Journal of Risk Model Validation **15**(4), 29–50
DOI: 10.21314/JRMV.2021.011

Research Paper

Evaluation of backtesting techniques on risk models with different horizons

Grigorios Kontaxis and Ioannis E. Tsolas

School of Applied Mathematical and Physical Sciences, National Technical University of Athens, Zografou Campus, 157 72 Zografou, Athens, Greece; emails: gregkont2011@hotmail.com, itsolas@central.ntua.gr

(Received March 5, 2021; revised October 19, 2021; accepted October 27, 2021)

ABSTRACT

In this study different value-at-risk (VaR) models, which are used to measure market risk, are analyzed under different estimation approaches (filtered historical simulation, extreme value theory and Monte Carlo simulation) and backtested with different techniques. The autoregressive-moving-average and generalized-autoregressive-conditional-heteroscedasticity models are used to estimate VaR. In particular, selected VaR functions, marginal distributions and different horizons are combined over a set of extreme probability levels using the time series of the Financial Times Stock Exchange 100 Index. Several backtesting techniques are examined in this research, such as Kupiec's proportion-of-failures test and Christoffersen's independence test. This study shows that, for short horizons, some approaches underestimate VaR. However, various models present violation estimates that almost converge to the desired ones, according to the confidence levels used. Further, nonoverlapping returns tend to yield satisfactory results for most models. The main conclusion of this study is that the horizon selection can affect the estimation, and consequently the backtesting, of VaR models in some cases.

Keywords: value-at-risk (VaR); backtesting; efficient model selection; risk estimation; risk modeling.

Corresponding author: G. Kontaxis

Print ISSN 1753-9579 | Online ISSN 1753-9587

1 INTRODUCTION

1.1 Regulatory requirements

The importance of risk management in financial institutions stems from the necessity to have a reserve of capital able to cover their financial obligations. The concept of financial risk comprises many different aspects of various types of risks. Risk in finance can be divided into credit, liquidity, operational, legal and market risks. Some of the most important risk elements include: credit risk, which arises when the counterparties are unable to fulfill their contractual obligations; liquidity risk, which refers to the inability of a financial institution to meet its payment obligations; and market risk, which is related to unexpected changes in prices over short time horizons and is the main research subject of this paper.

1.2 Risk metrics

One of the most important components of measuring risk is the selection of risk models that capture the complex characteristics of different time series and that produce adequate estimations to meet regulatory bodies' requirements. Time series can contain autocorrelation, heteroscedasticity, asymmetric tails on their returns, seasonality patterns and trends that require modeling capable of estimating coherent risk measures.

Value-at-risk (VaR) models are currently the most well-known risk tools. The most common means of verifying VaR models is backtesting. In this process actual profits and losses are compared with the projected VaR estimations, and the consistency and reliability of VaR calculations can be examined by employing various alternative approaches. Among the most well-known backtesting techniques are Kupiec's proportion of failures (POF) test (Kupiec 1995), which examines the frequency of losses above the VaR; the independence test (Christoffersen 1998); and the Basel test proposed by the Basel Committee on Banking Supervision (1996).

Risk management departments use various criteria and backtesting techniques to select the model to be adopted for risk estimation purposes. Most of the backtesting methodologies are based on the correctness of the obtained results. When the number of violations converges to the confidence level of the VaR, the model is considered to be efficient. Violations that are more or less frequent than the confidence level are indicative of an underestimation or overestimation of risk. While the overestimation of VaR can imply that a methodology is costly in terms of capital requirements, underestimation can lead to the default of a financial institution in cases of market turbulence. Thus, the convergence of the number of violations to the selected confidence level is always desirable.

The confidence level selection can vary significantly across financial institutions or portfolios. One of the most common risk levels (or probability levels) for the estimation of VaR is 1%. Another parameter of the VaR, in addition to the confidence level, is the time horizon. Most of the literature related to the estimation of risk metrics examines a one-day horizon. However, the Basel Committee regulations (Basel Committee on Banking Supervision 1996) require that VaR be reported for ten-day horizons for liquidity purposes (for the trading book). There are also situations such as long-term liabilities (pension funds, government bonds) in which yearly horizons may be desirable. VaR estimation at long-term horizons can be more complex than at daily horizons due to data availability issues.

It is common in risk modeling for many VaR methodologies to appear adequate under certain given settings. When these configurations are changed, however, the performance of the model may deteriorate. Most of the backtesting processes are designed to test risk models' reliance on specific conditions. As stated above, the Basel regulations (Basel Committee on Banking Supervision 1996) require financial institutions to provide ten-day VaR estimates (for the trading book) and additionally 250 days of the risk model's (VaR) backtesting, which amounts to 2500 observations (10 years) in cases in which a nonoverlapping method is used. This creates issues, due to data limitations, and thus financial institutions tend to use one-day VaR estimations for their backtesting analysis and ten-day overlapped returns for the VaR calculation; further, it is assumed that the backtesting findings on one-day time series apply to ten-day overlapped estimations as well. In this way, the financial institutions can values with limited data availability.

However, the overlapped data induce autocorrelation in the ten-day time series, and thus it is not correct to use the standard backtesting methods to evaluate a portfolio built by using this technique. Standard backtesting methods assume independent and identically distributed (iid) data, and this assumption is violated when the overlapped time series are created. Consequently, the backtesting performance of one-day VaR models should not be correlated with that of ten-day overlapped VaR estimations. In addition, most studies (see, for example, Sun *et al* 2009) suggest that overlapped returns create underestimated VaR estimates.

1.3 Contribution of the paper

The main purpose of this paper is to provide an in-depth analysis of risk metrics' effectiveness on longer horizons and to evaluate the performance of various VaR models using nonoverlapping ten-day returns and compare their behavior against VaR models with a one-day horizon. A variety of nonoverlapping data sets are used across numerous alternative VaR techniques to assess the models' compatibility on long forecast horizons. Specifically, the case investigated involves a financial institution with a long position on the Financial Times Stock Exchange 100

(FTSE 100) index. Several variants of autoregressive-moving-average generalized-autoregressive-conditional-heteroscedasticity (ARMA-GARCH) models are examined for a broad range of marginal distributions and three different estimation techniques (filtered historical simulation (FHS), extreme value theory (EVT) and Monte Carlo simulation). The findings of this study will provide evidence as to whether ten-day observations present similar behavior to one-day observations and whether the latter can be considered an appropriate representative sample for backtesting purposes on longer horizons. Although the literature has explored various techniques for one-day VaR estimation (Barone-Adesi and Giannopoulos 2001; McNeil and Frey 2000), the calculation and evaluation of the stability of risk metrics on longer horizons using alternative methodologies are yet to be examined.

1.4 Structure

The paper is structured as follows. In Section 2 the VaR metrics are presented and analyzed. In addition, the backtesting processes are described. Section 3 offers a detailed introduction to financial time series modeling, specifically using ARMA-GARCH models based on FHS, EVT and Monte Carlo methodologies. All of the various techniques are analyzed at different time horizons. In Section 4 backtesting techniques are introduced to assess the selected risk models' adequacy. Section 5 presents a detailed description of the empirical results. Finally, Section 6 states the conclusions of our research.

2 VALUE-AT-RISK ESTIMATION AND EVALUATION TECHNIQUES

Financial institutions use VaR to estimate the market risk of their exposures and to define their capital requirements. Agencies need a quantitative means of identifying risks related to market positions. VaR is a metric that can model adverse events, taking portfolio history or independent risk factors into account. VaR's numerous advantages include its simplicity and applicability. Moreover, it is applied to returns with different risk factors and portfolios. In contrast to prices, returns generally remain stationary, a property that is required in most VaR methodologies.

2.1 VaR

VaR is a risk measure of investment loss. It can be used to estimate investment losses under a given probability level, assuming normal conditions in the market, and for different periods, depending on the assets it is applied to. VaR is the quantile of a projected distribution of profits and losses over a time horizon. Mathematically, VaR is defined as the minimal potential loss that a portfolio can suffer in the $100a\%$ worst case with $a \in (0,1)$. Given that $a \in (0,1)$, the VaR at a level a, VaR_a, of the final net worth X with distribution P is given as follows (Artzner *et al* 1998):

$$\mathrm{VaR}_t^{1-a} = \sup[r \mid P(R_t \leqslant r) \leqslant a], \tag{2.1}$$

where r denotes the return of an asset and R_t is a random variable that represents the return of the asset at the end of period T. A large number of alternatives for estimating VaR exist. One of the most well-known processes is historical simulation, which follows a simple methodology but relies on numerous assumptions about the returns of the time series. Other alternatives include Monte Carlo simulation, FHS and EVT.

2.2 Backtesting

To evaluate the accuracy and validity of VaR estimates, the Basel Committee developed a statistical test called the traffic light test (Basel Committee on Banking Supervision 1996). According to the regulations, backtesting should be used on at least 250 estimates of VaR. Nonetheless, nearly all backtesting results require a significantly higher number of samples for accurate interpretations.

Most backtesting methodologies are based on the following process:

$$I_t^a = 1(R_t < \mathrm{VaR}_t^a), \tag{2.2}$$

where $t = T+1, \dots, T+n$, $1(\cdot)$ is the indicator function, T indicates the size of the VaR estimation sample and n indicates the number of one-step-ahead forecasts. The VaR is considered to be accurate when the number of failures corresponds to its confidence level. The correspondence rate between the number of failures and the confidence interval is evaluated by some statistical backtesting method, such as that of Kupiec or Christoffersen.

The well-known backtesting process proposed by Kupiec (1995) is based on the number of failures, which is assumed to follow a binomial distribution.

3 VALUE-AT-RISK METHODOLOGIES

VaR can be estimated with many alternative methodologies, each of which has some particular weaknesses and strengths. VaR estimation can be a difficult computational task due to the variety of factors that have an impact on its evaluation; for example, portfolio VaR depends on the contents of the portfolio, because complex instruments such as barrier options can affect its returns. The complexity of certain financial instruments, the portfolio size, the required confidence level and the available history of risk factors, along with the calculation speed, are some of the various factors that play an important role in VaR methodology decisions.

In this research the ARMA and GARCH models are used to capture the volatility clustering of the time series. In addition to the simple GARCH model, exponential GARCH (EGARCH) (Nelson 1991) and Glosten–Jagannathan–Runkle GARCH (GJR-GARCH) (Glosten *et al* 1993) are used to address asymmetry in the data. Further, a large number of distributions were examined for the GARCH model fitting. Thus, normal, Student t and skewed Student t distributions and the generalized error distribution (GED) were selected to represent the properties of the data. In contrast to the normal distribution, which is symmetric, the other distributions are more representative of time series with the fat tails and skewness so common to financial instruments.

ARMA and GARCH models were combined with three alternative methodologies to estimate VaR: the FHS model (Barone-Adesi and Giannopoulos 2001), which is based on using random draws with replacement from the standardized residuals; Monte Carlo simulation, a popular methodology in which an assumption about the residuals' distribution is made; and EVT (McNeil and Frey 2000), which assumes the distribution of the tails of the standardized residuals.

There are several alternative VaR techniques – such as historical simulation, bootstrapping, EVT with copulas and exponential weighted averages – that appear promising as well, but they are not tested in this study; the reasons for this are explained below.

Historical simulation makes assumptions about the symmetry of the data that are rarely true. Moreover, while the bootstrapping (resampling) technique tends to require a very large amount of data, it offers a relatively noncomplex methodology for VaR estimation. EVT combined with copulas is a complex and time-consuming method that most financial institutions prefer to avoid due to computational speed issues. Further, fitting copulas to insufficient data can lead to inaccurate predictions. Finally, the exponentially weighted moving average (EWMA) method stems from GARCH related models. GARCH models, which represent the evolution of EWMA models, are preferred for this research.

3.1 Modeling the one-step-ahead forecasts

One of the oldest and most popular VaR estimation approaches is historical simulation, which relies on the empirical distribution of the returns under the assumption that they are iid. Advantages of this method include its simplicity as well as the speed of the calculations. However, empirical quantiles are rarely good estimators of extreme quantiles, and the iid assumption is invalid for most financial time series.

Because the iid assumption is inadequate for many risk factors and asset prices, various alternative methodologies are presented in this section.

The following model of returns is assumed where $(R_t,\ t \in Z)$ is a strictly stationary time series representing daily observations of the negative log return on a financial asset price. We assume that the dynamics of R_t (McNeil and Frey 2000), which follow the standard stock price model for simulating the path of a stock price (the Black–Scholes model assumption), are given by

$$R_t = \mu_t + \varepsilon_t \sigma_t, \tag{3.1}$$

where μ_t and σ_t respectively represent the mean and standard deviation of the returns and ε_t refers to the standardized residuals with mean equal to 0 and standard deviation equal to 1.

In addition, based on (3.1), VaR can be estimated using the following formula:

$$\mathrm{VaR}_t^a = \mu_t + q_a \sigma_t, \tag{3.2}$$

where q_a describes the $100a\%$ quantile of $f(\varepsilon_t)$, or the density of the standardized residuals.

According to (3.2), three terms must be defined for VaR estimation. The first term, the conditional mean, can be assumed as an ARMA model, given by

$$\mu_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i R_{t-i} + \sum_{j=1}^{q} \theta_j a_{t-j}, \tag{3.3}$$

where $a_t = \varepsilon_t \sigma_t$, φ_i refers to the autoregressive parameters, θ_j describes the moving-average parameters and R_{t-i} represents the previous returns of the portfolio.

Further, many models derived from the GARCH family for conditional variance modeling exist in the literature. The most well known of these is GARCH(1,1) (Bollerslev 1986):

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{3.4}$$

where α_0, α_1, β_1 represent the estimated parameters. In addition, $\alpha_0 > 0$, $\alpha_1 \geqslant 0$, $\beta_1 \geqslant 0$ and $(\alpha_1 + \beta_1) < 1$; these factors imply stationarity.

The EGARCH (Nelson 1991) and GJR-GARCH (Glosten *et al* 1993) models were introduced to model data asymmetry. Hentschel (1995) proposed the following equation to model the leverage effect:

$$\frac{\sigma_t^\delta - 1}{\delta} = \alpha_0 + \alpha_1 \sigma_{t-1}^\delta g^v \varepsilon_{t-1} + \beta_1 \frac{\sigma_t^\delta - 1}{\delta}, \tag{3.5}$$

where $g(\varepsilon_t) = |\varepsilon_t - b| - c(\varepsilon_t - b)$. The function g is linear. It encompasses two parameters that define the "size effect" and the "sign effect" of the shocks on volatility. In addition, α_0, α_1, β_1 are the estimated parameters. Many different GARCH models have been derived from (3.5). The parameter δ represents the shape of the transformation, while the parameter v serves to transform the absolute value function $g(\cdot)$ (Hentschel 1995). EGARCH is among the most popular for modeling asymmetry. It is generated when the conditions $\delta = 0$, $v = 1$ and $b = 0$ apply to (3.5).

Another well-known GARCH model for modeling asymmetry is GJR-GARCH, which is derived from (3.5) when $\delta = 2$ and $v = 2$. The GJR-GARCH model assumes a specific parametric form of conditional heteroscedasticity.

The third term that should be determined from (3.2) is the quantile (q_a), which is estimated based on the distribution of the standardized residuals. While numerous alternative methodologies for this estimation can be found in the literature, this research focuses only on three specific methodologies.

The first methodology for obtaining the quantile is to assume a particular distribution for the standardized residuals. When a distribution is chosen randomly to create the standardized residuals and many different simulations are implemented to forecast price paths in the future, Monte Carlo simulation is used. The most popular distribution assumption is normality, due to the assumptions of the Black–Scholes stock price model. Nonetheless, most time series in the financial world exhibit fat tails and skewness. Consequently, Student t distributions, skewed Student t distributions and generalized Pareto distributions (GPDs) can be used in addition to making assumptions about the standardized residuals.

Alternatively, the quantile of the distribution can be estimated directly without particular assumptions. This methodology (McNeil and Frey 2000) is called extreme value theory. Unlike Monte Carlo simulation, EVT is applied on the tails of the distribution of the standardized residuals. EVT is based on the assumption that the distribution of the standardized residuals above a threshold u follow, for example, a GPD.

According to McNeil and Frey (2000), the quantile of the distribution of the standardized residuals can be obtained using the following equation:

$$\hat{q}_a = -\left(\hat{\varepsilon}_{(k+1)} + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{\alpha}{\kappa/T}\right)^{-\hat{\xi}} - 1\right)\right), \tag{3.6}$$

where the number of observations in the tail is fixed to be $N = \kappa$, $\kappa \ll T$, which yields a threshold at the $(k+1)$th-order statistic. Then, if $\hat{\varepsilon}_{(1)} \geqslant \cdots \geqslant \hat{\varepsilon}_{(T)}$ are the ordered standardized residuals, the threshold is $\hat{\varepsilon}_{(\kappa+1)}$ and the GPD is as follows (McNeil and Frey 2000):

$$\hat{\varepsilon}_{(1)} - \hat{\varepsilon}_{(k+1)} \geqslant \cdots \geqslant \hat{\varepsilon}_{(k)} - \hat{\varepsilon}_{(k+1)}.$$

Finally, Barone-Adesi and Giannopoulos (2001) proposed the bootstrapping methodology named filtered historical simulation. This technique does not make any assumptions about the distribution of the residuals. FHS uses random draws with replacement from the standardized residuals created using the conditional mean and variance estimated parameters. This method is called semiparametric because it incorporates the characteristics of the empirical distribution within a simulation process.

3.2 Modeling the ten-steps-ahead forecasts

All of the methodologies described in Section 3.1 can be used to estimate short (one-day) and long (ten-day) VaR horizons. Based on the Basel regulatory framework, most financial institutions require a ten-day VaR. Thus, (3.2) can be modified as follows:

$$\mathrm{VaR}^{a}_{t+10} = \mu_{t+10} + q_a \sigma_{t+10}. \tag{3.7}$$

As described in Section 3.1, Monte Carlo simulation can be used to assume a particular distribution for the standardized residuals, assuming one of the following specific distributions applies: normal, Student t, skewed Student t or generalized error.

In addition, the EVT method (McNeil and Frey 2000) can be implemented by applying it on the tails of the distribution of the standardized residuals, generated from nonoverlapping ten-day returns.

Bootstrap methods (such as FHS) can also be implemented on ten-day distributions. The Barone-Adesi and Giannopoulos (2001) method can use random draws with replacement from the standardized residuals created from the empirical distribution of real ten-day returns.

4 BACKTESTING METHODOLOGIES

Several backtesting methods are currently used to evaluate VaR models. The Basel Committee has specific rules about backtesting strategies. According to regulations, backtesting should be based on 250 one-day VaR estimates. As mentioned above, in this paper the Kupiec and Christoffersen methods of examining the stability of VaR models at different probabilities are implemented; the backtesting techniques are used to test VaR accuracy under two different horizons and two different probability levels.

Other backtesting techniques, such as the Basel traffic light test, are not analyzed in this study because one-tail tests examine only the underestimation of VaR estimates and not the overestimation.[1]

[1] See https://bit.ly/3CvystA.

4.1 Kupiec test

The Kupiec (1995) test is based on the number of failures of VaR estimates depending on the selected confidence level. As described above, the failures are defined as follows:

$$I_t^a = 1(R_t < \mathrm{VaR}_t^a). \tag{4.1}$$

The number of failures is assumed to follow a binomial distribution. Kupiec suggested testing the null hypothesis that the number of failures over several days converges to the selected confidence level by using the following likelihood ratio statistic:

$$\mathrm{LR} = 2\log\left[\left(1-\frac{x}{n}\right)^{n-x}\left(\frac{x}{n}\right)^{x}\right] - 2\log[(1-a)^{n-x}a^{x}], \tag{4.2}$$

where n is the number of trials, x is the number of failures and a is the selected quantile. Under the hypothesis, the likelihood ratio statistic has an asymptotic χ^2 distribution with one degree of freedom. The Kupiec test is more efficient on large samples with more than four years of observations.

4.2 Christoffersen test

Christoffersen (1998) developed a different model to evaluate VaR. This method examines whether the failures in VaR estimation are independent. The Christoffersen test is estimated by the following likelihood ratio statistic:

$$\mathrm{LR} = -2\ln\left(\frac{(1-\pi)^{n_{00}+n_{10}}(\pi)^{n_{01}+n_{11}}}{(1-\pi_0)^{n_{00}}(\pi_0)^{n_{01}}(1-\pi_1)^{n_{10}}(\pi_1)^{n_{11}}}\right), \tag{4.3}$$

where

$$\pi_0 = \frac{n_{01}}{n_{00}+n_{01}}, \qquad \pi_1 = \frac{n_{11}}{n_{10}+n_{11}}, \qquad \pi = \frac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}},$$

n_{00} is the number of periods with no failures followed by a period with no failures, n_{10} is the number of periods with failures followed by a period with no failures, n_{01} is the number of periods with no failures followed by a period with failures and n_{11} is the number of periods with failures followed by a period with failures. The asymptotic null distribution is a χ^2 distribution with one degree of freedom. Under the null hypothesis, the probabilities should be equal: $\pi_0 = \pi_1$. The test largely depends on the frequency with which consecutive exceedances are experienced. Only when failures of VaR estimations show independence can the test be successful.

TABLE 1 Descriptive statistics of FTSE 100 daily returns.

Mean	0.00013
Median	0.00053
Standard deviation	0.01193
Min	−0.09266
Max	0.09384
Skewness	−0.1468
Kurtosis	10.87109

5 DATA AND RESULTS

5.1 Data

In this paper the FTSE 100 index time series is used for a sample of 12 years (January 1, 2005–January 1, 2017). The FTSE 100 is a share index of the 100 companies listed on the London Stock Exchange with the highest market capitalization. The FTSE 100 data were collected from Bloomberg.

First, the log returns r_t were estimated from the prices P_t:

$$r_t = \frac{\ln(P_t)}{\ln(P_{t-1})}. \tag{5.1}$$

The returns were examined and some skewness and kurtosis were identified in their distribution. Thus, their characteristics converge more to a Student t or skewed Student t distribution than a normal distribution. The descriptive statistics are presented in Table 1.

5.2 Results

The first two years of the daily time series were used to calculate the initial VaR estimates. The rest of the data served to evaluate the performance of various VaR methodologies for two confidence levels (2.5% and 1%). Moreover, to analyze the performance of the models on longer horizons, the FTSE 100 daily returns were transformed to ten-day observations using nonoverlapping returns for the backtesting implementation.

In our analysis, six different ARMA-GARCH combinations, four different distributions (normal, Student's t, skewed Student's t and GED) and four different methodologies (FHS, EVT, Monte Carlo simulation and historical simulation) were employed to estimate our VaR calculations.

Kupiec (1995) and Christoffersen (1998) methods were implemented to give a better understanding of the tested models by analyzing their performance for two probability levels and horizons together; the Kupiec and Christoffersen null hypotheses

TABLE 2 Kupiec test for one-day VaR estimates on 99% and 97.5% significance levels. [Table continues on next page.]

Method	Model	Distribution	Kupiec test 99%	Kupiec test 97.5%
FHS	ARMA(0,0)-GARCH(1,1) 99%	Normal	0.197	0.778
FHS	ARMA(0,0)-GARCH(1,1) 99%	SD	0.593	0.589
FHS	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.593	0.589
FHS	ARMA(0,0)-GARCH(1,1) 99%	GED	0.359	0.681
FHS	ARMA(1,0)-GARCH(1,1) 99%	Normal	0.734	0.878
FHS	ARMA(1,0)-GARCH(1,1) 99%	SD	0.467	0.980
FHS	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.593	0.681
FHS	ARMA(1,0)-GARCH(1,1) 99%	GED	0.269	0.878
FHS	ARMA(0,0)-EGARCH(1,1) 99%	Normal	0.734	0.114
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SD	0.269	0.114
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.197	0.223
FHS	ARMA(0,0)-EGARCH(1,1) 99%	GED	0.593	0.144
FHS	ARMA(1,0)-EGARCH(1,1) 99%	Normal	0.359	0.089
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SD	0.467	0.069
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.593	0.223
FHS	ARMA(1,0)-EGARCH(1,1) 99%	GED	0.593	0.144
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	0.197	0.632
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	0.269	0.980
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.197	0.980
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	0.141	0.819
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	0.197	0.632
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	0.467	0.819
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.467	0.546
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	0.141	0.632
MC	ARMA(0,0)-GARCH(1,1) 99%	Normal	**0.000**	**0.002**
MC	ARMA(0,0)-GARCH(1,1) 99%	SD	0.066	**0.022**
MC	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.098	0.330
MC	ARMA(0,0)-GARCH(1,1) 99%	GED	**0.011**	0.053
MC	ARMA(1,0)-GARCH(1,1) 99%	Normal	**0.000**	**0.002**
MC	ARMA(1,0)-GARCH(1,1) 99%	SD	**0.011**	**0.012**
MC	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.141	0.466
MC	ARMA(1,0)-GARCH(1,1) 99%	GED	**0.004**	**0.022**
MC	ARMA(0,0)-EGARCH(1,1) 99%	Normal	**0.000**	**0.000**
MC	ARMA(0,0)-EGARCH(1,1) 99%	SD	**0.004**	**0.000**
MC	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.197	**0.030**
MC	ARMA(0,0)-EGARCH(1,1) 99%	GED	**0.004**	**0.001**

TABLE 2 Continued.

Method	Model	Distribution	Kupiec test 99%	Kupiec test 97.5%
MC	ARMA(1,0)-EGARCH(1,1) 99%	Normal	**0.000**	**0.000**
MC	ARMA(1,0)-EGARCH(1,1) 99%	SD	**0.002**	**0.000**
MC	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.467	**0.022**
MC	ARMA(1,0)-EGARCH(1,1) 99%	GED	**0.002**	**0.000**
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	**0.000**	**0.000**
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	**0.001**	**0.001**
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.098	0.180
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	**0.002**	**0.003**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	**0.000**	**0.000**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	**0.001**	**0.001**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.141	0.273
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	**0.001**	**0.001**
EVT	ARMA(0,0)-GARCH(1,1) 99%	Normal	0.098	0.819
EVT	ARMA(0,0)-GARCH(1,1) 99%	SD	0.141	0.919
EVT	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.098	0.819
EVT	ARMA(0,0)-GARCH(1,1) 99%	GED	0.098	0.819
EVT	ARMA(1,0)-GARCH(1,1) 99%	Normal	0.066	0.778
EVT	ARMA(1,0)-GARCH(1,1) 99%	SD	0.141	0.919
EVT	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.141	0.819
EVT	ARMA(1,0)-GARCH(1,1) 99%	GED	0.141	0.919
EVT	ARMA(0,0)-EGARCH(1,1) 99%	Normal	0.593	0.466
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SD	0.467	0.180
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.359	0.144
EVT	ARMA(0,0)-EGARCH(1,1) 99%	GED	0.467	0.394
EVT	ARMA(1,0)-EGARCH(1,1) 99%	Normal	0.359	0.466
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SD	0.359	0.330
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.359	0.330
EVT	ARMA(1,0)-EGARCH(1,1) 99%	GED	0.359	0.466
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	0.467	0.394
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	0.467	0.632
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.467	0.632
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	0.359	0.466
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	0.467	0.273
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	0.734	0.546
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.886	0.466
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	0.359	0.330

SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution. The bold p-values indicate a failure in the performing test.

TABLE 3 Kupiec test for nonoverlapping ten-day VaR estimates on 99% and 97.5% significance levels. [Table continues on next page.]

Method	Model	Distribution	Kupiec test 99%	Kupiec test 97.5%
FHS	ARMA(0,0)-GARCH(1,1) 99%	Normal	0.388	0.168
FHS	ARMA(0,0)-GARCH(1,1) 99%	SD	0.388	0.510
FHS	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.768	0.587
FHS	ARMA(0,0)-GARCH(1,1) 99%	GED	0.388	0.781
FHS	ARMA(1,0)-GARCH(1,1) 99%	Normal	0.388	0.306
FHS	ARMA(1,0)-GARCH(1,1) 99%	SD	0.388	0.903
FHS	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.768	0.781
FHS	ARMA(1,0)-GARCH(1,1) 99%	GED	0.388	0.781
FHS	ARMA(0,0)-EGARCH(1,1) 99%	Normal	0.166	0.781
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SD	0.768	0.781
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.733	0.320
FHS	ARMA(0,0)-EGARCH(1,1) 99%	GED	0.768	0.903
FHS	ARMA(1,0)-EGARCH(1,1) 99%	Normal	0.061	0.510
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SD	0.388	0.306
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.388	0.587
FHS	ARMA(1,0)-EGARCH(1,1) 99%	GED	**0.020**	0.086
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	—	0.320
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	—	0.320
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.733	0.781
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	—	0.320
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	—	0.781
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	0.273	0.903
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.273	0.781
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	—	0.781
MC	ARMA(0,0)-GARCH(1,1) 99%	Normal	0.388	0.587
MC	ARMA(0,0)-GARCH(1,1) 99%	SD	0.388	0.320
MC	ARMA(0,0)-GARCH(1,1) 99%	SSTD	—	**0.043**
MC	ARMA(0,0)-GARCH(1,1) 99%	GED	0.388	0.320
MC	ARMA(1,0)-GARCH(1,1) 99%	Normal	0.388	0.587
MC	ARMA(1,0)-GARCH(1,1) 99%	SD	0.388	0.320
MC	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.273	0.139
MC	ARMA(1,0)-GARCH(1,1) 99%	GED	0.388	0.587
MC	ARMA(0,0)-EGARCH(1,1) 99%	Normal	0.733	0.903
MC	ARMA(0,0)-EGARCH(1,1) 99%	SD	0.733	0.320
MC	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	—	**0.008**
MC	ARMA(0,0)-EGARCH(1,1) 99%	GED	0.733	0.587

TABLE 3 Continued.

Method	Model	Distribution	Kupiec test 99%	Kupiec test 97.5%
MC	ARMA(1,0)-EGARCH(1,1) 99%	Normal	0.768	0.781
MC	ARMA(1,0)-EGARCH(1,1) 99%	SD	0.273	**0.043**
MC	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.733	0.587
MC	ARMA(1,0)-EGARCH(1,1) 99%	GED	**0.020**	0.086
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	—	—
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	—	—
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	—	**0.008**
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	—	**0.008**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	—	**0.008**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	0.273	**0.008**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.273	**0.008**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	—	**0.043**
EVT	ARMA(0,0)-GARCH(1,1) 99%	Normal	0.388	0.587
EVT	ARMA(0,0)-GARCH(1,1) 99%	SD	0.388	0.320
EVT	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.768	0.139
EVT	ARMA(0,0)-GARCH(1,1) 99%	GED	0.388	0.320
EVT	ARMA(1,0)-GARCH(1,1) 99%	Normal	0.768	0.587
EVT	ARMA(1,0)-GARCH(1,1) 99%	SD	0.768	0.587
EVT	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.273	**0.043**
EVT	ARMA(1,0)-GARCH(1,1) 99%	GED	0.768	0.587
EVT	ARMA(0,0)-EGARCH(1,1) 99%	Normal	0.768	0.587
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SD	0.166	0.510
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.733	0.139
EVT	ARMA(0,0)-EGARCH(1,1) 99%	GED	0.768	0.781
EVT	ARMA(1,0)-EGARCH(1,1) 99%	Normal	0.768	0.587
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SD	0.166	0.781
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.733	0.139
EVT	ARMA(1,0)-EGARCH(1,1) 99%	GED	0.768	0.587
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	—	—
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	—	—
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.733	**0.043**
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	—	**0.008**
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	—	—
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	0.273	**0.043**
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.273	**0.008**
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	—	—

SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution. The bold p-values indicate a failure in the performing test.

TABLE 4 Independence test for one-day VaR estimates on 99% and 97.5% significance levels. [Table continues on next page.]

Method	Model	Distribution	Ind. test 99%	Ind. test 97.5%
FHS	ARMA(0,0)-GARCH(1,1) 99%	Normal	0.081	0.879
FHS	ARMA(0,0)-GARCH(1,1) 99%	SD	0.102	0.759
FHS	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.102	0.759
FHS	ARMA(0,0)-GARCH(1,1) 99%	GED	0.098	0.824
FHS	ARMA(1,0)-GARCH(1,1) 99%	Normal	0.097	0.222
FHS	ARMA(1,0)-GARCH(1,1) 99%	SD	0.102	0.579
FHS	ARMA(1,0)-GARCH(1,1) 99%	SSTD	**0.011**	0.455
FHS	ARMA(1,0)-GARCH(1,1) 99%	GED	0.091	0.545
FHS	ARMA(0,0)-EGARCH(1,1) 99%	Normal	0.705	0.281
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SD	0.370	0.281
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.289	0.474
FHS	ARMA(0,0)-EGARCH(1,1) 99%	GED	0.633	0.340
FHS	ARMA(1,0)-EGARCH(1,1) 99%	Normal	0.458	0.140
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SD	0.548	0.178
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.633	0.398
FHS	ARMA(1,0)-EGARCH(1,1) 99%	GED	0.633	0.302
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	0.289	0.879
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	0.370	0.883
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.289	0.883
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	0.218	0.944
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	0.289	0.879
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	0.548	0.826
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.548	0.826
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	0.218	0.879
MC	ARMA(0,0)-GARCH(1,1) 99%	Normal	**0.000**	**0.005**
MC	ARMA(0,0)-GARCH(1,1) 99%	SD	**0.046**	0.053
MC	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.198	0.622
MC	ARMA(0,0)-GARCH(1,1) 99%	GED	**0.014**	0.144
MC	ARMA(1,0)-GARCH(1,1) 99%	Normal	**0.000**	**0.007**
MC	ARMA(1,0)-GARCH(1,1) 99%	SD	**0.014**	**0.032**
MC	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.254	0.569
MC	ARMA(1,0)-GARCH(1,1) 99%	GED	**0.006**	0.053
MC	ARMA(0,0)-EGARCH(1,1) 99%	Normal	**0.000**	**0.000**
MC	ARMA(0,0)-EGARCH(1,1) 99%	SD	**0.015**	**0.000**
MC	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.316	0.087
MC	ARMA(0,0)-EGARCH(1,1) 99%	GED	**0.015**	**0.002**

TABLE 4 Continued.

Method	Model	Distribution	Ind. test 99%	Ind. test 97.5%
MC	ARMA(1,0)-EGARCH(1,1) 99%	Normal	**0.000**	**0.000**
MC	ARMA(1,0)-EGARCH(1,1) 99%	SD	**0.009**	**0.000**
MC	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.490	0.053
MC	ARMA(1,0)-EGARCH(1,1) 99%	GED	**0.009**	**0.001**
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	**0.000**	**0.001**
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	**0.003**	**0.002**
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.160	0.270
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	**0.005**	**0.009**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	**0.000**	**0.000**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	**0.003**	**0.002**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.218	0.547
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	**0.003**	**0.004**
EVT	ARMA(0,0)-GARCH(1,1) 99%	Normal	0.058	0.944
EVT	ARMA(0,0)-GARCH(1,1) 99%	SD	0.070	0.952
EVT	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.058	0.619
EVT	ARMA(0,0)-GARCH(1,1) 99%	GED	0.058	0.944
EVT	ARMA(1,0)-GARCH(1,1) 99%	Normal	**0.046**	0.879
EVT	ARMA(1,0)-GARCH(1,1) 99%	SD	0.070	0.952
EVT	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.070	0.944
EVT	ARMA(1,0)-GARCH(1,1) 99%	GED	0.070	0.952
EVT	ARMA(0,0)-EGARCH(1,1) 99%	Normal	0.527	0.764
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SD	0.490	0.405
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.439	0.340
EVT	ARMA(0,0)-EGARCH(1,1) 99%	GED	0.490	0.695
EVT	ARMA(1,0)-EGARCH(1,1) 99%	Normal	0.439	0.764
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SD	0.439	0.622
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.439	0.622
EVT	ARMA(1,0)-EGARCH(1,1) 99%	GED	0.439	0.764
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	0.548	0.695
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	0.548	0.722
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.548	0.722
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	0.458	0.764
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	0.548	0.547
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	0.705	0.658
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.755	0.764
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	0.458	0.622

SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution. The bold p-values indicate a failure in the performing test.

TABLE 5 Independence test for nonoverlapping ten-day VaR estimates on 99% and 97.5% significance levels. [Table continues on next page.]

Method	Model	Distribution	Ind. test 99%	Ind. test 97.5%
FHS	ARMA(0,0)-GARCH(1,1) 99%	Normal	0.646	0.256
FHS	ARMA(0,0)-GARCH(1,1) 99%	SD	0.646	0.619
FHS	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.923	0.779
FHS	ARMA(0,0)-GARCH(1,1) 99%	GED	0.646	0.787
FHS	ARMA(1,0)-GARCH(1,1) 99%	Normal	0.646	0.423
FHS	ARMA(1,0)-GARCH(1,1) 99%	SD	0.646	0.857
FHS	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.923	0.787
FHS	ARMA(1,0)-GARCH(1,1) 99%	GED	0.646	0.787
FHS	ARMA(0,0)-EGARCH(1,1) 99%	Normal	0.346	0.787
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SD	0.923	0.787
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.928	0.572
FHS	ARMA(0,0)-EGARCH(1,1) 99%	GED	0.923	0.857
FHS	ARMA(1,0)-EGARCH(1,1) 99%	Normal	0.150	0.619
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SD	0.646	0.423
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.646	0.779
FHS	ARMA(1,0)-EGARCH(1,1) 99%	GED	0.054	0.138
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	0.079	0.078
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	0.079	0.078
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.928	**0.033**
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	0.079	0.078
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	0.079	0.380
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	0.546	0.293
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.546	**0.033**
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	0.079	0.380
MC	ARMA(0,0)-GARCH(1,1) 99%	Normal	0.646	0.779
MC	ARMA(0,0)-GARCH(1,1) 99%	SD	0.646	0.572
MC	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.079	0.128
MC	ARMA(0,0)-GARCH(1,1) 99%	GED	0.646	0.572
MC	ARMA(1,0)-GARCH(1,1) 99%	Normal	0.646	0.779
MC	ARMA(1,0)-GARCH(1,1) 99%	SD	0.646	0.572
MC	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.546	0.322
MC	ARMA(1,0)-GARCH(1,1) 99%	GED	0.646	0.779
MC	ARMA(0,0)-EGARCH(1,1) 99%	Normal	0.928	0.857
MC	ARMA(0,0)-EGARCH(1,1) 99%	SD	0.928	0.572
MC	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.079	**0.030**
MC	ARMA(0,0)-EGARCH(1,1) 99%	GED	0.928	0.779

TABLE 5 Continued.

Method	Model	Distribution	Ind. test 99%	Ind. test 97.5%
MC	ARMA(1,0)-EGARCH(1,1) 99%	Normal	0.923	0.787
MC	ARMA(1,0)-EGARCH(1,1) 99%	SD	0.546	0.128
MC	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.928	0.177
MC	ARMA(1,0)-EGARCH(1,1) 99%	GED	0.054	0.138
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	0.079	**0.002**
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	0.079	**0.002**
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.079	**0.030**
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	0.079	**0.030**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	0.079	**0.030**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	0.546	**0.030**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.546	**0.030**
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	0.079	0.128
EVT	ARMA(0,0)-GARCH(1,1) 99%	Normal	0.646	0.779
EVT	ARMA(0,0)-GARCH(1,1) 99%	SD	0.646	0.572
EVT	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0.923	0.322
EVT	ARMA(0,0)-GARCH(1,1) 99%	GED	0.646	0.572
EVT	ARMA(1,0)-GARCH(1,1) 99%	Normal	0.923	0.779
EVT	ARMA(1,0)-GARCH(1,1) 99%	SD	0.923	0.779
EVT	ARMA(1,0)-GARCH(1,1) 99%	SSTD	0.546	0.128
EVT	ARMA(1,0)-GARCH(1,1) 99%	GED	0.923	0.779
EVT	ARMA(0,0)-EGARCH(1,1) 99%	Normal	0.923	0.779
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SD	0.346	0.619
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0.928	**0.022**
EVT	ARMA(0,0)-EGARCH(1,1) 99%	GED	0.923	0.787
EVT	ARMA(1,0)-EGARCH(1,1) 99%	Normal	0.923	0.779
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SD	0.346	0.787
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	0.928	**0.022**
EVT	ARMA(1,0)-EGARCH(1,1) 99%	GED	0.923	0.779
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	0.079	**0.002**
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	0.079	**0.002**
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0.928	0.128
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	0.079	**0.030**
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	0.079	**0.002**
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	0.546	0.128
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	0.546	**0.030**
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	0.079	**0.002**

SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution. The bold p-values indicate a failure in the performing test.

were considered to hold when the p-value was above 5%. The p-values of both tests on all the tested models are presented in Tables 2–5. The p-values in bold indicate a failure in the performing test.

For one-day VaR estimates, the Kupiec test accepts all the models that use FHS and EVT methods (Table 2). Meanwhile, the Kupiec test null hypothesis for the Monte Carlo methodology is rejected for most of the models tested. The same results are observed for both significance levels. Further, the Christoffersen test presents similar observations (Table 4); among the models that use FHS and EVT, all but two are accepted by the Christoffersen test. The Christoffersen test null hypothesis is rejected for most models that use the Monte Carlo method.

Further, the Kupiec test presents several models as efficient, even those using the Monte Carlo method, for ten-day VaR estimates (Table 3) and both confidence levels (99%, 97.5%). On the other hand, several failures of the test are observed for GJR-GARCH models, when each of the different VaR methods (FHS, EVT or Monte Carlo simulation) are used. In addition, the Christoffersen test (Table 5) suggests all the models as efficient for the 99% confidence level. However, failures are observed for GJR-GARCH models at the 97.5% confidence level.

It can be seen that the Kupiec and Christoffersen tests yielded several efficient models for the different horizons tested (one-day, ten-day nonoverlapped). The general conclusion is that, for daily observations, the EVT and FHS methodologies were deemed to be efficient in several trials, while the Monte Carlo method appears to fail for both backtesting techniques. On the other hand, all methodologies (FHS, EVT and Monte Carlo simulation) were deemed to be efficient when used on ten-day horizons; one exception to this was the GJR-GARCH model, which presented several failures for ten-day levels in both the Kupiec and Christoffersen tests.

According to the research outcomes, ten-day VaR estimates based on nonoverlapping portfolio returns can present similar behavior to one-day VaR calculations on many occasions. However, the Monte Carlo method that presents backtesting failures on daily horizons can be proven to be adequate based on Kupiec and Christoffersen tests when longer horizons were used. One reason for the above finding is the limited number of ten-day observations compared with daily ones, which might not be enough to judge the efficiency of the models.

6 CONCLUSIONS

The decision-making process involved in selecting suitable risk metrics is often difficult, depending on various circumstances. This paper focuses on the evaluation of market risk (ie, VaR) models over longer forecast horizons and examines the behavior of such models at different confidence levels. The results of this study show that changing the forecast horizon can affect the efficiency of some VaR models. Thus, the assumption of most financial institutions that backtesting results based on daily observations can be generalized for ten-day risk estimations might not always be accurate.

This finding is important and requires further testing concerning the assumptions that financial institutions make when VaR values are estimated on longer horizons. Further, this research found that the GJR-GARCH model presents several failures at ten-day levels in both the Kupiec and Christoffersen tests, in contrast to the other models that were used. Finally, the results suggest that some models' efficiency can be affected by the selection of the significance level.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1998). Coherent measures of risk. *Mathematical Finance* **9**(3), 203–228 (https://doi.org/10.1111/1467-9965.00068).

Barone-Adesi, G., and 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).

Basel Committee on Banking Supervision (1996). Supervisory framework for the use of backtesting in conjunction with the internal models approach to market risk capital requirements. Standards Document, Bank for Internatonal Settlements. URL: www.bis.org/publ/bcbs22.pdf.

Basel Committee on Banking Supervision (2016). Minimum capital requirments for market risk. Standards Document, Bank for International Settlements. URL: www.bis.org/bcbs/publ/d457.pdf.

Basel Committee on Banking Supervision (2017). Pillar 3 disclosure requirements: consolidated and enhanced framework. Standards Document, Bank for International Settlements. URL: www.bis.org/bcbs/publ/d400.pdf.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3), 307–327 (https://doi.org/10.1016/0304-4076(86)90063-1).

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review* **39**(4), 841–862 (https://doi.org/10.2307/2527341).

Danielsson, J., and De Vries, C. (2000). Value at risk and extreme returns. *Annales d'Economie et de Statistique* **60**, 239–270 (https://doi.org/10.2307/20076262).

Fissler, T. Z. (2016). Expected shortfall is jointly elicitable with value-at-risk: implications for backtesting. *Risk* **29**(1), 58–61.

Glosten, L., Jagannathan, R., and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**(5), 1779–1801 (https://doi.org/10.1111/j.1540-6261.1993.tb05128.x).

Hentschel, L. (1995). All in the family: nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics* **39**(1), 71–104 (https://doi.org/10.1016/0304-405X(94)00821-H).

Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives* **3**(2), 73–84 (https://doi.org/10.3905/jod.1995.407942).

McNeil, A., and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* **7**(3–4), 271–300 (https://doi.org/10.1016/S0927-5398(00)00012-8).

Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* **59**(2), 347–370 (https://doi.org/10.2307/2938260).

Sun, H., Nelken, I., Han, G., and Guo, J. (2009). Error of VaR by overlapped intervals. *Asia Risk* (April), 50–55. URL: www.risk.net/risk-management/1500264/error-var-overlapping-intervals.

Weber, S. (2006). Distribution invariant risk measures, information, and dynamic consistency. *Mathematical Finance* **16**(2), 419–441 (https://doi.org/10.1111/j.1467-9965.2006.00277.x).