PressAcademia
www.pressacademia.org/journals/jefa

JEFA
ISSN 2148-6697

Journal of Economics, Finance and Accounting YEAR 2023 VOLUME 10 ISSUE 3

COMPARISON OF THE ACCURACY OF MODELS IN FORECASTING VAR AND ES THROUGH TIME

DOI: 10.17261/Pressacademia.2023.1814
JEFA- V.10-ISS.3-2023(4)-p.158-169

Sukriye Tuysuz
Yeditepe University, International Finance Department, Istanbul, Turkiye.
sukriye.tuysuz@yeditepe.edu.tr, ORCID: 0000-0001-8391-6521

Date Received: May 7, 2023 **Date Accepted:** September 12, 2023

To cite this document
Tuysuz, S., (2023). Comparison of the accuracy of models in forecasting VAR and ES Through time. Journal of Economics, Finance and Accounting (JEFA), 10(1), 158-169.
Permanent link to this document: http://doi.org/10.17261/Pressacademia.2023.1814

ABSTRACT

Purpose- Identify the best model/method to accurately forecast the Value-at-Risk (VaR) and the Expected Shortfall (ES) of position.
Methodology- The dynamic of each retained return series was estimated with one of retained GARCH-type model combined with one of retained probability distributions (normal, fat-tailed, and skewed) in each retained sub-periods (window). In each window (sub-period), the 1-day ahead VaR and ES were forecasted by using the best selected GARCH-type model. More than 4000 1-day ahead VaR and ES were forecasted with each retained model/method. Based on 252-day rolling-window, forecasted VaR and ES with each retained model/method were backtested around 3750 times.
Findings- Our results revealed that the best fitting GARCH-specifications combined with skewed Student or GED distribution enable to accurately forecast VaR more often. However, the best methods based on the best fitting GARCH-specifications combined with the best fitting probability distribution do not improve the frequency of acceptance of the null hypothesis stating the accuracy of the method. The accuracy of models tends to deteriorate during crises periods.
Conclusion- Modeling and forecasting the dynamic of retained series with skewed probability distributions (skwed student or wked GED) improve the forecasting accuracy of a parametric or semi parametric model. A performan model in sample may not perform well o ut sample. Forecasted VaR should be complemented with Stressed VaR or ES.

Keywords: VaR, ES, parametric, semi-parametric, backtesting, probability distribution, rolling-window.
JEL Codes: G11, C10, C51

1. INTRODUCTION

It is very important for a financial institution to determine (forecast) the risks it faces. In the specific case of market risk, a possible method of measurement is the evaluation of losses likely to be incurred when the price of portfolio assets falls. The important growth of trading activity and the trading loss of well-known financial institutions during the first part of the 90s have led financial regulators and supervisory committee of banks to prescribe rules and methods for quantifying market risk and determining the regulatory capital necessary to cover this risk.

The most well-known method often used to determine market risk is the Value-at-Risk (VaR) measure. Initially proposed by Till Guldimann, in the late 1980s1, the VaR method gained importance since the Basel committee required the calculation of regulatory capital for market risk based on VaR. The Basel committee proposed a standard approach as well as an internal method in determining market risk. In Basel 2 and 3, the proposed internal method is based on the VaR at 99% over 10 days. The VaR enables to determine the worst loss of a position or portfolio over a given period for a given confidence level.

The VaR is also used to determine the trading limit, the risk-adjusted performance, the optimum portfolio as well as the optimum hedge ratio.

According to Artzner et al. (1997, 1999) a coherent risk measure should verify some axioms: 1) Translation invariance, 2) Subadditivity, 3) Positivity, 4) Homogeneity, 5) Relevance, and 6) Conservatism. These axioms are not verified by the VaR. Artzner et al. (1999) proposed a new measure of risk satisfying these axioms: the Expected Shortfall (ES) (also called conditional value at risk (CVaR), average value at risk (AVaR), and expected tail loss (ETL)). The ES measures the average loss in case the VaR is violated. The last regulation on market risk prescribed by the Basel Committee (Fundamental Review of Banking Book, FRTB) consists of determining market risk and then the required regulatory capital by using the ES and the VaR as well as the profit and loss attribution (P/L) as a mean of backtesting.

Since the 1996 Basel Amendment, several models/methods for forecasting the VaR have been proposed. These models are parametric, semi-parametric or non-parametric. Most existing studies based on parametric models in forecasted VaR and ES showed higher performance of asymmetric GARCH-type specifications (such as gjrGARCH, APARCH, TGARCH) as well as of skewed probability distributions (such as skewed Student-t, skewed GED...). Furthermore, it was also shown that the choice of the volatility model appears to be less relevant than the choice of the probability distribution. These findings were also found in studies based on semi-parametric methods using GARCH-type specification and probability distribution, such as Filtered Historical simulation, and conditional extreme value methods.

The purpose of this article is to compare the performance of some parametric and semi-parametric models in forecasting the VaR and ES of a single position in three types of assets (SP500, EURUSD and Gold) over a long period ranging from January 2000 to June 2018 covering stable and unstable economic and financial periods. The retained parametric methods are the AR-GARCH-type models. Four semi-parametric methods were also used: the AR-GARCH model combined with Monte-Carlo (AR-GARCH-MC), the Filtered Historical Simulation method (FHS), the FHS method combined with Bootstrapping (FHS-B) and the conditional EVT method (SEVT). For each retained return series, the 1-day ahead VaR and ES were forecasted by using a 500 days rolling window method. More than 4000 windows were considered and more than 4000 1-day ahead VaR and ES were forecasted for each retained return and for each model/method.

Existing studies using the rolling-window method retained the same GARCH-type specification and the same probability distribution in each window. The performance of a model may change over time. We accounted for this fact by using two approaches: Approach 1: In each window the dynamic of each retained series was evaluated with the best fitting GARCH specification for each retained probability distribution and the 1-day ahead VaR and ES were forecasted. Approach 2: For each series and each estimation window, the best fitting model was selected among all GARCH specifications combined with all retained probability distributions and the 1-day ahead VaR and ES were forecasted.

The performance of models in forecasting the VaR and ES should be checked. Most existing studies compared the performance of some models in forecasting the VaR/ES. However, in a large part of these studies the VaR and ES were forecasted over one period or a few periods. Furthermore, some did not backtest these forecasts and few backtested them only one time by using different lengths of forecasted VaR. The BCBS recommends backtesting a model by using 1 year (252 days) forecasted VaR. Furthermore, the performance of a model can change through time. A model can perform well in forecasting the VaR over a certain period and perform badly over another period. Owing that, our forecasted VaR were backtested more than 3750 times by using 252-days rolling-window approach with the conditional coverage (CC) test of Christoffersen (1998) and the Dynamic Quantile (DQ) test of Engle and Managanelli (2004). In the same way, the forecasted ESs were backtested by using the residual exceedance over the VaR test proposed by McNeil and Frey (2000). By forecasting and backtesting the VaR and ES over a long period, it is possible to check the performance of retained models during stable as well as unstable periods.

The methods used in forecasting the VaR and ES are presented in the second part of this paper. Backtesting methods are presented in section 3. A presentation and descriptive analysis of retained return series are done in the fourth section. This section presents and discusses also results obtained.

2. REVIEW OF MODELS

Value-at-Risk (VaR) is a quantitative tool used to measure the maximum potential loss in value (or in percentage) of a portfolio of assets over a defined period at a given level. Specifically, VaR construction requires a quantile estimate of the left tail (right tail) of the return's distribution (or profit/loss distribution) of a long position (or short position). The h-period ahead VaR and ES are determined as follows:

$$VaR_{t+h}(100-\alpha) = G^{-1}(\alpha) \quad (1)$$

$$VaR_{t+h}(100-\alpha) = \hat{\mu}_{t+h} + \hat{\sigma}_{t+h}F^{-1}(\alpha). \quad (2)$$

$$ES_{t+h}(\alpha)\frac{1}{\alpha}\int_0^{\alpha} VaR_{t+h}(100-\alpha). \quad (3)$$

Forecasting the h-period ahead VaR and ES at level α can be done by using parametric methods (eq. 2-3), non-parametric methods (eq. 1), and semi-parametric methods. The parametric method requires the specification of the cumulative probability distribution F as well as the h-period expected mean return $\hat{\mu}_{t+h}$ and the h-period expected volatility $\hat{\sigma}_{t+h}$. As for the non-parametric method, it is based on the empirical probability distribution G.

2.1. Parametric Methods

Parametric methods consist of modelling and estimating the dynamic of risk factors by using the appropriate volatility model and the appropriate probability distribution (F). Based on these estimations and on the relation between these risk factors and the considered financial instrument prices (or returns), profits and losses in value (or in return) of the position is determined and then the VaR and ES can be forecasted (equations 2-3).

Initially, RiskMetrics proposed the delta-normal model, consisting of forecasting the VaR by estimating the standard deviation of the portfolio or position by means of the sample estimate of variance appropriately augmented by the quantile of a normal distribution. In this approach, the conditional mean and volatility are modelled with an AutoRegressive and Exponentially Weighted Moving Average (AR-EWMA) model. Since the 1980s several models belonging mainly to the AR/ARMA - GARCH family were proposed in modelling the volatility. These GARCH-type models can be classified as classical, asymmetric and non-linear GARCH models. Empirical studies found that the EWMA model performs the worst in forecasting VaR compared to GARCH-type models (Gonzalez-Rivera et al., 2004; Huang and Lin, 2004; Niguez, 2008; Chen et al., 2011; Abad et al. 2014). Most of the existing empirical studies modelled the dynamic of financial securities and forecasted the VaR by using GARCH-type models (Mittnik and Paolella, 2000; Giot and Laurent, 2004; Haas et al., 2004; Angelidis et al., 2007; Bali and Theodossiou, 2007). Although there is no evidence of an overpowering model, the results obtained in these studies seem to indicate that asymmetric GARCH models produce better outcomes.

Regarding the probability distribution F, initially it was specified as a normal distribution. However, in the literature it is widely recognized that return distributions of financial assets are skewed and fat tailed. The excess kurtosis can be taken into account by probability distributions, such as t-Student (std) and GED (ged). As for the skewness, it is accounted by skewed probability distributions. Different probability distributions were proposed to account for both features (excess kurtosis and skewness); such as skewed t-Student (sstd); skewed GED (sged).

In the empirical literature different GARCH-type models with different distributions were retained in existing empirical studies (Giot and Laurent, 2003; Bali and Theodossiou, 2007; Bali et al., 2008). Most existing studies forecasting the VaR and ES by using different GARCH type models combined with classical and skewed probability distributions showed the higher performance of asymmetric GARCH models (such as gjrGARCH, APARCH, TGARCH) as well as of skewed probability distributions (such as skewed Student-t, skewed GED...). Moreover, it was also shown that the choice of the volatility model appears to be less relevant than the choice of the probability distribution (Giot and Laurent, 2003; Angelidis et al., 2004).

2.2. Non-Parametric and Semi-Parametric Methods

The Historical Simulation (HS) method is the simplest non-parametric approach. This method consists of approximating G (equation 1) with the empirical distribution of financial returns. The forecasted VaR corresponds to the α quantile of this empirical distribution. Although this method is very simple, it presents several drawbacks due mainly to the size of the considered window and the main assumption on which is it based; returns are iid (Down, 2002). Some solutions to these drawbacks were proposed. The main important solutions are: 1) forecasting the VaR and ES with the HS method by using the standardized return (Filtering returns) instead of raw returns (Filtered Historical Simulation (FHS)), and 2) forecasting the VaR and ES by applying the HS method to bootstrapped raw returns or standardized returns (HS-B and FHS-B).

2.2.1. Filtered Historical Simulation (FHS)

The Filtered Historical Simulation (FHS) method consists of standardizing returns ($r_{t,s} = (r_t - \mu_t)/\sigma_t$), where μ_t and σ_t are the mean and the standard deviation) and then forecasting the VaR as the α quantile of the empirical distribution of those standardized returns G($r_{t,s}$). FHS method requires the determination of the mean and variance (μ_t and σ_t) of the raw data in the first stage in order to filter those data. Unconditional or conditional mean and variance can be used. The unconditional method consists of determining the mean and variance of the raw data over the retained window. Regarding the conditional method, it consists of determining the mean and variance of the raw data by choosing first the appropriate model describing the best the dynamic of these data. Empirical studies employing this method modelled the conditional volatility with AR-GARCH-type models. These researchers considered different probability distributions, such as normal distribution, student distribution, skewed distributions, ... The importance of the choice of the probability distribution was empirically shown (Adcock et al., 2012). These authors showed that FHS models with skewed distributions produce more accurate VaR forecasts.

Another drawback of the HS approach is that the estimation of a quantile in the tails, as in the HS and FHS method, can be subject to important variance, mainly in case of reduced number of used observations in this estimation. In case of a sample composed with one-year daily observations, the quantile at confidence level 99% (corresponding to the forecasted VaR at 99%) will be the second or the third largest losses. This forecasted VaR depends on the realizations of the risk factors rather than on their probability distribution. Given this fact, the forecasted VaR with the quantile of the empirical distribution can be highly unstable, especially when the VaR is forecasted with a high confidence level by using only few available data. To overcome this instability problem bootstrapping can be combined with HS and with FHS.

2.2.2. Extreme Value Theory (EVT)

Another widely used semi-parametric method in forecasting the VaR and ES is the methods based on Extreme Value Theory (EVT). These methods deal with the study of extreme events, which are characterized by extreme deviations from the normal median of their probability distributions. More precisely, the EVT studies and models the behaviour of distributions in their extreme tails, therefore, could potentially give better estimates and forecasts on risk.

Among the methods based on EVT, the exceedances or peaks over threshold (POT) model is the most used in forecasting the VaR and the ES (Smith, 1989; Embrechts et al., 1997; McNeil and Frey, 2000; McNeil et al., 2005; Gilli and Këllezi, 2006). This method concentrates on returns (r_i) in the series that exceed a certain high threshold (u) and model these returns separately from the rest of the unknown distribution. Balkema and De Haan (1974) and Pickands (1975) showed that the generalized Pareto distribution (GPD) is the limiting distribution of the probability distribution of returns exceeding the threshold u by at most an amount y given that r exceeds the threshold u. Given this fact (this probability distribution) and the expression of the conditional probability distribution of returns exceeding the threshold u, the VaR and ES of a long position are determined as:

$$VaR_{t+h}(100-\alpha) = u + \frac{\beta}{\xi}\left(\left(\frac{n}{n_u}\alpha\right)^{-\xi} - 1\right), \tag{3}$$

$$ES_{t+h}(\alpha) = E(E|X > VaR_{t+h}(100-\alpha) = \frac{\widehat{VaR}_{\alpha}}{1-\hat{\xi}} + \frac{\hat{\beta}-\hat{\xi}_u}{1-\hat{\xi}}, \tag{4}$$

where ξ and β represent the shape parameter of the distribution and the scaling parameter, respectively. n and nu denote the total number of observations and the number of observations exceeding the threshold u. The choice of the threshold u is important for the performance of the GPD modelling. This threshold (u) value can be put directly at the lower 5% and 10% quantile of the distribution for the retained series. The optimal threshold can be determined with the graph of the sample mean excess (MEF), the Hill estimator, the Q − Q graph, the graphical bootstrap method (Embrechts et al., 1997), and the sample percentile approach (DuMouchel, 1983).

This presented GPD approach put emphasis on the tail of the distribution but does not recognize the fact that returns are no-i.i.d. One way to overcome this drawback is to combine the classical GPD approach with time-varying volatility models, such as ARCH and GARCH models, as suggested by McNeil and Frey (2000). These authors proposed to take into account the conditional heteroscedasticity in the data through the GARCH models and model the extreme tail behaviour through the EVT method. This method, first, filters different financial time series with GARCH-type models and then deduces the residuals. In the second step, the POT method is applied to these residuals.

In this conditional POT model, the h-period forecasted conditional VaR and conditional ES at level α are determined as:

$$\widehat{VaR}_{t+h}(100-\alpha) = \hat{\mu}_{t+h} + \sqrt{\hat{\sigma}_{t+h}}\widehat{VaR}(Z)_{t+h}(100-\alpha), \tag{5}$$

$$\widehat{ES}_{t+h}(\alpha) = \hat{\mu}_{t+h} + \sqrt{\hat{\sigma}_{t+h}}\widehat{ES}(Z)_{t+h}(\alpha), \tag{6}$$

where $\hat{\mu}_{t+h}$ and $\hat{\sigma}_{t+h}$ are the h-step forecasts for the conditional mean and variance, respectively. Regarding $\widehat{VaR}(Z)_{t+h}(100-\alpha)$ and $\widehat{ES}(Z)_{t+h}(\alpha)$, they are determined with POT model applied to standardized residuals (see equations 4 and 5).

Several authors forecasted the VaR and ES with EVT and conditional EVT models. Most of these authors showed that conditional EVT models produce more accurate VaR and ES forecasts. It was also revealed that the choice of the probability distribution in the conditional EVT models was more relevant than the choice of the GARCH-type models. Precisely, existing studies emphasized the higher accuracy of skewed distribution in forecasting the VaR and ES (Bali and Weinbaum, 2007; Bali and Theodossiou, 2007, 2008; Bali et al., 2008).

3. BACKTESTING

It is important to backtest the accuracy of the retained model in forecasting the VaR and ES. Backtesting VaR consists of comparing the forecasted/predicted losses (or returns) to the realized losses (or returns) and then testing the accuracy and effectiveness of the model. According to the Basel agreement, a model is acceptable when the proportion of exceptions/violations (losses greater than the VaR) is lower than α%, where α% denotes the level used to measure the 1-day ahead VaR and ES. Furthermore, exceptions should be independent: an exception today should not depend on whether an exception occurred on the previous day. Results of backtesting methods enable to select accurate models for VaR forecasting but also have an impact on the market risk capital requirement.

In the literature, different backtesting methods were proposed. Reviews of backtesting are provided by Campbell (2005) and Zhang and Nadarajah (2018). Among these methods, the most often used are the conditional coverage test introduced by Christoffersen (1998) and the Dynamic Quantile (DQ) test proposed by Engle and Managanelli (2004).

Based on the proportion of failure (POT) (Kupiec test) and the independence test of Christoffersen (1998), Christoffersen (1998) proposed the conditional coverage test enabling to test the correct failure rate as well as the independence of violations. The null hypothesis states the accuracy of the failure rate and the independence of exceptions. An exception occurs if the forecasted VaR is violated, meaning that the realized return (or losses) is lower than the VaR in return (or in value) (long position).

Engle and Managanelli (2004) proposed the Dynamic Quantile statistic (DQ) enabling to test the null hypothesis that the current VaR exceptions are uncorrelated with past exceptions (null hypothesis).

Although several backtesting methods for VaR were proposed, methods for backtesting ES are rare. One possible reason for this fact is that VaR has often been used more frequently compared to ES. Furthermore, backtesting ES is more difficult than that of the VaR. The earliest and the most used statistic is the residual exceedances over the VaR series r (representing the violations) which was proposed by McNeil and Frey (2000). McNeil and Frey (2000) proposed to test ES by directly using the residual exceedances over the VaR series r or by using standardized residuals (r_s)). McNeil and Frey (2000) proposed to test the null hypothesis stipulating that r (or r_s) has a zero mean, against the alternative that the mean of r (or r_s) is negative. This alternative hypothesis corresponds to the underestimation of the risk. Overestimation of the ES is limited to the difference between the ES and VaR. In this case, there is no violation and r (or r_s) assumes value zero. On the other hand, underestimation is theoretically unbounded.

4. EMPIRICAL ANALYSIS

4.1. Methodology

The purpose of this paper is to forecast the VaR and ES of position in different type of assets by using parametric and semi-parametric models over a long period. The retained assets are EURUSD currency exchange, Gold and SP500 stock index. The retained daily returns cover the period ranging from 4 January 2000 to 26 June 2018. This period covers periods of financial instability (IT crisis (2001-2003), subprime crisis (2007-2009) ...), economic recession, and of stability. By considering a long period it is then possible to check whether the retained models enable to correctly forecast the VaR and ES over different periods. A 500-day rolling window was employed to estimate the retained model and forecast 1-day ahead VaR and ES. Precisely, the first window covered the period ranging from 4 January 2000 to 3 December 2001 and the VaR as well as the ES for 4 December 2001 were forecasted. The last window included daily data from 26 July 2016 to 25 June 2018 and enabled to forecast the VaR and ES for the 26 June 2018. More than 4000 windows were considered and more than 4000 1-day ahead VaR and ES were forecasted for each retained return.

Retained parametric methods are classical AR-GARCH-type models (AR-GARCH-type). The conditional volatility of retained returns' series was modelled with different GARCH-type models: the classical GARCH model, the exponential GARCH model.

(eGARCH, Nelson (1991)), the Threshold GJR-GARCH model (gjrGARCH, Glosten et al. (1993)), and the Threshold GARCH (TGARCH, Zakoian (1994)). These models are presented in table 1. These models were estimated by assuming that the innovations follow a normal distribution, distribution accounting for the excess of kurtosis (Student T (std) and GED (ged)) as well as the skewed version of these probability distributions (skewed normal (snorm), skewed Student T (sstd) and skewed GED (sged)) which account to the skewness of financial returns' distributions. All these specifications were also used in the retained semi-parametric methods, which are the AR-GARCH type models combined with Monte-Carlo simulation (AR-GARCH-type-MC), the Filtered Historical simulation (FHS), the FHS combined with bootstrapping (FHS-B) and the conditional POT (SEVT). For this later method, threshold values are put directly at the lower 10% quantile of the distribution for the retained series.

Existing studies using rolling-window approach retained the same GARCH-type specification and the same probability distribution in each window. For a given series, the best-fitting GARCH specification and/or the best-fitting probability distribution can differ from one period to another period and mainly during stable and unstable periods. In order to account for these facts, two approaches were used in this article.

Table 1: Retained GARCH Type Models

Model	Formulations	Model	Formulations
GARCH	$\sigma_t^2 = w + \alpha.\varepsilon_{t-1}^2 + \beta.\sigma_{t-1}^2$	**gjrGARCH**	$\sigma_t^2 = w + \alpha.\varepsilon_{t-1}^2 + \gamma.\varepsilon_{t-1}^2 I(\varepsilon_{t-1} > 0) + \beta.\sigma_{t-1}^2$
eGARCH	$\ln(\sigma_t^2) = (1-\alpha_1)\alpha_0 + \alpha_1 \ln(\sigma_{t-1}^2) + g(\varepsilon_{t-1})$ $g(\varepsilon_{t-1}) = \theta.\varepsilon_{t-1} + \gamma(\lvert\varepsilon_{t-1}\rvert - \sqrt{\frac{2}{\pi}})$	**TGARCH**	$\sigma_t^2 = w + \alpha.\varepsilon_{t-1} + \gamma.\varepsilon_{t-1} I(\varepsilon_{t-1} > 0) + \beta.\sigma_{t-1}^2$

Approach 1: For each return series and each probability distribution, the best fitting GARCH specification in each window was selected based on AIC criterion. For each retained method (parametric and semi-parametric), the dynamic of each retained series was evaluated in each window with the best fitting GARCH specification for each retained probability distribution and the 1-day ahead VaR and ES were forecasted. As we retained 5 methods (1 parametric and 4 semi-parametric) and 6 probability distributions, in total 30 (5*6) series of around 4000 VaR and ES were forecasted for each retained returns' series.

Approach 2: For each series and each estimation window, the best fitting model was selected and estimated based on AIC among all GARCH specifications combined with all retained probability distributions and the 1-day ahead VaR and ES were forecasted. For each return series more than 5 series (5=1 parametric + 4 semi-parametric methods) of around 4000 VaR and ES were forecasted.

The accuracy of the retained models should be backtested. The Basel committee recommends backtesting a model by using 252-day forecasted VaR and ES. As more than 4000 1-day VaR and ES were forecasted for each series by using a specific model with a specific probability distribution, forecasted VaR and ES were backtested by using 252-days rolling-window approach. In sum, for each series each model was backtested around 3750 times. This method enables to determine the performance of retained models through time.

Table 2: Descriptive Statistics

	Gold				EURUSD				SP500			
	mean	std	skew	kurtosis	mean	std	skew	kurtosis	mean	std	skew	kurtosis
mean	0.0003	0.0111	-0.281	4.2493	0.0000	0.0060	0.0477	1.3067	0.0001	0.0110	-0.202	2.5417
std	0.0004	0.0023	0.7634	3.1793	0.0003	0.0011	0.2597	1.0358	0.0005	0.0046	0.3138	1.9580

std: standard deviation

4.2. Empirical Results

4.2.1. Descriptive Statistics

The summary of the descriptive statistics of each 500-day rolling window is represented in table 2. The average returns of retained series are very slightly positive. The returns of the Gold and of the SP500 index are slightly negatively skewed. As for the return of the currency exchange rates, it is slightly positively skewed. As displayed in this table, most of the time, the skewness of Gold returns are negative and the kurtosis is higher than 3.

Table 3: Best Fitting GARCH (Bfg) Specification per Distribution (In-Sample) (Percent %)

	norm	snorm	std	sstd	ged	sged	norm	snorm	std	sstd	ged	sged
	SP500						EURUSD					
egarch	**76.2**	**77**	**77.1**	**77.8**	**77.8**	**78.8**	**49.6**	**49.7**	**46.4**	**47.7**	**46.8**	**46.8**
sgarch	0.7	0.7	0.4	0.3	0.9	0.9	27	17.9	33	4.6	2.6	2.6
gjrgarch	4	2.7	3.3	3.4	2.5	2.7	3.8	3.4	4.6	15	17.7	17.7
tgarch	19.2	19.6	19.3	18.4	18.8	17.7	19.7	18.9	16	32.9	32.9	32.9
	Gold											
egarch	**60.7**	**53.7**	**44.5**	**45.2**	**49**	**47.1**						
sgarch	5.9	9.1	27.1	27.4	23.2	24.8						
gjrgarch	5.2	8.3	9.9	9	4.5	6.4						
tgarch	28.2	28.9	18.4	18.4	23.2	21.8						

4.2.2. Empirical Results Related to Approach 1

For each retained probability distribution and each return series the best fitting GARCHtype specification is determined for each retained 500-day window based on AIC. Table 3 depicts a summary of the best-fitting GARCH specification for each probability distribution. According to our results, asymmetric GARCH-type models (eGARCH, gjrGARCH and TGARCH) fit better retained series whatever the probability distribution. For instance, the EGARCH model combined with the normal probability distribution was the best fitting model for the SP500 series on 76,2% of retained windows. Among the retained asymmetric model, the EGARCH dominates the gjrGARCH and TGARCH model whatever the retained series and the probability distribution. The classical GARCH model fits better only on a few windows.

For each retained series and each retained probability distribution, the best fitting GARCH-type specification is used for forecasting 1-day ahead VaR and ES at each time (in each window) with each of the retained parametric and semi-parametric methods. The performing model is the one that is the most often accurate one; this means that the null hypothesis of model accuracy is the least rejected by backtesting methods. In this article the conditional coverage (CC) test of Christoffersen (1998) and the dynamic quantile (DQ) test of Engle and Managanelli (2004) are used.

For each method, the accuracy of VaR(99%) and VaR(95%) is checked at levels 1% and 5% for each series and each probability distribution combined with the best fitting GARCH type specification. Tables 4 - 6 represent the percent of time the null hypothesis stating the accuracy of the retained model in forecasting the VaR is rejected. Due to the large number of results (30=6*5 series of around 4000 forecasted VaR), these tables contain for each method and for each series only the results of the model rejecting the least the null hypothesis of model accuracy based on the conditional coverage test and the DQ test. Both these tests may sometimes lead to different results. For instance, based on the conditional coverage test at 1% level the null hypothesis is the least rejected (4.84% times) for the AR-GARCH model combined with skewed Student (sstd) distribution.

In the forecasted VaR(99%) compared to AR-GARCH models combined with other retained probability distributions for the SP500 return series. And according to the DQ test, the null hypothesis is the least rejected (15.75% times) for the AR-GARCH model combined with skewed GED (sged) distribution for the SP500 return series compared to AR-GARCH models combined with other retained probability distributions. For this return series, the VaR(95%) forecasted with the AR-GARCH model combined with skewed GED (sged) distribution is the most accurate based according to both backtesting tests at 1% compared to other AR-GARCH model combined with the other retained probability distribution.

The percentages of rejection of the null hypothesis of model accuracy are different in both backtesting methods. The null hypothesis is more rejected by the DQ test than the conditional coverage test, except in a few cases such as backtesting the VaR(95%) at 5%. These percentages of rejection depend also on the level of VaR and the level of backtesting method.

Models producing forecasted VaR which are the most often accurate ones are represented in bold in tables 4 - 6 whatever the method and the probability distribution. These tables reveal that methods using skewed probability distributions (skewed Student (sstd) and skewed GED (sged)) enable accurately forecast VaR more often followed by methods based on heavy-tailed probability distributions (Student (std) and GED).

Table 4: Backtesting VaR-SP500 (%)

	VaR(99%)	BT	at 1%	VaR(95%)	BT	at 1%	VaR(99%)	BT	at 5%	VaR(95%)	BT	at 5%
		CC	DQ		CC	DQ		CC	DQ		CC	DQ
AR-GARCH	sged	5.19	15.75	sged	2.73	6.05	sged	11.82	16.07	sged	12.31	11.28
	sstd	4.84	21.63									
	bf	5.14	15.53	bf	2.73	6.39	bf	14.40	15.95	bf	12.41	11.87
AR-GARCH-MC	sged	6.24	19.73	sged	3.29	6.05	sged	9.39	19.93	sged	12.17	10.99
	sstd	4.89	21.41	sstd	2.16	6.27						
	bf	6.24	23.45				bf	14.01	23.69	bf	12.71	12.66
SEVT	**std**	**3.24**	**14.55**	ged	0.42	5.53	ged	9.98	21.33	**sged**	**6.29**	**9.29**
							std	**11.28**	**15.75**			
	bf	4.52	28.78	bf	0.42	6.02	bf	16.00	30.10	bf	6.07	10.37
FHS	sstd	5.55	17.25	ged	2.16	4.67	sged	9.78	20.13	norm	5.87	10.81
				sstd	0.00	6.22	sstd	14.52	18.14	sstd	5.06	13.39
	bf	5.73	30.18	bf	2.70	4.60	bf	14.40	15.95	bf	4.55	9.09
FHS-B	sged	3.86	17.35	**sstd**	**0.00**	**4.37**	**sged**	**8.70**	18.65	ged	5.16	10.47
										sstd	**4.18**	**11.55**
	bf	3.86	23.45	bf	2.7	4.08	bf	10.00	25.76	bf	4.55	9.02

bf: best fitting model among all retained GARCH specifications and retained probability distributions.

Figures in this table represent the percentage of rejection of Ho. BT: backtesting.

The conditional EVT (SEVT) and Bootstrapped FHS (FHS-B) methods produce forecasted VaR of SP500 which are the most often accurate ones. Indeed, the null hypothesis of accuracy is the least rejected for one of these methods dependent on the level of VaR and the level of the backtest test (Table 4). For instance, 3.24% of more than 3750 backtested forecasted VaR(99%) of SP500 determined with the SEVT model based on Student probability (std) distribution leads to the rejection of the null hypotheses at 1% according to the conditional coverage test. Similar findings are observed for the EURUSD currency exchange (Table 5). Regarding the Gold, the results in Table 6 do not lead to the conclusion of a dominant method. Our findings related to SP500 and EURUSD are in line with the summary of the review of literature done by Abad et al. (2014). In their paper these authors did a review of VaR methodologies and findings. They resumed the findings of 24 papers using classical parametric, semi-parametric and non-parametric methods in forecasting VaR. In 83.3% of these papers, the EVT was selected as the best method in forecasting the VaR, followed by FHS.

For each retained series the observed returns and the forecasted VaR(99%) with the selected models are displayed in graphs 1, 2 and 3.[1] These graphs show that realized returns are almost higher than the forecasted VaR(99%), except a few times. Regarding the backtesting results, the p-values related to the CC test and DQ test of the retained models for forecasting VaR(99%) are displayed in graphs 4,5, and 6. These graphs reveal that the null hypothesis of accuracy is mainly rejected during unstable periods, such as the beginning of the subprime crisis (2007-2008). These plots show the dependence between exceptions.

Figure 1: Realized Return and VAR (99%) of SP500 with SEVT-Std and FHS-B-Sged Models

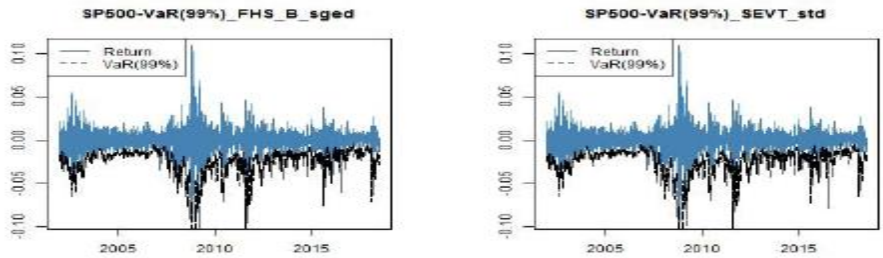


Table 5: Backtesting VaR-EURUSD (%)

	VaR(99%)	BT	at 1%	VaR(95%)	BT	at 1%	VaR(99%)	BT	at 5%	VaR(95%)	BT	at 5%
		CC	DQ		CC	DQ		CC	DQ		CC	DQ
AR-GARCH	sged	**0.00**	18.44	norm	2.56	5.02	sged	1.65	18.51	norm	10.28	7.79
				snorm	2.19	5.66				snorm	8.24	10.43
	bf	0.30	18.54	bf	3.17	7.97	bf	2.31	19.03			
AR-GARCH-MC	**sged**	**0.00**	18.44	norm	2.16	4.84	sged	1.67	18.44	norm	9.32	7.55
	bf	0.30	18.54	bf	2.24	7.77	bf	2.31	19.03	bf	7.47	13.62
SEVT	ged	0.30	18.39	**ged**	0.20	**0.34**	sged	1.65	18.51	**ged**	5.97	**3.25**
										snorm	**3.29**	7.62
	bf	0.30	18.39	bf	0.20	0.32	bf	2.26	18.59	bf	5.97	3.07
FHS	**ged**	0.00	18.24	norm	0.52	4.33	norm	0.02	18.37	ged	7.89	6.42
	norm	0.00	18.24				sged	1.65	18.29	sstd	4.38	6.47
	bf	0.00	18.49	bf	1.03	5.14				bf	7.89	6.25
FHS-B	**ged**	0.00	**18.15**	**norm**	0.00	2.41	**sged**	**0.02**	**18.29**	ged	7.72	5.83
				snorm	0.00	2.24				snorm	4.30	7.72
	bf	0.00	18.37	bf	0.61	4.5	bf	0.02	18.29	bf	7.72	5.75

bf: best fitting model among all retained GARCH specifications and retained probability distributions.

Figures in this table represent the percentage of rejection of Ho. BT: backtesting.

Table 6: Backtesting VaR-Gold (%)

	VaR(99%)	BT	at 1%	VaR(95%)	BT	at 1%	VaR(99%)	BT	at 5%	VaR(95%)	BT	at 5%
		CC	DQ		CC	DQ		CC	DQ		CC	DQ
AR-GARCH	std	0.00	23.81	sstd	0.00	6.88	std	2.24	31.33	sstd	1.79	15.48
	bf	6.19	37.77	bf	0.10	6.64	bf	6.27	39.54	bf	6.17	22.88
AR-GARCH-MC	sstd	0.05	39.89	sged	0.05	8.87	**std**	**2.04**	27.18	sstd	1.79	19.61
	std	**0.00**	**20.15**									
	bf	6.17	37.55	bf	1.38	9.90	bf	6.27	38.51	bf	6.34	28.70

[1] Due to space, only the p-value graphs and VaR(99%) graphs of the selected models are displayed. Others obtained graphs can be provided on demand.

SEVT	snorm	2.78	20.77	std	0.00	4.74	**snorm**	**5.51**	**21.33**	**ged**	**0.00**	**18.70**
	sstd	1.52	36.99				sstd	3.49	39.44	sstd	2.38	16.10
	bf	3.64	37.72	bf	0.00	8.04	bf	6.46	39.52	bf	0.05	19.69
FHS	snorm	2.38	20.59	ged	0.00	4.30	snorm	5.58	21.55	ged	0.71	15.21
	sstd	0.02	36.27				sstd	4.20	40.70	**sstd**	**1.35**	**12.02**
	bf	2.11	32.64	bf	0.00	3.76	bf	5.70	34.06	bf	0.71	14.33
FHS-B	norm	1.99	20.45	std	0.00	3.24	snorm	5.58	21.60	ged	0.29	14.60
	sstd	0.02	24.65				sstd	4.15	26.96	sstd	3.00	13.44
	bf	1.23	31.04	**bf**	**0.00**	**3.10**	bf	5.14	33.05	bf	0.00	13.64

bf: best fitting model among all retained GARCH specifications and retained probability distributions.

Figures in this table represent the percentage of rejection of Ho. BT: backtesting.

For the selected models, a summary of the results of backtesting ES are displayed in Table 7. For each series and for each selected model within each retained method, this table contains the percent of time the null hypothesis is rejected at 5% level with the 1-side and 2-side tests proposed by McNeil and Frey (2000). The null hypothesis states the nullity of the residuals between the observed losses (returns) and the predicted ES in case the VaR is exceeded. If a model accurately enables to forecast the ES then the average of the difference between the observed return and the predicted ES should be equal to 0. Compared to the percentages of rejection of the null hypothesis of accuracy in forecasting VaR, the percentages of rejection of the null hypothesis related to backtest of ES are higher. In all selected models, the percentages of rejection of the null hypothesis related to ES are between 0,3-0,4.

Figure 2: Realized Return and VaR (99%) of EURUSD with FHS-B-Sged Model

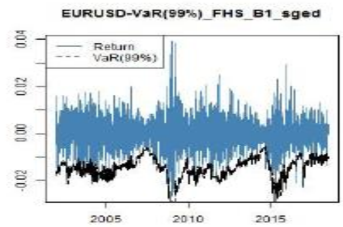


Figure 3: Realized Return and VaR (99%) of Gold with AR-GARCH-MC-Std And SEVT-Snorm Models

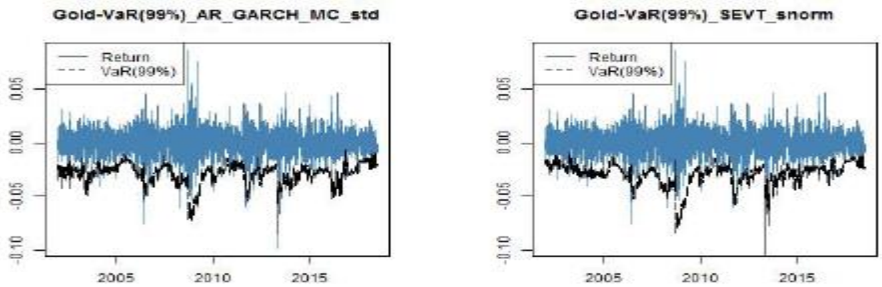


Table 7: Backtesting ES (%)

	SP500	er-1side	er-2side	EURUSD	er-1side	er-2side	Gold	er-1side	er-2side
AR-GARCH	sged	38.58	38.58	sged	39.64	39.64	std	39.30	37.18
	bf	42.54	42.54	bf	30.74	30.64	bf	39.54	38.61
AR-GARCH-MC	sged	39.74	38.29	sged	41.14	39.86	std	31.19	30.40
	bf	35.49	34.63	bf	33.19	33.10	bf	42.71	40.03
FHS	sged	43.60	43.60	norm	37.96	37.87	snorm	45.44	45.44
	sstd	37.33	36.27	sged	45	43.62	sstd	46.77	42.07

	bf	44.61	41.26	bf	37.72	36.22	bf	44.63	42.57
FHS-B	sged	45.56	45.56	sged	38.04	36.66	snorm	47.83	47.83
							sstd	46.77	40.89
	bf	40.08	40.08	bf	37.03	35.68	bf	44.8	38.76

bf: best fitting model among all retained GARCH specifications and retained probability distributions.

Figures in this table represent the percentage of rejection of Ho. er-1side: 1 side test and er-2side: 2 side test.

Figure 4: P-value of Conditional Coverage Test and DQ Test for VaR (99%) Tested at 5% for SP500 with SEVT-std and FHS-B-sged Models

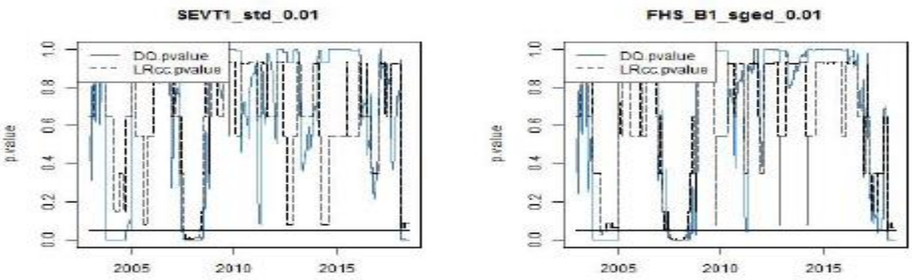


Figure 5: P-value of Conditional Coverage Test and DQ test for VaR(99%) Tested at 5% for Gold with SEVT-snorm and AR-GARCH-MC-std models

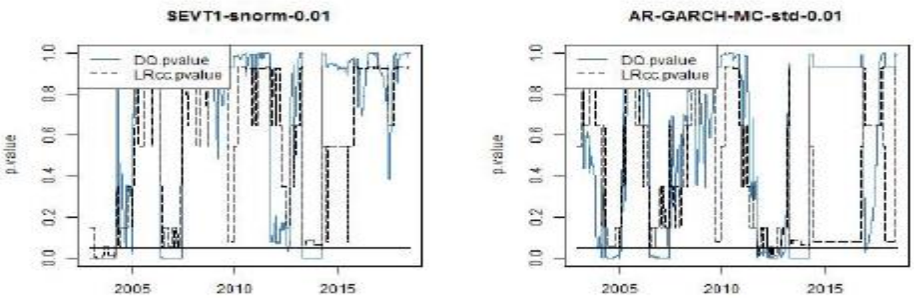
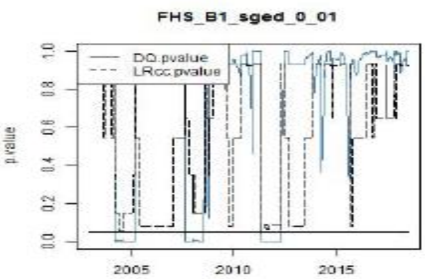


Figure 6: P-value of Conditional coverage test and DQ test for VaR(99%) tested at 5% for EURUSD with FHS-B-sged model



4.2.3. Empirical Results Related to Approach 2

The best way would be to forecast the VaR and the ES for each time with the GARCH type model combined with the probability distribution fitting the best retained series. Among the retained GARCH-type models and probability distribution, a summary of the best fitting specification for each series is presented in table 8. This table represents the percentages of each specification retained as the best fitting specification based on AIC criteria. According to our results, the GED probability distribution is the dominant distribution in case of the EURUSD and Gold series. As for the return of SP500, the skewed GED (sged) is the dominant distribution.

For each method, the dynamic of each retained series is evaluated with the retained best fitting model in each window and more than 4000 1-day ahead forecasted VaR and ES are determined. The percent of rejection of the null hypothesis among the 3750 backtests are represented with bf (bf=best fitting) in Tables 4 - 6. Surprisingly, the best-fitting models do not improve the frequency of the accuracy of forecasted VaR. Indeed, the percent of time the rejection of the null hypothesis of accuracy is not lower for best-fitting models compared to the best specification based on the probability distribution (approach 1).

In this article, for each probability distribution the best GARCH-type model was selected based on the AIC (in-sample). Similarly, the best-fitting model was also selected based on AIC determined in sample. The best-fitting model in-sample does not mean that this model is the best-fitting model out-sample. This fact can explain our backtesting results related to the best fitting model (approach 2) compared to the backtesting results related to the best model retained in approach 1.

Regarding the backtest of forecasted ES by using the best fitting model, the percent of time the null hypothesis of accuracy is rejected is important (see Table 7). However, these percentages are lower than the percentages of rejection of the null hypothesis of accuracy in forecasting ES with methods used in the first approach, except for a few methods.

5. CONCLUSION

The purpose of this article was to compare the performance of parametric and semiparametric models in forecasting the VaR and the ES through time. The performance of a model in modelling the dynamic of financial returns may change through time and then the accuracy of this model to forecast the VaR and ES may change too. To account for these facts, two approaches were considered.

Approach 1: For each return series and each probability distribution, the best fitting GARCH specification in each window was selected based on AIC criterion. For each retained method (parametric and semi-parametric), the dynamic of each retained series was evaluated in each window with the best fitting GARCH-specification for each retained probability distribution and the 1-day ahead VaR and ES were forecasted.

Approach 2: For each series and each estimation window, the best fitting model was selected and estimated based on AIC among all GARCH-specifications combined with all retained probability distributions and the 1-day ahead VaR and ES were forecasted.

Our results revealed that the best fitting GARCH-specifications combined with skewed Student or GED distribution enable to accurately forecast VaR more often. However, the best methods based on the best fitting GARCH-specifications combined with the best fitting probability distribution do not improve the frequency of acceptance of the null hypothesis stating the accuracy of the method. The accuracy of models tends to deteriorate during crises periods. This latter finding can be explained by the fact that the best fitting model in sample does not mean that this model is the best fitting model out-sample. Indeed, the best fitting GARCH-specification and the best fitting model were selected in-sample based on AC.

Our results revealed that the selected model among all retained methods/approaches enables to forecast accurately the realized loss most of the time but not all the time. Forecasted VaR with the selected model underestimates the realized losses over certain period mainly during unstable periods. This finding can explain the consideration of the Stressed VaR(99%) in addition of the VaR(99%) in the determination of the risk weighted assets (RWA) for market risk since Basel 2.5. In the Basel 2, the RWA is determined by using only the VaR(99%). The Stressed VaR (sVaR) should be determined by using the worst 1-year daily data occurred during the past 10 year.

Results related to backtesting ES with nostandardized version of McNeil and Frey (2000) test are not homogeneous. These results can also be due to the fact that backtesting ES is difficult compared to backtesting VaR. Due to this latter fact, in the FRTB approach the Basel committee did not prescribe a method enabling to directly backtest ES but this later is done by using Profit and Loss attribution test as well as backtesting VaR. Furthermore, these results can explain why the ES for the determination of the required capital for market risk in the FRTB is not determined as done in this article. In this article, the ES is determined as the average loss when the forecasted VaR is exceeded by using daily returns of the considered window. In the FRTB, different ES are determined and combined in the determination of the risk-weighted assets for market risk. One type of ES is determined by using current data as we did and one type by using data from the stressed periods.

REFERENCES

Abad, P., Benito, S. & Lopez C. (2014) A comprehensive review of Value at Risk methodologies. The Spanish Review of Financial Economics 12, 15-32.

Adcock, C., Eling, M. & Loperfido, N. (2012). Skewed distributions in finance and actuarial science: a review. The European Journal of Finance, 4, 1-29.

Angelidis, T., Benos, A. & Degiannakis, S. (2007). A robust VaR model under different time periods and weighting schemes. Review of Quantitative Finance and Accounting, 28, 187-201.

Artzner, P., Delbaen, F., Eber, J. & Heath, D. (1997). Thinking coherently. Risk, 10(11), 68-71.

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

Bali, T. G. & Theodossiou, P. (2008). Risk measurement performance of alternative distribution functions. The Journal of Risk and Insurance 75, 411-437.

Bali, T. G. & Weinbaum, D. (2007). A conditional extreme value volatility estimator based on high-frequency returns. Journal of Economic Dynamics and Control, 31, 361-397.

Bali, T., Mo, H. & Tang, Y. (2008). The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VAR. Journal of Banking and Finance, 32, 269-282.

Bali, T. & Theodossiou, P. (2007). A conditional-sgt-var approach with alternative GARCH models. Annals of Operations Research, 151(1), 241-267.

Balkema, A. A. & de Haan, L. (1974). Residual lifetime at great age. Annals of Probability, 2, 792-804.

Campbell, S. (2005), A review of backtesting and backtesting procedures. Finance and Economics Discussion Series. Washington, DC: Federal Reserve Board.

Chen, C., Gerlach, R., Lin, E. & Lee, W. (2011). Bayesian forecasting for financial risk management, pre and post the global financial crisis. Journal of Forecasting, 31(8), 661-687.

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

Down, K. (2002), Measuring market risk, John Wiley and Sons, Chichester.

DuMouchel, W. (1983). Estimating the stable index in order to measure tail thickness: a critique. Annals of Statistics, 11(3), 1019-1031.

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

Engle, R. & Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics, 22, 367-381.

Gilli, M. & Kezi, E. (2006). An application of extreme value theory for measuring financial risk. Computational Economics, 27(2), 207-228.

Giot, P. & Laurent, S. (2003). Value-at-risk for long and short trading positions. Journal of Applied Econometrics, 18, 641-663.

Giot, P. & Laurent, S. (2004). Modelling daily value-at-risk using realized volatility and arch type models. Journal of Empirical Finance, 11, 379-398.

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, 5, 1779-1801.

Gonzalez-Rivera, G., Lee, T. & Mishra, S. (2004). Forecasting volatility: a reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. International Journal of Forecasting, 20, 629-645.

Haas, M., Mittnik, S. & Paolella, M. (2004). Mixed normal conditional heteroskedasticity. Journal of Financial Econometrics, 2, 211-250.

Huang, Y. & Lin, B. (2004). Value-at-risk analysis for Taiwan stock index futures: fat tails and conditional asymmetries in return innovations. Review of Quantitative Finance and Accounting, 22, 79-95.

McNeil, A. & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. Journal of Empirical Finance, 7, 271-300.

McNeil, A., Frey, R. & Embrechts, P. (2005). Quantitative risk management: Concepts, techniques, and tools. Princeton Series in Finance. Princeton University Press.

Mittnik, S. & Paolella, M. (2000). Conditional density and value-at-risk prediction of Asian currency exchange rates. Journal of Forecasting, 19, 313-333.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach. Econometrica, 59, 347-370.

Niguez, T.-M. (2008). Volatility and VaR forecasting in the Madrid stock exchange. Spanish Economic Review, 10, 169-196.

Pickands, J. (1975). Statistical inference using extreme order statistics. Annals of Statistics, 3, 119-131.

Smith, R. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. Statistical Science, 4, 367-393.

Zakoian, J. M. (1994). Threshold heteroskedastic models. Journal of Economic Dynamics and Control, 18 (5), 931-955.

Zhang, Y. & Nadarajah, S. (2018). A review of backtesting for value at risk. Communications in Statistics Theory and Methods, 47 (15) ,3616-3639.