University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 29/2017 (258)

Is CAViaR model really so good in Value at Risk forecasting? Evidence from evaluation of a quality of Value-at-Risk forecasts obtained based on the: GARCH(1,1), GARCH-t(1,1), GARCH-st(1,1), QML-GARCH(1,1), CAViaR and the historical simulation models depending on the stability of financial markets

Mateusz Buczyński
Marcin Chlebus

Warsaw 2017

Is CAViaR model really so good in Value at Risk forecasting? Evidence from evaluation of a quality of Value-at-Risk forecasts obtained based on the: GARCH(1,1), GARCH-t(1,1), GARCH-st(1,1), QML-GARCH(1,1), CAViaR and the historical simulation models depending on the stability of financial markets

MATEUSZ BUCZYŃSKI
Faculty of Economic Sciences
University of Warsaw
e-mail: mateusz.buczynski@student.uw.edu.pl

MARCIN CHLEBUS
Faculty of Economic Sciences
University of Warsaw
e-mail: mchlebus@wne.uw.edu.pl

Abstract

In the literature, there is no consensus which Value-at-Risk forecasting model is the best for measuring a market risk in banks. In the study an analysis of Value-at-Risk forecasting models quality over varying economic stability periods for main indices from stock exchanges was conducted. The VaR forecasts from GARCH(1,1), GARCH-t(1,1), GARCH-st(1,1), QML-GARCH(1,1), CAViaR and historical simulation models in periods with contrasting volatility trends (increasing, constantly high and decreasing) for countries economically developed (the USA – S&P 500, Germany - DAX and Japan – Nikkei 225) and economically developing (China – SSE COMP, Poland – WIG20 and Turkey – XU100) were compared. The data samples used in the analysis were selected from period 01.01.1999 – 24.03.2017. To assess the VaR forecasts quality: excess ratio, Basel traffic light test, coverage tests (Kupiec test, Christoffersen test), Dynamic Quantile test, cost functions and Diebold-Marino test were used. Obtained results shows that the quality of Value-at-Risk forecasts for the models varies depending on a volatility trend. However, GARCH-st (1,1) and QML-GARCH(1,1) were found as the most robust models to the different volatility periods. The results shows, as well that the CAViaR model forecasts were less appropriate in the increasing volatility period. Moreover, no significant differences for the VaR forecasts quality were found for the developed and developing countries.

Keywords:

risk management, value at risk, GARCH, CAViaR, historical simulation, quality of model assessment

JEL:

G32, C52, C53, C58

DOI:

https://doi.org/10.26405/WP/WNE/2017/258/029

1. Introduction

Risk measurement in financial institutions is one of the key areas of the bank's activity, which is to ensure solvency and its good functioning. Since the solvency of a single bank can affect the stability of the entire banking system, national and international regulators place a strong emphasis on the high quality of models for measuring different types of risks in banks. The core of the regulatory framework has been developed by the Basel Committee on Banking Supervision. The European Union has established the rules set by the Basel Committee in force in the EU countries via the CRD IV Directive.

The market risk is one of the most important types of risk in banks, which is defined as a measurable uncertainty associated with changes in the interest rates, exchange rates and financial instruments price values. The CRD IV Directive imposes a number of rules on the banks on how to measure the market risk (the *Value-at-Risk* (VaR) measure should be used for measuring the market risk, the confidence level should be set at a minimum of 99%, the VaR forecast period should be at least 10 working days, and the period of historical observations used for forecasting is at least a year).

In addition, the market risk measurement model used in a bank should be subject to the process of backtesting, which compares the VaR forecasts with the actual rate of return from an asset. The aim of such comparison is to count the cases, in which the actual exceeding of the empirical rate of return over VaR took place. The basis for evaluating the model quality is the comparison of the calculated empirical number of exceedances with the assumed number of exceedances, which should not exceed 1% on the basis of the above described rules. For backtesting purposes the one-day VaR should be used.

The VaR can be defined as the maximum portfolio loss in a given period, assuming a level of confidence, assuming "normal" market conditions (Jorion, 2001). By simplifying, VaR determines the greatest possible loss, assuming a α significance level, when unexpected negative events do not appear in the analysed period. VaR can be presented by the formula (Abad, Benito and López 2014):

$$P(r_t < VaR_\alpha(t)|\Omega_{t-1}) = \alpha, \quad (1)$$

where r_t = return on the asset in period t,
$VaR_\alpha(t)$ = VaR at the α level in the period t,
Ω_{t-1} = set of information available in the period $t-1$.

In order to reduce uncertainty, many VaR estimation methods have been developed. The analysis of the quality of forecasts generated by diverse models, which characterise various approaches, is a very important element of the VaR assessment. This has led to an open discussion among researchers on the prevalence of VaR estimation approach over other ones.

The assessment of the VaR forecast quality and the method of their proper estimation have been broadly discussed in the literature for a long time, e.g. in Engle (2001, 2004); Tagilafichi (2003) or Engle & Manganelli (2001). Continuous development of new VaR estimation methods makes the topic still relevant and current. Abad et al. (2014) take a broad look at the VaR methodology. Their work summarises the results of many new research papers devoted to the topic of forecasting this measure. VaR estimation methods that they consider, among others, include the historical simulation, filtered historical simulation, RiskMetrics™, GARCH model class (including models with normal, t-Student and skewed t-Student distributions), CAViaR model, models based on Extreme Value Theory and the Monte Carlo method. The results presented by Abad et al. (2014) suggest that choosing the best and most versatile VaR model, which forecasts would always be no worse than the forecasts of all other models, regardless of the market conditions, is virtually impossible. However, the filtered historical simulation, models based on Extreme Value Theory and CAViaR model were assessed as the best when considered.

Researchers, aware that there is no single best model for predicting VaR, try to determine the conditions under which certain models predict the best. Examples can include the evaluation and categorization of models carried out, among others, in the work by McAleer, Jimenez-Martin and Perez-Amaral (2009) and Shams and Sina (2014). Researchers compared models in periods of varying volatility – before the crisis (where there was no high volatility) and after the outbreak of the crisis (where financial conditions are characterised by high volatility). The results of their research confirm that some models predict very well before the onset of the crisis, but with the increase in volatility, their prognostic quality drastically decreases. Others are more conservative during periods of relative calmness, but in the course of the crisis the number of errors made by these models is relatively low. In both studies, the GARCH(1,1) model generated good VaR forecasts before the crisis, but this changed when the instability of financial markets increased.

McAleer et al. (2009) showed that RiskMetrics™ was the best fitted model during crisis, while Shams and Sina (2014) recognized GARCH(1,1) and GJR-GARCH as a well forecasting models. In contrast to the results obtained by McAleer et al. (2009), the level of quality of forecasts generated by the RiskMetrics™ model was considered unsatisfactory by them. However, attention needs to be drawn to one difference in the samples, on which the study was conducted, i.e. the first one comes from the developed country (USA, S&P500), and the second one from the developing one (Iran, TSEM).

Summarizing the results presented, it can be stated that the selection of the VaR forecasting model should be based on a prior analysis of the period considered in terms of the expected volatility and the level of the financial market development, for which the forecast will be made.

The analysis of the literature discussed previously indicates that the researchers do not have a full consent in the evaluation of which models should be used during periods of calm, and which ones during turbulence. Therefore, it is reasonable to compare different VaR forecast models in different periods of stability, as well as among countries with different levels of economic development.

The main objective of the research was to investigate which models amongst benchmark models, such as: historical simulation, GARCH(1,1) models with normal, t-Student and skewed t-Student random error distributions and those pointed in Abad et al. (2014) as potentially the best: QML-GARCH(1,1) (similar model to filtered historical simulation) and CAViaR model are more robust to market conditions and forecast VaR better in the periods: before the crisis, during the crisis and after the financial crisis, for a series of returns from developing and developed countries.

The results of the forecasts of aforementioned models were compared to test which of them are better in forecasting than others. This comparison was carried out in two dimensions. First of all, forecasts were compared in three periods selected for different volatility dynamics, i.e. in periods of: growing volatility, constantly high volatility and decreasing volatility. Secondly, forecasts were made for two groups of countries: developed countries (Germany, the USA, Japan) and developing countries (Poland, Turkey, China). The rates of return, for which the forecasts were made, come from the largest stock exchange indices of these countries. The quality of forecasts was assessed based on the most popular tests used for verification of VaR forecast adequacy (Kupiec test, Christoffersen test, Basel traffic light test – binomial test and Dynamic Quantile test) and the analysis of the cost function (Caporin absolute cost function, Caporin company cost function and excessive cost function). In addition, the Diebold-Mariano test was used in order to compare the forecasts between models. So many criterions give an opportunity to look at the VaR forecasting quality from different perspectives: fulfilling regulatory requirements (Basel traffic light test), forecasting adequateness (Kupiec test, Christoffersen test and Dynamic Quantile test), effectiveness of methods (cost functions) and methods superiority (Diebold-Marino test).

It was decided that the study will compare the VaR forecast quality on the level of economic development of the country. For this purpose, six different indexes were used: DAX, S&P500, Nikkei 225, WIG20, SSE COMP, XU100. At the same time, the analysis is considered over periods of varying volatility, by examining how it affects the results of the models.

2. Value-at-Risk estimation models

The concept of VaR models is based on the assumption that the loss on a given asset, with a given probability α in the given period, assuming normal market conditions, will not exceed the projected VaR level. This means that for a given rates of return distribution, the VaR measure is the α quantile of that distribution.

Therefore, estimation of such a value can be approached by directly finding the inverse function of the distribution of the rates of return from the asset. Such a method is called the non-parametric method and involves estimating VaR, with no assumptions about the return distribution. This method is based on the assumption that the future rates of return will be sufficiently close to the historical ones.

Another approach is to estimate parameters for assumed theoretical distribution of rates of return, which follows the process:

$$r_t = \mu + \sigma_t \varepsilon_t, \varepsilon_t \sim \mathrm{iid}(0,1), \quad (2)$$

where μ = mean rate of return,
σ_t^2 = $E(\varepsilon_t^2 | \Omega_{t-1})$, variable variance,
ε_t = residual,
Ω_{t-1} = set of information available in the period $t-1$.

This method is called the parametric method. Methods combining both approaches are called semi-parametric methods. A broader perspective on VaR forecasting is presented in a description of the particular methods presented below.

2.1. Historical simulation

Historical simulation is one of the basic methods of VaR estimation (Dowd, 2002). The estimated empirical distribution of rates of return from the asset is used for calculations, and VaR is the α quantile of this distribution. Thus, historical simulation is one of the non-parametric methods. This method requires defining the "moving" window, i.e. determining how many historical periods should be considered while estimating VaR for the specific day. The width of the window is fixed and usually ranges from 6 months to 2 years (Engle & Manganelli, 2001).

2.2. Models of the GARCH class

The GARCH model (*Generalized Autoregressive Conditional Heteroskedasticity*), proposed by Bollerslev (1986) is a generalization of the ARCH process created by Engle (1982), in which the conditional variance is not only the function of lagged random errors, but also of lagged conditional variances. The standard GARCH model (p,q) can be written as:

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t \xi_t, \tag{3}$$

where r_t = rate of return of the asset in the period t,
μ_t = conditional mean (in the present study, we do not include the mean in the functional form of the GARCH model, due to the occurrence of the zero mean in a series of rates of return, see the chapter dedicated to empirical results),
ε_t = random error in the period t, which equals to the product of conditional standard deviation σ_t and the standardized random error ξ_t in the period t ($\xi_t \sim \mathrm{iid}(0,1)$.

In turn, the equation of conditional variance, in the GARCH(p,q) model can be written as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2, \tag{4}$$

where σ_t^2 = conditional variance in the period t,
ω = constant ($\omega > 0$),
α_i = weight of the random squared error in the period $t-1$,
β_i = weight of the conditional variance in the period $t-1$,
ε_{t-i}^2 = squared random error in the period $t-1$,
σ_{t-i}^2 = variance in the period $t-1$,
q = number of random error squares periods used in the functional form of conditional variance,
p = number of lagged conditional variances used in the functional form of conditional variance.

Therefore, the GARCH(1,1) model is a model, where $p = 1$ and $q = 1$, so the equation of conditional variance is as follows:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{5}$$

The assumptions of this model are: $\alpha_1 \geq 0$ and $\beta_1 \geq 0$, as well as the sum $\alpha_1 + \beta_1 \leq 1$, which ensures the series are covariance-stationary.

In order to estimate VaR from the GARCH model estimators, one can use the following formula (Abad et al., 2014):

$$VaR_\alpha(t) = \widehat{\mu_t} + k_{1-\alpha} * \sqrt{\hat{\sigma}_t^2}, \tag{6}$$

where $VaR_\alpha(t)$ = VaR at the α level in the period t,
$\widehat{\mu_t}$ = estimator of the conditional mean in the period t (in this study it equals 0),
$k_{1-\alpha}$ = α quantile from the assumed random error distribution,
$\hat{\sigma}_t^2$ = conditional variance estimator in the period t.

The study analysed the GARCH (1,1) models with random error distributions: normal, t-Student (GARCH-t), skewed t-Student (GARCH-ts) and QML-GARCH(1,1) (GARCH(1,1) with correction for the empirical error).

The QML-GARCH(1,1) (*Quasi-Maximum Likelihood GARCH*) model is based on the Bollerslev and Woolridge (1992) proof, who proved that in the GARCH model, the OLS estimator is consistent even if the random errors do not come from the normal distribution. This allows to use the GARCH process to standardize the residuals non-derived from the normal distribution. Based on this finding, Engle and Manganelli (1999) proposed the QML-GARCH(1,1) model. It involves the usage of the GARCH(1,1) model to estimate the conditional variance, and then to estimate the VaR value as the empirical distribution quantile of the standardized residuals of this model. This is a combination of the GARCH model with historical simulation for a series of standardized residuals. This model, in essence, is similar to the filtered historical simulation proposed by Barone-Adesi, Giannopoulos and Vosper (1999), but without bootstrapping standardised residuals.

2.3. The CAViaR model

A GARCH class model estimates the parameters of rates of return distribution, and then the distribution quantile is estimated on its basis. Engle and Manganelli (2004) proposed the CAViaR model (*Conditional Autoregressive Value-at-Risk*), which completely avoids the modelling of the rates of return distribution, thus directly modelling the distribution of the quantile. This concept is based on the financial stylized fact that there is a high autocorrelation in the variance of financial series. VaR is also strongly related to variance, so among it the autocorrelation should also be present. The basic specification of the CAViaR model is as follows:

$$VaR_{\alpha}(t) = \beta_0 + \sum_{i=1}^{q} \beta_i VaR_{\alpha}(t-i) + \sum_{j=1}^{r} \beta_j l\left(r_{t-j}\right), \tag{7}$$

where $VaR_{\alpha}(t)$ = VaR at the α level in the t period,
β_0 = model constant,
$\beta_i, \dots, \beta_q$ = weights of the lagged VaRs,
$VaR_{\alpha}(t-i)$ = lagged VaR*s* (included in the smoothing function),
$\beta_j, \dots, \beta_r$ = weights of lagged rates of return,
$l\left(r_{t-j}\right)$ = function of a finite number of rates of return (the function connecting VaR with the dataset).

Out of four specifications of the CAViaR process proposed by Engle and Manganelli (2004), this paper has analysed the *Indirect GARCH(1,1)*. The quantile is modelled in a similar way as the GARCH(1,1) models the variance. The VaR forecast from the CAViaR model can be written as follows:

$$VaR_{\alpha}(t) = \sqrt{\beta_0 + \beta_1 VaR_{\alpha}^2(t-1) + \beta_2 {r_{t-1}}^2}, \tag{8}$$

3. Testing the Value-at-Risk forecast quality

3.1. Performance tests

In order to perform a comprehensive quality assessment of VaR forecasts, in this paper a set of criterions to assess VaR forecast was performed. The idea was to check whether VaR forecasts obtained from the models fulfil regulatory requirements (excess ratio, Basel traffic light test), are adequate (Kupiec test, Christoffersen test and Dynamic Quantile test), are effective (cost functions) and are statistically better than others (Diebold-Marino test).

The excess ratio can be described by the formula:

$$ER = \frac{\sum_{t=1}^{N} 1_{r_t < VaR_{\alpha}(t)}}{N}, \tag{9}$$

where n = the number of VaR forecasts,
$1_{r_t < VaR_t}$ = the number of rates of return, for which the VaR forecast was larger than the real value of the rate of return on the same day.

The excess ratio can then be expressed as the percentage of model failure, which, for the correctly forecasting model, should (in theory) be equal the significance level, at which VaR was calculated.

The Basel traffic light test is created based on the excess ratio value. The assessment of VaR forecast quality is made based on the attribution of, respectively: green (no problems with the forecast quality, model considered valid), yellow (model supervision is recommended, warning zone) and red (model almost for sure generates the VaR forecasts of bad quality) zones. The lights are assigned on the basis of exceeding the next thresholds of *excess ratio*. The yellow zone begins at the point at which the cumulative binomial distribution (with 1% probability of success and N trials, where N is equal to a number of VaR forecasts) is greater or equal 0,95. Similarly, the red zone starts at the point where the value of the same distribution is greater or equal 0,9999. In order to find out which zone the analysed model will be assigned to, the following formula should be used:

$$F\left(1_{r_t<VaR_\alpha(t)}\right) = \sum_{k=0}^{1_{r_t<VaR_\alpha(t)}} \binom{N}{k} p^k (1-p)^{N-k} = \alpha \begin{cases} red\ zone & \alpha \geq 0{,}9999 \\ yellow\ zone & \alpha \geq 0{,}95 \\ green\ zone & \alpha < 0{,}95 \end{cases}, \quad (10)$$

where $F(\cdot)$ = distribution of variable from the binominal distribution,
N = number of VaR forecasts,
$1_{r_t<VaR_t}$ = number of VaR exceedances,
p = level of probability, for which VaR was estimated.

For the $p = 0{,}01$ and $N = 500$ the yellow zone begins with the 9. exceedance, for which the cumulative binomial distribution is 0,9689 ($>0{,}95$) and the red zone at the 15. exceedance, for which the cumulative distribution is 0,9999.

Another test that was used in the study is the Kupiec test (1995). Despite the rather clear results of the Basel traffic light test, it does not take into account the overestimation of the model, i.e. too high VaR forecasts, which will have a lower number of exceedances than assumed. For example, the model with zero exceedances will be qualified to the green zone, but it does not predict well. In the Kupiec test, the deviations from both sides of the assumed number of exceedances are taken into account, and the test statistics is built on this difference. It has a chi-squared distribution with one degree of freedom and is as follows:

$$LR_{UC} = 2\ln\left(\left(\frac{1-\hat{\alpha}}{1-\alpha}\right)^{N-X}\left(\frac{\hat{\alpha}}{\alpha}\right)^{X}\right) \sim \chi^2(1), \quad (11)$$

where α = assumed excess ratio,
$\hat{\alpha}$ = empirical excess ratio,
N = number of VaR forecasts,
X = number of VaR forecasts exceedances.

In the Kupiec test, we test the null hypothesis H_0: $\alpha = \hat{\alpha} = X/N$, i.e. the assumption theoretical and empirical excess ratio are equal. This test is used for checking the models both in terms of underestimations and overestimations.

For analysis, the test of conditional coverage that is the Christoffersen test (1998) was also used. It uses the test statistics of the Kupiec test and complements it with the statistic of the independence of VaR forecast exceedances test. As a result, this test is sensitive to sequences of subsequent exceedances of the VaR forecasts. The test statistics comes from the chi-squared distribution with two degrees of freedom and its formula is as follows:

$$LR_{CC} = LR_{UC} + LR_{IND} \sim \chi^2(2), \quad (12)$$

where LR_{UC} = statistic of the unconditional coverage test,
LR_{IND} = statistic of the VaR forecast independence test,
which comes from the chi-squared distribution with one degree of freedom, tests the null hypothesis about the independence of exceedances, against the alternative hypothesis that the exceedances are characterised by the first order Markov chain. It can be expressed as:

$$LR_{IND} = 2[\ln\left((1-\pi_{01})^{N_{00}}\pi_{01}^{N_{01}}(1-\pi_{11})^{N_{10}}\pi_{11}^{N_{11}}\right) - \ln((1-\hat{\alpha})^{N_{00}+N_{01}}\hat{\alpha}^{N_{00}+N_{01}})] \sim \chi^2(1), \quad (13)$$

where N_{ij} = number of observations, where the j conditions (0 – not an exceedance, 1 – exceedance) occurred after observing the i state in the previous observation (0 – not an exceedance, 1 - exceedance),

π_{01} = $\frac{N_{01}}{N_{01}+N_{00}}$, probability of exceedance provided the lack of exceedance in the previous period,

π_{11} = $\frac{N_{11}}{N_{11}+N_{10}}$, probability of exceedance provided the exceedance in the previous period,

$\hat{\alpha}$ = observed *excess ratio.*

The DQ test, proposed by Engle and Manganelli (2004), was the next test taken into account of the study. The aim of this test is to jointly check the occurrence of autocorrelation among the exceedances of the VaR forecasts and whether the number of exceedances agrees with the expectation. The null hypothesis of the DQ test is that all coefficients in a regression:

$$I_t = \beta_0 + \sum_{i=1}^{p} \beta_i I_{t-1} + \sum_{j=1}^{q} \mu_j X_j + \varepsilon_t, \quad (14)$$

where I_t = $\begin{cases} 1-\ \alpha\ for\ r_t < VaR_\alpha(t) \\ \ -\alpha\quad for\ r_t \geq VaR_\alpha(t) \end{cases}$,

X_j = all explanatory variables included in the information set while forecasting,

p = number of lags of the dependent variable,

q = number of lags of independent variables.

are zero, H0: $\beta_0 = \beta_1 = \cdots = \beta_p = \mu_1 = \cdots = \mu_q = 0$.

An alternative hypothesis is that at least one of the parameters of the above regression is significantly different from zero. Including any explanatory variable to the information set, we can estimate its effect on the occurrence of exceedances. The most common explanatory variable in DQ test includes the lagged VaR forecasts. The test statistics of the DQ test comes from the chi-squared distribution with the $p + q$ degrees of freedom and is as follows:

$$DQ = \frac{I_t' X (X'X)^{-1} X' I_t}{\alpha(1-\alpha)} \sim \chi^2(p+q), \quad (15)$$

where I_t = vector of exceedances, in the form as described above,

X = matrix, in which the columns are p lags of the exceedances vector and q lags of explanatory variables,

α = significance level of VaR forecasts.

In the study also the Diebold-Mariano (1995) test is used, which statistically evaluates which of two models has forecasts of a better quality. Statistic of the Diebold-Mariano test comes from the standard normal distribution, it can be written as follows:

$$DM = \frac{\mu(d)}{\sqrt{\frac{1}{T} * \sigma^2(d)}} \sim N(0,1), \quad (16)$$

where $\mu(d)$ = mean of a process d,

d = $e_1^2 - e_2^2$, difference of squared residuals of the forecasts of the 1. and 2. model, where the residual is understood as the difference between the realized rate of return, and the predicted level of VaR,

$\sigma^2(d)$ = variance of the process d,

T = number of forecast periods

The DM statistic assumes that the d process is stationary, and the null hypothesis of this test is as follows, H0: $E(d) = 0$, which means that both forecasts are equally good, with the alternative hypothesis $E(d) > 0$, which means that the first forecast is alternatively better than the second forecast.

3.2. Cost functions

The cost functions are a supplementary way of comparing VaR models. Unlike the statistic tests, they do not have a formal character. In general, the overall form of such a function can be presented as:

$$C_t = \begin{cases} f(VaR_\alpha(t), r_t) \; for \; r_t < VaR_\alpha(t) \\ g(VaR_\alpha(t), r_t) for \; r_t \geq VaR_\alpha(t) \end{cases}, \tag{17}$$

where r_t = realized rate of return in the period t,
$VaR_\alpha(t)$ = VaR forecast for the same period t.

However, it must be assumed that $f(VaR_\alpha(t), r_t) \geq g(VaR_\alpha(t), r_t)$, for the cost of exceedance to be always not smaller than the cost of its lack. Therefore, the smaller the value of the cost function, the better the VaR model.

The cost function representing the cost of exceedance is, for example, the Caporin cost function (2008), which can be presented as follows:

$$RC_t = \begin{cases} |VaR_\alpha(t) - r_t| \\ 0 \end{cases} for \begin{matrix} r_t < VaR_\alpha(t) \\ r_t \geq VaR_\alpha(t) \end{matrix}, \tag{18}$$

where r_t = realized rate of return in the period t,
$VaR_\alpha(t)$ = VaR forecast for the same period t.

The final result for the given model is the average cost $RC = \sum_{t=1}^{N} RC_t / N$, where N is the number of all VaR exceedances. The best model is the one, which has the smallest value of the average cost.

The above function focuses only on exceedances, i.e. it does not take into consideration penalties for too high capital protection. Sarma, Thomas and Shah (2003) introduced the concept of the firm's cost function, which also includes the cost of no exceedance. One of the forms of this function is the Caporin firm's cost function (2008), i.e. the above discussed Caporin function extended to the whole analysed period, i.e.:

$$FC_t = \begin{cases} |VaR_\alpha(t) - r_t| \\ |VaR_\alpha(t) - r_t| \end{cases} for \begin{matrix} r_t < VaR_\alpha(t) \\ r_t \geq VaR_\alpha(t) \end{matrix}, \tag{19}$$

where r_t = realized rate of return in the period t,
$VaR_\alpha(t)$ = VaR forecast for the same period t.

In the case of this function, Caporin (2008) also proposes the average cost to be the final result for the model $FC = \sum_{t=1}^{N} FC_t / N$.

The last cost function used in the study is the absolute excessive cost. This function pays more attention to the excessive cost of using the particular model than its precision. The absolute function of excessive cost can be presented as follows:

$$CAE_t = \begin{cases} |VaR_\alpha(t)| & r_t \geq VaR_\alpha(t) \wedge r_t \geq 0 \\ |VaR_\alpha(t) - r_t| & for \; r_t \geq VaR_\alpha(t) \wedge r_t < 0, \\ |r_t| & r_t < VaR_\alpha(t) \end{cases} \tag{20}$$

where r_t = realized rate of return in the period t,
$VaR_\alpha(t)$ = VaR forecast for the same period t.

The final result of the function is the average cost $\overline{CAE} = \frac{\sum_{t=1}^{N} CAE_t}{N}$, where N is the number of all VaR forecasts. The interpretation of $\overline{CAE}$ can be the measure of model conservatism. The more conservative models will be attributed with a high $\overline{CAE}$, and more liberal models with a low $\overline{CAE}$. This is caused by straining the average cost value with too high VaR forecasts in relation to the realized rates of return.

It should be emphasised that in the cost functions presented, costs are scaled by the total number of exceedances/forecasts, what can lead to a preference of a model that has a higher number of exceedances, but a lower average loss. Because of that, the cost functions should be treated as a second step in the assessment process for those models, which according to the previous tests, provides the VaR forecasts of high quality.

4. Results of the empirical study

4.1. Data

The quality assessment of VaR forecasts was conducted in three periods, selected so that different periods with respect to the volatility can be distinguished. The daily rates of return were calculated based on the formula $r_t = \ln(\frac{p_t}{p_{t-1}})$ and VaR forecasts were constructed on their basis (Dowd, 2002).

Each of these periods lasts 11 years, which equals about 2600 – 2750 observations. The periods last, respectively: from January 1, 1999 to December 21, 2009 (the in-sample lasts from the beginning of 1999 to the beginning of 2008) – this period was called the period of increasing volatility, because the volatility in the out-of-sample is growing in relation to the end of the in-sample; from January 1, 2001 to December 21, 2011 (the in-sample lasts from the beginning of 2001 to the beginning of 2010) – this period was called the period of constantly high volatility, because the end of the in-sample period and the out-of-sample period are characterised by high volatility; from March 25, 2006 to March 24, 2017 (the in-sample lasts from March 2006 to March 2015) – this period was called the period of decreasing volatility, because the volatility in the out-of-sample is smaller than at the end of the in-sample. In each out-of-sample there were 500 one-day ahead VaR forecasts calculated at the significance level of 1%, which is in line with the Basel Committee regulations and the CRD IV Directive.

In order to compare the quality of VaR forecasts depending on the level of economic development, three developed and three developing countries were selected for the analysis. For each country, the primary stock index was selected: German DAX, American S&P500, Japanese Nikkei 225 and Polish WIG20, Chinese SSE COMP, Turkish XU100. The distributions of the analysed indexes were characterised by leptokurtosis (kurtosis ranged from 5,5-14) and in most cases the left tail skewness (skewness was less than zero), detailed descriptive statistics are presented in table 4. In addition, it is worth mentioning that the mean of each analysed time series, used in the study, was statistically not different from zero ($p < 0{,}05$).

[Please, insert Table 4 about here]

During the estimation process for VaR forecasting the rolling window method was used, i.e. parameters of each model were estimated 500 times. For each estimation, the width of the window was constant and equal to $N - 500$, where N is the number of periods from each of the analysed samples. Calculations were made in R 3.4.0 and MATLAB R2017a.

In case of forecasting GARCH models, an aspect that must be included is the examination of the occurrence of the ARCH (autocorrelation of squares of random errors) effect. This was done using the LM test (*Lagrange Multiplier*) and the Q Ljung-Box test for 8 lags (the number of lags was calculated based on the formula: $p = ln(N)$, where N is the size of the sample, com. Tsay (2005)). Each stock index in the analysed time intervals was characterised by a strong ARCH effect. Moreover, regardless of the tested GARCH model, it was possible to eliminate the ARCH effect in squares of the standardized residuals from the model for a satisfactory number of time series (at least in 5 out of 6). Therefore, it can be inferred that the class of GARCH models may be used to analyse the studied time series and their forecasts have at least correct results.

In addition, the use of the constant in the model was omitted. This decision was made because (as previously shown) the average rate of return value in each case was very close to 0.

4.2.Analysis at the level of the studied periods

The obtained results are shown in tables 1-3. Table 1 illustrates the numbers of exceedances, empirical excess ratio and cost function results of the analysed models for each index, over the analysed periods. Table 2 illustrates the results of statistical tests used for assessing the VaR forecast quality (Kupiec, Christoffersen, DQ), and table 3 shows the results of the Diebold-Mariano test.

[Please, insert Tables 1 – 3 about here]

The analysis of the test results should begin with GARCH(1,1) model with a normal distribution. Almost in each analysis of the VaR forecasts' quality, this model is considered as the basic benchmark, hence it is a reference point, because in many studies from before the financial crisis, it allowed to obtain high quality VaR forecasts (e.g. Engle (2004), Alonso and Arcos (2006), Angelidis, Benos and Degiannakis (2004)). However, the results of our study show its imperfections.

This model in each case was characterised by a higher excess ratio than the expected 1%, and in most cases (4 out of 6 for each of the analysed periods) by a much higher number of exceedances (9 and more) than expected (equal 5, corresponding 1% of cases), which qualified the model at least to the yellow zone of the Basel traffic light test in these cases. Moreover, the excess ratio observed for this model in most cases was the highest among the analysed periods. The Diebold-Mariano test indicated that only in the first period and only historical simulation (4 cases) and the CAViaR model (2 cases) provide forecasts less accurate than the GARCH(1,1) model.

The model's assessment is undermined by the fact that it is impossible to indicate the period, or assets, for which the quality of the model's forecasts would be relatively good and stable. The deterioration in the quality of forecasts depending on the asset takes place in different periods of the financial stability. This ascertainment is confirmed by the results of the Kupiec, Christoffersen and DQ tests. Based on them, it is straightforward that this model is relatively good for the XU100 index in the growing volatility, the DAX index in the constantly high volatility, for the WIG20 index in the decreasing volatility, and for S&P500 in none of the analysed periods.

The obtained results of the quality of VaR forecasts from the GARCH(1,1) model are weak enough to assume that this model should not be considered as a potential model for VaR forecasting, therefore the cost function analysis is unfounded. The most likely cause for such bad results is the too liberal assumption of normal distribution.

The second model, which in many studies is treated as a benchmark VaR model, is historical simulation. The use of historical simulation in the period with increasing volatility leads to extremely poor quality of VaR forecasts. High, explosive volatility highlights one of the biggest drawbacks of this method – a very long period of assimilation to market changes. Providing strong fall of stock market condition occurs, historical simulation predictions will not react quickly enough to increasing volatility. In this period, the VaR forecasts obtained are characterised by the highest number of exceedances among the analysed models (for S&P500 *excess ratio* reached 6%!). Kupiec, Christoffersen and DQ tests confirm the low quality of historical simulation forecasts. For most indexes and studied periods, the null hypothesis should be rejected. Similarly, the Diebold-Mariano test confirms the low precision of forecasts with respect to the other models.

Historical simulation achieved completely opposite results in periods of constantly high decreasing volatility. In these periods, in most cases, *excess ratio* is less than 1%, and only in one case the model would be qualified to the yellow zone of the Basel traffic light test. The high quality of forecasts is also confirmed by the Kupiec, Christoffersen and DQ tests. In most cases, there is no reason to reject hypotheses of good quality of forecasts, and apart from the SSE COMP index in the III period, all cases of rejecting the null hypothesis result from excessive conservatism. From the prudential perspective, excessive conservatism is not a problem, but it translates into the cost of modelling – the cost function values are the highest among all analysed models. Based on the Diebold-Marino test, it should be concluded that historical simulation is not worse than all considered models.

By analysing the results of historical simulation obtained for periods of constantly high and decreasing volatility, it can be concluded that historical simulation in the period of high volatility provides a high quality forecast. In our opinion, the obtained results reveal exactly the same problem (only from the other side), as the results for the period of increasing volatility – a very long period of assimilation to the market changes. In this case, the forecasts obtained are extremely conservative, because the model is estimated in a very turbulent period.

In conclusion, historical simulation model provides the VaR forecast of poor quality and should not be considered as a VaR predicting model in any of the analysed periods.

For GARCH(1,1) with other than normal random error distributions (t-Student, skewed t-Student and empirical) and the CAViaR(1,1) model the quality of forecasts should be considered better than for the previously analysed models. The number of exceedances of individual models makes them qualified to the green zone of the Basel traffic light test in most cases. In most cases, the decision about qualification to a particular zone would be analogous for all models. The exception is the result for the SSE COMP index in the period of decreasing volatility, for which only the QML-GARCH model would be qualified to the green zone and the results for the CAViaR model in the period of growing volatility – for all indexes from the developed countries the yellow light was assigned (the green light was assigned to the remaining models).

The high quality of forecasts based on the discussed models is confirmed by the results of the Kupiec, Christoffersen and DQ tests. In the period of increasing volatility, the CAViaR model is clearly a worse model (Kupiec and DQ test). In the period of constant and decreasing volatility, relatively better test results were obtained by the GARCH(1,1) models with skewed t-Student distribution and the CAViaR model. Based on the Diebold-Mariano test, the best models during the increasing volatility are the QML-GARCH and GARCH-st models, in the period of constant volatility the GARCH-st model, and then QML-GARCH and CAViaR, and in the period of decreasing volatility, CAViaR, GARCH-st and QML-GARCH.

By analysing model costs, it is straightforward that among the analysed models, the GARCH-st model is relatively the most expensive model to be used.

In conclusion, the test results show that all models predict relatively well. This does not change the fact that each of them reveals some weaknesses. The greatest weakness of the CAViaR model are the results of the relatively low quality for the increasing volatility periods. Therefore, it seems that other models should be preferred to it, as they generate a lower risk of maladjustment of the model to increase in volatility.

Of the remaining three models, the GARCH-t model obtained the relatively worst results, which based on the Diebold-Mariano test was most often indicated as significantly worse than other models.

The choice between the GARCH-st and QML-GARCH model depends on the preferences concerning the quality of forecasts and costs of maintaining the model. The GARCH-st model provides slightly better forecasts, with slightly greater costs of using the model.

Comparing forecasts for developed and developing countries, it should be noted that their results are close to each other. It is difficult to point out many differences in the quality of forecasts for individual models depending on the level of development of the country.

The only phenomenon worth emphasising is the clearly poorer quality of VaR forecasts for historical simulation and the CAViaR model in the period of increasing volatility for the developed countries. In our evaluation, the most probable cause of this phenomenon is the relative stability of the developed countries. It made that there were relatively less events far from average in the available (before the increase in volatility) history, and for that reason the adaptation of the model to the new conditions on the developed markets took longer time.

5. Conclusions

Research evaluating the quality of VaR forecasts based on the market stability indicates that different models predict VaR best in periods of high and relatively low volatility.

The study attempted to assess the quality of VaR forecasts of the GARCH(1,1) models with the normal, t-Student and skewed t-Student distribution, the QML-GARCH(1,1) model, historical simulation and the CAViaR model. For the purposes of the study, the indexes were selected from the developed and developing countries: DAX (Germany), S&P500 (USA), Nikkei 225 (Japan) – developed countries and WIG20 (Poland), SSE COMP (China), XU100 (Turkey) – developing countries. The data samples used in the analysis were selected so that each of them had a different tendency of market volatility (increasing, constantly high and decreasing).

The GARCH(1,1) model with the skewed t-Student distribution and the QML-GARCH model were recognized as the best ones based on the obtained results. The distinguishing feature of both models is the ability to simultaneously model thick tails and distribution skewness (both phenomena were observed in the data). These results are consistent with the conclusions from the Abad et al. (2014) summary.

As in the aforementioned summary, the obtained results tell us to reject the GARCH(1,1) model and historical simulation from the group of well-forecasting VaR. This statement can be expanded to all three studied periods.

Contrary conclusions to Abad et al. (2014), based on our findings, should be drawn regarding the quality of VaR forecasts for the CAViaR model. In the studies discussed by Abad et al. (2014), the CAViaR model was considered to be a very well forecasting model. However, the results obtained in our study show that the CAViaR model fails while volatility increases and therefore cannot be regarded as a well predicting VaR model in each market situation. Moreover, the study results indicate the clearly lower quality of the CAViaR model among the developed countries in the period of increasing volatility.

The models, for which the VaR forecasts with the best quality were obtained in the study, are quite conservative in nature, so it is worth looking into the verification of the quality of VaR forecasts in the future for these models in the situation of relative stability of the market. Furthermore, it is worth comparing the quality of forecasts of the discussed models with the group of models from the EVT family (*Extreme Value Theory*) that is gaining reputation.

Bibliography

Abad, P., Benito, S., López, C. (2014). A comprehensive review of Value at Risk methodologies. *The Spanish Review of Financial Economics*. 12(1), 15-32.

Alonso, J. C., Arcos, M. A. (2006). Cuatro hechos estilizados de las series de rendimientos: Una ilustración para Colombia. *Estudios Gerenciales*. 22(100), 103-124.

Angelidis T., Benos A., Degiannakis S. (2004). The Use of GARCH Models in VaR Estimation. *Statistical Methodology*. 1(2), 105-128.

Barone-Adesi G., Giannopoulos K., Vosper L. (1999). VaR without correlations for nonlinear portfolios. *Journal of Futures Markets*. 19(5), 583–602.

Basle Committee on Banking Supervision. (1996). *Supervisory Framework for the Use of "Backtesting" In Conjunction with the Internal Models Approach to Market Risk Capital Requirements.*

Bollerslev T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics.* 31(3), 307-327.

Bollerslev T. (1987). Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics.* 69(3), 542-547.

Caporin M. (2008). Evaluating Value-at-Risk measures in presence of long memory conditional volatility. *Journal of Risk.* 10(3), 79-110.

Christoffersen P. (1998). Evaluating Interval Forecasts. *International Economic Review*. 39(4), 841–862.

Diebold F.X., Mariano R.S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*. 20(1), 134-144.

Dowd K. (2002). *Measuring Market Risk.* Chichester: John Wiley & Sons, Ltd.

Engle R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica.* 50(4), 987-1007.

Engle R. F. (2001). GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives*. 15(4), 157-168.

Engle R. F. (2004). Risk and volatility: Econometric models and financial practice. *The American economic review*. 94(3), 405-420.

Engle R. F., Manganelli S. (2001). Value at Risk models in finance. *European Central Bank Working Papers*. 75.

Engle R. F., Manganelli S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics.* 22(4), 367-381.

Jorion P. (2007). *Value at Risk – The New Benchmark for Managing Financial Risk.* New York: McGraw-Hill,.

Kupiec, P. (1995). Techniques for Verifying the Accuracy of Risk Management Models. *Journal of Derivatives.* 3(2), 73-84.

McAleer M., Jimenez-Martin J.-A., Perez-Amaral T. (2009). *Has the Basel II Accord Encouraged Risk Management During the 2008-2009 Financial Crisis?*

Tsay R.S. (2005). *Analysis of Financial Time Series*, Chicago: John Wiley & Sons, Ltd,

Sarma M., Thomas S., Shah A. Selection of Value-at-Risk Models. (2003). *Journal of Forecasting*. 22(4), 337-358.

Shams M., Sina A. (2014). Evaluating Market Risk Assessment through VAR Approach before and after Financial Crisis in Tehran Stock Exchange Market (TSEM). *Journal of Management and Sustainability*. 4(2), 134-146.

Su J.B., Hung J.C. (2011). Empirical analysis of jump dynamics, heavy-tails and skewness on value-at-risk estimation. *Economic Modelling.* 28(3), 1117-1130.

Tagliafichi R.A. (2003). *The estimation of Market VaR using GARCH models and a heavy tail distributions.* Paper presented at ASTIN Colloquium International Actuarial Association, Brussels.

Table 1. The number of exceedances, excess ratio, results of the Basel light test and the cost function values divided by the analysed models and periods, for each observed index

		PERIOD I (INC. VOL.)						PERIOD II (HIGH VOL.)						PERIOD III (DEC. VOL.)					
MODEL	**INDEKS**	**E**	**ER**	**LT**	**RC**	**FC**	**CAE**	**E**	**ER**	**LT**	**RC**	**FC**	**CAE**	**E**	**ER**	**LT**	**RC**	**FC**	**CAE**
GARCH(1,1)	WIG20	12	2,40%		0,78%	4,12%	3,52%	9	1,80%		2,76%	2,73%	2,36%	7	1,40%		1,72%	2,34%	2,01%
	XU100	6	1,20%		4,01%	5,34%	4,55%	8	1,60%		4,19%	3,90%	3,37%	10	2,00%		4,13%	3,38%	2,94%
	SSE COMP	16	3,20%		1,63%	5,39%	4,68%	10	2,00%		5,40%	3,18%	2,79%	13	2,60%		4,02%	4,12%	3,63%
	A2	11,33	2,27%		2,14%	4,95%	4,25%	*9*	*1,80%*		*4,12%*	*3,27%*	*2,84%*	*10*	*2,00%*		*3,29%*	*3,28%*	*2,86%*
	DAX	10	2,00%		2,38%	4,44%	3,80%	9	1,80%		1,01%	3,27%	2,79%	10	2,00%		4,22%	3,13%	2,69%
	Nikkei 225	7	1,40%		2,17%	4,80%	4,07%	7	1,40%		4,64%	3,19%	2,74%	14	2,80%		2,49%	3,41%	2,96%
	S&P500	14	2,80%		4,32%	4,37%	3,74%	12	2,40%		2,13%	2,84%	2,42%	7	1,40%		1,73%	2,07%	1,78%
	A1	10,33	2,07%		3,25%	4,53%	3,87%	*9,33*	*1,87%*		*3,38%*	*3,10%*	*2,65%*	*10,33*	*2,07%*		*2,82%*	*2,87%*	*2,48%*
	AA	10,83	2,17%		2,58%	4,74%	4,06%	*9,17*	*1,83%*		*3,82%*	*3,19%*	*2,75%*	*10,17*	*2,03%*		*3,05%*	*3,07%*	*2,67%*
GARCH-t(1,1)	WIG20	10	2,00%		0,99%	4,42%	3,82%	7	1,40%		2,98%	2,93%	2,56%	5	1,00%		1,87%	2,55%	2,21%
	XU100	4	0,80%		4,43%	5,76%	4,94%	3	0,60%		4,54%	4,21%	3,65%	7	1,40%		4,46%	3,64%	3,19%
	SSE COMP	5	1,00%		2,15%	6,05%	5,23%	9	1,80%		5,94%	3,56%	3,18%	11	2,20%		4,71%	4,63%	4,14%
	A2	6,33	1,27%		2,52%	5,41%	4,66%	*6,33*	*1,27%*		*4,49%*	*3,57%*	*3,13%*	*7,67*	*1,53%*		*3,68%*	*3,61%*	*3,18%*
	DAX	7	1,40%		2,45%	4,67%	4,01%	7	1,40%		1,11%	3,46%	2,96%	5	1,00%		4,44%	3,41%	2,95%
	Nikkei 225	7	1,40%		2,41%	5,11%	4,38%	6	1,20%		4,80%	3,35%	2,89%	10	2,00%		2,62%	3,61%	3,15%
	S&P500	7	1,40%		4,54%	4,73%	4,06%	12	2,40%		2,24%	3,05%	2,64%	6	1,20%		1,99%	2,28%	1,99%
	A1	7	1,40%		3,48%	4,84%	4,15%	*8,33*	*1,67%*		*3,52%*	*3,29%*	*2,83%*	*7*	*1,40%*		*3,02%*	*3,10%*	*2,70%*
	AA	6,67	1,33%		2,91%	5,12%	4,41%	7,33	1,47%		4,10%	3,43%	2,98%	*7,33*	*1,47%*		*3,35%*	*3,35%*	*2,94%*
GARCH-st(1,1)	WIG20	9	1,80%		1,08%	4,53%	3,92%	7	1,40%		3,04%	3,04%	2,67%	4	0,80%		1,97%	2,67%	2,33%
	XU100	4	0,80%		4,51%	5,87%	5,05%	3	0,60%		4,66%	4,41%	3,85%	7	1,40%		4,71%	3,86%	3,41%
	SSE COMP	4	0,80%		2,39%	6,37%	5,55%	9	1,80%		6,14%	3,72%	3,34%	10	2,00%		5,00%	4,91%	4,40%
	A2	5,67	1,13%		2,66%	5,59%	4,84%	*6,33*	*1,27%*		*4,62%*	*3,72%*	*3,29%*	*7*	*1,40%*		*3,89%*	*3,81%*	*3,38%*
	DAX	4	0,80%		2,60%	4,96%	4,29%	5	1,00%		1,30%	3,69%	3,18%	4	0,80%		4,66%	3,63%	3,16%
	Nikkei 225	7	1,40%		2,51%	5,28%	4,56%	4	0,80%		4,97%	3,53%	3,06%	10	2,00%		2,75%	3,83%	3,38%
	S&P500	5	1,00%		4,69%	4,95%	4,27%	9	1,80%		2,35%	3,25%	2,83%	5	1,00%		2,23%	2,48%	2,19%
	A1	5,33	1,07%		3,60%	5,07%	4,37%	*6*	*1,20%*		*3,66%*	*3,49%*	*3,02%*	*6,33*	*1,27%*		*3,21%*	*3,32%*	*2,91%*
	AA	5,5	1,10%		3,03%	5,33%	4,60%	6,17	1,23%		4,23%	3,61%	3,15%	*6,67*	*1,33%*		*3,55%*	*3,56%*	*3,14%*

QML-GARCH(1,1)	WIG20	9	1,80%		1,04%	4,59%	3,98%	6	1,20%		3,04%	3,07%	2,68%	4	0,80%		1,99%	2,64%	2,30%
	XU100	3	0,60%		4,57%	6,07%	5,24%	6	1,20%		4,47%	4,10%	3,57%	7	1,40%		4,52%	3,74%	3,29%
	SSE COMP	8	1,60%		1,82%	5,88%	5,09%	9	1,80%		5,76%	3,53%	3,14%	7	1,40%		4,95%	5,05%	4,50%
	A2	6,67	1,33%		2,48%	5,51%	4,77%	*7*	*1,40%*		*4,42%*	*3,56%*	*3,13%*	*6*	*1,20%*		*3,82%*	*3,81%*	*3,36%*
	DAX	6	1,20%		2,53%	4,79%	4,13%	6	1,20%		1,22%	3,53%	3,02%	5	1,00%		4,45%	3,41%	2,95%
	Nikkei 225	7	1,40%		2,36%	5,16%	4,44%	5	1,00%		4,82%	3,31%	2,85%	10	2,00%		2,69%	3,83%	3,37%
	S&P500	8	1,60%		4,48%	4,66%	3,99%	12	2,40%		2,24%	3,06%	2,65%	6	1,20%		2,12%	2,47%	2,19%
	A1	7	1,40%		3,42%	4,87%	4,18%	*7,67*	*1,53%*		*3,53%*	*3,30%*	*2,84%*	*7*	*1,40%*		*3,09%*	*3,24%*	*2,84%*
	AA	6,83	1,37%		2,86%	5,19%	4,48%	7,33	1,47%		4,06%	3,43%	2,98%	*6,5*	*1,30%*		*3,45%*	*3,52%*	*3,10%*
Historical simulation	WIG20	19	3,80%		1,35%	3,56%	3,01%	8	1,60%		3,79%	3,58%	3,22%	2	0,40%		3,88%	3,91%	3,56%
	XU100	3	0,60%		7,01%	6,71%	5,89%	1	0,20%		7,02%	5,62%	5,05%	2	0,40%		5,60%	4,94%	4,46%
	SSE COMP	19	3,80%		1,46%	4,88%	4,19%	1	0,20%		7,00%	5,18%	4,74%	13	2,60%		5,50%	5,79%	5,31%
	A2	13,67	2,73%		3,27%	5,05%	4,36%	*3,33*	*0,67%*		*5,94%*	*4,79%*	*4,33%*	*5,67*	*1,13%*		*4,99%*	*4,88%*	*4,45%*
	DAX	14	2,80%		4,66%	4,87%	4,29%	6	1,20%		3,28%	5,03%	4,55%	2	0,40%		6,03%	4,58%	4,09%
	Nikkei 225	23	4,60%		2,74%	4,45%	3,84%	2	0,40%		5,74%	4,59%	4,10%	4	0,80%		5,13%	4,94%	4,45%
	S&P500	30	6,00%		4,20%	3,76%	3,24%	4	0,80%		4,44%	4,09%	3,65%	0	0,00%		-	4,31%	4,00%
	A1	22,33	4,47%		3,47%	4,36%	3,79%	*4*	*0,80%*		*5,09%*	*4,57%*	*4,10%*	*2*	*0,40%*		*5,58%*	*4,61%*	*4,18%*
	AA	18	3,60%		3,35%	4,71%	4,08%	3,67	0,73%		5,60%	4,68%	4,22%	*3,83*	*0,77%*		*5,23%*	*4,74%*	*4,31%*
CAViaR	WIG20	10	2,00%		1,00%	4,52%	3,91%	3	0,60%		2,91%	3,11%	2,71%	5	1,00%		1,94%	2,71%	2,37%
	XU100	5	1,00%		4,44%	5,88%	5,08%	4	0,80%		4,84%	4,33%	3,79%	8	1,60%		4,32%	3,77%	3,33%
	SSE COMP	8	1,60%		2,45%	5,96%	5,17%	9	1,80%		6,03%	3,43%	3,04%	9	1,80%		4,82%	5,03%	4,50%
	A2	7,67	1,53%		2,63%	5,45%	4,72%	*5,33*	*1,07%*		*4,60%*	*3,63%*	*3,18%*	*7,33*	*1,47%*		*3,69%*	*3,84%*	*3,40%*
	DAX	11	2,20%		2,64%	4,34%	3,72%	8	1,60%		1,14%	3,45%	2,97%	5	1,00%		4,51%	3,36%	2,90%
	Nikkei 225	11	2,20%		2,14%	4,77%	4,07%	4	0,80%		4,57%	3,38%	2,92%	10	2,00%		2,82%	3,79%	3,33%
	S&P500	10	2,00%		4,36%	4,45%	3,79%	11	2,20%		2,26%	3,03%	2,61%	5	1,00%		2,14%	2,54%	2,25%
	A1	10,67	2,13%		3,25%	4,52%	3,86%	*7,67*	*1,53%*		*3,41%*	*3,29%*	*2,83%*	*6,67*	*1,33%*		*3,15%*	*3,23%*	*2,83%*
	AA	9,17	1,83%		2,88%	4,99%	4,29%	6,5	1,30%		4,12%	3,46%	3,01%	*7*	*1,40%*		*3,42%*	*3,53%*	*3,11%*

Note: This table above presents the number of exceedances, excess ratio, zones of the Basel light test, Caporin cost function, Caporin firm's cost and excessive cost function, divided by the analysed periods and models for each analysed index. The analysis was performed at 500 *Value-at-Risk* forecasts generated by each model. Abbreviations used in the table: E – number of exceedances, ER – excess ratio, BSLT – Basel light test, RC – Caporin cost function, FC – Caporin firm's cost function, CAE – excessive cost function, A2 – average of each result for developing countries, A1 – average of each result for developed countries, AA – average of the model for each period.
Source: own calculations.

Table 2. Test results: Kupiec, Christoffersen, DQ divided into the analysed models and periods, for each observed index

		PERIOD I (INC. VOL.)			PERIOD II (HIGH VOL.)			PERIOD III (DEC. VOL.)		
MODEL	**INDEKS**	**UC**	**CC**	**DQ**	**UC**	**CC**	**DQ**	**UC**	**CC**	**DQ**
GARCH(1,1)	WIG20	0,008	0,021	0,003	0,106	0,007	0,000	0,397	0,629	0,374
	XU100	0,663	0,844	0,958	0,215	0,008	0,000	0,048	0,114	0,040
	SSE COMP	0,000	0,000	0,000	0,048	0,114	0,014	0,003	0,008	0,000
	DAX	0,048	0,114	0,026	0,106	0,228	0,296	0,048	0,058	0,008
	Nikkei 225	0,397	0,629	0,738	0,397	0,149	0,005	0,001	0,003	0,000
	S&P500	0,001	0,003	0,000	0,008	0,021	0,002	0,397	0,006	0,000
GARCH-t(1,1)	WIG20	0,048	0,114	0,044	0,397	0,006	0,000	1,000	0,951	0,783
	XU100	0,641	0,870	0,919	0,331	0,021	0,000	0,397	0,629	0,435
	SSE COMP	1,000	0,106	0,000	0,106	0,228	0,060	0,020	0,051	0,011
	DAX	0,397	0,629	0,358	0,397	0,629	0,814	1,000	0,951	0,873
	Nikkei 225	0,397	0,629	0,681	0,663	0,142	0,002	0,048	0,058	0,002
	S&P500	0,397	0,629	0,400	0,008	0,021	0,001	0,663	0,004	0,000
GARCH-st(1,1)	WIG20	0,106	0,228	0,111	0,397	0,006	0,000	0,641	0,870	0,963
	XU100	0,641	0,870	0,913	0,331	0,021	0,000	0,397	0,629	0,413
	SSE COMP	0,641	0,059	0,000	0,106	0,228	0,054	0,048	0,114	0,083
	DAX	0,641	0,870	0,911	1,000	0,951	0,911	0,641	0,870	0,853
	Nikkei 225	0,397	0,629	0,681	0,641	0,059	0,000	0,048	0,058	0,002
	S&P500	1,000	0,951	0,742	0,106	0,228	0,063	1,000	0,106	0,000
QML-GARCH(1,1)	WIG20	0,106	0,228	0,109	0,663	0,004	0,000	0,641	0,870	0,962
	XU100	0,331	0,615	0,761	0,663	0,004	0,000	0,397	0,629	0,381
	SSE COMP	0,215	0,128	0,000	0,106	0,228	0,047	0,397	0,629	0,809
	DAX	0,663	0,844	0,584	0,663	0,844	0,845	1,000	0,951	0,872
	Nikkei 225	0,397	0,629	0,737	1,000	0,106	0,000	0,048	0,058	0,002
	S&P500	0,215	0,404	0,202	0,008	0,021	0,002	0,663	0,004	0,000
Historical simulation	WIG20	0,000	0,000	0,000	0,215	0,008	0,000	0,125	0,308	0,561
	XU100	0,331	0,615	0,845	0,028	0,091	0,341	0,125	0,308	0,594
	SSE COMP	0,000	0,000	0,000	0,028	0,091	0,343	0,003	0,007	0,000
	DAX	0,001	0,003	0,000	0,663	0,844	0,002	0,125	0,308	0,517
	Nikkei 225	0,000	0,000	0,000	0,125	0,004	0,000	0,641	0,870	0,971
	S&P500	0,000	0,000	0,000	0,641	0,870	0,902	-	-	-
CAViaR	WIG20	0,048	0,114	0,058	0,331	0,615	0,843	1,000	0,951	0,546
	XU100	1,000	0,951	0,956	0,641	0,059	0,000	0,215	0,404	0,263
	SSE COMP	0,215	0,128	0,000	0,106	0,228	0,028	0,106	0,228	0,317
	DAX	0,020	0,051	0,037	0,215	0,404	0,496	1,000	0,951	0,880
	Nikkei 225	0,020	0,051	0,023	0,641	0,059	0,000	0,048	0,058	0,001
	S&P500	0,048	0,114	0,029	0,020	0,051	0,014	1,000	0,106	0,000

Note: The table above presents the results of formal tests: Kupiec (unconditional coverage), Christoffersen (conditional coverage), and Dynamic Quantile for each analysed index divided by models in all analysed periods. Abbreviations used in the table: UC – p-value of the unconditional coverage test, CC – p-value of the conditional coverage test, DQ – p-value of the Dynamic Quantile test. The number of lags selected in the DQ test is 3.
Tests were performed at the 5% significance level. Green fields indicate p-values greater than 5%.
Source: own calculations.

Table 3. Diebold – Mariano test results divided into the analysed models and periods

	PERIOD I (INC. VOL.)							PERIOD II (HIGH VOL.)							PERIOD III (DEC. VOL.)						
	GARCH(1,1)	**GARCH-t(1,1)**	**GARCH-st(1,1)**	**QML-GARCH(1,1)**	**HS**	**CAViaR**	**RANKING**	**GARCH(1,1)**	**GARCH-t(1,1)**	**GARCH-st(1,1)**	**QML-GARCH(1,1)**	**HS**	**CAViaR**	**RANKING**	**GARCH(1,1)**	**GARCH-t(1,1)**	**GARCH-st(1,1)**	**QML-GARCH(1,1)**	**HS**	**CAViaR**	**RANKING**
GARCH(1,1))	-	0	0	0	4	2	5th	-	0	0	0	0	0	6th	-	0	0	0	0	0	6th
GARCH-t(1,1)	6	-	0	2	5	4	3rd	6	-	0	3	0	3	4th	6	-	0	0	0	1	5th
GARCH-st(1,1)	6	6	-	4	5	6	1st	6	6	-	5	0	5	2nd	6	6	-	4	0	2	2nd
QML-GARCH(1,1)	6	4	2	-	5	5	2nd	6	2	1	-	0	4	3rd	6	5	2	-	0	2	3rd
HS	1	1	1	1	-	1	6th	6	6	6	6	-	6	1st	6	6	6	5	-	5	1st
CAViaR	3	0	0	0	4	-	4th	6	2	0	2	0	-	5th	6	5	2	2	0	-	3rd

Note: The table above shows the Diebold-Mariano test results for the null hypothesis that the forecast are of equal goodness of fit and alternative hypothesis that the first tested forecast (rows) is better than the second one (columns). The values in the table indicate in how many cases one model (rows) was proven to be better than the other one (columns), i.e. the null hypothesis was rejected. In some cases tested models were proven to be equally good, hence in such situation no points were granted. HS is an abbreviation of historical simulation. Tests were carried out at the significance level of 5%. Columns marked as RANKING presents overall goodness of a forecasts of model in a data sample. Each of them ranks the models on the basis of averaged number of cases the model was proven better (the more cases model was better, the higher its position in the ranking).
Source: own calculations.

Table 4. Descriptive statistics of the analysed time series for each index

PERIOD I (INC. VOL.)								
INDEX	**Mean**	**Median**	**Min**	**Max**	**St. Dev.**	**JB**	**Kurtosis**	**Skewness**
WIG20	0,000	0,001	-0,085	0,068	0,015	756,54 (0,00)	5,53 (0,00)	-0,22 (0,00)
XU100	0,001	0,001	-0,200	0,178	0,027	2961,18 (0,00)	8,095 (0,00)	0,11 (0,00)
SSE COMP	0,000	0,001	-0,093	0,094	0,017	1592,81 (0,00)	6,80 (0,00)	-0,04 (0,00)
DAX	0,000	0,001	-0,089	0,108	0,017	2007,34 (0,00)	7,15 (0,00)	0,04 (0,00)
Nikkei 225	0,000	0,000	-0,121	0,132	0,016	4326,12 (0,00)	9,17 (0,00)	-0,29 (0,00)
S&P500	0,000	0,000	-0,095	0,110	0,014	6344,62 (0,00)	10,42 (0,00)	-0,10 (0,00)
PERIOD II (HIGH VOL.)								
INDEX	**Mean**	**Median**	**Min**	**Max**	**St. Dev.**	**JB**	**Kurtosis**	**Skewness**
WIG20	0,000	0,001	-0,083	0,061	0,014	856,42 (0,00)	5,66 (0,00)	-0,30 (0,00)
XU100	0,001	0,001	-0,200	0,127	0,023	4098,03 (0,00)	8,95 (0,00)	-0,26 (0,00)
SSE COMP	0,000	0,001	-0,093	0,094	0,017	1580,03 (0,00)	6,77 (0,00)	-0,12 (0,01)
DAX	0,000	0,001	-0,089	0,108	0,017	2097,39 (0,00)	7,24 (0,00)	0,01 (0,77)
Nikkei 225	0,000	0,000	-0,121	0,132	0,016	5183,64 (0,00)	9,74 (0,00)	-0,39 (0,00)
S&P500	0,000	0,001	-0,095	0,110	0,014	6606,68 (0,00)	10,56 (0,00)	-0,17 (0,00)
PERIOD III (DEC. VOL.)								
INDEX	**Mean**	**Median**	**Min**	**Max**	**St. Dev.**	**JB**	**Kurtosis**	**Skewness**
WIG20	0,000	0,000	-0,083	0,061	0,013	1776,57 (0,00)	6,82 (0,00)	-0,48 (0,00)
XU100	0,000	0,001	-0,111	0,121	0,017	1823,92 (0,00)	6,93 (0,00)	-0,28 (0,00)
SSE COMP	0,000	0,001	-0,093	0,090	0,018	1822,68 (0,00)	6,85 (0,00)	-0,62 (0,00)
DAX	0,000	0,001	-0,074	0,108	0,014	3775,52 (0,00)	8,69 (0,00)	-0,03 (0,00)
Nikkei 225	0,000	0,001	-0,121	0,132	0,016	6449,06 (0,00)	10,51 (0,00)	-0,49 (0,00)
S&P500	0,000	0,001	-0,095	0,110	0,013	13160,89 (0,00)	13,66 (0,00)	-0,34 (0,00)

Note: The table above presents descriptive statistics of rates of return from each of the discussed indexes in each period. P-values of tests are given in the parentheses. Abbreviations used in the table: JB – Jarque-Bera test for the normality of random variable. Kurtosis and skewness – test statistics of the kurtosis and skewness tests from the normal distribution.
Source: own calculations.

Faculty of Economic Sciences
University of Warsaw
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl