Journal of Operational Risk **17**(1), 51–79
DOI: 10.21314/JOP.2021.014

Research Paper

Evaluation of backtesting on risk models based on data envelopment analysis

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 June 8, 2021; revised September 5, 2021; accepted October 4, 2021)

ABSTRACT

In this study, different value-at-risk (VaR) models, which are used to measure market risk, are analyzed under different estimation approaches and backtested with an alternative strategy. The methodologies examined include filtered historical simulation, extreme value theory, Monte Carlo simulation and historical simulation. Autoregressive-moving-average and generalized-autoregressive-conditional-heteroscedasticity models are used to estimate VaR. 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. Data envelopment analysis, which investigates the efficiency of VaR models using a number of different parameters, is carried out in lieu of standard backtesting techniques. This study shows that, for short horizons, some approaches underestimate VaR. However, a sufficient number of models present violation estimates that almost converge to the desired ones. Surprisingly, aside from historical simulation and some extreme value theory models, overlapping returns tend to yield conservative ten-day

Corresponding author: G. Kontaxis

Print ISSN 1744-6740 | Online ISSN 1755-2710

VaR estimations for most models; in cases of nonoverlapping returns, the results are satisfactory.

Keywords: value-at-risk (VaR); backtesting; data envelopment analysis (DEA); efficient model selection; risk estimation.

1 INTRODUCTION

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. Market risk, which is related to unexpected changes in returns over short time horizons, is the main research focus of this paper.

One of the most important components of determining risk is the selection of sufficient 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 unpredictable losses, autocorrelation, heteroscedasticity and asymmetric tails on their returns that require modeling capable of estimating coherent and sufficient risk measures.

Value-at-risk (VaR) models are 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 projected VaR estimations, and the consistency and reliability of VaR calculations can be examined by means of various alternative approaches. Among the most well-known backtesting techniques are the proportion of failures (POF) test (Kupiec 1995), which examines the frequency of losses in excess of the VaR; the independence test (Christoffersen 1998); and the Basel traffic light test (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. However, if violations occur more or less frequently they are indicative of an underestimation or overestimation of risk. While 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 turbulence in the market. 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%. However, most financial institutions hold assets with very long maturity horizons for which a lower probability level, such is 0.05%, is preferable. 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 requires that VaR be reported for ten-day horizons for liquidity purposes. There are also situations in which yearly horizons may be desirable, eg, long-term liabilities (pension funds, government bonds). VaR estimation at long-term horizons can be more complex than at daily horizons due to data availability issues. Numerous techniques are presented in the literature for risk metric calculation with relatively small time series, each with several different advantages and disadvantages.

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 backtesting processes, such as Kupiec's test and the independence test, are designed to test risk models' reliance on specific conditions. An alternative approach, suggested by this paper, involves constructing efficient model sets using data envelopment analysis (DEA) (Charnes *et al* 1978) to provide an overall evaluation of market risk models. The main aim of this paper is to evaluate risk models utilizing the DEA methodology, taking different probability levels and horizons into account. The results of this process for specific configurations are examined to determine DEA's suitability for decision-making regarding risk metrics. One of the cases investigated involves a financial institution with a long position on the Financial Times Stock Exchange 100 (FTSE 100) equity index. Several variants of autoregressive-moving-average generalized-autoregressive-conditional-heteroscedasticity (ARMA-GARCH) models are examined for a broad range of marginal distributions and four different estimation techniques (filtered historical simulation (FHS), extreme value theory (EVT), Monte Carlo simulation and historical simulation).

The contributions of this paper are twofold. First, we employ the DEA methodology to examine its suitability for risk model decision-making – an issue that has not been adequately addressed in the existing literature – and its potential for providing an alternative, sophisticated means of evaluating market risk models' stability. Second, several market risk models are evaluated by combining selected VaR functions (ie, different ARMA-GARCH combinations), marginal distributions (normal, Student t and skewed Student t distributions and the generalized error distribution (GED)) and different horizons (one day and ten days) over a set of extreme probability levels. Although the literature has explored various techniques for one-day and ten-day VaR estimation (see, for example, Kontaxis and Tsolas 2021), the stability of risk metrics on longer horizons has yet to be calculated and evaluated using alternative methodologies. This research provides an in-depth analysis of risk metrics' effectiveness on longer horizons using various approaches and manipulations of data. Further, various nonoverlapping data were compared with overlapping data across numerous alternative VaR techniques to assess the models' compatibility on long forecast horizons and using alternative data aggregation techniques.

The paper is structured as follows. In Section 2 VaR risk metrics are presented and analyzed. In addition, 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 on different time horizons. In Section 4 DEA is introduced to assess the selected risk models' adequacy under various settings. Section 5 presents a detailed description of the empirical results. Section 6 provides the research conclusions.

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 be modeled on 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 (Kontaxis and Tsolas 2021).

VaR tends to focus on large price drops in the market. However, risk is influenced by deviations in future portfolio returns, and depending on the positions in the market even a positive unexpected jump can result in losses. An adequate estimation of risk metrics that eliminates over- or underestimations is thus always essential.

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 time 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 100% worst case with $a \in (0, 1)$ on some fixed time horizon:

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

where R refers to the returns of an asset portfolio and a refers to the confidence level. The most well-known process for VaR estimation 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 (Kontaxis and Tsolas 2021).

2.2 Backtesting

In order 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 (Kontaxis and Tsolas 2021) are based on the following process:

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

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.

Most backtesting methodologies are based on the above implication of accuracy. A well-known backtesting process proposed by Kupiec (1995) is based on the number of failures, which is assumed to follow a binomial distribution. However, this methodology should be implemented using many different confidence levels to determine the model's overall accuracy.

This paper does not examine standard backtesting methodologies. Instead, the DEA algorithm is used to analyze VaR accuracy for two different horizons (one-step and ten-steps ahead). The purpose of this study is to determine whether DEA is an adequate means of examining the efficiency of different VaR models that use various parameterizations, which is an aspect that standard backtesting techniques in the industry do not investigate.

3 VALUE-AT-RISK METHODOLOGIES

VaR can be estimated with many alternative methodologies, each of which have particular strengths and weaknesses. VaR estimation can be a difficult computational task due to the variety of factors that have an impact on its evaluation. 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 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) and Glosten–Jagannathan–Runkle GARCH (GJR-GARCH) are used to address asymmetry in the data. Further, a large number of distributions are examined for the GARCH model fitting. Thus, normal, Student t and skewed Student t distributions and the generalized error distribution (GED) are selected to represent the properties of the data. In contrast with 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 are combined with three alternative methodologies to estimate VaR: the FHS model presented by Barone-Adesi and Giannopoulos (2001), which is based on using random draws with replacements from the standardized residuals; Monte Carlo simulation, a popular methodology in which an assumption about the residuals' distribution is made; and EVT, which was developed by McNeil and Frey (2000) and makes an assumption about the distribution of the tails of the standardized residuals.

3.1 Modeling of 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 independent and identically distributed (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.

We assume the following model of returns:

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

where μ_t and σ_t represent the mean and the 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, VaR can be estimated using the following formula:

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

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 to be an ARMA model, given by

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

where $z_t = \varepsilon_t \sigma_t$, 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). EGARCH (Nelson 1991) is among the most popular for modeling asymmetry. EGARCH 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 last component 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, however, is normality. The cumulative distribution function for a normal distribution is

$$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x - \mu}{\sigma\sqrt{2}}\right)\right]. \tag{3.6}$$

Nonetheless, most time series in the financial world exhibit fat tails and skewness. Consequently, Student t distributions, skewed Student t distributions and the GED 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 follows, for example, a GPD:

$$G_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi y}{\beta}\right)^2 & \text{if } \xi \neq 0, \\ 1 - \exp\left(\dfrac{-y}{\beta}\right) & \text{if } \xi = 0, \end{cases} \tag{3.7}$$

where β represents the scale parameter, y the standardized returns and ξ the shape. According to McNeil and Frey (2000) the quantile of the distribution of the standardized residuals can be obtained by using the following equation:

$$\hat{q}_a = -\left(\hat{\varepsilon}_{(k+1)} + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{\alpha}{\kappa/T}\right)^{-\hat{\xi}} - 1\right)\right), \tag{3.8}$$

where the number of observations in the tail is fixed to be $N = \kappa$, $\kappa \ll T$, yielding 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:

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

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 of ten-steps-ahead forecasts

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

$$\mathrm{VaR}^a_{t+10} = \mu_{t+10} + q_a \sigma_{t+10}. \tag{3.9}$$

Monte Carlo simulation can create a ten-steps-ahead distribution of standardized residuals, assuming one of the following specific distributions applies: normal, Student t, skewed Student t or GED.

In addition, Danielsson and De Vries (2000) described a methodology for estimating ten-steps-ahead distributions based on the following:

$$\hat{q}_a = -(10)^{\hat{\xi}^{(\mathrm{H})}} \hat{\varepsilon}_{(\kappa+1)} \left(\frac{a}{k/T}\right)^{-\hat{\xi}^{(\mathrm{H})}}, \tag{3.10}$$

According to this method, VaR can be estimated using the term $1/\xi$, where ξ can be calculated from Hill (1975) estimators

$$\hat{\xi}^{(\mathrm{H})} = \frac{1}{K} \sum_{j=1}^{k} \log(\hat{\varepsilon}_{(k)}) - \log(\hat{\varepsilon}_{(k+1)}). \tag{3.11}$$

Bootstrap methods can also be used to estimate ten-steps-ahead distributions. Barone-Adesi and Giannopoulos (2001) proposed the simulation of ten-steps-ahead pathways based on the empirical distribution of the returns. This method formulates the conditional volatility, and the forecasts are obtained recursively.

3.3 Alternative methodologies in the literature

Several alternative VaR techniques – such as bootstrapping, EVT with copulas and exponential weighted averages – appear promising as well and could be tested with the backtesting method proposed in this study. 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. Copulas are not easily calculated through risk factors, but they are regarded as a sophisticated method by which to model VaR. 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 our research.

4 DATA ENVELOPMENT ANALYSIS

Several backtesting methods are currently used to evaluate VaR models. The Basel Committee has specific rules about backtesting strategies. According to the regulations, backtesting should be based on 250 one-day VaR estimates. As mentioned above, in this paper an alternative method (DEA) that examines the stability of VaR models under different probability levels is implemented; the DEA approach is used to test VaR accuracy for two different horizons.

DEA is a nonparametric method for estimating production frontiers. More specifically, it can be used for the empirical measurement of the productive efficiency of decision-making units (DMUs).

The Charnes–Cooper–Rhodes (CCR) DEA methodology was formally developed by Charnes *et al* (1978). According to DEA, for each DMU the virtual input and output weights are formed as follows:

$$\text{input} = u_1 x_{1o} + \cdots + u_m x_{mo}, \tag{4.1}$$

$$\text{output} = v_1 y_{1o} + \cdots + v_s y_{mo}. \tag{4.2}$$

Then, linear programming is used to determine the weight that maximizes the output/input ratio.

Optimal weights may vary for different DMUs. Thus, the weights in DEA are derived from the data rather than being fixed in advance. Each DMU is assigned a best set of weights, with values that can vary for different DMUs.

In the DEA method, the efficiency of each DMU is measured, and the needs of each DMU_j are evaluated after n optimizations. The following fractional programming model measures the relative efficiency score of an evaluated DMU, ie, DMU_o, $o \in 1, \dots, n$:

$$\text{maximize } \theta = \frac{v_1 y_{1o} + v_2 y_{2o} + \cdots + v_s y_{so}}{\rho_1 x_{1o} + \rho_2 x_{2o} + \cdots + \rho_m x_{mo}} \tag{4.3}$$

subject to

$$\begin{aligned} \frac{v_1 y_{1j} + \cdots + v_s y_{sj}}{\rho_1 x_{1j} + \cdots + \rho_m x_{mj}} &\leqslant 1, \quad j = 1, \dots, n, \\ \rho_1, \rho_2, \dots, \rho_m &\geqslant 0, \\ v_1, v_2, \dots, v_s &\geqslant 0, \end{aligned}$$

where the input weights (ρ_i, $i = 1, \dots, m$) and output weights (v_r, $r = 1, \dots, s$) are the variables.

The ratio of the output to the input should not exceed 1 for each DMU. The objective is to obtain weights ρ_i and v_r that maximize the DMU_o ratio.

Equation (4.3) can be transformed into the equivalent linear program using the transformation method presented by Charnes and Cooper (1962):

$$\text{maximize } \theta = \mu y_{1o} + \cdots + \mu y_{so} \tag{4.4}$$

subject to

$$\begin{aligned} k_1 \chi_{1o} + \cdots + k_m \chi_{mo} &= 1, \\ \mu_1 y_{1j} + \cdots + \mu_s y_{sj} &\leqslant k_1 \chi_{1j} + \cdots + k_m \chi_{mj}, \quad j = 1, \dots, n, \\ k_1, k_2, \dots, k_m &\geqslant 0, \\ \mu_1, \mu_2, \dots, \mu_s &\geqslant 0. \end{aligned}$$

TABLE 1 Descriptive statistics of FTSE 100 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

In this paper, the FTSE 100 index time series is used for a sample of 12 years. The FTSE 100 is a share index of the 100 companies listed on the London Stock Exchange with the highest market capitalization.

Initially, we estimated the log returns r_t from the prices P_t:

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

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

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 various horizons, the FTSE 100 daily returns were transformed to ten-day observations using overlapping returns (summing daily returns by changing only one day in each sample) and nonoverlapping returns for backtesting implementation. Most financial institutions prefer to use the overlapping method for ten-day VaR estimate backtesting. Cases in which the nonoverlapping method is generally selected are cases without data limitations. The Basel regulations require 250 days of VaR estimates for risk model backtesting, which amounts to 2500 observations (10 years) in cases in which the nonoverlapping method is used. One of the aims of this paper is to evaluate the performance of the various VaR models with regard to both overlapping and nonoverlapping ten-day returns.

In our analysis we employ six different ARMA-GARCH combinations, four different distributions (normal, Student t, skewed Student t and GED) and four different methodologies (FHS, EVT, Monte Carlo simulation and historical simulation) to estimate our VaR values.

Tables 2–4 depict the number of violations for the different VaR models used for one-day, overlapping ten-day and nonoverlapping ten-day observations. The violations are summarized for the significance levels of 99% and 97.5%. The assumed number of violations amounts to 25 and 63 for 99% and 97.5%, respectively. For the nonoverlapping ten-day returns, however, the assumed number of violations is 3 and 6 for the different confidence levels (99% and 97.5%) due to the aggregation of observations. As can be seen from the tables, the pattern of violations differs significantly among the different returns' horizons. For one-day returns, the number of violations tends to be higher than the assumed values, which implies an underestimation of the VaR; thus, the capital requirements will be lower than they should be. However, a sufficient number of models present a number of violations that nearly converges to the assumed number of violations.

In addition, in historical simulation and some EVT models, ten-day overlapping returns tend to create very conservative VaR estimations for most of the models examined in this research. This outcome is not typical, as most studies (see, for example, Sun *et al* 2009) suggest that overlapping returns create underestimated VaR values. In our study the overestimation can be related to a dependency on jump diffusion processes.

Finally, nonoverlapping ten-day returns present the expected number of violations in a large number of the models tested.

The difference between the assumed and the actual number of violations is not an appropriate indicator for a VaR model's accuracy. Thus, the DEA methodology was implemented to give a better understanding of the tested models by analyzing their performance under two probability levels together; the results are presented in Tables 5–7. Further, Figures 1–3 give a better understanding of the efficiency scores presented in those tables.

When the applying the DEA methodology, we considered differences between the observed and the assumed numbers of violations at particular significance levels as inputs, and we did not consider any explicit output. The efficiency scores summarize the performance of the models. High scores indicate the model's efficiency, while low scores reflect the inability to find weights that would enable the selected model to perform efficiently.

DEA analysis indicated a number of efficient models for each of the different tested horizons (one-day, ten-day overlapping and ten-day nonoverlapping). The EVT and FHS methodologies were deemed to be efficient in a number of trials, while the Monte Carlo method appeared to fail the DEA test in all cases.

More specifically, for one-day VaR models, DEA efficiency scores presented six efficient models, mainly those utilizing FHS and EVT methods. Meanwhile, DEA gave the models based on the Monte Carlo method the lowest scores among the various models tested. In contrast, DEA for ten-day VaR models with overlapping

TABLE 2 Total number of VaR violations and difference from assumed violations for one-day VaR estimates. [Table continues on next three pages.]

Method	Model	Distribution	Violations		Difference from assumed value	
			99%	97.5%	99%	97.5%
Historical simulation			38	79	13	16
FHS	ARMA(0,0)-GARCH(1,1) 99%	Normal	35	66	10	3
FHS	ARMA(0,0)-GARCH(1,1) 99%	SD	31	62	6	−1
FHS	ARMA(0,0)-GARCH(1,1) 99%	SSTD	33	63	8	0
FHS	ARMA(0,0)-GARCH(1,1) 99%	GED	34	66	9	3
FHS	ARMA(1,0)-GARCH(1,1) 99%	Normal	32	68	7	5
FHS	ARMA(1,0)-GARCH(1,1) 99%	SD	31	65	6	2
FHS	ARMA(1,0)-GARCH(1,1) 99%	SSTD	33	66	8	3
FHS	ARMA(1,0)-GARCH(1,1) 99%	GED	34	66	9	3
FHS	ARMA(0,0)-EGARCH(1,1) 99%	Normal	31	78	6	15
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SD	32	79	7	16
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	33	76	8	13
FHS	ARMA(0,0)-EGARCH(1,1) 99%	GED	32	78	7	15
FHS	ARMA(1,0)-EGARCH(1,1) 99%	Normal	34	80	9	17
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SD	31	78	6	15
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	31	74	6	11
FHS	ARMA(1,0)-EGARCH(1,1) 99%	GED	29	78	4	15

TABLE 2 Continued.

Method	Model	Distribution	Violations 99%	Violations 97.5%	Difference from assumed value 99%	Difference from assumed value 97.5%
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	35	70	10	7
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	33	67	8	4
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	34	67	9	4
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	35	68	10	5
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	34	70	9	7
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	34	67	9	4
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	32	70	7	7
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	35	68	10	5
MC	ARMA(0,0)-GARCH(1,1) 99%	Normal	49	95	24	32
MC	ARMA(0,0)-GARCH(1,1) 99%	SD	39	84	14	21
MC	ARMA(0,0)-GARCH(1,1) 99%	SSTD	35	73	10	10
MC	ARMA(0,0)-GARCH(1,1) 99%	GED	39	81	14	18
MC	ARMA(1,0)-GARCH(1,1) 99%	Normal	50	93	25	30
MC	ARMA(1,0)-GARCH(1,1) 99%	SD	41	86	16	23
MC	ARMA(1,0)-GARCH(1,1) 99%	SSTD	34	76	9	13
MC	ARMA(1,0)-GARCH(1,1) 99%	GED	43	83	18	20
MC	ARMA(0,0)-EGARCH(1,1) 99%	Normal	56	102	31	39
MC	ARMA(0,0)-EGARCH(1,1) 99%	SD	41	97	16	34
MC	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	32	83	7	20
MC	ARMA(0,0)-EGARCH(1,1) 99%	GED	43	94	18	31

TABLE 2 Continued.

Method	Model	Distribution	Violations 99%	Violations 97.5%	Difference from assumed value 99%	Difference from assumed value 97.5%
MC	ARMA(1,0)-EGARCH(1,1) 99%	Normal	56	107	31	44
MC	ARMA(1,0)-EGARCH(1,1) 99%	SD	45	103	20	40
MC	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	30	84	5	21
MC	ARMA(1,0)-EGARCH(1,1) 99%	GED	45	100	20	37
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	59	100	34	37
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	46	92	21	29
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	36	76	11	13
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	48	90	23	27
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	58	99	33	36
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	46	94	21	31
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	36	74	11	11
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	46	93	21	30
EVT	ARMA(0,0)-GARCH(1,1) 99%	Normal	35	68	10	5
EVT	ARMA(0,0)-GARCH(1,1) 99%	SD	34	67	9	4
EVT	ARMA(0,0)-GARCH(1,1) 99%	SSTD	34	69	9	6
EVT	ARMA(0,0)-GARCH(1,1) 99%	GED	34	67	9	4
EVT	ARMA(1,0)-GARCH(1,1) 99%	Normal	36	66	11	3
EVT	ARMA(1,0)-GARCH(1,1) 99%	SD	33	67	8	4
EVT	ARMA(1,0)-GARCH(1,1) 99%	SSTD	33	66	8	3
EVT	ARMA(1,0)-GARCH(1,1) 99%	GED	34	67	9	4

TABLE 2 Continued.

Method	Model	Distribution	Violations 99%	Violations 97.5%	Difference from assumed value 99%	Difference from assumed value 97.5%
EVT	ARMA(0,0)-EGARCH(1,1) 99%	Normal	29	76	4	13
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SD	30	77	5	14
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	32	75	7	12
EVT	ARMA(0,0)-EGARCH(1,1) 99%	GED	29	76	4	13
EVT	ARMA(1,0)-EGARCH(1,1) 99%	Normal	30	70	5	7
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SD	30	75	5	12
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	32	75	7	12
EVT	ARMA(1,0)-EGARCH(1,1) 99%	GED	30	73	5	10
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	31	73	6	10
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	29	70	4	7
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	31	68	6	5
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	31	73	6	10
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	32	73	7	10
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	30	69	5	6
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	29	70	4	7
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	32	73	7	10
Assumed value			25	63		

SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution.

TABLE 3 Total number of VaR violations and difference from assumed violations for overlapping ten-day VaR estimates. [Table continues on next three pages.]

Method	Model	Distribution	Violations		Difference from assumed value	
			99%	97.5%	99%	97.5%
Historical simulation			38	85	13	22
FHS	ARMA(0,0)-GARCH(1,1) 99%	Normal	9	28	−16	−35
FHS	ARMA(0,0)-GARCH(1,1) 99%	SD	9	27	−16	−36
FHS	ARMA(0,0)-GARCH(1,1) 99%	SSTD	10	26	−15	−37
FHS	ARMA(0,0)-GARCH(1,1) 99%	GED	9	26	−16	−37
FHS	ARMA(1,0)-GARCH(1,1) 99%	Normal	11	37	−14	−26
FHS	ARMA(1,0)-GARCH(1,1) 99%	SD	11	34	−14	−29
FHS	ARMA(1,0)-GARCH(1,1) 99%	SSTD	13	38	−12	−25
FHS	ARMA(1,0)-GARCH(1,1) 99%	GED	12	33	−13	−30
FHS	ARMA(0,0)-EGARCH(1,1) 99%	Normal	10	37	−15	−26
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SD	9	36	−16	−27
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	9	35	−16	−28
FHS	ARMA(0,0)-EGARCH(1,1) 99%	GED	9	37	−16	−26
FHS	ARMA(1,0)-EGARCH(1,1) 99%	Normal	10	38	−15	−25
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SD	9	35	−16	−28
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	10	36	−15	−27
FHS	ARMA(1,0)-EGARCH(1,1) 99%	GED	9	36	−16	−27

TABLE 3 Continued.

Method	Model	Distribution	Violations 99%	Violations 97.5%	Difference from assumed value 99%	Difference from assumed value 97.5%
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	11	33	–14	–30
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	8	33	–17	–30
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	10	33	–15	–30
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	11	32	–14	–31
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	11	35	–14	–28
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	12	34	–13	–29
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	8	34	–17	–29
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	9	34	–16	–29
MC	ARMA(0,0)-GARCH(1,1) 99%	Normal	10	33	–15	–30
MC	ARMA(0,0)-GARCH(1,1) 99%	SD	10	34	–15	–29
MC	ARMA(0,0)-GARCH(1,1) 99%	SSTD	8	31	–17	–32
MC	ARMA(0,0)-GARCH(1,1) 99%	GED	10	33	–15	–30
MC	ARMA(1,0)-GARCH(1,1) 99%	Normal	15	38	–10	–25
MC	ARMA(1,0)-GARCH(1,1) 99%	SD	12	34	–13	–29
MC	ARMA(1,0)-GARCH(1,1) 99%	SSTD	9	33	–16	–30
MC	ARMA(1,0)-GARCH(1,1) 99%	GED	12	35	–13	–28
MC	ARMA(0,0)-EGARCH(1,1) 99%	Normal	13	38	–12	–25
MC	ARMA(0,0)-EGARCH(1,1) 99%	SD	11	34	–14	–29
MC	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	8	33	–17	–30
MC	ARMA(0,0)-EGARCH(1,1) 99%	GED	12	37	–13	–26

TABLE 3 Continued.

Method	Model	Distribution	Violations 99%	Violations 97.5%	Difference from assumed value 99%	Difference from assumed value 97.5%
MC	ARMA(1,0)-EGARCH(1,1) 99%	Normal	8	37	−17	−26
MC	ARMA(1,0)-EGARCH(1,1) 99%	SD	8	35	−17	−28
MC	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	11	35	−14	−28
MC	ARMA(1,0)-EGARCH(1,1) 99%	GED	13	35	−12	−28
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	8	34	−17	−29
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	8	34	−17	−29
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	7	33	−18	−30
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	13	34	−12	−29
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	8	35	−17	−28
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	7	34	−18	−29
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	9	33	−16	−30
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	12	35	−13	−28
EVT	ARMA(0,0)-GARCH(1,1) 99%	Normal	30	102	5	39
EVT	ARMA(0,0)-GARCH(1,1) 99%	SD	31	101	6	38
EVT	ARMA(0,0)-GARCH(1,1) 99%	SSTD	31	98	6	35
EVT	ARMA(0,0)-GARCH(1,1) 99%	GED	32	102	7	39
EVT	ARMA(1,0)-GARCH(1,1) 99%	Normal	30	96	5	33
EVT	ARMA(1,0)-GARCH(1,1) 99%	SD	29	96	4	33
EVT	ARMA(1,0)-GARCH(1,1) 99%	SSTD	30	95	5	32
EVT	ARMA(1,0)-GARCH(1,1) 99%	GED	30	96	5	33

TABLE 3 Continued.

Method	Model	Distribution	Violations 99%	Violations 97.5%	Difference from assumed value 99%	Difference from assumed value 97.5%
EVT	ARMA(0,0)-EGARCH(1,1) 99%	Normal	10	52	–15	–11
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SD	8	49	–17	–14
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	7	50	–18	–13
EVT	ARMA(0,0)-EGARCH(1,1) 99%	GED	8	49	–17	–14
EVT	ARMA(1,0)-EGARCH(1,1) 99%	Normal	9	50	–16	–13
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SD	9	50	–16	–13
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	8	46	–17	–17
EVT	ARMA(1,0)-EGARCH(1,1) 99%	GED	9	50	–16	–13
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	10	52	–15	–11
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	10	50	–15	–13
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	11	52	–14	–11
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	11	53	–14	–10
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	10	48	–15	–15
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	9	42	–16	–21
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	11	46	–14	–17
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	11	43	–14	–20
Assumed value			25	63		

SD, Student *t* distribution. SSTD, skewed Student *t* distribution. GED, generalized error distribution.

TABLE 4 Total number of VaR violations and difference from assumed violations for nonoverlapping ten-day VaR estimates. [Table continues on next three pages.]

			Violations		Difference from assumed value	
Method	**Model**	**Distribution**	**99%**	**97.5%**	**99%**	**97.5%**
Historical simulation			8	16	5	10
FHS	ARMA(0,0)-GARCH(1,1) 99%	Normal	4	10	1	4
FHS	ARMA(0,0)-GARCH(1,1) 99%	SD	4	8	1	2
FHS	ARMA(0,0)-GARCH(1,1) 99%	SSTD	3	5	0	–1
FHS	ARMA(0,0)-GARCH(1,1) 99%	GED	4	7	1	1
FHS	ARMA(1,0)-GARCH(1,1) 99%	Normal	4	9	1	3
FHS	ARMA(1,0)-GARCH(1,1) 99%	SD	4	6	1	0
FHS	ARMA(1,0)-GARCH(1,1) 99%	SSTD	3	7	0	1
FHS	ARMA(1,0)-GARCH(1,1) 99%	GED	4	7	1	1
FHS	ARMA(0,0)-EGARCH(1,1) 99%	Normal	5	7	2	1
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SD	3	7	0	1
FHS	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	2	4	–1	–2
FHS	ARMA(0,0)-EGARCH(1,1) 99%	GED	3	6	0	0
FHS	ARMA(1,0)-EGARCH(1,1) 99%	Normal	6	8	3	2
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SD	4	9	1	3
FHS	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	4	5	1	–1
FHS	ARMA(1,0)-EGARCH(1,1) 99%	GED	7	11	4	5

TABLE 4 Continued.

Method	Model	Distribution	Violations 99%	Violations 97.5%	Difference from assumed value 99%	Difference from assumed value 97.5%
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	0	4	–3	–2
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	0	4	–3	–2
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	2	7	–1	1
FHS	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	0	4	–3	–2
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	0	7	–3	1
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	1	6	–2	0
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	1	7	–2	1
FHS	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	0	7	–3	1
MC	ARMA(0,0)-GARCH(1,1) 99%	Normal	4	5	1	–1
MC	ARMA(0,0)-GARCH(1,1) 99%	SD	4	4	1	–2
MC	ARMA(0,0)-GARCH(1,1) 99%	SSTD	0	2	–3	–4
MC	ARMA(0,0)-GARCH(1,1) 99%	GED	4	4	1	–2
MC	ARMA(1,0)-GARCH(1,1) 99%	Normal	4	5	1	–1
MC	ARMA(1,0)-GARCH(1,1) 99%	SD	4	4	1	–2
MC	ARMA(1,0)-GARCH(1,1) 99%	SSTD	1	3	–2	–3
MC	ARMA(1,0)-GARCH(1,1) 99%	GED	4	5	1	–1
MC	ARMA(0,0)-EGARCH(1,1) 99%	Normal	2	6	–1	0
MC	ARMA(0,0)-EGARCH(1,1) 99%	SD	2	4	–1	–2
MC	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	0	1	–3	–5
MC	ARMA(0,0)-EGARCH(1,1) 99%	GED	2	5	–1	–1

TABLE 4 Continued.

Method	Model	Distribution	Violations 99%	Violations 97.5%	Difference from assumed value 99%	Difference from assumed value 97.5%
MC	ARMA(1,0)-EGARCH(1,1) 99%	Normal	3	7	0	1
MC	ARMA(1,0)-EGARCH(1,1) 99%	SD	1	2	−2	−4
MC	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	2	5	−1	−1
MC	ARMA(1,0)-EGARCH(1,1) 99%	GED	7	11	4	5
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	0	0	−3	−6
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	0	0	−3	−6
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	0	1	−3	−5
MC	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	0	1	−3	−5
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	0	1	−3	−5
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	1	1	−2	−5
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	1	1	−2	−5
MC	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	0	2	−3	−4
EVT	ARMA(0,0)-GARCH(1,1) 99%	Normal	4	5	1	−1
EVT	ARMA(0,0)-GARCH(1,1) 99%	SD	4	4	1	−2
EVT	ARMA(0,0)-GARCH(1,1) 99%	SSTD	3	3	0	−3
EVT	ARMA(0,0)-GARCH(1,1) 99%	GED	4	4	1	−2
EVT	ARMA(1,0)-GARCH(1,1) 99%	Normal	3	5	0	−1
EVT	ARMA(1,0)-GARCH(1,1) 99%	SD	3	5	0	−1
EVT	ARMA(1,0)-GARCH(1,1) 99%	SSTD	1	2	−2	−4
EVT	ARMA(1,0)-GARCH(1,1) 99%	GED	3	5	0	−1

TABLE 4 Continued.

Method	Model	Distribution	Violations 99%	Violations 97.5%	Difference from assumed value 99%	Difference from assumed value 97.5%
EVT	ARMA(0,0)-EGARCH(1,1) 99%	Normal	3	5	0	–1
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SD	5	8	2	2
EVT	ARMA(0,0)-EGARCH(1,1) 99%	SSTD	2	3	–1	–3
EVT	ARMA(0,0)-EGARCH(1,1) 99%	GED	3	7	0	1
EVT	ARMA(1,0)-EGARCH(1,1) 99%	Normal	3	5	0	–1
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SD	5	7	2	1
EVT	ARMA(1,0)-EGARCH(1,1) 99%	SSTD	2	3	–1	–3
EVT	ARMA(1,0)-EGARCH(1,1) 99%	GED	3	5	0	–1
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	Normal	0	0	–3	–6
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SD	0	0	–3	–6
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	SSTD	2	2	–1	–4
EVT	ARMA(0,0)-GJR-GARCH(1,1) 99%	GED	0	1	–3	–5
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	Normal	0	0	–3	–6
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SD	1	2	–2	–4
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	SSTD	1	1	–2	–5
EVT	ARMA(1,0)-GJR-GARCH(1,1) 99%	GED	0	0	–3	–6
Assumed value			3	6		

SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution.

TABLE 5 CCR efficiency scores for one-day VaR models (in percent).

Method	Model	Normal	SD	SSTD	GED
FHS	ARMA(0,0)-GARCH(1,1)	57	100	100	69
FHS	ARMA(1,0)-GARCH(1,1)	73	100	71	70
FHS	ARMA(0,0)-EGARCH(1,1)	65	60	58	59
FHS	ARMA(1,0)-EGARCH(1,1)	42	65	65	66
FHS	ARMA(0,0)-GJR-GARCH(1,1)	50	67	63	64
FHS	ARMA(1,0)-GJR-GARCH(1,1)	55	61	68	64
MC	ARMA(0,0)-GARCH(1,1)	18	29	46	30
MC	ARMA(1,0)-GARCH(1,1)	17	26	42	26
MC	ARMA(0,0)-EGARCH(1,1)	14	24	48	22
MC	ARMA(1,0)-EGARCH(1,1)	13	19	72	20
MC	ARMA(0,0)-GJR-GARCH(1,1)	13	20	45	19
MC	ARMA(1,0)-GJR-GARCH(1,1)	14	19	42	20
EVT	ARMA(0,0)-GARCH(1,1)	54	61	57	61
EVT	ARMA(1,0)-GARCH(1,1)	53	68	55	68
EVT	ARMA(0,0)-EGARCH(1,1)	68	72	58	68
EVT	ARMA(1,0)-EGARCH(1,1)	100	75	57	57
EVT	ARMA(0,0)-GJR-GARCH(1,1)	67	100	100	67
EVT	ARMA(1,0)-GJR-GARCH(1,1)	60	81	81	63

Historical simulation received a score of 33. SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution.

TABLE 6 CCR efficiency scores for overlapping ten-day VaR models (in percent).

Method	Model	Normal	SD	SSTD	GED
FHS	ARMA(0,0)-GARCH(1,1)	59	57	58	57
FHS	ARMA(1,0)-GARCH(1,1)	72	68	79	70
FHS	ARMA(0,0)-EGARCH(1,1)	70	66	65	67
FHS	ARMA(1,0)-EGARCH(1,1)	70	65	68	66
FHS	ARMA(0,0)-GJR-GARCH(1,1)	67	60	65	66
FHS	ARMA(1,0)-GJR-GARCH(1,1)	69	61	62	64
MC	ARMA(0,0)-GARCH(1,1)	65	66	59	65
MC	ARMA(1,0)-GARCH(1,1)	86	71	59	72
MC	ARMA(0,0)-EGARCH(1,1)	79	68	61	75
MC	ARMA(1,0)-EGARCH(1,1)	66	63	70	64
MC	ARMA(0,0)-GJR-GARCH(1,1)	62	62	69	73
MC	ARMA(1,0)-GJR-GARCH(1,1)	63	60	62	72
EVT	ARMA(0,0)-GARCH(1,1)	63	75	79	71
EVT	ARMA(1,0)-GARCH(1,1)	100	100	86	84
EVT	ARMA(0,0)-EGARCH(1,1)	82	83	81	72
EVT	ARMA(1,0)-EGARCH(1,1)	84	81	80	73
EVT	ARMA(0,0)-GJR-GARCH(1,1)	100	100	79	71
EVT	ARMA(1,0)-GJR-GARCH(1,1)	81	82	83	80

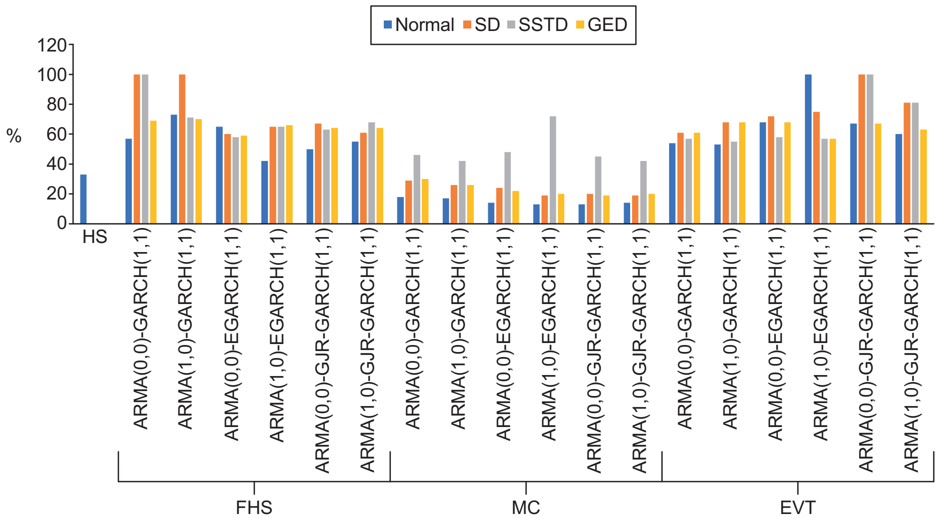
Historical simulation received a score of 83. SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution.

TABLE 7 CCR efficiency scores for nonoverlapping ten-day VaR models (in percent).

Method	Model	Normal	SD	SSTD	GED
FHS	ARMA(0,0)-GARCH(1,1)	75	85	100	80
FHS	ARMA(1,0)-GARCH(1,1)	72	84	100	82
FHS	ARMA(0,0)-EGARCH(1,1)	81	100	75	100
FHS	ARMA(1,0)-EGARCH(1,1)	67	65	63	67
FHS	ARMA(0,0)-GJR-GARCH(1,1)	68	67	65	68
FHS	ARMA(1,0)-GJR-GARCH(1,1)	67	66	65	70
MC	ARMA(0,0)-GARCH(1,1)	63	62	63	58
MC	ARMA(1,0)-GARCH(1,1)	63	66	65	68
MC	ARMA(0,0)-EGARCH(1,1)	58	63	65	59
MC	ARMA(1,0)-EGARCH(1,1)	61	63	61	62
MC	ARMA(0,0)-GJR-GARCH(1,1)	58	62	45	19
MC	ARMA(1,0)-GJR-GARCH(1,1)	57	61	64	64
EVT	ARMA(0,0)-GARCH(1,1)	75	100	100	80
EVT	ARMA(1,0)-GARCH(1,1)	72	100	85	82
EVT	ARMA(0,0)-EGARCH(1,1)	81	85	75	87
EVT	ARMA(1,0)-EGARCH(1,1)	67	65	63	67
EVT	ARMA(0,0)-GJR-GARCH(1,1)	59	60	62	61
EVT	ARMA(1,0)-GJR-GARCH(1,1)	55	56	58	57

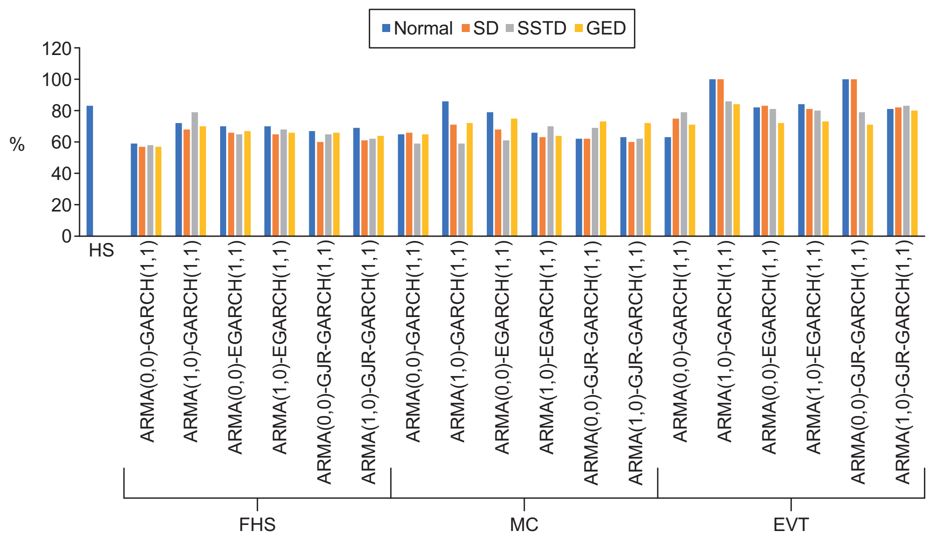
Historical simulation received a score of 10. SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution.

FIGURE 1 CCR efficiency scores for one-day VaR models.



SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution.

FIGURE 2 CCR efficiency scores for overlapping ten-day VaR models.



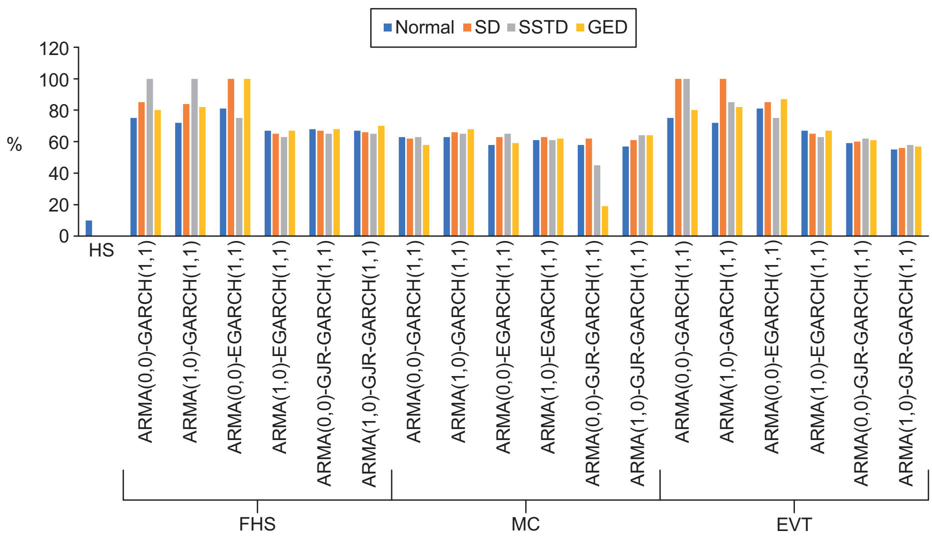
SD, Student *t* distribution. SSTD, skewed Student *t* distribution. GED, generalized error distribution.

returns indicated only four efficient models, all of which stem from the EVT method; this finding indicates that the FHS method is not efficient when overlapping ten-day returns are used. In addition, DEA efficiency scores presented seven models as efficient – mostly those using FHS and EVT methods – for ten-day VaR estimates using nonoverlapping returns. The above observations indicate that the EVT and FHS methods present some efficient VaR estimates for one-day and ten-day (nonoverlapping) horizons. However, when ten-day VaR is calculated with overlapping observations, the FHS method does not produce any efficient models, while EVT still produces a number of accurate VaR estimates.

6 CONCLUSIONS

The decision-making process involved in selecting suitable risk metrics is often difficult, depending on various circumstances. In this paper we focused on the VaR evaluation of market risk models for different significance levels and forecast horizons. The results suggest that some models are more efficient than others for various significance levels. Further, we observed that, when the forecast horizon changes, other models dominate in terms of efficiency under different confidence levels. Another finding was that the choice of overlapping or nonoverlapping observations

FIGURE 3 CCR efficiency scores for nonoverlapping ten-day VaR models.



SD, Student t distribution. SSTD, skewed Student t distribution. GED, generalized error distribution.

can seriously affect the efficiency of VaR estimation on longer horizons. These findings are clearly important and require further testing with regard to the assumptions that financial institutions make when VaR values are estimated for longer horizons.

Further, according to the research outcomes, DEA can be a suitable alternative and computationally efficient method for the selection of accurate risk methodologies. Unlike most of the commonly adopted backtesting techniques, which focus only on a specific quantile of the distribution tail, DEA examines the stability of the models under different parameterizations. Theoretically, a model could be superior compared with others in all circumstances; however, the complexity of financial markets is so vast that no model can achieve that. Thus, DEA may serve as an alternative backtesting methodology for assessing different risk methodologies for a range of significance levels and may offer better insight into a model's efficiency in risk management.

DECLARATION OF INTEREST

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

REFERENCES

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 International Settlements. URL: www.bis.org/publ/bcbs22.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).

Charnes, A., and Cooper, W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly* **9**(3-4), 181–186 (https://doi.org/10.1016/0377-2217(78)90138-8).

Charnes, A., Cooper, W., and Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research* **2**(6), 429–444 (https://doi.org/10.1016/0377-2217(78)90138-8).

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).

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).

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**(5), 1163–1174 (https://doi.org/10.1214/aos/1176343247).

Kontaxis, G., and Tsolas, I. E. (2021). Evaluation of backtesting techniques on risk models with different horizons. *The Journal of Risk Model Validation* **15**(4), 1–22 (https://doi.org/10.21314/JRMV.2021.011).

Kresta, A., and Tichý, T. (2016). Selection of efficient market risk models: backtesting results evaluation with DEA approach. *Computers and Industrial Engineering* **102**(5), 331–339 (https://doi.org/10.1016/j.cie.2016.07.017).

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 heteroskedasticity 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 overlapping intervals. *Asia Risk* (April), 50–55. URL: www.risk.net/risk-management/1509219/error-var-overlapping-intervals.