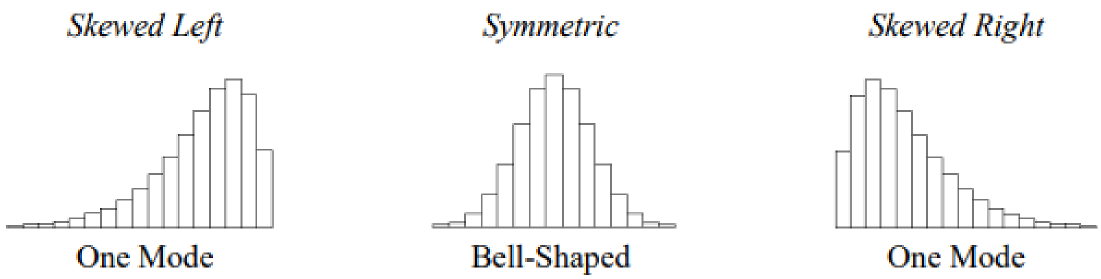
	DJI 30	**S&P 500**	**NASDAQ**
Summary Statistics			
Mean	0.0002	0.0002	0.0002
Median	0.0005	0.0006	0.0009
Maximum	0.1076	0.1096	0.1325
Minimum	-0.1384	-0.1277	-0.1315
Std. Deviation	1.21%	1.26%	1.60%
Std. Deviation (annualized)	19.18%	19.93%	25.42%
Skewness	-0.3767	-0.3934	-0.1317
Kurtosis	16.0191	13.9423	9.6197
Normality Test			
Jarque-Bera	37 428	26 488	9 659
P-value	(0.0000)	(0.0000))	(0.0000)
Autocorrelation Test			
Ljung-Box *(log-returns)*	19 374	21 108	35 715
P-value	(0.0000)	(0.0000)	(0.0000)
Ljung-Box *(squared log-returns)*	8 586	10 442	13 306
P-value	(0.0000)	(0.0000)	(0.0000)
Heteroskedasticity Test			
Arch Test on *(log-returns)*	1 623	1 583	1 220
P-value	(0.0000)	(0.0000)	(0.0000)
Arch Test on *(squared log-returns)*	1 003	935	367
P-value	(0.0000)	(0.0000)	(0.0000)

Distributions of daily log-returns are often found to be leptokurtic. This suggests that financial returns are often characterized by distributions with higher peaks and fatter tails than the normal distribution (Antoniou et al., 2004). The summary statistics and statistical tests presented in **Table 2** show that the returns distribution for the three indexes express an identical behavior. This same behavior is demonstrated by different statistical results (e.g., skewness, kurtosis, unconditional mean, normality tests, etc.).

Skewness is a measurement of the asymmetry of a distribution. A distribution with right (or positive) skewness is longer on the right side of its peak, see **Figure 6**. On the other hand, a distribution with left (or negative) skewness is longer on the left side of its peak. The normal distribution has a skewness of zero (Doane & Seward, 2011).

Figure 6

Graphical representation of left-skewed, symmetric and right-skewed distributions. Source: (Doane & Seward, 2011)



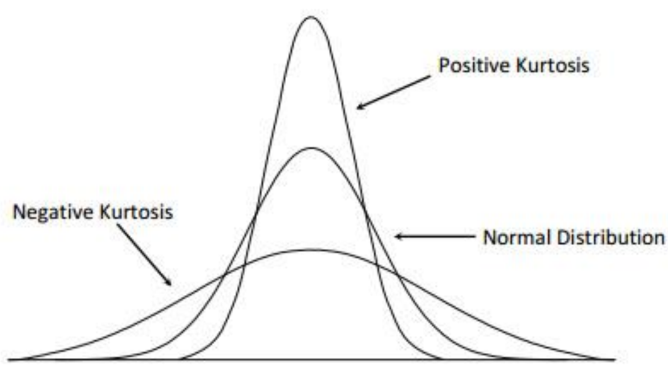
Skewness measures across all three indexes, namely for the DJI 30 (-03767), S&P 500 (-0.3934) and NASDAQ (-0.1317), suggests that their distributions of returns are not very far from being symmetric. (Note that, as a rule, if skewness is between -1 and 1, the distribution is highly skewed. If between -1 and – 0.5 it considered to be moderately skewed and if between -0.5 and 0.5 is approximately symmetric.) This is particularly more evident for the NASDAQ index which has the lowest skewness of all indexes, with a measurement relatively close to zero.

Furthermore, kurtosis represents a measure of how tailed the distribution of the data is. A distribution with high kurtosis has a sharp peak and short tails. A distribution with low or negative kurtosis has a low peak and long tails, see **Figure 7**. The normal distribution has a kurtosis of three. (Patitsas et al., 2020).

Within the underling historical data, kurtosis measures are considerably high, suggesting that the distributions of returns for the three indexes (the DJI 30 (16.0191), the S&P 500 (13.9423) and the NASDAQ (9.6197)) express significantly heavier tails than the normal distribution (Note that, kurtosis higher than three indicates that the distribution is heavier than the normal distribution).

Figure 7

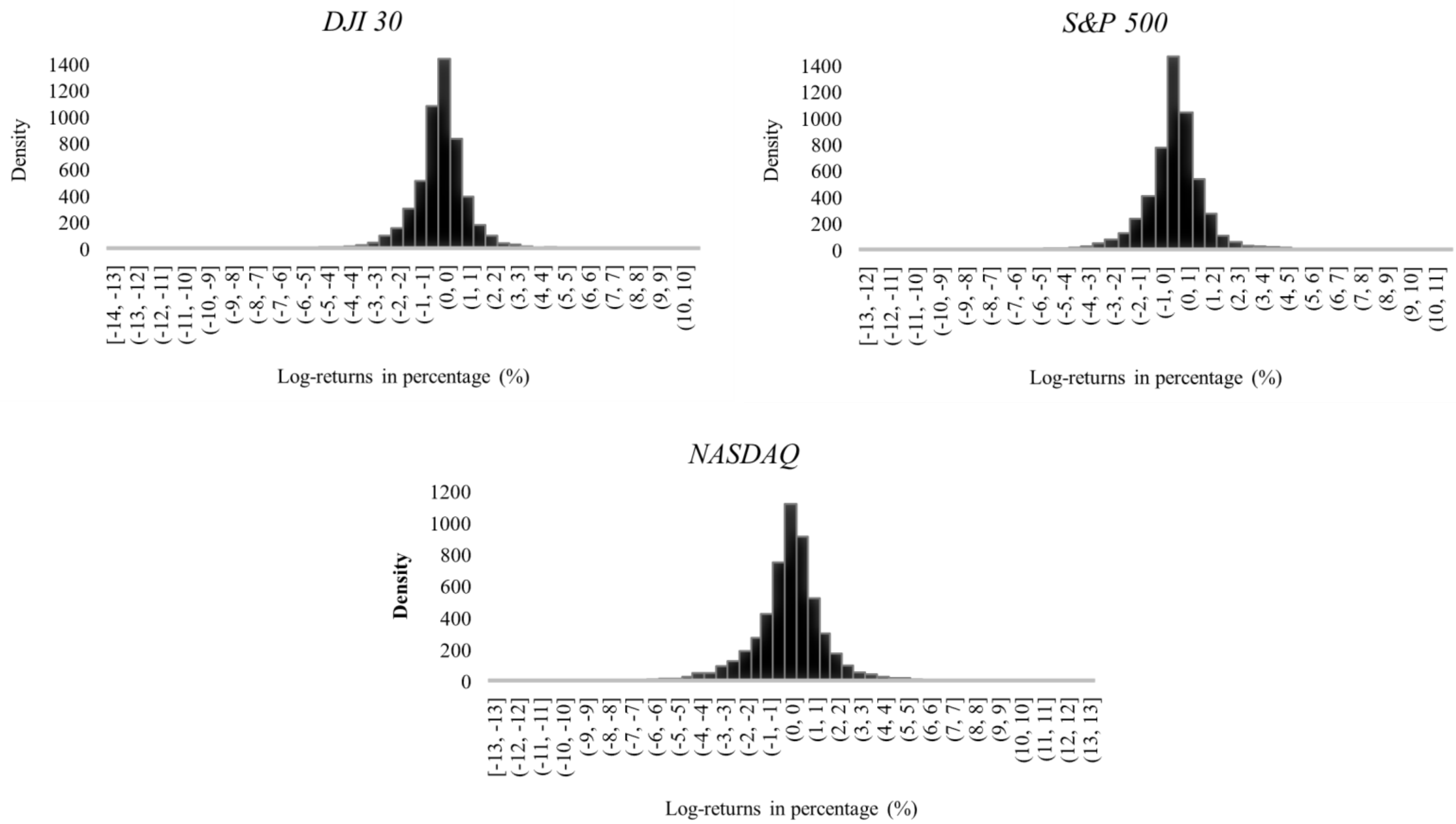
Graphical comparison of distributions with negative and positive skewness against the normal distribution. Source: (Patitsas et al., 2020)



Moreover, the unconditional mean of the distribution of returns (0.02% for the three indexes) is highly concentrated around the value of zero percent which further emphasizes the leptokurtic behavior of returns. These particular characteristics of returns can be analyzed with more detail in **Figure 8**.

Figure 8

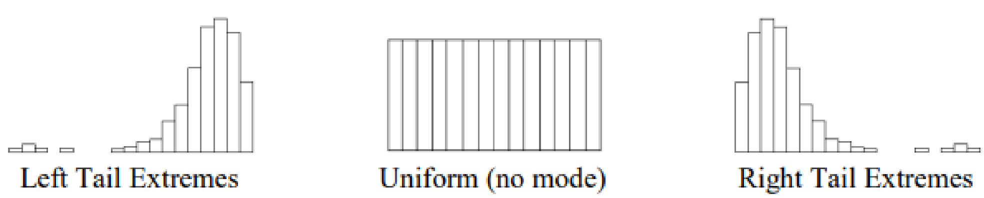
Histograms with the distribution of daily log-returns for the period between January 3, 2000, and Dec 30, 2020, for the three indexes



Heavy tails or extreme returns denote the likelihood to encounter significant deviations from the mean, which is much greater than in the case of the normal distribution, see **Figure 9**.

Figure 9

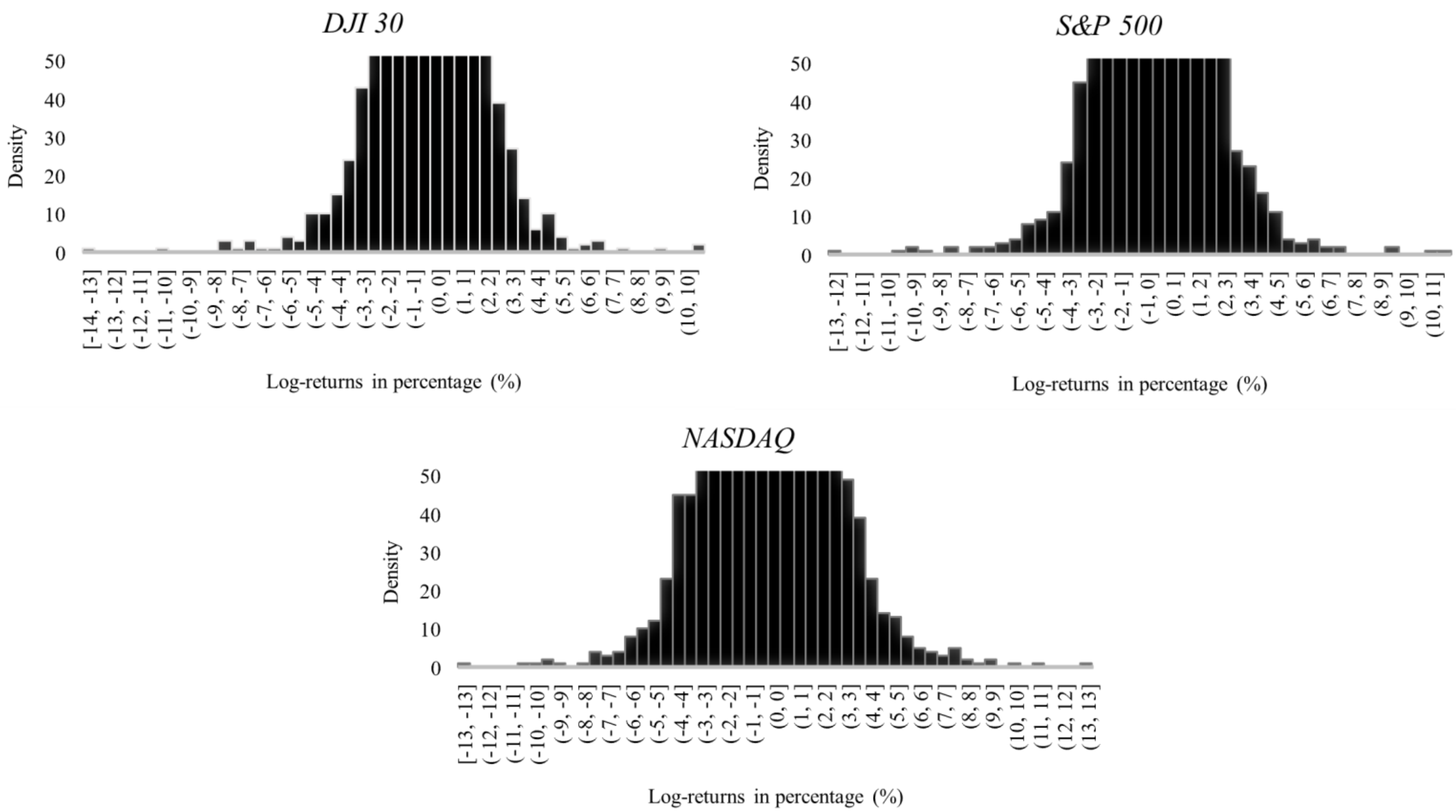
Multiple histograms representing a heavy left-tailed distribution, a uniform distribution and a heavy right-tailed distribution. Source: (Doane & Seward, 2011)



Once again, within the considered data, if the distribution of returns is amplified, see **Figure 10**, it becomes apparent the presence of extreme negative and positive returns across all three indexes. To illustrate, the DJI 30 index is shown to have a prevalent distribution of returns between]-7%, 5%]. Despite that, it are also reported seventeen business days where its returns deviated significantly from the mean of its distribution. On the left side of the distribution there are about eight observations which can be classified as extreme negative returns, which are lower than -7%. Moreover, on the right side of the distribution there are also about nine observations that can be classified as extreme positive returns, which are larger than 5%. This behavior is consistent across all three indexes.

Figure 10

Histograms with the distribution of log-returns and extreme returns for the period between January 3, 2000, and Dec 30, 2020, for the three indexes



Given these characteristics of financial returns and as already documented by previous studies, see Angelidis et al. (2006), Giot and Laurent (2004) and Lu et al. (2009), modeling returns using a normal distribution assumption is, in general, likely to yield very disappointing VaR forecasts. As noted, this occurs because the distribution of financial returns is highly skewed, with a high kurtosis and with heavy tails and extreme returns. As a result, the normal distribution is highly likely to not be suitable for fitting these returns' distributions.

With this being said, the previous argument is also supported by the results of the Jarque-Bera (JB) normally test, which applies a statistic to test whether the distribution of returns follows a normal distribution. The JB statistic equals zero when the distribution is normal, i.e., with zero skewness and kurtosis of three. The null hypothesis, H_0, states that the distribution is normal. Large values of skewness and, kurtosis values far from three, lead to the rejection of the null hypothesis (Yap & Sim, 2011).

For the considered historical period, the JB statistics rejects the null hypothesis, therefore indicating that there is enough statistical evidence to suggest that returns of all three indexes are not normally distributed.

The highest and lowest reported log-returns vary between, 13.25% (NASDAQ) and 10.76% (DJI 30), and between -13.84% (DJI 30) and -12.77% (S&P 500), respectively. As such, the DJI 30 index has been the index with the highest 1-day loss reported in the past 20 years.

Given the volatile nature of technological stocks, the NASDAQ has the highest daily volatility of 1.60%, which annualizes to a volatility of 25.42%. Moreover, the lowest volatility is expressed by the DJI 30 (1.21%, daily and 19.18% annualized) which does not differ significantly from the S&P 500's volatility.

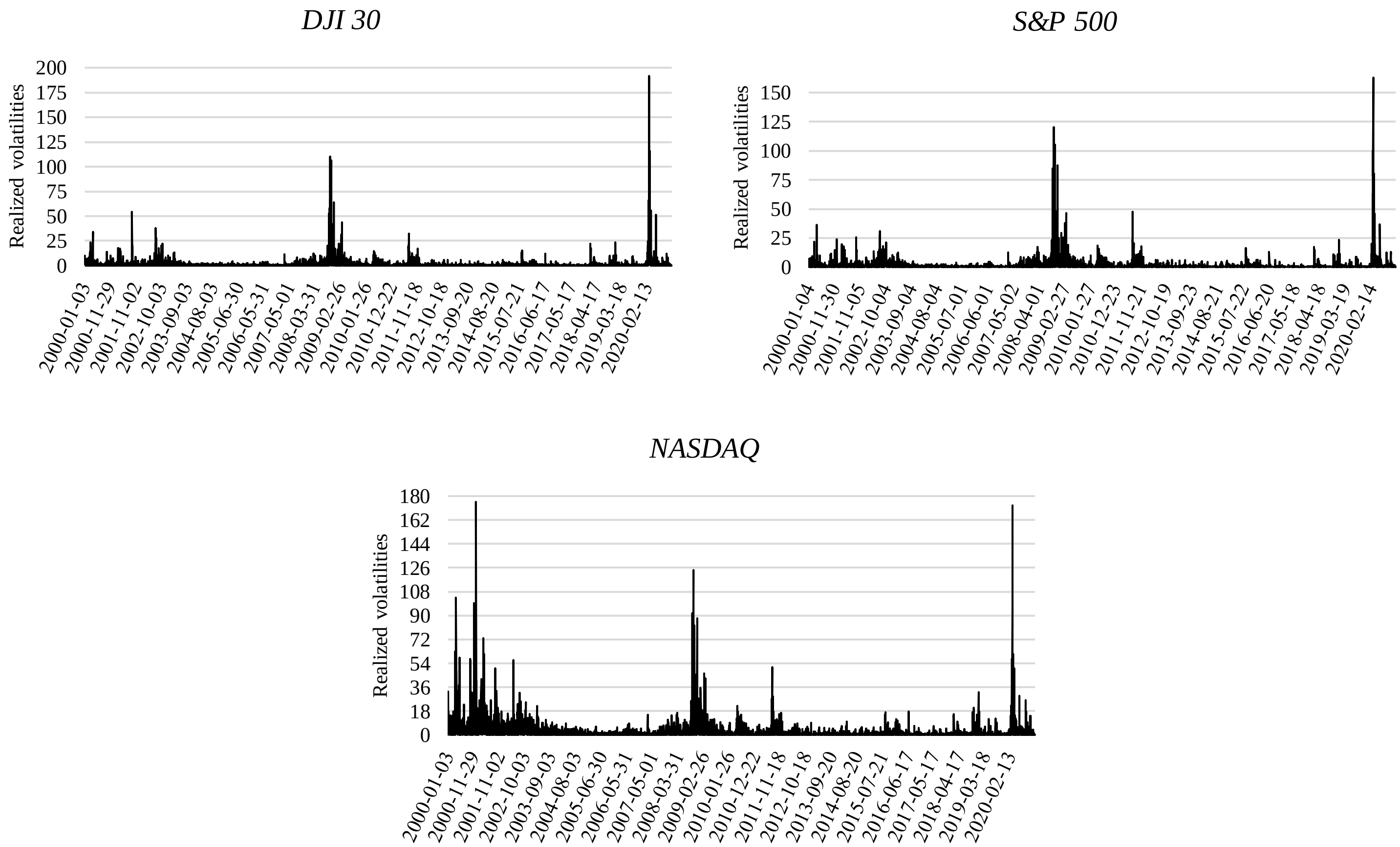
Figure 11 presents the daily realized volatilities (RV) of returns for all three indexes. These are computed as the square of each individual log-return,

$$RV_n = u_n^2$$

Where u_n^2 represents the squared log-return for period n.

Figure 11

Stock indexes realized volatility for the period between January 3, 2000, and Dec 30, 2020



Notably, the variance of returns is non-constant throughout time. This suggests that the distribution of log-returns is also affected by heteroskedasticity. In fact, there are three major moments throughout this historical period where volatility spiked considerably. These periods of increased volatility coincide with three of the major financial crisis that occurred during the 20th century: 1) the Dot-com bubble in the early 2000's, 2) the Great Financial Crisis, from 2007 until 2009 and 3) the COVID-19 pandemic recession of 2020. Moreover, it is noticeable, a significant, but less evident, spike in volatility during the year of 2011. This period of increased volatility coincides with the downgrading of the US sovereign debt performed by the credit rating agency Standard and Poor's. This period was called Black Monday.

Furthermore, the heteroskedastic nature of the indexes stock returns is substantiated by the results of the ARCH test also known as the Lagrange Multiplier (LM) test. This statistical test assesses whether if the null hypothesis, H_0, holds. The null hypothesis assumes that the residuals of a time series of returns exhibit no ARCH effects. The residuals are computed by fitting the series of returns to an ARCH (1) model. In other words, this hypothesis states that the residuals of a series are homoscedastic (Engle, 1982). If the results of the ARCH test are statistically significant the null hypothesis is rejected.

Moreover, the Ljung-Box test is also applied to test if the residuals of a series of returns express the absence of serial autocorrelation. In other words, the test tries to assess if the residuals of a series of returns are independent or not. As such, the null hypothesis H_0 states the residuals of returns are independent and not correlated. The residuals are computed by fitting the series of returns to an ARMA (1,1) model. (Ljung & Box, 1978). If the results of the Ljung-Box test are statistically significant the null hypothesis is rejected.

For this analysis, the ARCH tests reject the null hypothesis revealing that there is enough statistical evidence to suggest that the distribution of the three indexes suffers from a certain level of heteroskedasticity within its residuals and squared residuals. The Ljung-Box test also shows that there is enough statistical evidence to suggest that the residuals of each series are not independently distributed. As a result, the different series of indexes also contain some autocorrelation across its residuals.

After reviewing the outcome of these statistical tests, it can be concluded that to remove heteroskedasticity from the residuals of the distribution one might consider applying a volatility model. Additionally, an autoregressive process also ought to be considered to remove the autocorrelation across the residuals of each series.

In the study, the forecasting ability of different VaR models is compared using two distinct forecasting periods: a stable period (from 03/01/2000 to 31/12/2018 to train each model and a period between 02/01/2019 to 31/12/2019 as a 251 day forecasting horizon. This results in a full sample period of 5030 days between 03/01/2000 to 31/12/2019, with 4778 days used for estimation and 251 used for out-of-sample forecasting); and a volatile period (from 03/01/2000 to 31/12/2019 to train each model and a period between 02/01/2020 to 30/12/2020 as a 252 day forecasting horizon. This results in a full sample period of 5282 days between 03/01/2000 to 30/12/2020, with 5030 used for estimation and 252 used for out-of-sample forecasting).

Table 3, **Panel A** and **Panel B**, present some summary statistics for both the volatile and stable forecasting periods, respectively.

Table 3.

Summary statistics of the distribution of daily log- returns for the periods between January 2, 2020, and Dec 30, 2020 (Panel A) and January 2, 2019, and Dec 30, 2019 (Panel B), for the three indexes

	DJI 30	*S&P 500*	*NASDAQ*
Panel A. *Volatile Forecasting Period*			
Mean	0.0002	0.0005	0.0014
Median	0.0014	0.0024	0.0042
Maximum	0.1076	0.0897	0.0893
Minimum	-0.1384	-0.1277	-0.1315
Std. Deviation	2.34%	2.19%	2.26%
Std. Deviation (annualized)	37.02%	34.67%	35.78%
Skewness	-0.8127	-0.8514	-0.9820
Kurtosis	8.9155	8.3349	7.2239
Panel B. *Stable Forecasting Period*			
Mean	0.0008	0.0010	0.0012
Median	0.0010	0.0010	0.0014
Maximum	0.0324	0.0338	0.0417
Minimum	-0.0309	-0.0302	-0.0354
Std. Deviation	0.79%	0.79%	0.99%
Std. Deviation (annualized)	12.46%	12.49%	15.67%
Skewness	-0.6550	-0.6285	-0.4969
Kurtosis	3.1624	3.1519	2.4047

The statistical results presented in **Table 3** show that both forecasting periods display different characteristics. While the distribution of log-returns during the stable forecasting period presents some resemblance with normal distribution, the same does not apply for the volatile forecasting period.

Panel A shows that the forecasting period of 2020 is exceptionally volatile, with a much larger dispersion of returns relative to the mean. For instance, daily log-returns fluctuated from a low of -13.84% to a peak of 10.76% for the DJI 30 index. Similarly, daily and annualized volatilities are significantly higher than the average of the 20-year historical period (recall **Table 2**). Skewness is also much higher (more negative) than not only the historical period, but also higher than the normal distribution. Although kurtosis is not as substantial as the kurtosis of the 20-year historical period, these measurements are still considerably high. This suggests the presence of longer and fatter tails, i.e., negative returns, on the left side of the distribution and also higher peaks in the distribution. On the other hand, **Panel B** shows that for the stable forecasting period there is significantly less volatility, with a lower dispersion of returns throughout the historical period. For instance, the largest daily-log return registered during the year of 2019 for the S&P 500 was of 3.38% while the highest loss was of -3%. Moreover, the daily and annual volatilities are much lower than the historical average. Kurtosis statistics are also much closer to the values of the normal distribution.

The purpose of using two samples is two-fold. Firstly, the use of the first sample allows to provide an answer to the main research question of the study which aims at identifying which of the VaR models can accurately predict the losses incurred during the 2020 pandemic recession. Secondly, the other sample allows to assess the applicability of these models during periods of less volatility. This way, it is possible to validate the robustness and reliability of results across different time periods.

3.2. MODEL FRAMEWORK

This study compares the performance of different VaR models relative to their out-of-sample one-step-ahead predictive ability. These VaR models include different parametric, non-parametric and semi-parametric approaches.

For parametric methods, the 1-day VaR is forecasted using several volatility models, whose mean equation of log-returns are modeled under an autoregressive process, AR (1). These methods include the ARCH (1,1), EWMA (1,1), GARCH (1,1), GJR GARCH (1,1) and EGARCH (1,1) models. For each of these processes, volatility is fitted and forecasted using four different empirical distributions, two symmetric, the normal distribution (N) and the student t-distribution (STD), and two skewed, the skewed student t-distribution (SSTD) and the skewed generalized error distribution (SGED). See **Figure 12, Figure 13** and **Figure 14** for a graphical comparison of the normal distribution against the other three.

Figure 12

Plot of the standard normal density (blue) and student t-densities with 0.5, 1 and 30 degrees of freedom. Source: (Kumari & Tan, 2013)

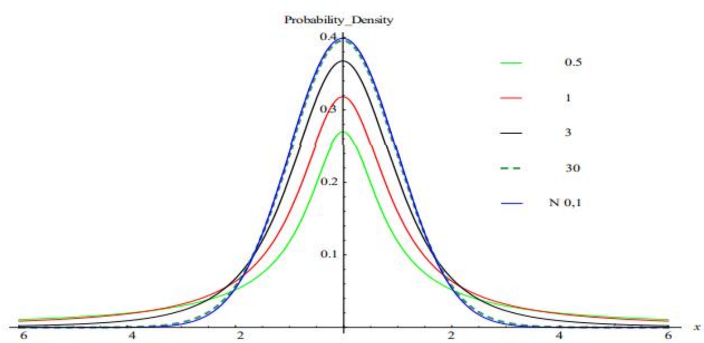


Figure 13

Plot of the normal distribution and skewed student t-distribution, where k represents the skewness parameter and v the kurtosis parameter. Source: (Scharnagl et al., 2015)

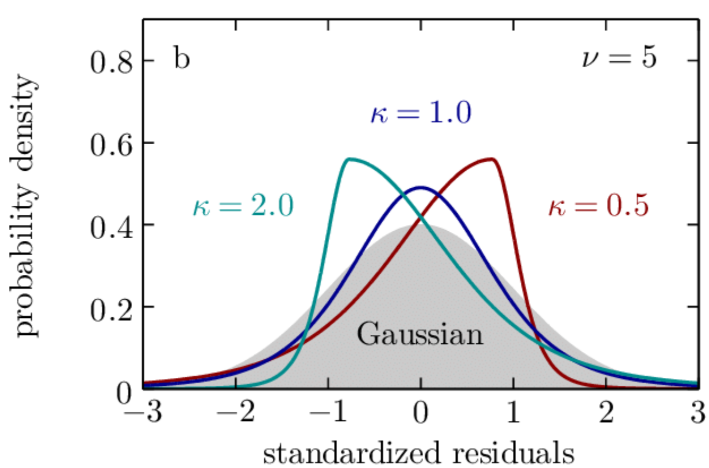
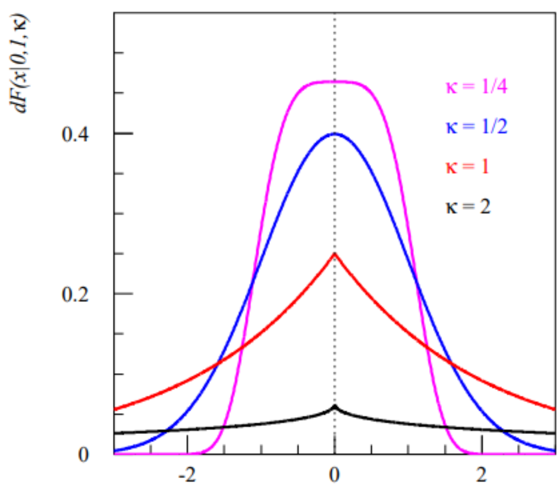


Figure 14

Plot of the normal distribution (in blue, with k=1/2) and skewed generalized error distribution, where k represents the skewness parameter. Source: (Giller, 2013)



The 1-day VaR forecasts under the different parametric assumptions are computed with a significance level, $\dot{\alpha}$, of 1%, see **Equation 26**.

$$VaR_{n+1|n}^{(\dot{\alpha})} = \mu_{n+1|n} + F^{-1}{}_{(\dot{\alpha})}\sigma_{n+1|n}$$

Equation 26. Specification of the 1-day VaR forecast using a parametric VaR model.

Where $\mu_{n+1|n}$ and $\sigma_{n+1|t}$ are the forecasts for the mean and standard deviation for the day $n + 1$, given information of day n, and $F^{-1}{}_{(\dot{\alpha})}$ is the qth quantile ($q = 1 - \dot{\alpha}$) of the respective distribution function F.

For non-parametric methods the 1-day 99% VaR is forecasted using two different approaches, the Historical Simulation (HS) and BRW simulation.

VaR is also forecasted using different semi-parametric methods, which include the Volatility Weighted Historical Simulation (VWHS), the Filtered Historical Simulation (FHS), Extreme Value Theory using both the Block Maxima (EVT-BM) and Peak Over Threshold (EVT-POT) approaches and two Monte Carlo Simulation (MCS) methods based on a geometric brownian motion (GBM) process and a normal distribution process (N).

3.2.1. Parametric models

For the parametric methods, the forecasting process of the 1% daily-VaR is derived in two steps. First the in-sample estimation for each model and then the out-of-sample forecasting of VaR for both the volatile and stable forecasting periods.

During the in-sample estimation process, the models are fitted to each of the four pre-determined distributions. After estimating the models' parameters, these are examined relative to their positivity constraints. This examination is performed to validate the model's adequacy for further analysis. The maximization of the log-likelihood function is used as the criteria for estimating each of the model's parameters.

Table 23, **Table 24** and **Table 25**, shown in the **Appendix** section, present a summary of the estimated parameters for each of the parametric models and indexes. Heteroskedasticity, autocorrelation statistical tests and goodness of fit statistics are also presented in **Table 26**. Given the similarity of the historical range used for estimation of both periods, it is only presented the estimation parameters for the period of 3 January 2000 until 31 December 2019. The log-likelihood ratio is used to assess the goodness of fit of the models.

The log-likelihood ratio increases proportionally to the ability of the method to model non-linear features of financial returns. That is, as a model becomes more sophisticated from a theoretical perspective, the higher is its goodness of fit, and vice-versa. For instance, the ARCH process, the least sophisticated parametric model, with a skewed generalized error distribution (ARCH-SGED) performs the worst at fitting the historical returns, having the lowest log-likelihood ratio of 5,204.21 for the DJI 30, 3,517.51 for the S&P 500 and 6,870.74 for the NASDAQ index. Comparably, the ARCH process with normal distribution (ARCH-N) also performs poorly. On the other hand, the EGARCH process, one of the most sophisticated parametric models, with a skew generalized error distribution (EGARCH-SGED) presents the highest goodness of fit statistic across all indexes with a log-likelihood ratio of 16,791.38 for the DJI 30, 16,604.21 for the S&P 500 and 15,241.8 for the NASDAQ.

Furthermore, models with skewed distribution assumptions are shown to fit the data better than models with normal and student t-distribution assumptions. For instance, within each family of parametric methods (i.e., EWMA, GARCH, GJR GARCH and EGARCH processes), the skewed distributional assumptions always report the highest goodness of fit ratios. Moreover, amongst skewed distributions, the models with a skewed generalized error distribution assumption provide the highest goodness of fit across most models. On the other hand, when comparing models with the same process, the ones based on a normal distribution assumption produce the lowest goodness of fit statistics. This suggests that the normal distribution performs poorly at fitting the historical log-returns. Note that no information is presented for both the ARCH-N and ARCH-STD models for the S&P 500 index. This happens because both models failed to reach an optimal log-likelihood criteria. As such, its parameters were not estimated.

Parametric models are disregarded from further analysis if these fail to comply with two main criteria. First, the model must be able to remove heteroskedasticity from its squared residuals and, secondly, the model must also be able to comply with the theoretically imposed positivity constraints for each parameter. As a result, models which reject the null hypothesis and show evidence of the presence of heteroskedasticity in its squared residuals, are therefore disregarded from further analysis. Also, models whose positivity constrains do not comply with theoretical standards should also be disregarded. The statistically significant tests, at a 0.05% level, are highlighted in bold.

Table 26 shows that the ARCH models with student t-distribution (ARCH-STD), skewed student t-distribution (ARCH-SSTD) and skewed generalized error distribution (ARCH-SGED) reject the null hypothesis of both ARCH tests. This happens because the models failed to remove heteroskedasticity from their residuals and squared residuals. As a result, these are disregarded from further analysis. On the other hand, other models, namely the EWMA with normal distribution (EWMA-N), is capable of removing heteroskedasticity from its squared residuals. However, it fails to remove heteroskedasticity from its residuals. As such, the EWMA-N process, despite not being capable of removing the whole heteroskedasticity from the residuals of the series of returns, it is still considered to be appropriate for further analysis.

The second criteria is related to the model's positivity constraints. For a detailed overview of each model's particular specificities regarding its positivity constraints see Abad et al. (2014).

An example of a volatility model which fails to meet these parameter positivity constraints is the EGARCH process. As mentioned, the EGARCH is a process that models the leverage effects in financial returns through a leverage parameter called gamma, γ, which is expected to be negative. As such, it is shown that, for all EGARCH models, the leverage parameter is estimated to be positive, therefore contradicting theoretical assumptions. As a result, the model must be rejected and also disregarded from further analysis.

After the previous filtering process, it was performed a 252 and 251 day rolling forecast of volatilities and means for the volatile and stable forecasting periods, respectively. Then, the 1-day 99% VaR is calculated from those estimates using **Equation 26**.

3.2.2. Non-parametric models

For both non-parametric approaches, the 1-day 99% VaR is forecasted as the 1% quantile of the empirical distribution of historical past returns. Thereafter, this forecast is used as the effective VaR risk measure for the following 252 and 252 days, for both the volatile and stable periods, respectively.

Also, to select the 1% quantile of the empirical distribution of returns, with the HS approach it are assigned equal weights to each past observation while with the BRW HS it are assigned exponentially decreasing weights to past observations. As suggest by Hull (2018)'s textbook, a lambda, or weight, of 0.94 was considered.

3.2.3. Semi-parametric models

Considering that the Volatility Weighted Historical Simulation (VWHS) standardizes returns to account for volatility changes, it was necessary to select a parametric model to estimate volatilities. A GARCH (1,1) process with skewed generalized error distribution was chosen for this purpose. After standardizing each of the historical returns, the 1-day 99% VaR is assigned as the quantile of the empirical distribution. Then, this result is used as the effective VaR risk measure for the following forecasting period.

Likewise, the same GARCH (1,1) specification was used to forecast volatilities for the Filtered Historical Simulation approach. Additionally, a bootstrapping procedure, each with 20 observations, was performed 10,000 times to produce various 1-day VaR forecasts. The most conservative simulation was considered.

To produce the 1-day 99% VaR forecast using both MCS methods, 10,000 simulations were randomly generated. Thereafter, the mean for these VaR forecasts was selected as the 1-day VaR effective risk measure.

Furthermore, to find each of the block maxima under the EVT-Block Maxima (EVT-BM) approach, the data was split into 45 different blocks for both the stable and volatile forecasting periods. The selected block maxima were then fitted to a Generalized Extreme Value (GEV) distribution. Afterwards, a 1-day 99% VaR risk measure was produced using the estimated parameters. This amount was also selected as the final daily VaR risk measure for the respective forecasting periods.

Lastly, for the EVT- Peaks over Threshold (EVT-POT) approach, a threshold of -5% and -2% was selected to filter extreme losses for both the volatile and stable forecasting periods, respectively. Then, a Generalized Pareto Distribution (GPD) distribution, based on a maximum likelihood criterion, was used to fit these extreme observations. Again, the estimated parameters were used to produce the 1-day 99% VaR forecast. This amount was also set as the 1-day VaR measurement for the following year.

Note that most of the intermediary decisions applied in the development of each of the models, such as for example, the selection of the number of bootstrapping iterations to include, the number of observation within each iterations used for the FHS approach, or the threshold percentage used for the EVT-POT model and, given the lack of a more general criteria, was based on an arbitrary trial and error process which aimed at identifying which combination of these parameters provided the most satisfactory VaR forecasts.

3.3. EVALUATION FRAMEWORK

The main purpose of this study is to compare the forecasting ability of different VaR models. As such, it is necessary to select an appropriate metric to evaluate the quality of produced forecasts. Given that VaR is inherently unobservable, it must be monitored and consistent with the actual realized returns of the historical period. Hence, a model to be considered adequate must not only generate reliable and statistically accurate forecasts but it also must be preferred to other models.

Accordingly, the performance of the different VaR methodologies is compared using a two-staged selection approach. For the first stage two statistical accuracy tests are applied. As the name itself implies, these accuracy tests (unconditional coverage test and conditional coverage test) are applied to assess the level of accuracy of each model. On a second stage, the remaining models are evaluated using quantitative loss functions. These loss functions (Lopez loss function and Quantile loss function) are applied to assess the relative performance of each of the models. To assure the reliability of results, if any of the models fails to comply with these tests, it is disregarded from further analysis.

3.3.1. First stage: accuracy tests

The first step involves testing the statistical accuracy of the different VaR models. To achieve this, it is applied two backtesting procedures, an unconditional and a conditional coverage tests. A model is only considered appropriate for the second stage of analysis if not rejected by any of these tests.

3.3.1.1. Unconditional coverage test

Consider $E = \sum_{t=1}^{T} I_n$ as the number of days over a specific time period where the registered portfolio, or financial instrument's loss was larger than the VaR forecast for a day n . I_n, in **Equation 27**, represents the exceedance indicator variable, which assumes the value of one when an exceedance occurs, i.e., when the historical log-return is lower than the VaR forecast for a day n $(u_{n+1} < VaR(\dot{\alpha})_{n+1})$.

$$I_{n+1} = \begin{cases} 1, & if\ u_{n+1} < VaR(\dot{\alpha})_{n+1} \\ 0, & if\ u_{n+1} \geq VaR(\dot{\alpha})_{n+1} \end{cases}$$

Equation 27. Specification of the Kupiec's unconditional coverage test.

E represents the number of observed exceedances within the sample.

Kupiec (1995)'s unconditional coverage accuracy test states that, given a constant probability of exceedances, the number of VaR exceedances, E, will follow a binomial distribution $E \sim B(N, \dot{\alpha})$. Thus, a VaR model, to be considered accurate, should yield an unconditional coverage ($\hat{\dot{\alpha}} = \sum_{n=1}^{N} I_E / T$) equal to the significance level of $\dot{\alpha}$.

Under the null hypothesis, H_0, it is tested whether if the number of exceedances over the number of days of a sample, T, is equal to a predetermined significance level of $\dot{\alpha}$ (H_0: $\hat{\dot{\alpha}} = \dot{\alpha}$). The unconditional likelihood ratio statistic is computed the following way:

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

The test follows a χ^2 distribution with one degree of freedom, represented by χ. Despite its direct usability, the test can reject a model for both high and low failures, which is not a good outcome. As such, it can sometimes perform very poorly.

3.3.1.2. Conditional coverage test

A more complete test was introduced by Christoffersen (1995). The author proposes the use of a likelihood ratio statitistic to test the join assumption of both the uncoditional coverage test and the independence of failures. The conditional likelihood ratio statistic for the conditional coverage test is a function of the sum of the ratio arising from the unconditional coverage test LR_{uc}, plus the ratio from a serial independence test of exceedences, LR_{ind} where the $LR_{cc} = LR_{uc} + LR_{ind}$.

This test is a significant improvement over the previous one since it accounts for any existing conditionality in the VaR forecasts. Accordingly, if volatilities were shown to be low in specific periods and high in others the forecast should be able to respond to this clustering event regardless of the underlying distributional assumption.

If a sequence of values is independent, then the probability to observe, or not, a VaR violation in the following period should be equal. Thereby, this statistic can reject a VaR model that generates either too many or too few clustered violations. Nonetheless, to perform accurately, the model needs several hundred of observations. The likelihood ratio statistic for this test is computed using the following expression:

$$LR_{cc} = -2ln[(1-p)^{T-E}p^{E}] + 2\ln[(1-\pi_{01})^{n00}\pi_{01}^{n01}(1-\pi_{11})^{n10}\pi_{11}^{n11}] \sim\chi^2(2)$$

Where n_{ij}corresponds to the number of observations with the value i followed by j for $i,j = 0{,}1$ and

$$\pi_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$$

represent the respective probabilities. The values of $i,j = 1$ denote an exception, while $i,j = 0$ indicate the opposite. If the sequence of values is independent, then the probabilities of observing, or not, a VaR violation in the following period must be equal ($i.e.$, $\pi_{01} = \pi_{11} = p$).

3.3.2. Second stage: loss function tests

In the second and final stage of this study it is examined the magnitude of losses experienced when a VaR exceedance occurs. As well as before, it are only considered models that were not rejected by any of the previously imposed statistical tests.

Financial institutions cannot afford to allocate an inadequate amount of capital to face imminent risks and losses. The mismanagement of capital might result into two different outcomes. In case of overestimation of risk and losses, the financial instituiton might be overcharged by institutional regulators. On the other hand, in case of understimation of capital requirements, the financial institution may be left without enough capital to cover its losses, which might force the liquidation of other assets to cover these requirements (Angelidis et al., 2004).

As such, and given these possible outcomes, it is of upmost importance for a financial institution to produce not only accurate (without significantly underestimating the actual incurred losses), but also adequate (without significantly overestimating the actual incurred losses) VaR forecasts.

3.3.2.1. Lopez loss function

To assess the magnitude of losses experienced when an VaR exceedance occurs, Lopez (1998) suggests measuring the accuracy of a VaR model by summing all the differences between the observed returns and the VaR forecasts when an exceedance occurs. This measurement is called the Lopez loss function (LLF). According to the author, a penalty variable is defined as in **Equation 28**.

$$\Psi_{n+1} = \begin{cases} 1 + (u_{n+1} - VaR_{n+1|n})^2, & if\ u_{n+1} < VaR_{n+1|n} \\ 0, & if\ u_{n+1} \geq VaR_{n+1|n} \end{cases}$$

Equation 28. Specification of the Lopez's loss function penalty score.

Where u_{n+1} represents the observed log-return for day $n+1$ and $VaR_{n+1|n}$ denotes the VaR forecast for that same period.

According to this penalty variable, a model will be penalized only when an exceedance takes place. For this reason, a model is preferred to another if it yields a lower penalty score. The total penalty assigned to each model is defined as the sum of all individual penalties assigned to the model, $\Psi = \sum_{n=1}^{N} \Psi_n$.

This test is very similar to the Kupiec (1995)'s test. However, this one assigns a penalty for both the cumulative number of exceedances and the magnitude of the losses incurred in these exceedances. The larger the difference, the larger the penalty added to the model.

3.3.2.1.Quantile loss function

On the other hand, the Quantile Loss function (QLF) is applied to account for both the magnitude of the distances of losses experienced when an exceedance occurs, and the magnitude of the distances when an exceedance does not occur (Angelidis et al., 2006). Its penalty variable is specified in **Equation 29**.

$$\Psi_{n+1} = \begin{cases} (u_{n+1} - VaR_{n+1|n})^2, & if\ u_{n+1} < VaR_{n+1|n} \\ [Percentile(u, 100p)_1^N - VaR_{n+1|n}]^2, & if\ u_{n+1} \geq VaR_{n+1|n} \end{cases}$$

Equation 29. Specification of the Quantile Loss function penalty score.

Given this loss function, a model's penalty score increases either by 1) the distance between the VaR forecast for day $n+1$, VaR_{n+1}, and the future 100th percentile of the empirical return distribution, $Percentile(u, 100p)_1^N$, if the observed return is larger than the VaR forecast for the same period; or 2) by the excess loss, if the observed return is lower than the forecasted VaR. The preferred model is the one which minimizes the QLF.

The QLF is probably one of the most effective ways of assessing a model's accuracy and adequacy. This loss function penalizes models both for the underestimation and overestimation of real incurred losses. It also presents meaningful improvements over the LLF penalty scores. While the LLF only evaluates the accuracy of each of the models, the QLF is capable of identifying from the most accurate models which are the ones that also produce the most adequate VaR forecasts. To clarify this point, a model might produce very accurate VaR forecasts, with low exceedance rates, but still be considered an unsatisfactory model for forecasting VaR. This happens because the VaR forecasts might be so high that model deeply overestimates the actual losses incurred during the forecasting period. This will lead to a mismanagement of capital as described before.

4. Empirical results and discussion

4.1. First stage: VaR accuracy tests

Within this section, it is presented the accuracy statistical measures for each of the different VaR models.

Table 4 displays the number of exceedances and exceedance rates obtained for each of the parametric methods during the volatile forecasting period. **Table 5** shows the results of the two statistical tests. **Table 6** and **Table 7** present the same results obtained for the stable forecasting period. Across the two tables for each period, four different metrics are presented. First, the total number of exceedances, second, the exceedance rate (ER) in **Table 4** and **Table 6**, and lastly, the p-values of two accuracy statistical tests, the unconditional coverage test (LR_{uc}) and the conditional coverage test (LR_{cc}) in **Table 5** and **Table 7**. ER represents the number of VaR exceedances over the number of daily forecasts. The statistical tests of the models whose VaR forecasts are considered to be inaccurate are highlighted in bold. A model is rejected and considered inaccurate when the null hypothesis is rejected. The null hypothesis tests whether if the "VaR forecasts are accurate". Both statistical accuracy tests were performed with a significance level ($\dot{\alpha}$), of 0.05 and 0.01. A VaR model is considered accurate if not rejected by both tests. The tests are also performed at a significance level of 0.01 to address the limitations of the Kupiec (1995)'s which tends to penalize models that have a low exceedance rate. **Table 5** and **Table 7** present the p-values of the accuracy tests at a 0.05 and 0.01 significance levels for the two forecasting periods.

Table 4

Total number of 1-day VaR exceedances and exceedance rates for parametric models (volatile forecasting period)

Model	*DJI 30*		*S&P 500*		*NASDAQ*		*Average*	
(Volatile period)	***#***	***ER***	***#***	***ER***	***#***	***ER***	***#***	***ER***
ARCH-N	21	8.33%	-	-	16	6.35%	18.5	7.34%
EWMA – N	12	4.76%	12	4.76%	11	4.37%	11.7	4.63%
EWMA-STD	12	4.76%	10	3.97%	10	3.97%	10.7	4.23%
EWMA-SSTD	11	4.37%	10	3.97%	9	3.57%	10.0	3.97%
EWMA-SGED	11	4.37%	10	3.97%	9	3.57%	10.0	3.97%
GARCH-N	11	4.37%	10	3.97%	9	3.57%	10.0	3.97%
GARCH-STD	10	3.97%	8	3.17%	8	3.17%	8.7	3.44%
GARCH-SSTD	9	3.57%	7	2.78%	8	3.17%	8.0	3.17%
GARCH-SGED	9	3.57%	7	2.78%	8	3.17%	8.0	3.17%
GJR GARCH-N	8	3.17%	7	2.78%	10	3.97%	8.3	3.31%
GJR GARCH-STD	8	3.17%	6	2.38%	6	2.38%	6.7	2.64%
GJR GARCH-SSTD	8	3.17%	6	2.38%	6	2.38%	6.7	2.64%
GJR GARCH-SGED	7	2.78%	5	1.98%	6	2.38%	6.0	2.38%

Notes: # represents the number of 1-day VaR exceedances and ER denotes the exceedance rate (i.e., $ER = \#/N$, where N is the total number of days of the forecasting horizon).

Table 5

P-values of the accuracy statistical tests for parametric models (volatile forecasting period)

Model	*DJI 30*				*S&P 500*				*NASDAQ*			
(Volatile period)	LR_{uc}		LR_{cc}		LR_{uc}		LR_{cc}		LR_{uc}		LR_{cc}	
$\dot{\alpha}$	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*
ARCH-N	**0.026**	**0.000**	**0.022**	**0.000**	-	-	-	-	0.345	**0.000**	0.215	**0.000**
EWMA – N	0.861	**0.000**	0.851	**0.000**	0.861	**0.000**	0.851	**0.001**	0.637	**0.000**	0.705	**0.000**
EWMA-STD	0.861	**0.000**	0.851	**0.000**	0.436	**0.000**	0.516	**0.001**	0.436	**0.000**	0.488	**0.001**
EWMA-SSTD	0.637	**0.000**	0.705	**0.000**	0.436	**0.000**	0.516	**0.001**	0.274	**0.001**	0.393	**0.005**
EWMA-SGED	0.637	**0.000**	0.705	**0.000**	0.436	**0.000**	0.516	**0.001**	0.274	**0.001**	0.393	**0.005**
GARCH-N	0.637	**0.000**	0.705	**0.000**	0.436	**0.000**	0.516	**0.001**	0.274	**0.001**	0.393	**0.005**
GARCH-STD	0.436	**0.000**	0.488	**0.001**	0.155	**0.006**	0.280	0.017	0.155	**0.006**	0.280	0.017
GARCH-SSTD	0.274	**0.001**	0.393	**0.005**	0.078	0.020	0.174	0.054	0.155	**0.006**	0.280	0.017
GARCH-SGED	0.274	**0.001**	0.393	**0.005**	0.078	0.020	0.174	0.054	0.155	**0.006**	0.280	0.017
GJR GARCH-N	0.155	**0.006**	0.280	0.017	0.078	0.020	0.174	0.054	0.436	**0.000**	0.516	**0.001**
GJR GARCH-STD	0.155	**0.006**	0.280	0.017	**0.034**	0.061	0.092	0.150	**0.034**	0.061	0.092	0.150
GJR GARCH-SSTD	0.155	**0.006**	0.280	0.017	**0.034**	0.061	0.092	0.150	**0.034**	0.061	0.092	0.150
GJR GARCH-SGED	0.078	0.020	0.174	0.054	**0.013**	0.166	**0.041**	0.346	**0.034**	0.061	0.092	0.150

Notes: $\dot{\alpha}$ represents the p-value for each of the accuracy tests in percentage (%). LR_{uc} and LR_{cc} denote the unconditional and conditional coverage tests, respectively.

Table 6

Total number of 1-day VaR exceedances and exceedance rates for parametric models (stable forecasting period)

Model	*DJI 30*		*S&P 500*		*NASDAQ*		*Average*	
(Stable period)	***#***	***ER***	***#***	***ER***	***#***	***ER***	***#***	***ER***
ARCH-N	5	2.00%	3	1.20%	11	4.40%	6.3	2.53%
EWMA – N	7	2.80%	7	2.80%	6	2.40%	6.7	2.67%
EWMA-STD	7	2.80%	6	2.40%	4	1.60%	5.7	2.27%
EWMA-SSTD	4	1.60%	4	1.60%	4	1.60%	4.0	1.60%
EWMA-SGED	7	2.80%	4	1.60%	4	1.60%	5.0	2.00%
GARCH-N	7	2.80%	6	2.40%	4	1.60%	5.7	2.27%
GARCH-STD	7	2.80%	4	1.60%	4	1.60%	5.0	2.00%
GARCH-SSTD	7	2.80%	4	1.60%	4	1.60%	5.0	2.00%
GARCH-SGED	5	2.00%	4	1.60%	3	1.20%	4.0	1.60%
GJR GARCH-N	7	1.60%	4	1.60%	4	1.60%	5.0	1.60%
GJR GARCH-STD	4	1.60%	4	1.60%	4	1.60%	4.0	1.60%
GJR GARCH-SSTD	4	1.60%	4	1.60%	4	1.60%	4.0	1.60%
GJR GARCH-SGED	4	1.60%	4	1.60%	4	1.60%	4.0	1.60%

Notes: # represents the number of 1-day VaR exceedances and ER denotes the exceedance rate (i.e., $ER = \#/N$, where N is the total number of days of the forecasting horizon).

Table 7

P-values of the accuracy statistical tests for parametric models (stable forecasting period)

Model	*DJI 30*				*S&P 500*				*NASDAQ*			
(Stable period)	***LR**uc*		**LRcc**		***LR**uc*		**LRcc**		***LR**uc*		**LRcc**	
$\dot{\alpha}$	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*
ARCH-N	**0.014**	0.164	**0.043**	0.343	**0.001**	0.763	**0.004**	0.921	0.197	**0.000**	0.677	**0.000**
EWMA – N	0.083	0.019	0.181	0.053	0.083	0.019	0.181	0.053	**0.037**	0.060	0.097	0.148
EWMA-STD	0.083	0.019	0.181	0.053	**0.037**	0.060	0.097	0.148	**0.004**	0.384	**0.016**	0.642
EWMA-SSTD	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642
EWMA-SGED	0.083	0.019	0.181	0.053	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642
GARCH-N	0.083	0.019	0.181	0.053	**0.037**	0.060	0.097	0.148	**0.004**	0.384	**0.016**	0.642
GARCH-STD	0.083	0.019	0.181	0.053	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642
GARCH-SSTD	0.083	0.019	0.181	0.053	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642
GARCH-SGED	**0.014**	0.164	**0.043**	0.343	**0.004**	0.384	**0.016**	0.642	**0.001**	0.763	**0.004**	0.164
GJR GARCH-N	0.083	0.019	0.181	0.053	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642
GJR GARCH-STD	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642
GJR GARCH-SSTD	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642
GJR GARCH-SGED	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642	**0.004**	0.384	**0.016**	0.642

Notes: $\dot{\alpha}$ represents the p-value for each of the accuracy tests in percentage (%). LR_{uc} and LR_{cc} denote the unconditional and conditional coverage tests, respectively.

Table 4 shows that the ARCH-N process reports the highest VaR exceedances across all parametric models, with an average ER of 7.34% across all three indexes. As a result, the null hypothesis is rejected for both tests (for the DJI 30 index), see **Table 5**. Within the models that share the same volatility process, the ones based on a normal distribution assumption perform worse than the ones based on the remaining distributions. For instance, amongst EWMA models, the EWMA-N performs the worst for two of the three indexes, with an ER of 4.76% and 4.37% for the S&P 500 and NASDAQ indexes, respectively. Similarly, the GARCH-N model is also the worst performer amongst all other GARCH processes across all three indexes. Similar conclusions can also be reported for the GJR-GARCH process. The GJR-GARCH models based on student t-distribution always outperform the ones based on a normal distribution.

There is also enough evidence to suggest that models based on skewed distributions produce more accurate VaR forecasts than models based on symmetrical distributional assumptions. For instance, within the EWMA, GARCH and GJR-GARCH volatility processes, the ones based on skewed distributions always outperform the ones based on symmetrical distributions. For the EWMA and GARCH processes, there is no meaningful difference between the accuracy of models based on skewed student t-distribution and skew generalized error distribution. For the GJR-GARCH family the model with skewed generalized error distribution outperforms the model with skewed student t-distribution.

Additionally, the accuracy of the models gradually increases as the model is able to capture more non-linear features of returns. The EWMA model outperforms the ARCH process, the GARCH outperforms the EWMA and the GJR-GARCH outperforms the GARCH. Nonetheless, despite being considered less theoretically sophisticated, it is found that models such as the GARCH and EWMA, also provide accurate VaR forecasts with relatively low VaR exceedance rates.

The GJR GARCH-SGED is the best performing model with an average ER of 2.38% across all three indexes.

Ultimately, there is not enough evidence to suggest that the models perform better on and index than another. Given the results of these statistical tests, all models, but the ARCH-N model, are fit for the second evaluation stage.

Table 6 and **Table 7** present similar statistics for the stable forecasting period. When compared with the forecasts reported during the volatile period, it is noticeable a considerable increase in accuracy across all parametric approaches. This results show the significant decrease of reported VaR exceedances during this historical period.

The overall relative performance of each individual model does also change significantly between both periods. For instance, during the stable forecasting period, the EWMA-SSTD outperforms all the other GARCH models, across 2 out of the 3 indexes. Also, the GJR GARCH-STD, GJR GARCH-SSTD and GJR GARCH-SGED report equivalently accurate VaR forecasts, with an average ER of 1.6 across all three indexes. Lastly, the VaR forecasts for the ARCH-N model also improve significantly for 2 of the 3 indexes. As such, the ARCH-N model reports the lowest ER (of 1.2%) of all parametric models for the S&P 500 index.

Table 8 and **Table 9** show the accuracy results and statistical tests obtained for the non-parametric methods during the volatile forecasting period. A similar table structure is used to describe the statistical findings.

Table 8

Total number of 1-day VaR exceedances and exceedance rates for non-parametric models (volatile forecasting period)

Model	*DJI 30*		*S&P 500*		*NASDAQ*		*Average*	
(Volatile period)	***#***	***ER***	***#***	***ER***	***#***	***ER***	***#***	***ER***
HS	13	5.16%	14	5.56%	9	3.57%	12	4.76%
BRW HS	47	18.65%	50	19.84%	47	18.65%	48	19.05%

Notes: # represents the number of 1-day VaR exceedances and ER denotes the exceedance rate (i.e., $ER = \#/N$, where N is the total number of days of the forecasting horizon).

Table 9

P-values of the accuracy statistical tests for non-parametric models (volatile forecasting period)

Model	*DJI 30*				*S&P 500*				*NASDAQ*			
(Volatile period)	LR_{uc}		LR_{cc}		LR_{uc}		LR_{cc}		LR_{uc}		LR_{cc}	
$\dot{\alpha}$	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*
HS	0.897	**0.000**	0.919	**0.000**	0.680	**0.000**	0.890	**0.000**	0.280	**0.001**	0.336	**0.004**
BRW HS	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Notes: $\dot{\alpha}$ represents the p-value for each of the accuracy tests in percentage (%). LR_{uc} and LR_{cc} denote the unconditional and conditional coverage tests, respectively.

Table 10

Total number of 1-day VaR exceedances and exceedance rates for non-parametric models (stable forecasting period)

Model	*DJI 30*		*S&P 500*		*NASDAQ*		*Average*	
(Stable period)	***#***	***ER***	***#***	***ER***	***#***	***ER***	***#***	***ER***
HS	0	0.00%	0	0.00%	0	0.00%	0	0.00%
BRW HS	0	0.00%	0	0.00%	0	0.00%	0	0.00%

Notes: # represents the number of daily VaR exceedances and ER denotes the exceedance rate (i.e., $ER = \#/N$, where N is the total number of days of the forecasting horizon).

Table 11

P-values of the accuracy statistical tests for non-parametric models (stable forecasting period)

Model	*DJI 30*				*S&P 500*				*NASDAQ*			
(Stable period)	***LR_{uc}***		***LRcc***		***LR_{uc}***		***LRcc***		***LR_{uc}***		***LRcc***	
$\dot{\alpha}$	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*
HS	-	-	-	-	-	-	-	-	-	-	-	-
BRW HS	-	-	-	-	-	-	-	-	-	-	-	-

Notes: $\dot{\alpha}$ represents the p-value for each of the accuracy tests in percentage (%). LR_{uc} and LR_{cc} denote the unconditional and conditional coverage tests, respectively.

Table 8 shows that for the volatile forecasting period, the BRW HS reports very inaccurate VaR forecasts across all three indexes, with an average ER of 19.05%. As a result, the null hypothesis is rejected for both accuracy tests at the two significance levels. On the other hand, the HS approach performs relatively well, providing reasonably accurate VaR forecasts. These conclusions contradict previous literature findings which suggest that the BRW simulation generally outperforms the HS, see Boudoukh et al. (1998) and Pritsker (2006).

As noted in **Section 2.4.2.2**, the BRW HS computes VaR as the $\dot{\alpha}$ percent quantile of the empirical distribution of returns by assigning exponentially decreasing weights based on the recency of the historical log-return. For that reason, recent returns have a larger impact on the VaR forecasts than past returns. This behavior leads to the underestimation of the impact of past crisis which, as a consequence, results in less conservative 1-day VaR forecasts. On the other hand, the HS approach, which assigns equal weights to each historical observation, reports more conservative, and therefore accurate VaR forecasts. As a result, given its poor performance, the BRW HS fails to pass to the second evaluation stage.

Table 10 and **Table 11** present the accuracy results and statistical tests for the non-parametric models during the stable forecasting period, respectively. Once again, both models produce more reliable VaR forecasts during the stable period than during the volatile period. In fact, both of the models improved significantly, disclosing no VaR exceedances across all three indexes.

Table 12 and **Table 13** present the results obtained from the accuracy results and statistical test for the semi-parametric models. A similar table structure as before is used to express each model's statistical results.

Table 12

Total number of 1-day VaR exceedances and exceedance rates for semi-parametric models (volatile forecasting period)

Model	*DJI 30*		*S&P 500*		*NASDAQ*		*Average*	
(Volatile period)	***#***	***ER***	***#***	***ER***	***#***	***ER***	***#***	***ER***
VWHS	6	2.38%	6	2.38%	5	1.98%	5.7	2.25%
FHS	6	2.38%	3	1.19%	3	1.19%	4.0	1.59%
MCS-N	21	8.33%	20	7.94%	15	5.95%	18.7	7.41%
MCS-GBM	26	10.32%	24	9.52%	19	7.54%	23.0	9.13%
EVT-BM	10	3.97%	9	3.57%	19	7.54%	12.7	5.03%
EVT-POT	2	0.79%	0	0.00%	0	0.00%	0.7	0.26%

Notes: # represents the number of 1-day VaR exceedances and ER denotes the exceedance rate (i.e., $ER = \#/N$, where N is the total number of days of the forecasting horizon).

Table 13

P-values of the accuracy statistical tests for semi-parametric models (volatile forecasting period)

Model	*DJI 30*				*S&P 500*				*NASDAQ*			
(Volatile period)	LR_{uc}		***LRcc***		LR_{uc}		***LRcc***		LR_{uc}		***LRcc***	
$\dot{\alpha}$	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*
VWHS	**0.035**	0.061	**0.032**	0.051	**0.035**	0.060	**0.032**	0.051	**0.013**	0.164	**0.042**	0.343
FHS	**0.035**	0.061	**0.032**	0.051	**0.001**	0.763	**0.004**	0.921	**0.001**	0.763	**0.004**	0.921
MCS-N	**0.025**	**0.000**	0.052	**0.000**	**0.046**	**0.000**	0.075	**0.000**	0.491	**0.000**	0.784	**0.000**
MCS-GBM	**0.001**	**0.000**	**0.002**	**0.000**	**0.003**	**0.000**	**0.011**	**0.000**	0.081	**0.000**	0.201	**0.000**
EVT-BM	0.445	0.045	0.523	0.060	0.280	0.109	0.336	0.095	0.081	**0.000**	0.201	**0.000**
EVT-POT	**0.000**	0.737	**0.001**	0.930	-	-	-	-	-		-	-

Notes: $\dot{\alpha}$ represents the p-value for each of the accuracy tests in percentage (%). *LRuc* and *LRcc* denote the unconditional and conditional coverage tests, respectively.

Table 14

Total number of 1-day VaR exceedances and exceedance rates for semi-parametric models (stable forecasting period)

Model	***DJI 30***		***S&P 500***		***NASDAQ***		***Average***	
(Stable period)	***#***	***ER***	***#***	***ER***	***#***	***ER***	***#***	***ER***
VWHS	0	0.00%	0	0.00%	0	0.00%	0.0	0.00%
FHS	0	0.00%	0	0.00%	0	0.00%	0.0	0.00%
MCS-N	3	1.19%	2	0.79%	0	0.00%	1.7	0.66%
MCS-GBM	5	1.98%	5	1.98%	5	1.98%	5.0	1.98%
EVT-BM	0	0.00%	0	0.00%	5	1.98%	1.7	0.66%
EVT-POT	0	0.00%	0	0.00%	0	0.00%	0.0	0.00%

Notes: # represents the number of 1-day VaR exceedances and ER denotes the exceedance rate (i.e., $ER = \#/N$, where N is the total number of days of the forecasting horizon).

Table 15

P-values of the accuracy statistical tests for semi-parametric models (stable forecasting period)

Model	***DJI 30***				***S&P 500***				***NASDAQ***			
(Stable period)	***LR_{uc}***		***LRcc***		***LR_{uc}***		***LRcc***		***LR_{uc}***		***LRcc***	
$\dot{\alpha}$	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*	*0.05*	*0.01*
VWHS	-	-	-	-	-	-	-	-	-	-	-	-
FHS	-	-	-	-	-	-	-	-	-	-	-	-
MCS-N	**0.001**	0.758	**0.004**	0.931	0.000	0.742	**0.001**	0.932	-	-	-	-
MCS-GBM	**0.014**	0.162	**0.044**	0.346	**0.014**	0.162	**0.044**	0.346	**0.014**	0.162	**0.044**	0.346
EVT-BM	-	-	-	-	-	-	-	-	**0.014**	0.162	**0.044**	0.346
EVT-POT	-	-	-	-	-	-	-	-	-	-	-	-

Notes: $\dot{\alpha}$ represents the p-value for each of the accuracy tests in percentage (%). *LRuc* and *LRcc* denote the unconditional and conditional coverage tests, respectively.

Table 12 shows that both MCS methods produce very inaccurate VaR forecasts with a large number of VaR exceedances. As such, the null hypothesis of both tests is rejected for the DJI 30 and S&P 500 indexes. Notably, the remaining VaR models, despite rejecting the null hypothesis at a 0.05 significance level, report very accurate VaR forecasts. Their adequacy for the next stage of evaluation is justified by the results of the tests performed with a 0.01 significance level, which do not reject the null hypothesis.

The EVT-POT is by far the best performing model across all semi-parametric methods, reporting a percentage of exceedances very close to zero across all three indexes. The FHS follows in second place with also very accurate 1-day VaR forecasts for all indexes. These models present an average ER of 0.26% and 1.59%, respectively. Similar findings are also reported in past research, see Bao et al. (2006), Barone-Adesi and Giannopoulos (2000) and Gilli and Këllezi (2006). On the contrary, the EVT-BM, despite producing relatively accurate VaR forecasts, performs significantly worse than the other semi-parametric alternatives.

Table 14 shows that all semi-parametric models provide considerably more accurate VaR forecasts across all three indexes during the stable forecasting period. In fact, the VWHS, FHS, and EVT-POT models report a zero percent exceedance rate across all three indexes. Despite the relative improvement in performance, both MCS approaches still report the least accurate VaR forecasts.

4.1.1. Accuracy tests: comprehensive synopsis

The results obtained for both accuracy and statistical tests suggest that semi-parametric methods are, without a doubt, the best models at predicting 1-day VaR for the volatile forecasting period of 2020. For instance, the EVT-POT approach clearly outperforms all others, reporting the lowest VaR exceedances. The second best performing model is the FHS which produces very satisfactory and accurate 1-day VaR forecasts across all three indexes. Moreover, the VWHS also represents a good alternative to forecast VaR. This approach not only reports very accurate forecasts but also outperforms all the remaining historical simulation approaches.

Both non-parametric approaches perform very poorly. For instance, the BRW HS is the second worst performer across all methods with forty-seven 1-day VaR exceedances for both the DJI 30 and NASDAQ indexes and fifty VaR exceedances for the S&P 500 index. Similarly, and despite much more accurate, the HS also reports a substantial number of VaR exceedances.

Despite the overall supremacy of semi-parametric methods over all others, most parametric approaches are still able to provide reasonably accurate VaR forecasts. The GARCH model generally outperforms the EWMA model, producing slightly more accurate 1-day VaR forecasts, with an average exceedance rate well below 5% across all three indexes. The GJR GARCH-SGED model is the best performing model amongst all parametric approaches, producing the most accurate VaR forecasts, with the lowest exceedance rates, across 2 out of the 3 indexes.

A graphical illustration of the performance of the best VaR models during the forecasting period of 2020 is shown in **Figure 15** (EVT-POT), **Figure 16** (FHS) and **Figure 17** (GJR GARCH-SGED) in the **Appendix** section.

Given the lack of meaningful financial losses in the year of 2019, it comes with no surprise that all models perform significantly better in the stable forecasting period. Nonetheless, the overall relative performance of all models does change significantly between both periods. For instance, the ARCH-N which produces unreliable forecasts during the volatile forecasting period, performs significantly better in the stable period, even outperforming some of more theoretically sophisticated methodologies. Additionally, both non-parametric methods are also significantly more accurate during this period.

4.2. SECOND STAGE: VAR LOSS FUNCTIONS

The models that are qualified for the second evaluation stage are analyzed in this section. The parametric models: EWMA-N, EWMA-STD, EWMA-SSTD, EWMA-SGED, GARCH-N, GARCH-STD, GARCH-SSTD, GARCH-SGED, GJR GARCH-N, GJR GARCH-STD, GJR GARCH-SSTD, GJR GARCH-SGED, non-parametric models: HS, and semi-parametric models: VWHS, FHS, EVT-BM, EVT-POT are considered for the volatile forecasting period. All models are considered for the stable forecasting period. The following tables present the penalty scores produced by both loss functions for each of the models.

First, it is examined the penalties produced by the Lopez (1998) loss function (LLF). This loss function punishes a model for both the cumulative number of exceedances and for the magnitude of those exceedances. As such, it allows to assess the relative accuracy of each of the VaR models. On a second phase, it is examined the penalties produced by the Quantile Loss function (QLF). This loss function, on the other hand, penalizes models from not only the magnitude of the distance of the VaR exceedances but also for the distance when the empirical observed return is larger than the one implied by the model. This loss function is considered to be superior because it allows to assess whether if an accurate model is capable of producing reliable and adequate VaR forecasts, without meaningful overestimations of the actual the incurred risk.

Table 16.

Loss function penalty scores for parametric models (volatile forecasting period)

Model	*DJI 30*		*S&P 500*		*NASDAQ*	
(Volatile period)	***LLF***	***QLF***	***LLF***	***QLF***	***LLF***	***QLF***
EWMA – N	12.003	0.669	12.003	0.580	11.008	0.630
EWMA-STD	12.002	0.786	10.002	0.670	10.002	0.718
EWMA-SSTD	11.002	0.893	10.002	0.788	9.007	0.818
EWMA-SGED	11.001	0.929	10.001	0.824	9.002	0.847
GARCH-N	11.002	0.636	10.002	0.553	9.003	0.592
GARCH-STD	10.002	0.823	8.002	0.731	8.002	0.755
GARCH-SSTD	9.001	0.911	7.001	0.822	8.001	0.847
GARCH-SGED	9.001	0.897	7.001	0.807	8.001	0.842
GJR GARCH-N	8.003	0.701	7.003	0.601	10.003	0.629
GJR GARCH-STD	8.003	0.862	6.003	0.766	6.002	0.787
GJR GARCH-SSTD	8.002	0.987	6.002	0.894	6.002	0.926
GJR GARCH–SGED	7.002	0.973	5.002	0.882	6.002	0.922

Table 17

Loss function penalty scores for parametric models (stable forecasting period)

Model	*DJI 30*		*S&P 500*		*NASDAQ*	
(Stable period)	***LLF***	***QLF***	***LLF***	***QLF***	***LLF***	***QLF***
ARCH-N	5.001	0.263	3.000	0.418	11.000	0.304
EWMA – N	7.001	0.099	7.000	0.100	6.001	0.156
EWMA-STD	7.001	0.115	6.000	0.115	4.001	0.175
EWMA-SSTD	7.000	0.128	4.000	0.132	4.000	0.201
EWMA-SGED	7.000	0.134	4.000	0.140	4.000	0.210
GARCH-N	7.000	0.099	6.000	0.101	4.001	0.158
GARCH-STD	7.000	0.122	4.000	0.127	4.000	0.186
GARCH-SSTD	5.000	0.138	4.000	0.143	4.000	0.212
GARCH-SGED	5.000	0.138	4.000	0.144	3.000	0.215
GJR GARCH-N	7.000	0.103	4.000	0.103	4.000	0.160
GJR GARCH-STD	5.000	0.123	4.000	0.123	4.000	0.184
GJR GARCH-SSTD	4.000	0.142	4.000	0.144	4.000	0.217
GJR GARCH-SGED	4.000	0.142	4.000	0.144	4.000	0.219

Table 16 presents the penalty scores of the two loss functions (LLF and QLF) for each parametric model during the volatile forecasting period.

As implied by the specification of the LLF test, see **Equation 26.**, models with a higher cumulative number of exceedances produce higher penalty scores. As a result of that, the EWMA-N is the worst performing model, reporting the highest penalty scores across all three indexes. Its LLF penalty scores range from 11.008 (NASDAQ) to 12.003 (DJI30, S&P500). Also, the EWMA-STD is the second worst performer across all parametric approaches. The model reports LLF penalty scores between 10.002 (NASDAQ, S&P 500) and 12.002 (DJI 30). Note that, despite presenting the same number of exceedances for the DJI 30 index, the EWMA-STD reports a lower LLF penalty score than the EWMA-N. These are of 12.003 and 12.002, respectively. Although their difference in penalty scores is residual, this suggests that the EWMA-STD model is capable of producing slightly more accurate VaR forecasts, with a lower magnitude of VaR exceedances.

On the contrary, the loss function results suggest that the GJR GARCH-SGED is the best performing model, reporting the lowest LLF penalty scores across each individual index. The model's penalties range between 6.002 (NASDAQ) to 7.002 (DJI30).

Within the family of EWMA and GARCH models, the processes with skewed generalized error distribution and skewed student t-distribution produce equivalently satisfactory 1-day VaR forecasts, with identical penalty scores and therefore, equivalent performance. Similar conclusions can be reported for the GJR GARCH models with student t-distribution and skewed student t-distribution, which also have identical LLF penalty scores.

Furthermore, the LLF penalty scores imply that, despite reporting the same number of exceedances, the EWMA-SGED model slightly outperforms the GARCH-N. On the other hand, the GARCH-N model outperforms the EWMA-SSTD for the NASDAQ index, with penalty scores of 9.007 and 9.003, respectively.

Despite being considered one of the most accurate parametric models, with the lowest LLF penalty scores, the GJR GARCH-SGED reports one of the highest QLF penalties across all volatility methods and indexes. Likewise, the GJR GARCH-SSTD also reports very high QLF penalties.

Moreover, even though it was reported before that the EWMA-SSTD and EWMA-SGED have an equivalent performance, it can be found that this is not true. The QLF results suggest that the EWMA-SSTD model is slightly more adequate than the EWMA-SGED for forecasting VaR. This conclusion is justified by the model's lower QLF penalty scores. The EWMA-SSTD reports QLF penalty scores which range between 0.788 (S&P 500) to 0.893 (DJI 30) while the EWMA-SGED penalty scores range between 0.824 (S&P 500) to 0.929 (DJI 30). Similarly, the GARCH-SGED also outperforms the GARCH-SSTD, reporting lower penalty scores. Moreover, the GJR GARCH-STD also outperforms the GJR GARCH-SSTD reporting lower penalty scores and, therefore, more adequate VaR forecasts. The model's penalty scores range between 0.766 (S&P 500) to 0.862 (DJI 30) and 0.895 (S&P 500) to 0.987 (DJI 30), respectively.

Table 17 shows that all parametric models perform significantly better during the stable forecasting period, reporting much lower LLF penalty scores across all three indexes. Moreover, the contribution of the actual magnitude of exceedances, i.e., without the unitary contribution of the number of VaR exceedances, is also significantly lower. This suggests that during this period these models are not only more accurate, producing a lower amount of 1-day VaR exceedances, but also more reliable, with lower penalties arising from the underestimation of the actual incurred losses.

The EWMA-N model is the worst performing model, reporting LLF penalty scores that range between 6.001 (NASDAQ) and 7.001 (DJI 30). The ARCH-N process reports the highest LLF penalty scores, of

11.001 for the NASDAQ index, across all parametric models. Despite that, the model outperforms all others for the S&P 500 index, reporting the lowest LLF penalty scores of 3.000. Comparably to what was reported for the volatile forecasting period, the EWMA-SSTD and EWMA-SGED also have an equivalent performance. On the other hand, despite reporting the same number of VaR exceedances, the GJR GARCH-SSTD outperforms the GJR GARCH-STD with a significantly lower LLF penalties. Ultimately, both the GJR GARCH-SSTD and GJR GARCH-SGED produce the most accurate VaR forecasts across all models, both with LLF penalty scores that range between 4.001 (DJI 30, S&P 500) and 4.002 (NASDAQ).

Furthermore**,** the results of the QLF penalty scores further emphasize the premise that parametric models produce significantly more adequate VaR forecasts during the stable forecasting period. This happens because the models' QLF penalty scores are shown to be significantly lower.

As such, the same parametric VaR models, not only produce more accurate but also less overestimated, and more reliable, VaR forecasts during the stable forecasting period. For instance, the GJR GARCH-SGED, one of best performing models across both periods, produces QLF penalty scores that range between of 0.142 (DJI30) and 0.169 (NASDAQ). These are, on average, 82% lower in the stable period than in the volatile period. Moreover, amongst the EWMA family of models, the EWMA-STD is the preferred one, producing the lowest average QLF penalty scores, between 0.115 (DJI 30) and 0.175 (NASDAQ). Likewise, amongst the family of GARCH processes, the GARCH-STD also reports the lowest penalty scores. The QLF penalty scores also suggest that the GJR GARCH-STD is the best performing parametric model for forecasting 1-day VaR during the stable forecasting period of 2019.

Table 18 presents the penalty scores (LLF and QLF) for the non-parametric methods during the volatile forecasting period. Similarly, **Table 19** shows the same penalty scores for the stable period.

Table 18

Loss function penalty scores for non-parametric models (volatile forecasting period)

Model	*DJI 30*		*S&P 500*		*NASDAQ*	
(Volatile period)	***LLF***	***QLF***	***LLF***	***QLF***	***LLF***	***QLF***
HS	13.023	0.268	14.016	0.287	9.011	0.479

Table 19

Loss function penalty scores for non-parametric models (stable forecasting period)

Model	*DJI 30*		*S&P 500*		*NASDAQ*	
(Stable period)	***LLF***	***QLF***	***LLF***	***QLF***	***LLF***	***QLF***
HS	0.000	0.267	0.000	0.302	0.000	0.497
BRW HS	0.000	0.247	0.000	0.271	0.000	0.376

Table 19 shows that the performance of both non-parametric models increased significantly during the stable forecasting period.

Resulting from the absence of 1-day VaR exceedances during this period, both non-parametric models report no LLF penalty scores. Also, the BRW HS, which was discarded from further analysis during the volatile forecasting period, performs much better in the stable period.

Despite the LLF penalty scores suggesting that, from a standpoint of accuracy alone, there are no differences in the performance of both models, the results of the QLF penalty scores indicate the opposite. The BRW HS clearly outperforms the HS reporting lower QLF penalty scores. Recall that, during the volatile forecasting period the HS was found to be substantially more accurate.

In closing, **Table 20** shows the penalty scores (LLF and QLF) obtained for the semi-parametric methods during the volatile forecasting period.

Table 20

Loss function penalty scores for semi-parametric models (volatile forecasting period)

Model	***DJI 30***		***S&P 500***		***NASDAQ***	
(Volatile period)	***LLF***	***QLF***	***LLF***	***QLF***	***LLF***	***QLF***
VWHS	6.014	0.548	6.011	0.545	5.010	0.596
FHS	6.011	0.968	3.007	1.165	3.008	1.072
EVT-BM	10.018	0.406	9.013	0.416	19.02	0.224
EVT-POT	2.003	1.744	0.000	29.804	0.000	41.889

Table 21

Loss function penalty scores for semi-parametric models (stable forecasting period)

Model	***DJI 30***		***S&P 500***		***NASDAQ***	
(Stable period)	***LLF***	***QLF***	***LLF***	***QLF***	***LLF***	***QLF***
VWHS	0.000	0.588	0.000	0.558	0.000	0.633
FHS	0.000	2.395	0.000	2.893	0.000	2.825
MCS-N	3.000	0.161	2.000	0.182	0.000	0.318
MCS-GBM	5.000	0.111	5.000	0.125	5.000	0.217
EVT-BM	0.000	0.340	0.000	0.377	5.000	0.196
EVT-POT	0.000	1.660	0.000	28.625	0.000	6.234

Once again, given the absence of 1-day VaR exceedances, the EVT-POT is shown to be best performing, and therefore, most accurate model across all semi-parametric approaches, with LLF penalty scores between 0.000 (NASDAQ, S&P 500) and 2.003 (DJI 30). The FHS is the second best performing model with LLF penalty scores between 3.007 (S&P 500) and 6.011 (DJ I30). Despite not as accurate as these two semi-parametric models, the VWHS also reports very satisfactory VaR forecasts. Although both report the same amount of VaR exceedances for the DJI 30 index, the lower LLF penalty scores suggest that FHS produces more adequate VaR forecasts than the VWHS. Also, the EVT-BM is the worst performer amongst all semi-parametric methods, presenting the highest penalty scores across all indexes.

Moreover, the QLF penalty scores presented in **Table 20** also show that most of the models, despite reporting very accurate forecasts, also produce highly overestimated VaR forecasts. For instance, the EVT-POT model produces zero 1-day VaR exceedances across two of the three indexes, being classified as the most accurate of all models applied in this study. However, its QLF penalty scores, which range between 1.744 (DJI 30) and 41.889 (NASDAQ), suggest that there is a meaningful overestimation of the actual incurred risk. The same occurs for the FHS which also overestimates the actual incurred risk, reporting an QLF penalty scores between 2.395 (DJI 30) and 2.893 (S&P 500).

Additionally, the VWHS provides reasonably adequate VaR forecasts, reporting much lower QLF penalty scores, between 0.968 (DJI 30) and 1.165 (S&P 500), than the other semi-parametric approaches. Notably, the QLF penalty scores also suggest that the VWHS is more appropriate for forecasting VaR for the DJI 30 index than the FHS. While the VWHS reports an QLF score of 0.548 the FHS reports a penalty score of 0.968.

Furthermore, there is a significant improvement of the performance of the semi-parametric methods during the stable forecasting period, see **Table 21**. All models report a significantly lower amount of 1-day VaR exceedances.

The VWHS performs equivalently as well as the FHS. Both models report zero 1-day VaR exceedances, with LLF penalty scores of zero across all three indexes. Likewise, the EVT-POT maintains a VaR exceedance rate of zero percent across all of three indexes. The EVT-BM improves significantly also reporting zero VaR exceedances across two of the three indexes.

At last, both MCS methods also perform significantly better in the stable forecasting period. However, despite the improvement, these two approaches are still heavily outperformed by the remaining semi-parametric methods.

During the stable forecasting period, see **Table 21**, there is also a notable improvement in the adequacy of most semi-parametric approaches. For instance, both the EVT-POT and EVT-BM produce significantly lower penalty scores, reporting QLF penalty scores which range between 1.660 (DJI 30) and 28.625 (S&P 500), and, 0.196 (NASDAQ) and 0.377 (S&P 500), respectively.

On the other hand, despite more accurate, the FHS reports a higher QLF penalty score during the stable forecasting period. The VHWS performs equivalently well in both periods and it is considered to one of the most adequate of all semi-parametric models. The model produces the lowest average QLF scores across all three indexes. However, the EVT-BM outperforms all semi-parametric models for both the DJI 30 and S&P 500 indexes, with QLF scores of 0.340 and 0.377, respectively.

4.2.1. Loss functions: comprehensive synopsis

By examining the results derived from both loss functions it can be noticed a notable contrast between the performance of all models across indexes and forecasting periods. The LLF focuses solely on measuring the cumulative number of losses and the magnitude of those VaR exceedances, which disregards overestimations of the actual incurred risk. On the other hand, the QLF focuses on not only the magnitude of the distance of the VaR exceedances but also on the distance when the empirical return is larger than the one implied by the model. Thus, the QLF is capable of accounting for both VaR overestimations and underestimations of the real incurred risk. As a result, the QLF can be viewed as more granular and effective metric to evaluate and compare each model's VaR forecasts.

For the volatile forecasting period, the EVT-POT is the preferred model across all tested methodologies. Given the absence of VaR exceedances, its LLF penalty scores are very close to zero. Nonetheless, the QLF penalty scores suggest that, despite incredibly accurate, the model produces highly overestimated VaR forecasts, with much larger penalties.

The FHS follows the EVT-POT model as the second most accurate approach for the volatile forecasting period. The FHS reports the second lowest average LLF penalty scores across all three indexes. Despite this, the extra layer of conservativeness in the 1-day VaR forecasts is also penalized by the QLF scores. These penalty scores are substantially larger for the FHS than for the other semi-parametric, and even parametric and non-parametric, alternatives.

Overall, there is a general tendency for more conservative and accurate (with lower exceedance rates) VaR models to produce less adequate (or overestimated) forecasts. Despite that, some of the semi-parametric models, namely the VWHS, FHS and EVT-POT are proven to be the best at forecasting VaR. Other parametric models, namely the GJR GARCH-STD, GJR GARCH-SSTD and GJR GARCH-SGED also produce satisfatory forecasts. Lastly, there is evidence that suggests that the HS approach is the worst model for forecasting VaR during the volatile forecasting period, producing the least accurate forecasts, with the largest LLF penalty scores, across all three indexes.

In addition, the VaR forecasts produced by each model are also proven to be more adequate during the stable forecasting period. This is expected given the significantly lower amount of losses observed during the historical period, which results in lower penalties arising from both loss functions.

The BRW simulation, which did not pass to the second stage of evaluation for the volatile forecasting period, is the model which reports the most accurate and adequate VaR forecasts across all three indexes. The HS closely follows, reporting the second lowest QLF penalty scores. Moreover, the EVT-BM is the third best performing model for two of the three indexes producing not only accurate VaR forecasts, with zero 1-day VaR exceedances, but also adequate forecasts, with low QLF penalty scores. The VWHS also reports satisfatory 1-day VaR forecasts.

Once again, despite also very accurate, most of the semi-parametric approaches are shown to overestimate the actual incurred losses. For instance, both the EVT-POT and FHS methods report zero VaR exceedances, however, the their QLF penalty scores are significantly higher than the other semi-parametric alternatives.

5. Conclusion

The purpose of this study is twofold. First it aims at providing a detailed overview of the currently available, and most commonly used methods for forecasting Value at Risk (VaR). Secondly, it seeks to identify, from some of these models, the ones that are the most capable of capturing the market losses incurred during the 2020 COVID-19 pandemic recession. This analysis is perfomed for three of the most well known american indexes, the DJI30, the S&P 500 and the NASDAQ.

To achieve the second point, it is compared the performance of different VaR models. 1) parametric models: ARCH, EWMA, GARCH, GJR GARCH, EGARCH, measured using two symmetric and two skewed distributions, 2) non-parametric models: BRW HS and HS and 3) semi-parametric models: VWHS, FHS, MCS under a normal distribution and geometric brownian motion assumptions, EVT-BM and EVT-POT. The models' performance is compared using a two-staged backtesting procedure. In the first stage it is applied two accuracy statistical tests that analyze the model's capacity of producing 1-day VaR forecasts which are not exceeded by the real incurred losses. As such, in the first stage is it analyzed the model's accuracy. After that, in the second stage, it is tested, using two different loss functions, the capacity of the models to produce adequate 1-day VaR forecasts, without meaninful overestimation nor underestimations of the real incurred risk. The best model is the one that combines both of the most satisfatory accuracy and adequacy statistical results. This statistical analysis is performed during two different forecasting historical periods, a volatile (year 2020) and stable period (year 2019).

A statistical comparison of the accuracy and loss function tests indicates that the semi-parametric methods produce the most accurate VaR forecasts for the volatile forecasting period, with the EVT-POT being considered the best performing model for this period. Most parametric models also produce reasonable 1-day VaR forecasts throughout the volatile forecasting period, with the GJR GARCH-SGED model reporting the most accurate results amongst all parametric approaches. On the contrary, the HS approach produces the least accurate VaR forecasts.

Despite that, for the volatile period, the Quantile Loss functions (QLF) penalty scores indicate that there is a general tendency for the most conservative and accurate VaR models to produce less adequate forecasts. That means that as the level of conservativeness increases, the overestimation of VaR also occurs more often.

The same statistical tests indicate that all models perform signifcantly better during the stable forecasting period, producing not only more accurate but also adequate VaR forecasts during this period. Unexpectedly, for the stable forecasting period, despite reporting the same level of VaR exceedances as other models, the BRW HS displays the most reliable VaR forecasts, reporting the lowest QLF penalty scores across all three indexes. The HS approach also performs incredibly well during this forecasting period.

This study's conclusions highlight important issues. First, there is enough evidence to conclude that the normal distribution assumption is inadequate to model financial return's distributions. As such, it is also proven in this study that if the distribution of returns displays heavier tails than the normal distribution, the results can improve substantially if modelled using other symetric distributions (e.g., the student t-distribuiton), or skewed distributions (e.g., skewed student t-distribution and skewed generalized error distribution). Second, VaR processes which are capable of modeling more non-linear features of returns also produce the most accurate VaR forecasts. Finally, it is found that most of the parametric approaches, despite theorectically less atractive, are still able to perform at least as well as other more sophisticated VaR models, reporting both reasonably accurate and adequate 1-day VaR forecasts.

The applied VaR models perform significantly better in the stable forecasting period than in the volatile forecasting period. Moreover, the relative performance of the models also changes from one period to the other. As such, it becomes impractical to underline a single flawless and ideal model that should always be applied for 1-day VaR forecasting purposes. To ilustrate, the EVT-POT is often considered to be the best model for forecasting VaR under stressed market conditions. Although that is true for the volatile forecasting period of this study, it is found find that, during the stable forecasting period, there are less attractive methodologies, at least from a theorectical prespective, that provide more optimal VaR forecasts.

This all to say that it is impossible to have a "one-size fits all" approach in regards to forecasting VaR. Regardless of the methodology used and the complexity of the model itself, the reported results and conclusions will always be heavily dependent on the dataset, the volatily of the underlying period, the statistical assumptions made and the financial instrument itself.

With this is mind, this study concludes that the EVT-POT approach performs the best at forecasting losses during the 2020 COVID-19 pandemic recession. Nontheless, there is always a tradeoff that must be considered. The more conservative the model is, the least adequate the forecast will become. Thus, banks will have to keep more capital aside to deal with these potential risks. This might become useful if the period ahead is highly volatile with meaninful losses, however, if not the case, banks will end up with a lot of tied-up capital which might lead to not only the loss of potential investment opportunies but also to operational inefficiency.

This being said, it is our duty as risk managers to perform an effective due dilgence process, test different alternatives and scenarios and, most importantly, always be aware that what happened in the past might not always repeat itself in the future.

6. Limitations and recommendations for future work

Despite the coherent results, this study also contains potential limitations. First and foremost, the diversity of VaR models studied by researchers is endless. There is a broad testing of different VaR models, with different (sometimes small) adaptations and distributional assumptions, throughout the literature. As a consequence, it becomes impossible to apply all the existing VaR models from a practical perspective. With this being said, it is recognized the possibility that there might be other VaR models, which were not addressed in this study, that also report satisfactory or even better forecasts.

Furthermore, VaR was forecasted for three of the most well-known american indexes which were proven to have very similar statistical properties, such as mean, volatility, skewness and kurtosis. Therefore, it would be of value to compare the performance of different indexes across other geographical regions.

Also, the consideration of a portfolio composed of only stocks is unrealistic in a real world scenario. Banks often manage complex portfolios of financial instruments, often resorting to the use of financial derivatives to mitigate risk. As such, the replication of this study using a complex portfolio ought to be a very interesting exercise.

Lastly, there is an interesting dichotomy that must be considered. The VaR model's capacity to avoid financial losses was used as the main metric to evaluate its performance. However, it must be acknowledged that as more conservative the VaR forecast becomes, the more capital the bank has to hold aside to deal with potential losses. Thus, banks might potentially be losing investments opportunities whose rate of return could exceed the actual incurred losses, if the VaR forecasts were substantially lower. As a result, when evaluating a VaR model, and as a suggestion for future work, one might perform a more detailed risk management analysis that accounts for this added opportunity cost.

7. Bibliography

Abad, P., & Benito, S. (2013). A detailed comparison of value at risk estimates. *Mathematics and Computers in Simulation*, *94*(1), 258–276. https://doi.org/10.1016/j.matcom.2012.05.011

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. In *Journal of Financial Economics* (Vol. 61, Issue 1). https://doi.org/10.1016/S0304-405X(01)00055-1

Angelidis, T., Benos, A., & Degiannakis, S. (2004). The use of GARCH models in VaR estimation. *Statistical Methodology*, *1*(1–2), 105–128. https://doi.org/10.1016/j.stamet.2004.08.004

Angelidis, T., Benos, A., & Degiannakis, S. (2006). A robust VaR model under different time periods and weighting schemes. *Review of Quantitative Finance and Accounting*, *28*(2), 187–201. https://doi.org/10.1007/s11156-006-0010-y

Antoniou, I., Ivanov, V. V., Ivanov, V. V., & Zrelov, P. V. (2004). On the log-normal distribution of stock market data. *Physica A: Statistical Mechanics and Its Applications*, *331*(3–4), 617–638. https://doi.org/10.1016/j.physa.2003.09.034

Balkema, a a, & de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability*, *2*(5), 347–370. https://doi.org/10.1214/aop/1176996548

Balvers, R., Wu, Y., & Gilliand, E. (2000). Mean Reversion across National Stock Markets and Parametric Contrarian Investment Strategies. *The Journal of Finance*, *55*(2), 745–772.

Bao, Y., Lee, T. H., & Saltoglu, B. (2006). Evaluating predictive performance of value-at-risk models in emerging markets: A reality check. *Journal of Forecasting*, *25*(2), 101–128. https://doi.org/10.1002/for.977

Barone-Adesi, G., & Giannopoulos, K. (2000). *Filtering Historical Simulation: Backtest Analysis*.

Barone-Adesi, G., & Giannopoulos, K. (2001). Non-parametric VaR techniques. Myths and realities. *Economic Notes*, *30*(2), 167–181. https://doi.org/10.1111/j.0391-5026.2001.00052.x

Barone-Adesi, G., Giannopoulos, K., & Vosper, L. (1998). VaR without Correlations for Nonlinear Portfolios. *Journal of Futures Markets*, *19*, 583–602.

Basu, S. (2011). Comparing simulation models for market risk stress testing. *European Journal of Operational Research*, *213*(1), 329–339. https://doi.org/10.1016/j.ejor.2011.02.023

Bekiros, S. D., & Georgoutsos, D. A. (2005). Estimation of Value-at-Risk by extreme value and conventional methods: A comparative evaluation of their predictive performance. *Journal of International Financial Markets, Institutions and Money*, *15*(3), 209–228. https://doi.org/10.1016/j.intfin.2004.05.002

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heroskedasticity. *Journal of Econometrics*, *45*(1–2), 267–290. https://doi.org/10.1016/0304-4076(90)90101-X

Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics*, *69*(3), 542. https://doi.org/10.2307/1925546

Bollerslev, T., Kretschmer, U., Pigorsch, C., & Tauchen, G. (2008). A discrete-time model for daily

S&P500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics*, *150*(2), 151–166. https://doi.org/10.1016/j.jeconom.2008.12.001

Boudoukh, J., Richardson, M., & Whitelaw, R. F. (1998). The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk. *Risk*, *11*(5), 64–67.

Bradley, B. O., & Taqqu, M. S. (2001). Financial Risk and Heavy Tails. *Handbook of Heavy Tailed Distributions in Finance*, 35–103. https://doi.org/10.1016/b978-044450896-6.50004-2

Brooks, C. (2008). *Econometrics for Finance* (Cambridge University Press (ed.); 2nd ed.). Cambridge University.

Chang, C. lin, Jiménez-Martín, J. Á., McAleer, M., & Pérez-Amaral, T. (2011). Risk management of risk under the Basel Accord: forecasting value-at-risk of VIX futures. *Managerial Finance*, *37*(11), 1088–1106. https://doi.org/10.1108/03074351111167956

Chen, C. W. S., Gerlach, R., Lin, E. M. H., & Lee, W. C. W. (2011). Bayesian Forecasting for Financial Risk Management, Pre and Post the Global Financial Crisis. *Journal of Forecasting*, *31*(8), 661–687. https://doi.org/10.1002/for.1237

Christoffersen, P. F. (1995). Evaluating Interval Forecasts. *The Journal of Derivatives*, *21*(2), 841–862.

Danielsson, J., & Vries, C. G. D. E. (2000). Value-at-Risk and Extreme Returns. *Annales d'Économie et de Statistique*, *60*, 239–270. https://doi.org/10.2307/20076262

Doane, D. P., & Seward, L. E. (2011). Measuring skewness: A forgotten statistic. *Journal of Statistics Education*, *19*(2). https://doi.org/10.1080/10691898.2011.11889611

Ener, E., Baronyan, S., & Ali Mengütürk, L. (2012). Ranking the predictive performances of value-at-risk estimation methods. *International Journal of Forecasting*, *28*(4), 849–873. https://doi.org/10.1016/j.ijforecast.2011.10.002

Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987. https://doi.org/10.2307/1912773

Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, *15*(4), 157–168. https://doi.org/10.1257/jep.15.4.157

Engle, R., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, *22*(4), 367–381. https://doi.org/10.1198/073500104000000370

Estrella, A., Hendricks, D., & Kambhu, J. (1994). The Price Risk of Options Positions. *Quarterly Review FRB New York*, *19*(2), 27–43.

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2), 180–190. https://doi.org/10.1017/S0305004100015681

Gençay, R., Selçuk, F., & Ulugülyağci, A. (2003). High volatility, thick tails and extreme value theory in value-at-risk estimation. *Insurance: Mathematics and Economics*, *33*(2), 337–356. https://doi.org/10.1016/j.insmatheco.2003.07.004

Giller, G. L. (2013). A Generalized Error Distribution. *SSRN Electronic Journal*, *November*. https://doi.org/10.2139/ssrn.2265027

Gilli, M., & Këllezi, E. (2006). An application of extreme value theory for measuring financial risk.

Computational Economics, *27*(2–3), 207–228. https://doi.org/10.1007/s10614-006-9025-7

Giot, P., & Laurent, S. (2003). Value-at-Risk for long and short trading positions. *Journal of Applied Econometrics*, *18*(6), 641–663. https://doi.org/10.1002/jae.710

Giot, P., & Laurent, S. (2004). Modelling daily Value-at-Risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, *11*(3), 379–398. https://doi.org/10.1016/j.jempfin.2003.04.003

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, *48*(5), 1779–1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x

Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The Annals of Mathematics*, *44*(3), 423. https://doi.org/10.2307/1968974

Gomes, T., & Khan, N. (2011). Strengthening Bank Management of Liquidity Risk: The Basel III Liquidity Standards. *Bank of Canada Financial System Review*, *December*, 35–42.

He, X. Z., Li, K., & Wang, C. (2016). Volatility clustering: A nonlinear theoretical approach. *Journal of Economic Behavior and Organization*, *130*, 274–297. https://doi.org/10.1016/j.jebo.2016.07.020

Holton, G. A. (2002). History of Value at Risk VaR. *Working Paper*.

Huang, A. Y. H. (2009). A value-at-risk approach with kernel estimator. *Applied Financial Economics*, *19*(5), 379–395. https://doi.org/10.1080/09603100701857906

Hull, J. (2018). *Risk Management and Financial Institutions* (Wiley (ed.); 5th ed.).

Hull, J., & White, A. (1998). Incorporating volatility updating into the historical simulation method for value-at-risk. *The Journal of Risk*, *1*(1), 5–19. https://doi.org/10.21314/jor.1998.001

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, *81*(348), 158–171. https://doi.org/10.1002/qj.49708134804

Kuester, K., Mittnik, S., & Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, *4*(1), 53–89. https://doi.org/10.1093/jjfinec/nbj002

Kumari, S. N., & Tan, A. (2013). Characterization of Student ' s T-Distribution with Some Application to Finance. *Mathematical Theory and Modeling*, *3*(10), 1–10. https://www.researchgate.net/publication/259632093

Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives*, *3*(2), 73–84. https://doi.org/10.3905/jod.1995.407942

Lazar, E., Pan, J., & Wang, S. (2022). *On the estimation of Value-at-Risk and Expected Shortfall at extreme levels*.

Li, M. Y. L., & Lin, H. W. W. (2004). Estimating value-at-risk via Markov switching ARCH models - An empirical study on stock index returns. *Applied Economics Letters*, *11*(11), 679–691. https://doi.org/10.1080/1350485042000236539

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. In *Biometrika* (Vol. 65, Issue 2, pp. 297–303). https://doi.org/10.1093/biomet/65.2.297

Lopez, J. A. (1998). Methods for Evatualing Value-at-Risk Estimates. *Federal Reserve Bank of New York*.

Lu, C., Wu, S. C., & Ho, L. C. (2009). Applying VaR to REITs: A comparison of alternative methods. *Review of Financial Economics*, *18*(2), 97–102. https://doi.org/10.1016/j.rfe.2008.03.001

Marimoutou, V., Raggad, B., & Trabelsi, A. (2009). Extreme Value Theory and Value at Risk: Application to oil market. *Energy Economics*, *31*(4), 519–530. https://doi.org/10.1016/j.eneco.2009.02.005

Markowitz, H. M. (1952). Portfolio Selection. *The Theory and Practice of Investment Management: Asset Allocation, Valuation, Portfolio Construction, and Strategies, Second Edition*, *60*, 45–78. https://doi.org/10.1002/9781118267028.ch3

Moody's Analytics, I. (2011). *Regulation Guide : An Introduction Banking regulations : An Introductory Framework*.

Morgan, J. . (1996). RiskMetricsTM—Technical Document. *Trust Company of New York: New*, 2–296. http://down.cenet.org.cn/upfile/10/2005516145842129.pdf

Nelson, D. B. (1991). *Conditional Heteroskedasticity in Asset Returns: A New Approach* (pp. 347–370).

Ñíguez, T. M. (2008). Volatility and VaR forecasting in the Madrid Stock Exchange. *Spanish Economic Review*, *10*(3), 169–196. https://doi.org/10.1007/s10108-007-9030-6

Nozari, M., Raei, S. M., Jahangiri, P., & Bahramgiri, M. (2010). A comparison of heavy-tailed VaR estimates and Filtered Historical Simulation : Evidence from emerging markets. *International Review of Business Research Papers*, *6*(4), 347–359.

Patitsas, E., Berlin, J., Craig, M., & Easterbrook, S. (2020). Evidence that computer science grades are not Bimodal. *Communications of the ACM*, *63*(1), 91–98. https://doi.org/10.1145/3372161

Pederzoli, C. (2006). Stochastic volatility and GARCH: A comparison based on UK stock data. *European Journal of Finance*, *12*(1), 41–59. https://doi.org/10.1080/13518470500039121

Pérignon, C., & Smith, D. R. (2009). The level and quality of Value-at-Risk disclosure by commercial banks. *Journal of Banking and Finance*, *34*(2), 362–377. https://doi.org/10.1016/j.jbankfin.2009.08.009

Philippe, J. (2001). *Value at Risk: The New Benchmark for Managing Financial Risk* (Vol. 53, Issue 9).

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, *3*(1). https://doi.org/10.1214/aos/1176343003

Pritsker, M. (2006). The hidden dangers of historical simulation. *Journal of Banking and Finance*, *30*(2), 561–582. https://doi.org/10.1016/j.jbankfin.2005.04.013

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reeves, J. J., Marimoutou, V., Raggad, B., Trabelsi, A., Bekiros, S. D., Georgoutsos, D. A., & Byström, H. N. E. (2005). Estimation of Value-at-Risk by extreme value and conventional methods: A comparative evaluation of their predictive performance. *International Journal of Forecasting*, *15*(2), 519–530. https://doi.org/10.1016/j.intfin.2004.05.002

Roy, A. D. (1952). Econometrics. *Handbook of Computational Statistics: Concepts and Methods: Second Edition*, *20*(3), 1061–1094. https://doi.org/10.1007/978-3-642-21551-3__35

Sašžiković, S., & Aktan, B. (2011). Decay factor optimisation in time weighted simulation - evaluating var performance. *International Journal of Forecasting*, *27*(4), 1147–1159.

https://doi.org/10.1016/j.ijforecast.2010.09.007

Scharnagl, B., Iden, S., Durner, W., Vereecken, H., & Herbst, M. (2015). Inverse modelling of in situ soil water dynamics: accounting for heteroscedastic, autocorrelated, and non-Gaussian distributed residuals. In *Hydrology and Earth System Sciences Discussions* (Vol. 12). https://doi.org/10.5194/hessd-12-2155-2015

Shaik, M., & Padmakumari, L. (2022). Value-at-risk (VaR) estimation and backtesting during COVID-19: Empirical analysis based on BRICS and US stock markets. *Investment Management and Financial Innovations*, *19*(1), 51–63. https://doi.org/10.21511/imfi.19(1).2022.04

Sharma, M. (2012). Evaluation of Basel III revision of quantitative standards for implementation of internal models for market risk. *IIMB Management Review*, *24*(4), 234–244. https://doi.org/10.1016/j.iimb.2012.09.001

Shephard, N. (2007). *Stochastic Volatility*. *2005*, 149–164. https://doi.org/10.1201/9781584889694.ch16

Su, E., & Knowles, T. W. (2006). Asian Pacific stock market volatility modeling and value at risk analysis. *Emerging Markets Finance and Trade*, *42*(2), 18–62. https://doi.org/10.2753/REE1540-496X420202

Tolikas, K., Koulakiotis, A., & Brown, R. A. (2007). Extreme risk and value-at-risk in the German stock market. *European Journal of Finance*, *13*(4), 373–395. https://doi.org/10.1080/13518470600763737

Yap, B. W., & Sim, C. H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, *81*(12), 2141–2155. https://doi.org/10.1080/00949655.2010.520163

Žiković, S., & Aktan, B. (2009). Global financial crisis and VaR performance in emerging markets: A case of EU candidate states - Turkey and Croatia. *Zbornik Radova Ekonomskog Fakultet Au Rijeci*, *27*(1), 149–170.

8. Appendix

Table 22

Overview of research papers that compare different VaR methodologies

Research Papers	**Parametric**				**Non-Parametric**		**Semi-Parametric**				
	ARCH	*EWMA*	*GARCH*	*SV*	*HS*	*BRW*	*VWHS*	*FHS*	*CAViaR*	*EVT*	*MCS*
(Abad & Benito, 2013)		X	X		X					X	X
(Abad et al., 2014)	X	X	X	X	X	X	X	X	X	X	X
(Adesi et al., 2000)								X			
(Angelidis et al., 2004)			X								
(Angelidis et al., 2006)	X		X					X		X	
(Baillie et al., 1996)	X		X								
(Bao et al., 2006)		X			X			X	X	X	X
(Barone-Adesi & Giannopoulos, 2001					X			X			
(Barone-Adesi et al., 1998)								X			
(Barone-Adesi et al., 1999)								X			
(Basu, 2011)					X	X	X				X
(Bekiros & Georgoutsos, 2005)	X	X	X		X					X	X
(Bollerslev, 1986)			X								
(Bollerslev, 1987)			X								
(Boudoukh et al., 1998)		X			X	X		X			
(Brooks, 2008)											
(Chen et al., 2011)		X	X							X	X
(Danielsson & Vries, 2000)		X	X		X					X	
(Ener et al., 2012)		X	X		X			X	X	X	X
(Gençay et al., 2003)		X	X		X					X	
(Gilli & Këllezi, 2006)										X	
(Giot & Laurent, 2003)	X	X									
(Guermat & Harris, 2002)		X	X								
(Huang, 2009)			X		X					X	X
(Hung et al., 2008)			X								
(J. Hull & White, 1998)						X	X				
(Kuester et al., 2006)		X	X		X			X	X	X	
(Li & Lin, 2004)	X		X								
(Linsmeier & Pearson, 2000)					X						X
(Lu et al., 2009)	X	X			X						
(Marimoutou et al., 2009)		X	X		X			X		X	
(Morgan, 1996)		X									
(Ñíguez, 2008)	X	X	X								
(Nozari et al., 2010)								X		X	
(Pederzoli, 2006)	X		X	X							
(Pérignon & Smith, 2009)					X						
(Pritsker, 2006)					X	X		X			
(R. Engle, 2001)	X		X								
(R. F. Engle & Manganelli, 2004)									X		
(R. F. Engle, 1982)	X										
(Reeves et al., 2005)					X					X	
(Sašžiković & Aktan, 2011)	X	X	X		X	X				X	
(Su & Knowles, 2006)		X	X								
(Tolikas et al., 2007)										X	
(Yu et al., 2010)			X						X		

Notes: The GARCH column accounts for both the classical GARCH process and other variants of that same process (e.g., EGARCH, GJR GARCH).

Table 23

Estimated parameters for each parametric model (Dow Jones Industrial Average)

Model	***DJI 30***															
	μ		***AR1***		***ω***		***α***		***β***		***γ***		***ε***		***υ***	
ARCH-N	0.015	(0.0000)	-0.725	(0.0003)	0.000	-	0.885	(0.0004)	-	-	-	-	-	-	-	-
ARCH-STD	0.001	(0.0001)	-0.050	(0.0159)	0.000	-	0.426	(0.0538)	-	-	-	-	-	-	3.154	(0.1686)
ARCH-SSTD	0.000	(0.0001)	-0.057	(0.0159)	0.000	-	0.413	(0.0515)	-	-	-	-	0.933	(0.0162)	3.188	(0.1711)
ARCH-SGED	0.000	(0.0000)	-0.07	(0.0000)	0.000	-	0.050	(0.0000)	-	-	-	-	1.000	(0.0000)	2.000	(0.0000)
EWMA – N	0.001	(0.0001)	-0.050	(0.0159)	0.000	-	0.072	(0.0051)	0.927	(0.0051)	-	-	-	-	-	-
EWMA-STD	0.001	(0.0001)	-0.051	(0.0141)	0.000	-	0.076	(0.0070)	0.923	(0.0069)	-	-	-	-	7.018	(0.5372)
EWMA-SSTD	0.001	(0.0001)	-0.061	(0.0141)	0.000	-	0.078	(0.0070)	0.921	(0.0068)	-	-	0.909	(0.0160)	7.374	(0.6043)
EWMA-SGED	0.000	(0.0001)	-0.064	(0.0142)	0.000	-	0.076	(0.0069)	0.923	(0.0067)	-	-	0.910	(0.0145)	1.334	(0.0339)
GARCH-N	0.001	(0.0001)	-0.048	(0.0153)	0.000	(0.0000)	0.117	(0.0113)	0.868	(0.0116)	-	-	-	-	-	-
GARCH-STD	0.001	(0.0001)	-0.051	(0.0143)	0.000	(0.0000)	0.111	(0.0205)	0.884	(0.0187)	-	-	-	-	6.369	(0.3593)
GARCH-SSTD	0.001	(0.0001)	-0.061	(0.0143)	0.000	(0.0000)	0.108	(0.0207)	0.885	(0.0194)	-	-	0.906	(0.0164)	6.837	(0.3723)
GARCH-SGED	0.000	(0.0001)	-0.063	(0.0138)	0.000	(0.0000)	0.109	(0.0419)	0.881	(0.0420)	-	-	0.910	(0.0173)	1.346	(0.0656)
GJR GARCH-N	0.000	(0.0001)	-0.048	(0.0149)	0.000	(0.0000)	0.000	(0.0018)	0.885	(0.0059)	0.187	(0.0117)	-	-	-	-
GJR GARCH-STD	0.000	(0.0001)	-0.049	(0.0142)	0.000	(0.0000)	0.000	(0.0092)	0.892	(0.0117)	0.185	(0.0261)	-	-	7.264	(0.7596)
GJR GARCH-SSTD	0.000	(0.0001)	-0.056	(0.0144)	0.000	(0.0000)	0.000	(0.0164)	0.892	(0.0111)	0.185	(0.0407)	0.883	(0.0171)	7.728	(0.8696)
GJR GARCH-SGED	0.000	(0.0001)	-0.056	(0.0141)	0.000	(0.0000)	0.000	(0.0015)	0.890	(0.0015)	0.184	(0.0113)	0.889	(0.0164)	1.416	(0.0403)
EGARCH-N	0.000	(0.0001)	-0.039	(0.0167)	-0.271	(0.0014)	-0.142	(0.0048)	0.971	(0.0002)	0.158	(0.0087)	-	-	-	-
EGARCH-STD	0.000	(0.0001)	-0.043	(0.0132)	-0.204	(0.0025)	-0.142	(0.0078)	0.979	(0.0001)	0.149	(0.0079)	-	-	7.122	(0.7086)
EGARCH-SSTD	0.000	(0.0001)	-0.050	(0.0133)	-0.220	(0.0029)	-0.142	(0.0076)	0.977	(0.0001)	0.149	(0.0074)	0.884	(0.0172)	7.588	(0.7903)
EGARCH-SGED	0.000	(0.0001)	-0.051	(0.0111)	-0.245	(0.0065)	-0.141	(0.0087)	0.974	(0.0007)	0.152	(0.0021)	0.890	(0.0151)	1.415	(0.0304)

Notes: The standard deviation for each model's parameters is presented inside parentheses (round brackets).

Table 24

Estimated parameters for each parametric model (S&P 500)

Model	***S&P 500***															
	μ		***AR1***		***ω***		***α***		***β***		***γ***		***ε***		***υ***	
ARCH-N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ARCH-STD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ARCH-SSTD	0.000	(0.0001)	-0.062	(0.0156)	0.000	(0.0000)	0.437	(0.0574)	-	-	-	-	0.929	(0.016)	3.011	(0.1564)
ARCH-SGED	0.000	(0.0000)	-0.074	(0.0000)	0.000	(0.0000)	0.05	(0.0000)	-	-	-	-	1.000	(0.0000)	2.000	(0.0000)
EWMA – N	0.000	(0.0001)	-0.059	(0.0150)	0.000	-	0.066	(0.0047)	0.933	(0.0047)	-	-	-	-	-	-
EWMA-STD	0.001	(0.0001)	-0.059	(0.0140)	0.000	-	0.075	(0.0070)	0.924	(0.0069)	-	-	-	-	6.868	(0.5171)
EWMA-SSTD	0.001	(0.0001)	-0.072	(0.0140)	0.000	-	0.078	(0.0070)	0.921	(0.0068)	-	-	0.897	(0.0157)	7.264	(0.5899)
EWMA-SGED	0.000	(0.0001)	-0.075	(0.0137)	0.000	-	0.075	(0.0102)	0.924	(0.0108)	-	-	0.897	(0.0135)	1.317	(0.0419)
GARCH-N	0.001	(0.0001)	-0.058	(0.0153)	0.000	(0.0000)	0.114	(0.0105)	0.870	(0.0109)	-	-	-	-	-	-
GARCH-STD	0.001	(0.0001)	-0.059	(0.0142)	0.000	(0.0000)	0.111	(0.0189)	0.887	(0.0168)	-	-	-	-	6.164	(0.3419)
GARCH-SSTD	0.001	(0.0001)	-0.071	(0.0142)	0.000	(0.0000)	0.109	(0.0202)	0.887	(0.0185)	-	-	0.893	(0.0162)	6.706	(0.3419)
GARCH-SGED	0.001	(0.0001)	-0.074	(0.0136)	0.000	(0.0000)	0.110	(0.0418)	0.881	(0.0416)	-	-	0.896	(0.0170)	1.332	(0.0639)
GJR GARCH-N	0.000	(0.0001)	-0.057	(0.0150)	0.000	(0.0000)	0.000	(0.0024)	0.883	(0.0063)	0.190	(0.0116)	-	-	-	-
GJR GARCH-STD	0.000	(0.0001)	-0.055	(0.0141)	0.000	(0.0000)	0.000	(0.0101)	0.889	(0.0154)	0.193	(0.0258)	-	-	7.095	(0.7627)
GJR GARCH-SSTD	0.000	(0.0001)	-0.065	(0.0142)	0.000	(0.0000)	0.000	(0.0093)	0.888	(0.0114)	0.197	(0.0242)	0.864	(0.0167)	7.748	(0.8629)
GJR GARCH-SGED	0.000	(0.0001)	-0.067	(0.0135)	0.000	(0.0000)	0.000	(0.0061)	0.885	(0.0080)	0.197	(0.0091)	0.869	(0.0150)	1.409	(0.0405)
EGARCH-N	0.000	(0.0001)	-0.054	(0.0142)	-0.257	(0.0017)	-0.154	(0.0076)	0.9724	(0.0002)	0.145	(0.0056)	-	-	-	-
EGARCH-STD	0.000	(0.0001)	-0.052	(0.0137)	-0.185	(0.005)	-0.156	(0.0100)	0.981	(0.0005)	0.136	(0.0132)			6.946	(0.7090)
EGARCH-SSTD	0.000	(0.0004)	-0.062	(0.0168)	-0.209	(0.0569)	-0.159	(0.0393)	0.978	(0.0071)	0.138	(0.1711)	0.864	(0.0295)	7.576	(4.0658)
EGARCH-SGED	0.000	(0.0001)	-0.064	(0.0136)	-0.237	(0.0028)	-0.157	(0.0077)	0.975	(0.0000)	0.142	(0.008)	0.869	(0.0154)	1.408	(0.0380)

Notes: The standard deviation for each model's parameters is presented inside parentheses (round brackets).

Table 25

Estimated parameters for each parametric model (NASDAQ)

Model	NASDAQ															
	μ		AR1		ω		α		β		γ		ε		υ	
ARCH-N	0.012	(0.0000)	-0.949	(0.0004)	0.000	(0.0000)	0.939	(0.0036)	-	-	-	-	-	-	-	-
ARCH-STD	0.001	(0.0001)	-0.011	(0.0161)	0.000	(0.0000)	0.595	(0.0699)	-	-	-	-	-	-	3.047	(0.1606)
ARCH-SSTD	0.001	(0.0002)	-0.023	(0.0162)	0.000	(0.0000)	0.559	(0.0638)	-	-	-	-	0.922	(0.0162)	3.104	(0.1640)
ARCH-SGED	0.000	(0.0000)	-0.033	(0.0162)	0.000	(0.0000)	0.050	(0.0000)	-	-	-	-	1.000	(0.0000)	2.000	(0.0001)
EWMA – N	0.001	(0.0001)	-0.026	(0.0147)	0.000	-	0.060	(0.0047)	0.939	(0.0047)	-	-	-	-	-	-
EWMA-STD	0.001	(0.0001)	-0.026	(0.0141)	0.000	-	0.069	(0.0067)	0.930	(0.0066)	-	-	-	-	8.298	(0.7451)
EWMA-SSTD	0.001	(0.0002)	-0.045	(0.0150)	0.000	-	0.070	(0.0091)	0.929	(0.0096)	-	-	0.880	0.0151	8.909	(0.9106)
EWMA-SGED	0.001	(0.0001)	-0.050	(0.014)	0.000	-	0.067	(0.0063)	0.932	(0.0062)	-	-	0.876	(0.0145)	1.430	(0.0371)
GARCH-N	0.001	(0.0001)	-0.027	(0.0149)	0.000	(0.0000)	0.097	(0.0105)	0.891	(0.0113)	-	-	-	-	-	-
GARCH-STD	0.001	(0.0001)	-0.026	(0.0143)	0.000	(0.0000)	0.095	(0.0182)	0.902	(0.0174)	-	-	-	-	7.873	(0.607)
GARCH-SSTD	0.001	(0.0001)	-0.044	(0.0144)	0.000	(0.0000)	0.093	(0.0173)	0.903	(0.0169)	-	-	0.878	(0.0165)	8.633	(0.7221)
GARCH-SGED	0.001	(0.0002)	-0.049	(0.0147)	0.000	(0.0000)	0.094	(0.1026)	0.898	(0.1060)			0.876	(0.0306)	1.449	(0.1437)
GJR GARCH-N	0.000	(0.0001)	-0.021	(0.0147)	0.000	(0.0000)	0.005	(0.0027)	0.902	(0.0025)	0.151	(0.0115)	-	-	-	-
GJR GARCH-STD	0.001	(0.0001)	-0.020	(0.0143)	0.000	(0.0000)	0.002	(0.0115)	0.907	(0.0245)	0.156	(0.0313)	-	-	8.791	(1.3143)
GJR GARCH-SSTD	0.000	(0.0001)	-0.035	(0.0144)	0.000	(0.0000)	0.000	(0.0082)	0.909	(0.0170)	0.158	(0.0254)	0.854	(0.017)	9.671	(1.4128)
GJR GARCH-SGED	0.000	(0.0001)	-0.039	(0.0142)	0.000	(0.0000)	0.001	(0.0072)	0.906	(0.0148)	0.159	(0.0269)	0.853	(0.0155)	1.503	(0.0455)
EGARCH-N	0.000	(0.0001)	-0.022	(0.0128)	-0.174	(0.0012)	-0.112	(0.0055)	0.9801	(0.0002)	0.1271	(0.0050)	-	-	-	-
EGARCH-STD	0.001	(0.0001)	-0.020	(0.0144)	-0.120	(0.0020)	-0.120	(0.0071)	0.987	(0.0000)	0.128	(0.0074)	-	-	8.558	(0.9714)
EGARCH-SSTD	0.001	(0.0001)	-0.036	(0.0140)	-0.132	(0.0017)	-0.123	(0.0083)	0.985	(0.0004)	0.125	(0.0108)	0.848	(0.0161)	9.709	(0.8012)
EGARCH-SGED	0.000	(0.0001)	-0.039	(0.0136)	-0.150	(0.0072)	-0.123	(0.0082)	0.983	(0.0009)	0.124	(0.0096)	0.845	(0.0147)	1.506	(0.0394)

Notes: The standard deviation for each model's parameters is presented inside parentheses (round brackets).

Table 26

Heteroskedasticity, autocorrelation tests and goodness of fit statistics for each parametric model

Models	***DJI 30***					***S&P 500***					***NASDAQ***				
	LM	LM^2	$Q_{(200)}$	$Q^2_{(200)}$	***Log Likelihood***	LM	LM^2	$Q_{(200)}$	$Q^2_{(200)}$	***Log Likelihood***	LM	LM^2	$Q_{(200)}$	$Q^2_{(200)}$	***Log Likelihood***
ARCH-N	0.9919	1.0000	**0.0000**	**0.0000**	3229.34	-	-	-	-	-	**0.0004**	0.9997	**0.0000**	1.0000	-13849.17
ARCH-STD	**0.0000**	**0.0000**	**0.0000**	**0.0000**	16205.19	-	-	-	-	-	**0.0000**	**0.0000**	**0.0000**	**0.0000**	14555.80
ARCH-SSTD	**0.0000**	**0.0000**	**0.0000**	**0.0000**	16213.26	**0.0000**	**0.0000**	**0.0000**	**0.0000**	15973.73	**0.0000**	**0.0000**	**0.0000**	**0.0000**	14566.54
ARCH-SGED	**0.0000**	0.9995	**0.0000**	**0.0000**	5204.21	**0.0000**	**0.0000**	**0.0000**	**0.0000**	3517.51	**0.0000**	**0.0000**	**0.0000**	**0.0000**	6870.74
EWMA-N	**0.0487**	1.0000	**0.0198**	**0.0001**	16485.20	**0.0043**	1.0000	**0.0000**	**0.0001**	16260.31	**0.0035**	0.9975	**0.0043**	**0.0000**	15002.46
EWMA-STD	0.1117	1.0000	0.2463	**0.0260**	16644.04	0.0596	1.0000	**0.0002**	**0.0006**	16435.79	0.0745	0.9999	**0.0220**	**0.0000**	15106.78
EWMA-SSTD	0.1524	1.0000	**0.0312**	**0.0260**	16658.81	0.0904	1.0000	**0.0003**	**0.0008**	16455.11	0.0926	0.9999	**0.0295**	**0.0000**	15131.56
EWMA-SGED	0.1171	1.0000	0.2307	**0.0290**	16664.07	0.0521	1.0000	**0.0002**	**0.0005**	16461.36	**0.0476**	0.9997	**0.0199**	**0.0000**	15134.49
GARCH-N	0.5331	0.9997	0.1900	0.2245	16568.38	0.1715	0.9996	**0.0000**	**0.0192**	16349.81	0.0710	0.9485	**0.0032**	**0.2041**	15062.23
GARCH-STD	0.5278	0.9999	0.6621	0.1924	16675.06	0.2446	1.0000	**0.0038**	**0.0157**	16463.60	0.1627	0.9954	0.1572	0.0863	15127.94
GARCH-SSTD	0.5279	0.9999	0.6165	0.1957	16688.72	0.2395	1.0000	**0.0029**	**0.0151**	16481.97	0.1462	0.9937	0.1495	0.0715	15150.59
GARCH-SGED	0.5393	0.9999	0.4895	0.2254	16699.83	0.2270	1.0000	**0.0010**	**0.0203**	16495.69	0.1236	0.9860	0.0894	0.1351	15160.21
GJR GARCH-N	0.7233	0.9998	0.0781	0.5868	16676.32	0.2047	0.9999	**0.0000**	0.0596	16464.79	0.0833	0.9949	**0.0000**	**0.2804**	15139.81
GJR GARCH-STD	0.6953	0.9999	0.4825	0.6118	16759.49	0.2392	1.0000	**0.0009**	0.0861	16552.07	0.1809	0.9982	**0.0214**	0.5097	15190.06
GJR GARCH-SSTD	0.7192	0.9999	0.4423	0.6365	16780.41	0.2016	1.0000	**0.0006**	0.0879	16581.15	0.1819	0.9985	**0.0235**	0.4973	15222.60
GJR GARCH-SGED	0.7265	0.9999	0.3001	0.6439	16786.44	0.1705	1.0000	**0.0001**	0.0883	16589.68	0.1414	0.9978	**0.0038**	0.4904	15230.61
EGARCH-N	0.5100	0.9861	0.0157	0.1790	16681.06	0.1103	0.9998	**0.0000**	**0.0045**	16477.08	0.0867	1.0000	**0.0000**	**0.0470**	15142.93
EGARCH-STD	0.2936	0.8172	0.3658	0.1351	16767.22	0.1425	0.9985	**0.0012**	**0.0106**	16569.71	0.4149	1.0000	0.1152	0.0589	15202.16
EGARCH-SSTD	0.2970	0.8256	0.2801	0.1532	16787.67	0.1127	0.9988	**0.0005**	**0.0105**	16598.66	0.3376	1.0000	0.0736	**0.0421**	15237.13
EGARCH-SGED	0.4316	0.9288	0.1312	0.2042	16791.38	0.1410	0.9998	**0.0000**	**0.0122**	16604.21	0.2297	1.0000	0.0099	0.0707	15241.80

Notes: *LM* stands for the Lagrange Multiplier heteroskedasticity test applied for the model's residuals. LM^2 denotes the Lagrange Multiplier heteroskedasticity test applied to the model's squared residuals. $Q_{(200)}$ is the Ljung-Box autocorrelation test applied to the model's residuals, using 200 lag observations. $Q^2_{(200)}$ corresponds to the Ljung-Box autocorrelation test applied to the model's squares residuals using 200 lag observations.

Figure 15

Graphical representation of the backesting results for the EVT-POT VaR forecasts

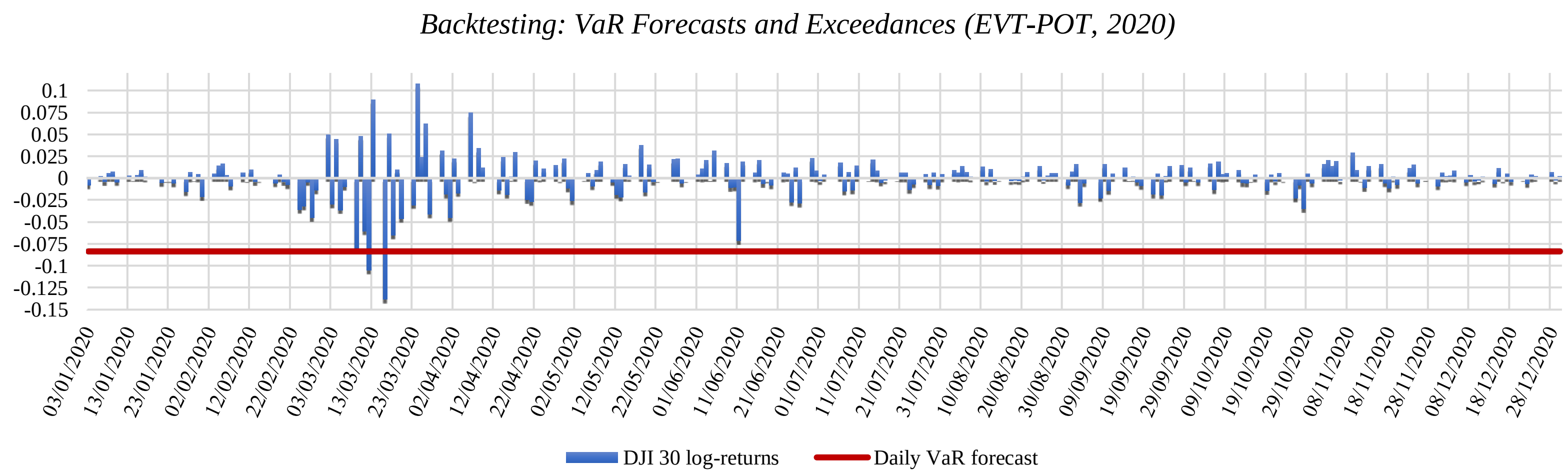


Figure 16

Graphical representation of the backesting results for the FHS VaR forecasts

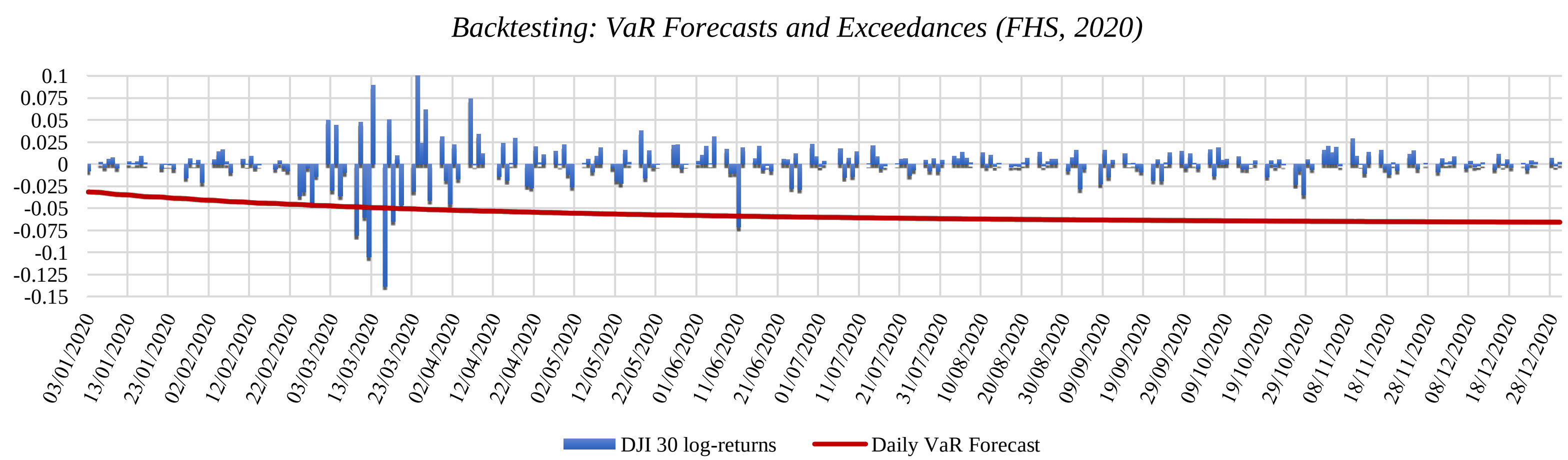


Figure 17

Graphical representation of the backesting results for the GJR GARCH-SGED VaR forecasts