University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 10/2021 (358)

HCR & HCR-GARCH – Novel Statistical Learning Models for Value at Risk Estimation

Michał Woźniak
Marcin Chlebus

Warsaw 2021

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

HCR & HCR-GARCH – novel statistical learning models for Value at Risk estimation

Michał Woźniak*, Marcin Chlebus

University of Warsaw, Faculty of Economic Sciences
** Corresponding author: mj.wozniak9@student.uw.edu.pl*

Abstract: Market risk researchers agree that an ideal model for Value at Risk (VaR) estimation does not exist, different models performance strongly depends on current economic circumstances. Under the conditions of sudden volatility increase, such as during the global economic crisis caused by the Covid-19 pandemic, no classical VaR model worked properly even for the group of the largest market indices. Therefore, the aim of the article is to present and formally test three novel statistical learning models for VaR estimation: HCR, HCR-GARCH and HCR-QML-GARCH, which, by considering additional volatility term (due to time context and statistical moments), should be able to perform well in times of market turbulence. In the benchmark procedure we compare the 1% and 2.5% one-day-ahead VaR forecasts obtained with the above models against the estimates of classical methods like: Historical Simulation, KDE, Modified Cornish-Fisher Expansion, GARCH(1,1) with varied distributions, RiskMetrics™, EVT and QML-GARCH. Four periods that vary in terms of market volatility: 2006-9, 2008-11, 2014-17 and mid-2016 to mid-2020 for six different stock market indexes: DAX, WIG 20, MOEX, S&P 500, Nikkei and SHC are selected. Models quality is tested from two perspectives: fulfilling regulatory requirements and forecasting adequateness. Obtained results show that HCR-GARCH outperforms other models during periods of sudden increased volatility in the markets. At the same time, HCR-QML-GARCH liberalizes the conservative estimates of HCR-GARCH and allows its use under moderate volatility, without any major loss of quality in times of crisis.

Keywords: Value at Risk, Hierarchical Correlation Reconstruction, GARCH, Standardized Residuals

JEL codes: G32, C52, C53, C58

Working Papers contain preliminary research results. Please consider this when citing the paper. Please contact the authors to give comments or to obtain revised version. Any mistakes and the views expressed herein are solely those of the authors

1. Introduction and literature review

Global economic recessions, like the 2008 financial crisis or the 2020 stock market crash (COVID-19 recession), emphasized the importance of adequate risk management by financial institutions. Such organizations distinguish the following main sources of risk: market risk, credit risk and operational risk (Bank for International Settlements 2006). The main subject of our interest in this research is market risk, which determines the risk of losses on financial investments caused by adverse movements in interest rates, foreign exchange rates and prices of financial instruments (Risk.net 2020).

Currently, one of the most popular measure of uncertainty in the market risk is Value at Risk (VaR) (Mostafa et al. 2017). It can be defined as maximum expected portfolio loss over a given period, at assumed confidence level, under normal market conditions (Jorion 2001). Importantly, VaR has been adopted to determine capital requirements by both banking and insurance institutions, respectively due to international regulatory frameworks such as Basel III and Solvency II (Corlosquet et al. 2013). The main advantages of VaR are its ease of interpretation and the ability to compare portfolios risk using it (Szubzda and Chlebus 2019). However, several researchers demonstrate the strong limitations of VaR measure, for instance: it does not control scenarios exceeding VaR and it is not a coherent measure (e.g., Rockafellar and Uryasev 2002, Sarykalin et al. 2008). Consequently, an alternative risk measure with superior statistical properties (which address most VaR issues) - Expected Shortfall (ES) was developed by Rockafellar and Uryasev (2002). ES is defined as expected return on the portfolio in the scenarios worse than the VaR. As result, ES measure will replace VaR as Basel IV becomes effective (PricewaterhouseCoopers 2016). Nevertheless, the study of VaR estimation is still very much in order. First and foremost, VaR is more robust than ES due to putting less weight on extreme tail events (Lux et al. 2020), moreover, by definition, a reliable VaR forecasting is useful for ES estimation and even though VaR is going to be replaced by ES, it will be still a core element of backtesting process for instance for trading desks (PricewaterhouseCoopers 2016).

Generally, market risk researchers agree that it is impossible to establish one best model for estimating Value at Risk regardless of market conditions (Abad et al. 2014). Critical to the performance of the model are the characteristics of both the financial instrument and the period under consideration (Chlebus and Buczyński 2018). Abad et al. (2014) cross-analyzed the results of many significant studies on VaR estimation. They take into account such well-established statistical models in the business world as well as in academia like: historical simulation (HS), kernel density estimation (KDE), RiskMetrics™, generalized autoregressive conditional heteroskedasticity (GARCH) model class (with different standardized error distributions), filtered historical simulation (FHS), conditional autoregressive value at risk (CAViaR), Extreme Value Theory (EVT) - Block Maxima Models (BMM), Extreme Value Theory - Peaks Over Threshold models (POT), Monte Carlo models etc. One of the main conclusions of their study is that the semiparametric approaches like EVT or FHS appear to be the best for VaR estimation in terms of performance and stability. More recent statistical studies largely support the above conclusion about the superiority of semiparametric models over other approaches (e.g, Taylor 2019, Patton et al. 2019, Wang and Zhao 2016, Yang and Hamori 2020, Avdulaj and Barunik 2017, Gerlach and Wang 2020). In counterpoint, there are also studies that indicate that classical parametric models are still the most stable and accurate (e.g., Buczyński and Chlebus 2018, Buczyński and Chlebus 2019). For purely nonparametric models, there is not much scientific publication nowadays (not including machine learning approaches). Usually the above-mentioned publications (presenting novel models) benchmark themselves to nonparametric models such as KDE or HS.

The development of machine learning (ML) also has important implications for market risk. An increasing number of proposed models for VaR and ES estimation use supervised learning approaches (mostly in semiparametric or nonparametric context). In the case of shallow ML, an immensely popular model used in this field is the Support Vector Machine (SVM) (Vapnik 1995). Its interesting adaptation was presented by Lux et al. (2020) who created the hybrid SVR-GARCH-KDE model. In this estimator mean and volatility are forecasted via SVR where the volatility model is motivated by GARCH formulation. What is more, they apply KDE on standardized residuals to smooth them. SVR-GARCH-KDE hybrid performs competitive to classical econometric models like GARCH. Other approaches using SVR to estimate the uncertainty metrics considered above have been proposed by: Khan (2011), Radović et al. (2015) and Xu et al. (2016). They argue that their methods are in competition with classical models. In line with intuition, models based on neural networks also play a strong role in successful VaR and ES estimation, for instance: Nguyen et al. (2019) proposed long short-term memory stochastic volatility (LSTM-SV) model, Li et al. (2020) developed Bayesian LSTM model for VaR and ES joint forecasting, Arimond et al. (2020) tested among others LSTM-CNN models in VaR regime switching approach, Arian et al. (2020) proposed generative Encoded Value-at-Risk model which is based on Variational Auto-Encoders. In addition, the first scientific papers based on reinforcement learning are being created (e.g., Banhudo 2019).

Interestingly, most mentioned above statistical but also machine learning models that use GARCH or its ideas (architecture), assume that the distribution of returns standardized by conditional means and conditional standard deviations is independent and identically distributed (i.i.d.). However, there is ample empirical evidence that the distribution of such a variable need not be i.i.d. (Abad et al. 2014) and models may relax this conventional assumption. On this basis, a strand of higher-order conditional time-varying moments models was created. A significant contribution to its development was made, among others, by Hansen (1994) who proposed Autoregressive Conditional Density (ACD) Estimation model. This model considers the problem of estimation the full parameters of a skewness t-generalized distribution (SGT) by imposing a quadratic law of motion on the conditioning information. This idea was adapted by Bali et al. (2008), who proposed novel GARCH model based on the SGT with time-dependent parameters. To be more specific, they allowed higher-order conditional moment parameters of the SGT density to depend on the past data using an autoregressive process. They empirically showed that time-varying conditional volatility, skewness, tail-thickness, and peakedness parameters of the SGT density are significant, thus SGT-GARCH model with time-varying skewness and kurtosis gave better results that their static version in VaR sense. Remarkably similar approach was proposed by Ergun and Jun (2010). They utilized ARCD-GARCH model, which is stated on skewness t-student distribution with a time-varying skewness and kurtosis parameter. They also found that such dynamic approach is better that static one in VaR estimations. An additional example based on the same idea can be the work of Polanski and Stoja (2010). Importantly, the authors of the above approaches, in our view, have not provided sufficiently reliable and extensive validation of their models. Nevertheless, based on their tests one can conclude that they were not able to outperform, among others, classical parametric approaches such as GARCH with skewness t-student distribution.

In 2018 Duda (2018) proposed the Hierarchical Correlation Reconstruction (HCR) statistical learning model that among other things allows to estimate very accurate the probability distribution function (pdf) of a time series at any point in time using orthonormal polynomials and historical information (so-called context), where the degree of the polynomial reflects the moment (e.g. 0 - constant, 1 - expected value, 2 - variance, 3 - skewness, 4 - kurtosis, etc.). This approach fits perfectly with the idea of higher-order conditional time-varying moments and HCR potentially can extend current state of development of this area (in semiparametric and machine learning way). We therefore decide to use HCR to create the semiparametric HCR-GARCH model, where the standardised residuals from the GARCH are modelled using HCR. At the same time, we examine whether HCR will also be able to forecast VaR independently at a satisfactory accuracy level. Thus, we also try to develop the area of nonparametric VaR models. A more detailed description of HCR (including its applications) is described in subsection 2.2.1.1.

The aim of the study is to present, empirically test and formally benchmark three novel statistical learning Value at Risk models: nonparametric HCR which extends the capabilities of classical non-parametric models by including a time component in the current forecasted distribution of financial returns; semiparametric HCR-GARCH which use the HCR approach to model the distribution of standardized residuals from the GARCH model; and semiparametric hybrid HCR-QML-GARCH which combines the properties of the HCR-GARCH and QML-GARCH models through a simple average.

The benchmark procedure compares the 1% and 2.5% one-day-ahead VaR forecasts obtained with the above models against the estimates of classical VaR methods, depending on the time period and the stock index. In the empirical study we considered the following classical VaR models based on the literature review: Historical Simulation, Kernel Density Estimator, Modified Cornish-Fisher Expansion, GARCH(1,1) with varied distributions, RiskMetrics™, Peaks Over Threshold, Quasi-Maximum Likelihood GARCH. The models are compared over four periods that vary in terms of market volatility: 2006-9, 2008-11, 2014-7 and mid-2016 to mid-2020. In the analysis, we consider six major stock market indices for selected countries (differentiating the selection by the level of economic development of the countries and their geographical location): DAX (Germany), WIG 20 (Poland), MOEX (Russia), S&P 500 (USA), Nikkei (Japan) and SHC (China). The models quality is tested from two perspectives: fulfilling regulatory requirements (Basel traffic light test) and forecasting adequateness (Kupiec Proportion of Failures test, Christoffersen Conditional Coverage test and Dynamic Quantile test).

The main hypothesis verified in this paper is (1) Whether HCR-GARCH can outperform in the formal backtesting procedure classical Value at Risk models. The additional research questions are (2) Will HCR-GARCH model works well in a period of sudden increased/decreased volatility - in particular during the 1st wave of the COVID pandemic? (3); Will the HCR model be able to outperform VaR models in its class (non-parametric models) during backtesting procedure? and (4) Does the hybrid HCR-QML-GARCH model perform better than the single HCR-GARCH and HCR-QML-GARCH models.

The remainder of the paper is organized as follows. Section 2 covers the methodology used in the research, including VaR: measure, estimation models and backtesting framework. In particular, three novel VaR models: HCR, HCR-GARCH, HCR-QML-GARCH are presented here. Section 3 contains the description and assumptions of the empirical study. Section 4 presents the results of the empirical research. Finally, our conclusions are given in Section 5.

2. Methodology

2.1 Value at Risk measure

The VaR measure is defined as the maximum expected loss over a given time horizon t under normal market conditions, at a given level of confidence α (Jorion 2001). More formally, VaR can be represented by the formula:

$$\mathrm{F}(\mathrm{VaR}_{a}) = \Pr(r_t < \mathrm{VaR}_{a}(t)|\Omega_{t-1}) = \alpha, \tag{1}$$

where $F(r)$ is the cumulative distribution function, r_t denotes i.i.d. financial return and Ω_{t-1} symbolize the information set available at time $t-1$ (Abad et al. 2014).

2.2 Value at Risk estimation models

Referring to Abad et al. (2014), we distinguish following VaR estimation methods: (2.2.1) nonparametric models, (2.2.2) parametric models and (2.2.3) semiparametric models. Below we describe the methods used in our study.

2.2.1 Nonparametric models

The nonparametric VaR models do not assume rates of return distribution. They are mainly based on historical realizations of returns (Abad et al. 2014).

Hierarchical Correlation Reconstruction

In 2017 Duda (2017) proposed the Rapid Parametric Density Estimation (RPDE) algorithm which allows for very inexpensive and accurate (approaching real distribution) modeling of density distribution as linear combination of chosen functions e.g. polynomials, sines/cosines or fouriers. Whereas Hierarchical Correlation Reconstruction (HCR) algorithm developed by Duda (2018) is directly extension of RPDE by modelling multivariate densities using the idea of decomposition of multiple variables into correlations between various subsets of variables. For now, HCR finds application in many statistical and economic problems among others:

predicting probability distribution of values in time series (Duda 2019), modelling joint probability distribution of yield curve parameters (Duda and Snarska 2018), and evaluating credibility of income data (Duda and Szulc 2020). Importantly, Duda (2019) points out that HCR combines the advantages of classical statistical and machine learning models. Classical statistical approaches allow for highly controllable and interpretable modelling, while they have a relatively small number of parameters, which affects the quality of their estimates. However, in the case of machine learning models (e.g. neural networks) the number of degrees of freedom in the models increases, causing a drastic increase in computational complexity, while the interpretability decreases. By contrast, HCR simultaneously ensures that model parameters: have an inexpensive direct formula, controllable accuracy, are unique and independently calculated and each has specific interpretation. The first HCR's application mentioned above is crucial from the perspective of this study, so we fully focus on its mathematical implementation and properties.

In general, HCR in time series environment allows to estimate the probability density function ϱ as linear combination of polynomial by taking into account two hyperparameters: time component (context) and any number of moments (degree of polynomial) given by the researcher. This is done in the following procedure (for convenience we change the mathematical notation in this section).

Let us assume that $\{r^t\}_{t=1..n_0}$ denotes i.i.d. stationary variable of n_0 financial returns. As a first step, this variable should be normalized to nearly uniform probability distribution on [0, 1], what allows, inter alia to compactify the tails and normalize further coefficients. We perform this stage using Laplace distribution, which cumulative distribution function (cdf) is given by:

$$G(r) = \frac{1}{2}[1 + \mathrm{sgn}(r - a)(1 - \exp(-|r - a|/b))], \tag{2}$$

where maximum likelihood estimation of parameters is a = median of r and b = mean of $|r - a|$. Thus, output normalizing transformation is equal to $x^t = G(r^t)$.

Next, we take d succeeding values of $\{x^t\}$, which come from nearly uniform distribution on $[0,1]^d$ (if they are statistically independent). We treat $d - 1$ as a so-called earlier context. In fact, statistical dependencies in our time series are described by their distortion from uniform distribution and we would like to model them via joint density estimation for d succeeding values of x as polynomial. To achieve it, let us define object $x_i^t = x^{t-i+1}$ for $i =$

$1, \ldots, \mathrm{d}$ and $\mathrm{t} = 1, \ldots \mathrm{n}$, where $n = n_0 - d + 1$. They form $\boldsymbol{x}^t = \{x_i^t\}_{i=1\ldots d} \in [0,1]^d$ vectors containing value with its earlier context. Then, we derive following vector sequence: $\{\boldsymbol{x}^t\}_{i=1\ldots d} \subset [0,1]^d$ which is directly modeled.

Finally, we can formulate density ρ as polynomial by utilizing orthonormal basis (orthonormal Legendre polynomials as a function *f*):

$$\rho(\boldsymbol{x}) = \sum_{\boldsymbol{j} \in \{0\ldots m\}^d} q_{\boldsymbol{j}} f_{\boldsymbol{j}}(\boldsymbol{x}) = \sum_{j_1 \ldots j_d = 0}^{m} q_{\boldsymbol{j}} f_{j_1}(x_1) * \ldots * f_{j_d}(x_d), \tag{3}$$

where m denotes a degree of polynomial and $a_{\boldsymbol{j}}$ is a coefficient. For such orthonormal functions mean-square optimization allows for inexpensive calculation of coefficients (average of this function over the sample):

$$q_{\boldsymbol{j}} = \frac{1}{n} \sum_{t=1}^{n} f_{\boldsymbol{j}}(\boldsymbol{x}^t) \tag{4}$$

The orthonormal basis in such approach has $|B| = (m+1)^d$ functions (coefficients), the number of which determines the computational complexity of the algorithm. Duda (2019) notices that too high number of coefficients leads to overfitting. Therefore, the responsible selection of these two hyperparameters is crucial to address classical bias-variance trade-off. In our study, this is done on the ground of empirical verification of model quality for successive m and d pairs in the grid search procedure.

In the next computation step, we utilize ρ to predict probability density function ϱ of the current time series point basing on the earlier context: by substituting earlier context to the polynomial and normalizing the remaining polynomial to integrate to 1.

The next necessary stage is application of the calibration procedure ϕ on the pdf ϱ, because unfortunately ϱ can very rarely be negative in some intervals (the higher $|B|$ the greater the likelihood of such events), which requires its transformation to a low positive value. We choose the simplest calibration solution here: $\phi(\varrho) = max(\varrho, v)$, where v is precalculated parameter using historical simulation algorithm at the α level on $\{r^t\}$ (see the subsection 2.2.1.2). Such transformation may damage probabilistic normalization, hence calibration step also includes division by $\int_0^1 \phi(\varrho(x))dx$.

Finally, we have to remove normalization from ϱ with Laplace distribution, thus inversed Laplace cdf is applied on ϱ's horizontal axis and ϱ's vertical axis is multiplied by Laplace pdf.

On the basis of the estimated ϱ we want to calculate its α quantile for risk evaluation purposes. This procedure is straight forward. First, the empirical cumulative distribution function $\mathrm{T}(x)$ must be determined based on ϱ. Then we inverse $\mathrm{T}(x)$ to obtain $\mathrm{T}^{-1}(x)$ and we utilize this quantile function to formulate α quantile estimator for any time series x:

$$HCR_{\alpha}(x) \;=\; \mathrm{T}^{-1}(x;\, \alpha). \tag{5}$$

In such environment Value at Risk is naturally estimated as:

$$\mathrm{VaR}_{\alpha}(t) = HCR_{\alpha}(r_t). \tag{6}$$

Historical Simulation

Historical Simulation is a basic and one of the simplest methods to forecast VaR. Here $\mathrm{VaR}_{\alpha}$is denoted as α quantile of the estimated empirical distribution of rates of return over the chosen time window (Abad et al. 2014):

$$\mathrm{VaR}_{\alpha}(t) = q_{\alpha} \tag{7}$$

Kernel Density Estimator

Kernel Density Estimator let to generalize population's probability density function based on finite data sample by solving data smoothing problem (Silverman 1986). If we assume that $(x_1, x_2, \ldots, x_n)$ is a sample of i.i.d. realization of random variable with an absolutely continuous cumulative distribution function *F*, then the kernel density estimator $\hat{f}_h(x)$ of f_h(x) is defined as:

$$\widehat{f_h}(x) = \frac{1}{hn}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right), \tag{8}$$

where h is a smoothing parameter called bandwidth with $h > 0$ and K is a non-negative function called kernel function. *K* might be represented by vast range of functions, for instance: Gaussian, Epanechnikov, Rectangular, Triangular, Biweight, Cosine, Optcosine etc. As prior research suggests kernel's type has small impact on the results of the KDE in the VaR estimation environment (Huang 2009). Taking it into account and the fact that Epanechnikov kernel is optimal in the sense of mean square error (Epanechnikov 1969), we choose Epanechnikov kernel to further calculations which formula is:

$$K(u) = \frac{3}{4}(1 - u^2)1_{\{|u|<1\}}. \tag{9}$$

On the other hand, *h* is a crucial hyperparameter for KDE, because large *h* would over-smooth the pdf and small *h* would under-smooth the estimation. There are many well developed methods to select bandwidth: plug-in formula of Sheather and Jones (1991), Silverman's rule of thumb (1986), Scott's rule of thumb (1992) etc. We decide to use the first approach mentioned above due to Jones et al. (1996) recommendations. This method solves the fixed-point equation:

$$h = \left[\frac{R(\psi)}{nR\left(\hat{f}''_{g(h)}\right)\left(\int x^2\,\psi(x)dx\right)^2}\right]^{\frac{1}{5}}, \tag{10}$$

where KDE's mean integrated square error (MISE) is $R(\psi) = \int \psi^2(x)dx$. Finally, Value at Risk is estimated here using following equation:

$$\mathrm{VaR}_\alpha(t) = \hat{F}_h^{-1}(x; \alpha), \tag{11}$$

where $\hat{F}_h^{-1}(x)$ is inverse function of KDE's cdf $\hat{\mathrm{F}}_\mathrm{h}(x)$ which is obtained by integrating pdf $\widehat{f}_h(x)$ (Alemany et al. 2012).

Modified Cornish-Fisher Expansion

Modified Cornish-Fisher Expansion estimator of the α quantile of non-Gaussian distribution allows to formulate VaR using following equitation (Cavenaile and Lejeune 2012):

$$\mathrm{VaR}_\alpha(t) = \mu_t + \sigma_t\, z_{CF,\alpha}, \tag{12}$$

where μ_t denotes conditional mean of the financial returns, σ_t is the conditional standard deviation of the returns and $z_{CF,\alpha}$ is the Cornish-Fisher approximation of the α quantile of the distribution. If we assume that: z_α is the α quantile of a standard normal distribution, *S* the standardized skewness and *K* excess kurtosis, then $z_{CF,\alpha}$is expressed by:

$$z_{CF,\alpha} = z_\alpha + \frac{1}{6}(z_\alpha^2 - 1)S + \frac{1}{24}(z_\alpha^3 - 3z_\alpha)K - \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)S^2 \tag{13}$$

2.2.2 Parametric models

Parametric VaR models assume parameters, the most often distribution of the returns. They fit probability curves to the data and then infer based on them (Abad et al. 2014).

Generalized Autoregressive Conditional Heteroskedasticity

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model introduced by Bollerslev (1986) generalizes the volatility modeling approach of Engle's (1982) ARCH process. The standard GARCH (p, q) regression model is denoted as:

$$r_t = \mu_t + \epsilon_t = \mu_t + \mathrm{h_t}\xi_t, \tag{14}$$

where μ_t is conditional mean of returns, ϵ_t is random error which is equal to the product of conditional standard deviation of returns $\mathrm{h_t}$ and the standardized i.i.d. random error ξ_t. Bollerslev's GARCH assumes that conditional variance is the function of lagged random errors and lagged conditional variances:

$$h_t^2 = \omega + \sum_{i=1}^{q} \zeta_i \epsilon_{t-1}^2 + \sum_{i=1}^{p} \eta_i h_{t-1}^2 \,, \tag{15}$$

where ω is an intercept greater than 0, ζ_i is a coefficient of the random squared error and η_i is a coefficient of the conditional variance, q and p denote respectively: number of random squared periods; number of shifted conditional variance periods used in the functional form of conditional variance. In this article we consider one of the simplest and most efficient (Buczyński and Chlebus 2019) models from the GARCH family, i.e. GARCH (1,1) with different random error distributions like: Gaussian (GARCH-n), t-Student (GARCH-t), skewed t-Student (GARCH-st) and Generalized Error Distribution (GARCH-ged).

One can estimate VaR using GARCH via following formula:

$$Va\mathrm{R}_\alpha(\mathrm{t}) = \widehat{\mu_t} + \sqrt{\widehat{h}_t^2} * k_\alpha, \tag{16}$$

where $\widehat{\mu_t}$ is the forecast of conditional mean, $\widehat{h}_t^2$ is the forecast of conditional variance and $k_\alpha$is the α quantile of the assumed underlying distribution of the financial series.

RiskMetrics™

The RiskMetrics™ model proposed by Morgan (1996) assumes that financial returns follow a Gaussian distribution. Thus, VaR is calculated as:

$$\mathrm{VaR}_\alpha(\mathrm{t}) = \mu_t + \sigma_t G^{-1}(\alpha), \tag{17}$$

where μ_t denotes conditional mean of the financial returns, σ_t is the conditional standard deviation of the returns and $G^{-1}(\alpha)$ is the quantile of the standard normal distribution.

2.2.3 Semiparametric models

The semiparametric VaR models combine the nonparametric approach with the parametric method (Abad et al. 2014).

Peaks Over Threshold model

Peaks Over Threshold (POT) models come from Extreme Value Theory (EVT) family, which is focused on modeling the tail of the data distribution by making assumption of sample built on extrema possessed from gathered historical information (Abad et al. 2014, Szubzda and Chlebus 2019). POT method allows for obtaining extreme observations by taking cut-off threshold denoted by θ, where all data points below it form a dataset used to pdf's tail estimation. Specifically, in this study we opt for use fully Parametric POT models based on the Generalized Pareto Distribution (GPD) (Embrechts et al. 1999). The GPD cdf is as follows:

$$GPD(r;\varsigma,\varpi,\rho) = \begin{cases} 1-\left(1+\varsigma\dfrac{r-\varpi}{\rho}\right)^{(-1/\varsigma)} & if\varsigma \neq 0 \\ 1-exp\left(-\dfrac{r-\varpi}{\rho}\right) & if\varsigma = 0 \end{cases}, \tag{18}$$

where ς, ϖ, ρ respectively are the parameters of: shape, location and scale. Moreover, let us define loss level above θ: $y_i = r_i - \theta$, assume that N_θ are the number of datapoints greater than θ and N is the total number of the analyzed financial returns. Naturally, the distribution of excess losses over the θ is defined as:

$$F_\theta(y) = P(r-\theta < y | r > \theta) = \frac{F(y+\theta)-F(\theta)}{1-F(\theta)} \tag{19}$$

Now we adopt the fact, that the distribution of observations being above the θ follows the $GPD(x;\varsigma,\varpi,\rho)$, thus the cdf of returns is given by:

$$F(r) = F(y+\theta) = [1-F(\theta)]GPD(r;\varsigma,\varpi,\rho) + F(\theta). \tag{20}$$

Taking it into account, VaR is expressed by:

$$\mathrm{VaR}_\alpha(\mathrm{t}) = \theta + \frac{\hat{\rho}}{\hat{\varsigma}}\left[[N/N_\theta(1-\alpha)]^{-\hat{\varsigma}} - 1\right], \tag{21}$$

To determine θ we use mean excess plot (Embrechts et al. 1997).

Quasi-Maximum Likelihood GARCH

Quasi-Maximum Likelihood (QML) GARCH model proposed among others by Engle and Manganelli (1999) allows to estimate VaR using classical GARCH environment except k_α

component, which is computed here by the empirical distribution quantile of GARCH's standardized residuals ξ_t. In other words, it is combination of the GARCH model with Historical Simulation approach for ξ_t. Summarizing, VaR is estimated via equation:

$$VaR_{\alpha}(\mathrm{t}) = \widehat{\mu_t} + \sqrt{\hat{h}_t^2} * q_{\alpha}(\xi_t). \tag{22}$$

HCR-GARCH

The HCR-GARCH is another novel Value at Risk forecasting model proposed by us in this article. It is a stacking hybrid of an econometric (GARCH) and machine learning (HCR) approaches and is empirically inspired by the QML GARCH. The whole idea is based on the usage of HCR to model GARCH's standardized distribution of residuals in the VaR environment, in contrast to QML GARCH, where empirical distribution is used to obtain α quantile. Thus, the VaR model can be calculated using the following equation:

$$VaR_{\alpha}(\mathrm{t}) = \widehat{\mu_t} + \sqrt{\hat{h}_t^2} * HCR_{\alpha}(\xi_t), \tag{23}$$

where $\widehat{\mu_t}$ is the GARCH's forecast of conditional mean, $\hat{h}_t^2$ is the GARCH's forecast of conditional variance and $HCR_{\alpha}(\xi_t)$ is the α quantile of the GARCH's standardized distribution of residuals ξ_t modelled via HCR algorithm. In this way, we consciously break the statistical assumption about the independence of the residual's component, because the estimated distribution of ξ_t includes a time component which is responsible for increasing the conservativeness of the VaR model. However, we strongly believe that the hybrid forecast constructed like that better suits to market realities, especially under conditions of increased variance in the markets. In this study, we base the HCR-GARCH estimation on the GARCH-n model. However, as we verified empirically, the choice of the intermediate distribution in this case is not crucial and the normal distribution can be substituted by e.g. skewed t-Student without significant differences in the HCR-GARCH results.

HCR-QML-GARCH

Finally, we decide to introduce the HCR-QML-GARCH, which is a simple ensemble model based on arithmetic average of HCR-GARCH and QML-GARCH. Here, VaR is computed like:

$$VaR_{\alpha}(\mathrm{t}) = \widehat{\mu_t} + \sqrt{\hat{h}_t^2} * \frac{HCR_{\alpha}(\xi_t) * q_{\alpha}(\xi_t)}{2}. \tag{24}$$

It seems that this model should address the trade-off between the liberality (QML-GARCH) and conservatism (HCR-GARCH) of the VaR forecast.

2.3 Backtesting

In our research we utilize popular and well establish backtesting framework (Buczyński and Chlebus 2018), which comprehensively evolves the quality of VaR models. First it controls regulatory requirements fulfillment (excess ratio and Basel traffic light test), secondly it checks adequacy of forecasts (Kupiec Proportion of Failures test, Christoffersen Conditional Coverage test and Dynamic Quantile test). For convenience, let I_t be VaR exceedance as an indicator function (Costanzino and Curran 2018):

$$\mathrm{I_t} = \begin{cases} 1 \text{ if } r_t < \mathrm{VaR_a(t)} \\ 0 \text{ if } r_t \geq \mathrm{VaR_a(t)} \end{cases} \tag{25}$$

2.3.1 Excess ratio and Basel traffic light test

Excess ratio α is expressed as:

$$\widehat{\alpha} = \frac{1}{N}\sum_{t=1}^{N} I_t\,, \tag{26}$$

where $\hat{\alpha}$ denotes observed excess ratio and N is the number of observations (number of VaR forecasts). According to theory, the excess ratio for the correctly specified and forecasting model should be equal to the significance level, at which VaR is calculated.

The Basel traffic light test (Basle Committee on Banking Supervision 1996, Costanzino and Curran 2018) is based on excess ratio α and it classifies model's responses into three zones: green (indicates that there is no problem with the quality or accuracy of the model), yellow (indicates potential problems with the model, recommends its close supervision) and red (indicates strictly a problem with a model). The zone's assignment is made using following formula (Buczyński and Chlebus 2018):

$$F(I_t) = \sum_{k=0}^{I_t} \binom{N}{K} p^K (1-p)^{N-K} = \alpha \begin{cases} red\ zone & \alpha \geq\ 0.9999 \\ yellow\ zone & \alpha \geq\ 0.95\,, \\ green\ zone & \alpha <\ 0.95 \end{cases} \tag{27}$$

where F is cumulative distribution of variable from the binomial distribution, p denotes level of probability for which VaR is estimated and K is number of VaR breaches.

2.3.2 Kupiec Proportion of Failures test (UC)

The Kupiec (1995) Proportion of Failures test (also known as unconditional coverage test) examines how many times a VaR is breached over a given time window. The null hypothesis assumes that theoretical and empirical excess ratios are equal ($\alpha = \hat{\alpha}$). The test statistics takes following form:

$$LR_{POF} = 2ln\left(\left(\frac{1-\hat{\alpha}}{1-\alpha}\right)^{N-K}\left(\frac{\hat{\alpha}}{\alpha}\right)^{K}\right) \sim \chi^2(1). \tag{28}$$

Importantly, the POF Kupiec test, unlike Basel traffic light test, considers the overestimation of the model (Buczyński and Chlebus 2018).

2.3.3 Christoffersen Conditional Coverage test (CC)

The Christoffersen (1998) Conditional Coverage test verifies the frequency of the VaR breaches occurrence as well as their independence**.** Its test statistics is as follows:

$$LR_{cc} = LR_{POF} + LR_{IND} \sim \chi^2(x), \tag{29}$$

where LR_{IND} is a likelihood ratio of the VaR forecast independence test which takes null hypothesis about the independence of exceedances, against the alternative hypothesis that the exceedances are characterized by the first order Markov chain. It is denoted as (Kaszyński et al. 2020):

$$LR_{IND} = -2ln\left(\frac{\pi_{*0}^{n_{00}+n_{10}}\pi_{*1}^{n_{01}+n_{11}}}{\pi_{00}^{n_{00}}\pi_{01}^{n_{01}}\pi_{10}^{n_{10}}\pi_{11}^{n_{11}}}\right) \sim \chi^2(1), \tag{30}$$

where $n_{i,j}$is the number of observations, $\pi_{ij} = n_{ij}n_{ij}/\sum_j n_{ij}$ and $\pi_{*j} = \sum_i n_{ij} / \sum_{i,j} n_{ij}$.

2.3.4. Dynamic Quantile test (DQ)

The Engle and Manganelli (2004) Dynamic Quantile test utilizes the linear regression model to jointly verify the presence of autocorrelation among VaR breaches. Let us define following term (Buczyński and Chlebus 2018):

$$J_t = \begin{cases} 1-\alpha & if \quad r_t < \mathrm{VaR_a(t)} \\ -\alpha & if \quad r_t \geq \mathrm{VaR_a(t)} \end{cases}, \tag{31}$$

which is crucial part of the following regression:

$$J_t = \beta_0 + \sum_{i=1}^{p} \beta_i J_{t-1} + \sum_{j=1}^{q} \mu_j X_j + \epsilon_t\,, \tag{32}$$

where X denotes all exogenous variables included in the forecasting model, p is the number of lags of dependent variable and q is the number of lags of independent variables. The null hypothesis assumes nullity of coefficients in above regression: $\beta_0 = \beta_1 = \ldots = \beta_p = \mu_1 =$

$\ldots = \mu_q = 0$, i.e. independence of the violation in t to all past violations (Kaszyński et al. 2020). The Wald test statistic is as follows:

$$DQ_{cc} = \frac{J_t' M (M'M)^{-1} M' J_t}{\alpha(1-\alpha)} \sim \chi^2(p+q), \tag{33}$$

where M is a matrix, in which the columns are p lags of the exceedance vector and q lags of the explanatory variables.

3. Empirical study setting

The basic assumption of the evaluation methodology is to meet the requirements of the Basel III, i.e. to forecast one-day-ahead VaR using a rolling-window and to assume significance levels of 1% (the current value of the metric) and 2.5% (the value imposed by Basel IV for the VaR needed to calculate Expected Shortfall metric) (PricewaterhouseCoopers 2016). Furthermore, we adopt that VaR is calculated based on daily logarithmic financial returns:

$$r_t = ln(p_t) - ln(p_{t-1}), \tag{34}$$

where p_t is the price of the asset at the end of the trading day t. For the purpose of the study, we assume that in the rolling-window training sample (in the sample) covers always 500 observations, which is nearly two stock market years. This is not fully consistent with the recommendation of the optimal cut-off point by Buczyński and Chlebus (2020), which is around 1000 for conservative models, however according to their research, the smallest reasonable number of observations in the training sample is exactly 500. Generally, such a threshold addresses the trade-off between the conservativeness of the model and its computational complexity (crucial factor during experiments). In the case of the test sample (out of the sample), we also take 500 observations, which naturally gives us a satisfactory inference period. To recall, in the one-step-ahead approach such a setup implies the need to estimate 500 VaR models each time.

The experiment covers a set of main stock indexes from three European and three global (non-European) countries like: DAX (Germany), WIG 20 (Poland), MOEX (Russia), S&P 500 (USA), Nikkei 225 (Japan) and SHC (China). Countries are selected to represent both developed (Germany, USA, Japan) and developing economies (Poland, Russia, China) to examine the quality of VaR forecasts depending on this factor.

In addition to spatial variation, the study also distinguishes four periods characterized by different volatility in the markets during last 15 years. Importantly, this difference also applies to in-sample and out-of-the-sample sets. Period I lasts from January 2006 to December

2009, so the training set has low volatility and the test set is full of variation's turbulence associated with the global economic crisis. Period II runs from January 2008 to December 2011. Here both in the sample and out of the sample have high volatility, with a decreasing trend over time. Period III covers the time from January 2014 to December 2017 and is characterized by tranquil volatility throughout the series. Period IV considers the first wave of the Covid-19 pandemic, thus it lasts from September 2016 to September 2020. Its characteristics are remarkably similar to Period I, but here the outbreak of volatility is much more sudden. From the perspective of the study, this is a key period.

To better understand the statistical properties of the distributions of returns for the analyzed six stock indices in the four periods, we present Table 1. As the Jarque-Bera and Anderson-Darling tests indicate, in no case do returns follow a normal distribution. Most often we deal with left skewed distributions (skewness less than zero) and leptokurtosis (excess kurtosis greater than 0), which is in line with left long tail expectations. Furthermore, always the mean value from the series is statistically not different from 0. This implies a practical lack of need to estimate the value of the average return in VaR models, which we also do. Finally, it is worth noting that the sample of stock indices is representative, due to the high heterogeneity in the values of the descriptive statistics presented.

Table 1. Descriptive statistics of the financial returns per each analyzed stock index and period

PERIOD I									
Index	**Mean**	**Median**	**Min**	**Max**	**St. Dev.**	**Skewness**	**Excess Kurtosis**	**JB**	**AD**
DAX	0.000	0.001	-0.074	0.108	0.017	0.207	7.215	0.000	0.000
S&P 500	0.000	0.001	-0.095	0.110	0.017	-0.210	8.208	0.000	0.000
Nikkei 225	0.000	0.000	-0.121	0.132	0.019	-0.382	7.326	0.000	0.000
SHC	0.001	0.003	-0.093	0.090	0.022	-0.444	2.071	0.000	0.000
WIG 20	0.000	0.000	-0.083	0.061	0.016	-0.339	2.043	0.000	0.000
MOEX	0.000	0.001	-0.207	0.252	0.031	-0.052	11.930	0.000	0.000
PERIOD II									
Index	**Mean**	**Median**	**Min**	**Max**	**St. Dev.**	**Skewness**	**Excess Kurtosis**	**JB**	**AD**
DAX	0.000	0.000	-0.073	0.108	0.018	0.213	4.679	0.000	0.000
S&P 500	0.000	0.001	-0.095	0.110	0.018	-0.219	5.755	0.000	0.000
Nikkei 225	-0.001	0.000	-0.121	0.132	0.020	-0.487	7.163	0.000	0.000
SHC	-0.001	0.000	-0.080	0.090	0.020	-0.150	2.578	0.000	0.000
WIG 20	0.000	0.000	-0.083	0.061	0.016	-0.360	2.624	0.000	0.000
MOEX	0.000	0.000	-0.207	0.252	0.029	0.020	13.959	0.000	0.000
PERIOD III									
Index	**Mean**	**Median**	**Min**	**Max**	**St. Dev.**	**Skewness**	**Excess Kurtosis**	**JB**	**AD**

DAX	0.000	0.001	-0.071	0.049	0.012	-0.362	2.413	0.000	0.000
S&P 500	0.000	0.000	-0.040	0.038	0.008	-0.411	3.104	0.000	0.000
Nikkei 225	0.000	0.001	-0.083	0.074	0.013	-0.173	5.245	0.000	0.000
SHC	0.000	0.001	-0.089	0.056	0.015	-1.324	7.425	0.000	0.000
WIG 20	0.000	0.000	-0.058	0.030	0.009	-0.626	3.952	0.000	0.000
MOEX	0.000	0.000	-0.114	0.051	0.012	-0.802	9.784	0.000	0.000

PERIOD IV

Index	Mean	Median	Min	Max	St. Dev.	Skewness	Excess Kurtosis	JB	AD
DAX	0.000	0.001	-0.131	0.104	0.013	-1.083	19.934	0.000	0.000
S&P 500	0.000	0.001	-0.128	0.090	0.013	-1.172	22.238	0.000	0.000
Nikkei 225	0.000	0.001	-0.063	0.077	0.012	-0.076	6.630	0.000	0.000
SHC	0.000	0.001	-0.080	0.056	0.011	-0.701	6.692	0.000	0.000
WIG 20	0.000	0.000	-0.135	0.056	0.012	-1.917	22.370	0.000	0.000
MOEX	0.000	0.001	-0.087	0.074	0.011	-1.121	14.717	0.000	0.000

JB and AD denote p-values for the Jarque-Bera and Anderson-Darling normality of distribution tests, respectively.

In addition, we verify the presence of Autoregressive Conditional Heteroskedasticity (ARCH) effects using the ARCH-LM test and clustering of variances using the Ljung-Box test in the analyzed time-series. At the 5% significance level, we find that returns are characterized by conditional heteroskedasticity and are subject to autocorrelation (in quadrature). Such statistical properties suggests the use of GARCH models. At the same time, we diagnose our GARCH models for autocorrelation and ARCH effect for standardized residuals using Weighted Ljung-Box Test on Standardized Squared Residuals and Weighted ARCH LM Tests. Based on these, we find that not all GARCH models are well specified i.e., standardized residuals are subject to autocorrelation. This reinforces our belief in the validity of using the HCR-GARCH approach, where we make practical use of this negative artefact of the GARCH model.

All calculations in the study are performed in the R and Python programming languages, which are point-preferred due to the availability of statistical and machine learning libraries. Importantly, the HCR algorithm is fully implemented in R.

4. Results

In this section, we first look for the best configuration of HCR and HCR-GARCH models on an independent validation period (4.1) and then compare them with other classical models (4.2).

4.1 HCR and HCR-GARCH model setting and tuning

Models based on the HCR algorithm require tuning of two hyperparameters: maximum polynomial degree (m) and historical context (d-1). We choose to solve this problem using grid search according to the best practices of statistical learning in the Value at Risk domain (Lux, 2020). The search of the hyperparameter space is performed on an independent validation set (January 1999 - December 2003) per: model class (HCR, HCR-GARCH), VaR level (1%, 2.5%) and index (DAX, S&P, Nikkei, SHC, WIG, MOEX). The adopted evaluation criterion is sequentially: p-value of the Conditional Coverage (CC) test, p-value of the Unconditional Coverage (UC) test, p-value of the Dynamic Quantile (DQ) test. The higher the p-value, the better the ranking of a given model (Lux, 2020). If the same p-values are reached, the simpler model (lower value of hyperparameters) is selected.

Our validation set description is presented in Table 2. This is a period of alternating calm and perturbation in the markets, so it seems appropriate to establish the best models (ideally the hyperparameters should be readjusted after every few trading days completed, but for computational reasons this idea was abandoned).

Table 2. Descriptive statistics of the financial returns per analyzed stock indexes in validation period

PERIOD 0 - validation									
Index	**Mean**	**Median**	**Min**	**Max**	**St. Dev.**	**Skewness**	**Excess Kurtosis**	**JB**	**AD**
DAX	0.000	0.000	-0.089	0.076	0.018	-0.024	1.729	0.000	0.000
S&P 500	0.000	-0.001	-0.060	0.056	0.014	0.151	1.118	0.000	0.000
Nikkei 225	0.000	-0.001	-0.072	0.072	0.016	0.059	1.404	0.000	0.000
SHC	0.000	0.000	-0.079	0.094	0.015	0.658	6.015	0.000	0.000
WIG 20	0.000	0.000	-0.085	0.068	0.015	-0.019	2.046	0.000	0.000
MOEX	0.002	0.002	-0.134	0.141	0.027	0.196	2.539	0.000	0.000

JB and AD denote p-values for the Jarque-Bera and Anderson-Darling normality of distribution tests, respectively.

Based on the literature review presented, we assume that *m* can take values at least from 1 (expected value) to 4 (kurtosis). We additionally decide to check performance of the 5^{th} moment. We arbitrarily select a historical context (*d*-1) on a range from 1 to 6 (thus we consider more than 1 trading week). It seems that even maximal value of *m* and *d*-1 should not significantly affect the overtraining of the models (Duda 2018). Furthermore, from our perspective, an important criterion for selecting the solution space is the computational complexity of the model (we try not to model more than 2500 coefficients).

In Tables 3 and 4 we present the validation results for the HCR and HCR-GARCH models, respectively. While all selected models passed the Traffic Light Test, not all models exceeded a p-value of 10% for the tests analyzed (they proved to be too conservative on the validation set). Furthermore, it can be observed that both models overwhelmingly prefer *m* equal to 2. However, for the historical context, it is hard to find a clear pattern at the quantile and index level. The selected hyperparameters are applied in the final models (see subsection 4.2). The HCR-QML-GARCH models are produced according to the HCR-GARCH recommendations for hyperparameters.

Table 3. The best HCR models in the tuning period per index and quantile

m	*d-1*	**Index**	**Quantile (%)**	**ER & LT**	**CC**	**UC**	**DQ**
5	2	SHC	1.0	1.000	0.951	1.000	0.997
4	3	SHC	2.5	2.400	0.730	0.885	0.030
2	4	DAX	1.0	1.000	0.951	1.000	0.000
2	3	DAX	2.5	2.200	0.648	0.661	0.005
4	3	Nikkei	1.0	1.000	0.951	1.000	0.001
2	6	Nikkei	2.5	2.800	0.555	0.673	0.562
2	6	WIG	1.0	0.200	0.008	0.028	0.524
2	5	WIG	2.5	1.000	0.003	0.015	0.329
2	6	MOEX	1.0	0.200	0.008	0.028	0.524
2	1	MOEX	2.5	0.200	0.000	0.000	0.029
2	5	S&P	1.0	0.200	0.951	1.000	0.001
2	5	S&P	2.5	0.800	0.555	0.673	0.562

ER & LT denote the excess ratio and color present the Light Test result. CC, UC, DQ, indicate the p value of the corresponding test.

Table 4. The best HCR-GARCH models in the tuning period per index and quantile

m	*d-1*	**Index**	**Quantile (%)**	**ER & LT**	**CC**	**UC**	**DQ**
2	1	SHC	1.0	0.400	0.095	0.125	0.772
2	4	SHC	2.5	1.600	0.134	0.168	0.757
2	1	DAX	1.0	0.200	0.008	0.028	0.524
2	1	DAX	2.5	0.200	0.000	0.000	0.029
2	6	Nikkei	1.0	0.400	0.095	0.125	0.772
2	1	Nikkei	2.5	0.200	0.000	0.000	0.029
2	1	WIG	1.0	0.000	0.000	0.002	0.000
2	6	WIG	2.5	0.200	0.000	0.000	0.029
2	6	MOEX	1.0	0.400	0.095	0.125	0.772
5	2	MOEX	2.5	1.000	0.003	0.015	0.329
2	3	S&P	1.0	0.400	0.095	0.125	0.772
2	1	S&P	2.5	0.800	0.000	0.005	0.207

ER & LT denote the excess ratio and color present the Light Test result. CC, UC, DQ, indicate the p value of the corresponding test.

4.2 Model comparison

The results are presented for each Value at Risk quantile separately. Table 5 presents the evaluation of the 1% VaR estimation of the financial returns for each analyzed model, stock index and period. Analogous results are prepared in Table 6 for 2.5% VaR. To compare models outcome we utilize following metrics: excess ratio, Basel traffic light test, Unconditional Coverage test, Conditional Coverage test and Dynamic Quantile test.

Initially, we focus on the analysis of the 1% VaR results. First, we examine how the classical models within each class (due to parametricity) performed. In the case of nonparametric models, it can be easily concluded that no model performs well in all periods analyzed. This is indicated by all evaluation metrics, i.e. multiple VaR exceedances, yellow and red lights in the Basel traffic light test and notorious rejection of the null hypotheses for the statistical tests undertaken. Moreover, the models performed significantly worse for indices from highly developed countries. One can say that the only period in which nonparametric models relatively return meaningful estimates is period III, which is consistent with intuition, i.e., nonparametric models perform poorly in periods of sudden increased volatility due to the problem of long adjustment to current market realities.

Remarkably similar conclusions can be drawn for the RiskMetrics™ model, which, as the simplest parametric approach in our research, only do well in Period III. In contrast, classical parametric models from the GARCH family perform well in periods I, II and III. However, none of them turn out to be sufficient in period IV, where volatility has increased very unexpectedly due to the onset of the COVID-19 pandemic. Nevertheless, undoubtedly the best in this class is the GARCH-st model thanks to its Basel traffic light, UC and CC tests performance. On the other hand, the GARCH-n model perform the worst. One can say that for parametric GARCH the indices from developing countries cause the most problems.

In the case of classical semiparametric models, the conclusions are twofold. The POT model behaves similar to the GARCH-st during the first three periods, however the GARCH-st appears to have slightly better performance in the period IV. In the case of QML-GARCH, it can be clearly stated that this model is the most conservative compared to those analyzed above, which means that in periods I, II and III it does not exceed enough to meet the UC, CC, DQ test requirements, but at the same time it does not go beyond the green light of the Basel test. In the case of period IV, it does relatively well and generates yellow light only once for the DAX index. However, at the same time it passes the DQ test only for the S&P and SHC indexes. It is difficult to unequivocally choose the best model from the above-described classical approaches to VaR estimation, however, it seems that the clear leaders are GARCH-st and QML-GARCH.

Taking above into account, we examine the results of the models from the HCR family. The nonparametric HCR model proposed by us performs generally very poorly on all indexes: in the case of periods I and IV, it records numerous VaR exceedances which are expressed, inter alia, by yellow and red lights in the Basel test. While in periods II and III it is too conservative, for instance UC and CC tests indicates this fact. Clearly, the HCR model is not capable to outperform models in its class. In the case of our target HCR-GARCH model, it turns out to be even more conservative than QML-GARCH, which leads to green light in the Basel traffic light test for all indexes in all periods. In the periods I, II and III, the estimates implement too few VaR exceedances to classify the model as good according to the UC and CC model interpretation. For the DQ test, the HCR-GARCH model performs slightly better in these periods than the GARCH-st and GARCH-QML. For period IV, HCR-GARCH is by far the leader among all the approaches discussed. Only for MOEX it fails the CC test and for WIG and DAX the DQ test. The last model we propose, which is HCR-QML-GARCH performs better on periods I, II, III than QML-GARCH and HCR-GARCH, while it still collectively performs worse than GARCH-st at this time. Period IV for HCR-QML-GARCH converges to the QML-GARCH model in terms of evaluation metrics.

We then analyze the results of the models for 2.5% VaR at the same convention as above. As in the case of 1% VaR, nonparametric models do not cope at all in times of sudden increased volatility, i.e. in periods I and IV, generating almost only yellow and red lights in the Basel traffic light test. In the case of periods with decreasing and low volatility, they are too conservative, for instance see UC and CC tests outcomes. The indices of highly developed countries turn out to be the most difficult to estimate: S&P, DAX and Nikkei.

As before, RiskMetrics™ has generally failed in VaR estimations and we can state that it is the weakest model in this study. The classic GARCH-n/t/ged parametric models perform well only in the third period of the analysis. At other times, they generate numerous Basel yellow lights and fail UC, CC and DQ tests. Again, the GARCH-st model performs very well in the I, II and III period, while in the around pandemic and pandemic times (period IV) it stops working properly, e.g. see outcome of Basel traffic light, UC, CC and DQ tests.

For semiparametric models, the POT model performs less well than the GARCH-st described above, e.g. generates more yellow traffic lights and fails the UC and CC tests more often. The QML-GARCH model again proves to be too conservative for the first three periods, as it fails the UC and CC tests in most cases, and never exceeds the green light in the Basel traffic lights test. At the same time, this model does well in period IV, generating only green lights, but for some indices it does not pass the CC and DQ tests. It seems that again the best models are QML-GARCH and GARCH-st.

For models with the family HCR results are varied. The simplest nonparametric HCR model performs very much like Historical Simulation and KDE, in other words it does not work properly. HCR-GARCH shows very similar prognostic properties as QML-GARCH, i.e. it is very conservative in all periods, and in the IV period it records relatively good quality of estimation (especially for the DAX, Nikkei, WIG and S&P indices). However, QML-GARCH appears to be more universal. In the case of the HCR-QML-GARCH hybrid model, it can be concluded that it performs better jointly than HCR-GARCH and QML-GARCH due to the generally better results of UC, CC and DQ tests in all periods.

Based on the above results, one can conclude that an ideal VaR model does not exist. The HCR-GARCH mode which we have proposed produces the expected results in terms of periods of sudden increased volatility, while it does not work very well in periods of relative calm in the markets, as it retains its conservatism. A good solution to this problem turns out to be the HCR-QML-GARCH hybrid, which liberalizes the conservatism of HCR-GARCH. In the case of our nonparametric model, HCR has failed and is unable to contribute anything groundbreaking to non-parametric VaR estimation. Furthermore, we were unable to detect an unambiguous relationship between the estimation quality of HCR family models and the origin of the stock market index (level of state development). It is worth mentioning that in times of very low and moderate volatility, GARCH-st undeniably proved to be the best model.

Table 5. Results of 1% VaR estimation of the financial returns per each analyzed model, stock index and period

		PERIOD I				**PERIOD II**				**PERIOD III**				**PERIOD IV**			
model	**index**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**
Historical Simulation	SHC	0.8	0.641	0.783	0.992	0.2	0.028	0.008	0.524	0	0.002			1.2	0.663	0.766	0.005
	DAX	2	0.048	0.016	0	1.8	0.106	0.025	0	0.2	0.028	0.008	0.524	3.4	0	0	0
	Nikkei	2.8	0.001	0	0	0.6	0.331	0.013	0	0.8	0.641	0.783	0	1.8	0.106	0.025	0
	WIG	1.8	0.106	0.061	0.019	1.8	0.106	0	0	0.8	0.641	0.783	0.992	2.8	0.001	0	0
	MOEX	2.8	0.001	0	0	0.8	0.641	0.783	0	0.6	0.331	0.323	0.772	2	0.048	0.001	0
	S&P	3.6	0	0	0	1	1	0.951	0	0.6	0.331	0.323	0.772	2.6	0.003	0	0
KDE Epanechnikov	SHC	0.8	0.641	0.783	0.992	0	0.002			0	0.002			1.2	0.663	0.766	0.005
	DAX	2	0.048	0.016	0	1.6	0.215	0.186	0	0.2	0.028	0.008	0.524	3.4	0	0	0
	Nikkei	2.8	0.001	0	0	0.6	0.331	0.013	0	1	1	0.951	0.001	2	0.048	0.008	0
	WIG	1.6	0.215	0.186	0.018	1.8	0.106	0	0	0.8	0.641	0.783	0.992	2.4	0.008	0	0
	MOEX	2.8	0.001	0	0	0.6	0.331	0.386	0.935	0.4	0.125	0.071	0.524	2	0.048	0.001	0
	S&P	3.8	0	0	0	0.8	0.641	0.783	0	0.6	0.331	0.323	0.772	2.6	0.003	0	0
Modified Cornish Fisher	SHC	0.8	0.641	0.783	0.992	0	0.002			0.4	0.125	0.095	0	1	1	0.951	0.001
	DAX	1	1	0.951	0	1.4	0.397	0.438	0	0.2	0.028	0.008	0.524	2.8	0.001	0	0
	Nikkei	2.2	0.02	0.002	0	0.4	0.125	0.001	0	0.8	0.641	0.783	0	1	1	0.106	0
	WIG	1.4	0.397	0.438	0.012	1.6	0.215	0	0	0.2	0.028	0.008	0.524	2	0.048	0.001	0
	MOEX	1.4	0.397	0.103	0.009	0.8	0.641	0.783	0	0	0.002			0.4	0.125	0.095	0
	S&P	2.6	0.003	0	0	0.8	0.641	0.783	0	0.2	0.028	0.008	0.524	1.6	0.215	0.186	0
GARCH-n (1,1)	SHC	2	0.048	0.008	0.001	2	0.048	0.016	0.168	2.2	0.02	0.003	0.008	1.8	0.106	0.025	0
	DAX	2	0.048	0.016	0.364	2	0.048	0.016	0.001	0.6	0.331	0.386	0.935	3.4	0	0	0
	Nikkei	1.6	0.215	0.059	0	0.8	0.641	0.053	0	1.8	0.106	0.061	0.374	3.8	0	0	0
	WIG	2.4	0.008	0.001	0.015	1.6	0.215	0.003	0	0.8	0.641	0.783	0.992	3	0	0	0
	MOEX	3	0	0	0	1.4	0.397	0.438	0.012	1	1	0.87	0.991	2	0.048	0.001	0
	S&P	2.6	0.003	0	0	2.6	0.003	0	0	1.4	0.397	0.417	0.973	2.6	0.003	0	0.004
GARCH-st (1,1)	SHC	0.4	0.125	0.095	0.772	1.6	0.215	0.186	0.64	1	1	0.951	0.001	1.4	0.397	0.103	0
	DAX	0.8	0.641	0.783	0.992	0.8	0.641	0.783	0.992	0.6	0.331	0.386	0.935	2	0.048	0.016	0
	Nikkei	1.2	0.663	0.766	0.005	1	1	0.106	0.001	1.2	0.663	0.766	0.974	1.6	0.215	0.186	0.018
	WIG	1.6	0.215	0.186	0.64	1.2	0.663	0.004	0	0.6	0.331	0.386	0.935	2.2	0.02	0.002	0
	MOEX	1.6	0.215	0.186	0.018	1.2	0.663	0.766	0.005	0.8	0.641	0.693	0.935	1.8	0.106	0.025	0
	S&P	0.6	0.331	0.386	0.935	1.2	0.663	0.766	0.005	0.8	0.641	0.783	0.992	1.2	0.663	0.766	0.996

		PERIOD I				PERIOD II				PERIOD III				PERIOD IV			
model	**index**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**
GARCH-t (1,1)	SHC	1.2	0.663	0.129	0.003	1.8	0.106	0.061	0.374	1.6	0.215	0.186	0.018	1.6	0.215	0.059	0
	DAX	1.4	0.397	0.438	0.971	1.2	0.663	0.766	0.974	0.6	0.331	0.386	0.935	2.4	0.008	0.001	0
	Nikkei	1.4	0.397	0.438	0.012	1	1	0.106	0.001	1.2	0.663	0.766	0.974	2.2	0.02	0.003	0.008
	WIG	2.2	0.02	0.003	0.058	1.2	0.663	0.004	0	0.6	0.331	0.386	0.935	2.8	0.001	0	0
	MOEX	2.2	0.02	0.003	0.008	1.2	0.663	0.766	0.005	1	1	0.87	0.991	2	0.048	0.001	0
	S&P	1	1	0.951	0.997	1.8	0.106	0.061	0	0.8	0.641	0.783	0.992	2.2	0.02	0.003	0.159
GARCH-ged (1,1)	SHC	1.2	0.663	0.129	0.003	1.8	0.106	0.061	0.374	1.6	0.215	0.186	0.018	1.6	0.215	0.059	0
	DAX	1.4	0.397	0.438	0.971	1.2	0.663	0.766	0.974	0.6	0.331	0.386	0.935	2.6	0.003	0	0
	Nikkei	1.4	0.397	0.438	0.012	0.8	0.641	0.053	0	1.2	0.663	0.766	0.974	2.2	0.02	0.003	0.008
	WIG	2	0.048	0.016	0.168	1.2	0.663	0.004	0	0.6	0.331	0.386	0.935	3	0	0	0
	MOEX	2.2	0.02	0.003	0.008	1.2	0.663	0.766	0.005	1	1	0.87	0.991	1.8	0.106	0.025	0
	S&P	1	1	0.951	0.997	1.8	0.106	0.061	0	0.8	0.641	0.783	0.992	2.2	0.02	0.003	0.159
RiskMetrics	SHC	3.2	0	0	0	0.8	0.641	0.783	0.992	1	1	0.951	0	3	0	0	0
	DAX	4.4	0	0	0	2.4	0.008	0	0	0.4	0.125	0.095	0.772	5	0	0	0
	Nikkei	5.2	0	0	0	0.6	0.331	0.013	0	2.2	0.02	0.003	0.008	3.2	0	0	0
	WIG	3.4	0	0	0	2.2	0.02	0	0	1.2	0.663	0.766	0.996	3.6	0	0	0
	MOEX	3.6	0	0	0	1.8	0.106	0.025	0	0.8	0.641	0.693	0.935	2.2	0.02	0	0
	S&P	5.8	0	0	0	1.8	0.106	0.025	0	1	1	0.87	0	6.2	0	0	0
POT	SHC	0.8	0.641	0.053	0	1	1	0.951	0.997	1.4	0.397	0.438	0.012	1.2	0.663	0.129	0
	DAX	1.2	0.663	0.766	0.996	0.8	0.641	0.783	0.992	0.2	0.028	0.008	0.524	2.8	0.001	0	0
	Nikkei	1.2	0.663	0.766	0.974	0.6	0.331	0.013	0	1.2	0.663	0.766	0.974	1.8	0.106	0.061	0.019
	WIG	1	1	0.951	0.997	1	1	0.106	0.001	0.4	0.125	0.095	0.772	2.8	0.001	0	0
	MOEX	1.6	0.215	0.186	0.018	1.2	0.663	0.766	0.005	0.8	0.641	0.693	0.935	1.8	0.106	0.002	0
	S&P	0.6	0.331	0.386	0.935	1.4	0.397	0.438	0.012	0.8	0.641	0.783	0.992	1.6	0.215	0.186	0.854
QML-GARCH (1,1)	SHC	0.8	0.641	0.053	0	0.4	0.125	0.095	0.772	0.8	0.641	0.783	0	0.8	0.641	0.783	0.992
	DAX	0.8	0.641	0.783	0.992	0.2	0.028	0.008	0.524	0.2	0.028	0.008	0.524	2	0.048	0.016	0
	Nikkei	0.4	0.125	0.095	0.772	0.4	0.125	0.001	0	0.8	0.641	0.783	0.992	1.6	0.215	0.186	0.018
	WIG	0.2	0.028	0.008	0.524	0	0.002			0	0.002			1	1	0.951	0
	MOEX	1.4	0.397	0.438	0.012	1.2	0.663	0.766	0.005	0.4	0.125	0.071	0.524	1.2	0.663	0.129	0
	S&P	0.6	0.331	0.386	0.935	0.8	0.641	0.783	0	1	1	0.951	0.997	1.4	0.397	0.438	0.97

		PERIOD I				PERIOD II				PERIOD III				PERIOD IV			
model	**index**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**
HCR	SHC	0.6	0.331	0.386	0.935	0	0.002	0	0	0.2	0.028	0.008	0.524	1.6	0.215	0.186	0
	DAX	2.4	0.008	0	0	1.4	0.397	0.438	0	0.2	0.028	0.008	0.524	2.4	0.008	0	0
	Nikkei	3.4	0	0	0	0.2	0.028	0.008	0.524	0.4	0.125	0.095	0.772	2.4	0.008	0.001	0
	WIG	1.6	0.215	0.186	0.018	0.8	0.641	0.783	0.992	0.2	0.028	0.008	0.524	2	0.048	0.008	0
	MOEX	3.4	0	0	0	0	0.002	0	0	0	0.002	0	0	2	0.048	0.008	0
	S&P	4.2	0	0	0	1.2	0.663	0.766	0	1	1	0.951	0.997	6	0	0	0
HCR-GARCH (1,1)	SHC	0.8	0.641	0.053	0	0.4	0.125	0.095	0.772	0.6	0.331	0.386	0.935	0.8	0.641	0.783	0.992
	DAX	0.6	0.331	0.386	0.935	0	0.002			0.2	0.028	0.008	0.524	1.6	0.215	0.186	0
	Nikkei	0	0.002	0	0	0.2	0.028	0.008	0.524	0.4	0.125	0.095	0.772	0.6	0.331	0.386	0.935
	WIG	0.2	0.028	0.008	0.524	0.4	0.125	0.095	0.772	0.4	0.125	0.095	0.772	1	1	0.951	0
	MOEX	0.4	0.125	0.095	0.772	0.2	0.028	0.008	0.524	0	0.002	0	0	0.4	0.125	0.095	0.772
	S&P	0.4	0.125	0.095	0.772	0.2	0.028	0.008	0.524	1	1	0.951	0.997	1.4	0.397	0.438	0.97
HCR-QML-GARCH (1,1)	SHC	0.8	0.641	0.053	0	0.4	0.125	0.095	0.772	0.8	0.641	0.783	0	0.8	0.641	0.783	0.992
	DAX	0.8	0.641	0.783	0.992	0.4	0.125	0.095	0.772	0.2	0.028	0.008	0.524	2	0.048	0.016	0
	Nikkei	0.4	0.125	0.095	0.772	1.2	0.663	0.129	0.003	0.4	0.125	0.095	0.772	1	1	0.951	0.001
	WIG	1	1	0.951	0.997	0.6	0.331	0.386	0.935	0.4	0.125	0.095	0.772	1.2	0.663	0.766	0
	MOEX	1	1	0.951	0.997	0.8	0.641	0.783	0	0.6	0.331	0.323	0.772	1	1	0.951	0.001
	S&P	0.6	0.331	0.386	0.935	0.8	0.641	0.783	0	1	1	0.951	0.997	1.4	0.397	0.438	0.97

ER & LT denote the excess ratio and color present the Basel traffic light test result. CC, UC, DQ, indicate the p value of the corresponding test, where green means exceeding 10% p-value.

Table 6. Results of 2.5% VaR estimation of the financial returns per each analyzed model, stock index and period

		PERIOD I				PERIOD II				PERIOD III				PERIOD IV			
model	**index**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**
Historical Simulation	SHC	3	0.487	0.467	0.56	0.8	0.005	0	0.207	0.8	0.005	0	0.004	3.8	0.083	0.046	0.069
	DAX	6	0	0	0	3	0.487	0.001	0	1	0.015	0	0	5.4	0	0	0
	Nikkei	5.4	0	0	0	0.8	0.005	0	0.003	2.4	0.885	0.552	0.429	4.2	0.026	0.004	0.001
	WIG	4.6	0.007	0	0	2.2	0.661	0	0	1.6	0.168	0.042	0.109	5.8	0	0	0
	MOEX	4.8	0.003	0	0	2	0.458	0.243	0.002	0.8	0.005	0	0.119	3.8	0.083	0.021	0
	S&P	5.8	0	0	0	2	0.458	0.243	0	1	0.015	0.002	0.004	6.2	0	0	0
KDE Epanechniko	SHC	3	0.487	0.467	0.56	0.8	0.005	0	0.207	0.8	0.005	0	0.004	3.8	0.083	0.046	0.069
	DAX	5.8	0	0	0	3	0.487	0.001	0	0.4	0	0	0.062	5.2	0.001	0	0
	Nikkei	5.2	0.001	0	0	0.8	0.005	0	0.003	2.2	0.661	0.648	0.606	4.2	0.026	0.004	0.001

		PERIOD I				PERIOD II				PERIOD III				PERIOD IV			
model	**index**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**
	WIG	4.4	0.014	0	0	2	0.458	0	0	1.6	0.168	0.042	0.109	5	0.002	0	0
	MOEX	4.8	0.003	0	0	2	0.458	0.243	0.002	0.8	0.005	0	0.119	3.4	0.221	0.067	0
	S&P	5.8	0	0	0	1.6	0.168	0.042	0	1	0.015	0.002	0.004	6	0	0	0
Modified Cornish Fisher	SHC	3.4	0.221	0.192	0.471	1	0.015	0.003	0.329	1	0.015	0.003	0	2.2	0.661	0.648	0.6
	DAX	4.4	0.014	0.002	0	3.2	0.336	0.001	0	0.8	0.005	0	0.004	4	0.048	0.01	0.001
	Nikkei	4.6	0.007	0	0	0.4	0	0	0	2.4	0.885	0.552	0.429	3.2	0.336	0.098	0.003
	WIG	3.8	0.083	0.021	0.001	2.2	0.661	0	0	0.8	0.005	0	0.207	4	0.048	0	0
	MOEX	2.4	0.885	0.552	0.429	2	0.458	0.243	0.002	0.6	0.001	0	0.119	1.6	0.168	0.002	0
	S&P	5.8	0	0	0	1.8	0.292	0.115	0	1	0.015	0.002	0.004	4.8	0.003	0	0
GARCH-n (1,1)	SHC	3.8	0.083	0.021	0.029	2.8	0.673	0.555	0.712	3.6	0.139	0.1	0.378	3	0.487	0.467	0.56
	DAX	4	0.048	0.008	0.108	4.4	0.014	0.001	0.001	1.6	0.168	0.134	0.757	5.2	0.001	0	0
	Nikkei	4.6	0.007	0.001	0.005	2.6	0.887	0.619	0.701	2.8	0.673	0.555	0.7	5.2	0.001	0	0
	WIG	3.4	0.221	0.192	0.569	3.2	0.336	0.098	0.105	2.4	0.885	0.73	0.918	4.6	0.007	0	0
	MOEX	4	0.048	0.008	0.105	3.2	0.336	0.23	0.077	2.4	0.885	0.734	0.924	2.6	0.887	0.117	0
	S&P	4.6	0.007	0	0.012	4.2	0.026	0.007	0.002	1.8	0.292	0.277	0.755	4	0.048	0.019	0.081
GARCH-st (1,1)	SHC	2.2	0.661	0.406	0.276	2.4	0.885	0.73	0.686	2.2	0.661	0.648	0.6	2.6	0.887	0.619	0.481
	DAX	3.2	0.336	0.23	0.647	3.2	0.336	0.23	0.007	1.8	0.292	0.283	0.852	4.2	0.026	0.004	0.002
	Nikkei	2.8	0.673	0.583	0.517	2	0.458	0.243	0.468	2	0.458	0.475	0.907	4.4	0.014	0.001	0.002
	WIG	2.8	0.673	0.583	0.542	2.6	0.887	0.117	0.033	2	0.458	0.475	0.847	3.6	0.139	0.04	0.001
	MOEX	3.4	0.221	0.121	0.429	3.2	0.336	0.23	0.077	2	0.458	0.469	0.85	2.6	0.887	0.117	0
	S&P	3.4	0.221	0.121	0.483	3	0.487	0.384	0.087	1.6	0.168	0.129	0.624	3.2	0.336	0.23	0.654
GARCH-t (1,1)	SHC	3.8	0.083	0.021	0.029	2.8	0.673	0.583	0.571	2.8	0.673	0.555	0.7	2.8	0.673	0.583	0.542
	DAX	4.2	0.026	0.003	0.003	4	0.048	0.008	0	2.2	0.661	0.648	0.928	4.8	0.003	0	0.002
	Nikkei	4.2	0.026	0.007	0.006	2.4	0.885	0.552	0.664	2.4	0.885	0.73	0.918	5.2	0.001	0	0
	WIG	3.2	0.336	0.321	0.606	3	0.487	0.123	0.081	2.4	0.885	0.73	0.921	4.6	0.007	0	0
	MOEX	3.6	0.139	0.056	0.302	3.2	0.336	0.23	0.077	2.4	0.885	0.734	0.924	2.8	0.673	0.131	0.001
	S&P	4.2	0.026	0.003	0.103	3.6	0.139	0.056	0.076	1.8	0.292	0.277	0.755	4	0.048	0.01	0.016
GARCH-ged (1,1)	SHC	3.4	0.221	0.067	0.008	2.4	0.885	0.73	0.686	2.6	0.887	0.691	0.712	2.8	0.673	0.583	0.542
	DAX	3.8	0.083	0.023	0.202	3.6	0.139	0.056	0	1.6	0.168	0.134	0.757	4.2	0.026	0.004	0.002
	Nikkei	3.4	0.221	0.192	0.125	2.4	0.885	0.552	0.664	2.2	0.661	0.648	0.928	5	0.002	0	0.001
	WIG	3.2	0.336	0.321	0.606	2.8	0.673	0.131	0.056	1.6	0.168	0.134	0.757	4.4	0.014	0.001	0
	MOEX	3.6	0.139	0.056	0.302	3.2	0.336	0.23	0.077	2.4	0.885	0.734	0.924	2.6	0.887	0.117	0

model	index	PERIOD I ER & LT	PERIOD I UC	PERIOD I CC	PERIOD I DQ	PERIOD II ER & LT	PERIOD II UC	PERIOD II CC	PERIOD II DQ
	S&P	4	0.048	0.008	0.18	3.4	0.221	0.121	0.09
RiskMetrics	SHC	5.4	0	0	0	2	0.458	0.475	0.49
	DAX	6.4	0	0	0	4	0.048	0	0
	Nikkei	6.4	0	0	0	0.8	0.005	0	0.003
	WIG	5.4	0	0	0	3	0.487	0	0
	MOEX	6.8	0	0	0	2.6	0.887	0.117	0.002
	S&P	7.8	0	0	0	3.2	0.336	0.015	0
POT	SHC	2.2	0.661	0.406	0.005	1.8	0.292	0.283	0.852
	DAX	3.6	0.139	0.056	0.334	2.8	0.673	0.555	0
	Nikkei	3	0.487	0.467	0.563	1.2	0.039	0.002	0.046
	WIG	2.8	0.673	0.583	0.547	2.2	0.661	0.052	0.011
	MOEX	3	0.487	0.384	0.645	2.2	0.661	0.648	0.6
	S&P	3.4	0.221	0.121	0.483	2.6	0.887	0.691	0.051
QML-GARCH (1,1)	SHC	1.6	0.168	0.042	0.208	1.8	0.292	0.283	0.852
	DAX	2.2	0.661	0.648	0.902	1.2	0.039	0.013	0.475
	Nikkei	1.2	0.039	0.013	0.052	0.6	0.001	0	0
	WIG	1.4	0.086	0.009	0.005	0.4	0	0	0
	MOEX	2.6	0.887	0.691	0.712	1.2	0.039	0.013	0.052
	S&P	1.8	0.292	0.283	0.354	2	0.458	0.475	0.007
HCR	SHC	3.8	0.083	0.019	0.071	1.2	0.039	0.013	0.475
	DAX	4.6	0.007	0	0	2.6	0.887	0.117	0
	Nikkei	6.4	0	0	0	1.4	0.086	0.012	0.109
	WIG	3.6	0.139	0.056	0.085	1.8	0.292	0.115	0.088
	MOEX	4	0.048	0	0	1.2	0.039	0.002	0.006
	S&P	8.6	0	0	0	2.8	0.673	0.583	0
HCR-GARCH (1,1)	SHC	1	0.015	0	0.014	1.4	0.086	0.049	0.625
	DAX	0.8	0.005	0	0.207	0.6	0.001	0	0.119
	Nikkei	0.6	0.001	0	0.119	0.6	0.001	0	0
	WIG	0.8	0.005	0	0.207	0.6	0.001	0	0.119
	MOEX	1.4	0.086	0.049	0.119	1.2	0.039	0.013	0.052
	S&P	1.4	0.086	0.049	0.119	1.4	0.086	0.049	0
HCR	SHC	1.4	0.086	0.012	0.109	2	0.458	0.475	0.907

model	index	PERIOD III ER & LT	PERIOD III UC	PERIOD III CC	PERIOD III DQ	PERIOD IV ER & LT	PERIOD IV UC	PERIOD IV CC	PERIOD IV DQ
	S&P	1.6	0.168	0.129	0.624	3.6	0.139	0.056	0.098
RiskMetrics	SHC	1.2	0.039	0.013	0.003	4	0.048	0.019	0.002
	DAX	1.2	0.039	0.002	0.002	6.6	0	0	0
	Nikkei	3	0.487	0.467	0.074	6.4	0	0	0
	WIG	2	0.458	0.243	0.332	5	0.002	0	0
	MOEX	0.8	0.005	0	0.119	3	0.487	0.123	0
	S&P	1.8	0.292	0.099	0.027	8.2	0	0	0
POT	SHC	2.4	0.885	0.73	0.678	2.4	0.885	0.552	0.387
	DAX	1.2	0.039	0.013	0.475	4.2	0.026	0.004	0.002
	Nikkei	1.6	0.168	0.134	0.757	5	0.002	0	0.001
	WIG	1	0.015	0.003	0.329	4.2	0.026	0.001	0
	MOEX	1.8	0.292	0.099	0.207	2.6	0.887	0.117	0
	S&P	1.6	0.168	0.129	0.624	3.4	0.221	0.121	0.552
QML-GARCH (1,1)	SHC	2	0.458	0.475	0.486	1.8	0.292	0.115	0.088
	DAX	0.6	0.001	0	0.119	3	0.487	0.384	0.001
	Nikkei	1.6	0.168	0.134	0.757	3.2	0.336	0.23	0.079
	WIG	0	0			2	0.458	0.025	0
	MOEX	0.6	0.001	0	0.062	2.4	0.885	0.087	0
	S&P	1	0.015	0.003	0.329	2.4	0.885	0.73	0.645
HCR	SHC	1	0.015	0.003	0	3.8	0.083	0.023	0.07
	DAX	0.6	0.001	0	0.119	4.8	0.003	0	0.001
	Nikkei	1.6	0.168	0.129	0.624	5	0.002	0	0
	WIG	0.4	0	0	0.062	4	0.048	0.01	0
	MOEX	0.4	0	0	0.029	2	0.458	0.025	0
	S&P	1.8	0.292	0.283	0.002	8.4	0	0	0
HCR-GARCH (1,1)	SHC	2	0.458	0.475	0.486	1.4	0.086	0.012	0.012
	DAX	0.2	0	0	0.029	2.4	0.885	0.73	0.026
	Nikkei	1.4	0.086	0.049	0.625	2	0.458	0.475	0.007
	WIG	0.4	0	0	0.062	1.6	0.168	0.134	0.001
	MOEX	0.4	0	0	0.029	1.2	0.039	0.013	0.052
	S&P	1	0.015	0.003	0.329	2.2	0.661	0.648	0.899
HCR	SHC	2	0.458	0.475	0.486	2	0.458	0.243	0.17

		PERIOD I				PERIOD II				PERIOD III				PERIOD IV			
model	**index**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**	**ER & LT**	**UC**	**CC**	**DQ**
	DAX	2.2	0.661	0.648	0.902	1.2	0.039	0.013	0.475	0.6	0.001	0	0.119	2.8	0.673	0.555	0
	Nikkei	1.2	0.039	0.013	0.052	0.6	0.001	0	0	1.6	0.168	0.134	0.757	3.2	0.336	0.23	0.079
	WIG	1.2	0.039	0.013	0.475	1.4	0.086	0.012	0.109	1	0.015	0.003	0.329	3.2	0.336	0.321	0.001
	MOEX	2.2	0.661	0.648	0.6	1.6	0.168	0.134	0.222	0.6	0.001	0	0.062	2.4	0.885	0.087	0.017
	S&P	1.8	0.292	0.283	0.354	2	0.458	0.475	0.007	1	0.015	0.003	0.329	2.4	0.885	0.73	0.645

ER & LT denote the excess ratio and color present the Basel traffic light test result. CC, UC, DQ, indicate the p value of the corresponding test, where green means exceeding 10% p-value.

5. Conclusion

The main purpose of this study was to present and empirically test three novel statistical learning Value at Risk models: nonparametric HCR, semiparametric HCR-GARCH and hybrid semiparametric HCR-QML-GARCH. The benchmark procedure compares the 1% and 2.5% one-day-ahead VaR forecasts obtained with the above models against the estimates of classical VaR methods, depending on the time period and the stock index. The major hypothesis verified in this paper was that HCR-GARCH model can outperform in the formal backtesting procedure classical VaR models. According to results, HCR-GARCH model outperforms classical VaR models under conditions of sudden increased volatility in the markets which emerges as a consequence of the Covid-19 pandemic. In the case of low, moderate and gradually increasing volatility, the GARCH-st model proved to be the best and universal model. This conclusion is in line with the results, inter alia, Abad et al. (2014), Buczyński and Chlebus (2018) and Buczyński and Chlebus (2019). Generally, HCR-GARCH works the best in extreme market conditions. The most basic VaR model proposed by us HCR is not able to outperform models in its class of nonparametric approaches. Despite the innovation of estimating the density with a strong consideration of time, in most cases HCR converges to the results presented by classical nonparametric models such as: Historical Simulation or KDE. Additionally, we check whether hybrid HCR-QML-GARCH model perform better than the single HCR-GARCH and HCR-QML-GARCH models. On the basis of the backtesting carried out, we can conclude that this hybrid performs better than single models in almost all cases. Which coincides with Lux's conclusion (2020). This approach makes the HCR-GARCH model more universal, i.e. it can be applied under conditions of moderate volatility.

We state that our contribution to literature is twofold. First, we provide added value to the strand of literature on the semiparametric Value at Risk models by introducing HCR-GARCH and HCR-QML-GARCH models. Second, our study contributes to the analysis of the performance of the classical Value at Risk models in the Covid-19 period.

We are aware of the significant limitations of this study. First, for the HCR model, one can test: (1) a different method to normalize the time series e.g. empirical cumulative distribution function or a different distribution than Laplace (2) a more sophisticated method to deal with negative density forecasting (3) a different basis than polynomial e.g. sine/consinus etc. Another improvement could be obtained using different algorithms of models ensembling for example regime switching approach may be applied. An interesting extension of hybrid models would be the combination of the HCR-GARCH semiparametric model with the most promising parametric models, e.g. GARCH-st. What is more, for models benchmarking purpose, it is worth using more State of The Art models in the field of machine learning and more stock indexes.

References

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. https://doi.org/10.1016/j.srfe.2013.06.001

Alemany R., Bolancé C., Guillen M., 2012. Nonparametric Estimation of Value-at-Risk. XARXA de Referencia en Economia Aplicada. https://dx.doi.org/10.2139/ssrn.2162585

Arian H., Moghimi M., Tabatabaei E., Zamani S., 2020. Encoded Value-at-Risk: A Predictive Machine for Financial Risk Management. arXiv preprint arXiv:2011.06742.

Arimond A., Borth D., Hoepner A. G., Klawunn M., Weisheit S., 2020. Neural Networks and Value at Risk. Michael J. Brennan Irish Finance Working Paper Series Research Paper, (20-7). arXiv preprint arXiv: 2005.01686.

Avdulaj K., Barunik J., 2017. A semiparametric nonlinear quantile regression model for financial returns. Studies in Nonlinear Dynamics & Econometrics 21(1), 81-97. https://doi.org/10.1515/snde-2016-0044

Bali T. G., Mo H., Tang Y., 2008. The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR. Journal of Banking & Finance 32(2), 269-282. https://doi.org/10.1016/j.jbankfin.2007.03.009

Banhudo G., 2019. Adaptive value-at-risk policy optimization: a deep reinforcement learning approach for minimizing the capital charge. University Institute of Lisbona. http://hdl.handle.net/10071/19197

Bank for International Settlements, 2006. International convergence of capital easurement and capital standards. Basel, Switzerland.

Basle Committee on Banking Supervision, 1996. Supervisory Framework For The Use of Back-Testing in Conjunction With The Internal Models Approach to Market Risk Capital Requirements. Basel, Switzerland.

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

Buczyński M., Chlebus M., 2018. Comparison of Semi-Parametric and Benchmark Value-At-Risk Models in Several Time Periods with Different Volatility Levels. e-Finanse. 14(2), 67-82. https://doi.org/10.2478/fiqf-2018-0013

Buczyński M., Chlebus M., 2019. Old-fashioned parametric models are still the best. A comparison of Value-at-Risk approaches in several volatility states. Journal of Risk Model Validation 14 (2) https://doi.org/10.21314/JRMV.2020.222

Buczyński M., Chlebus M., 2020. Size does matter. A study on the required window size for optimal quality market risk models. Working Papers 2020-09, Faculty of Economic Sciences, University of Warsaw.

Cavenaile L., Lejeune T., 2012. A Note on the Use of Modified Value-at-Risk. The Journal of Alternative Investments 14(4), 79–83. https://doi.org/10.3905/jai.2012.14.4.079

Christoffersen P., 1998. Evaluating Interval Forecasts. International Economic Review, 39(4), 841–862. https://doi.org/10.2307/2527341

Corlosquet M., Janssen J., Manca R., 2013. Use of Value-at-Risk (VaR) Techniques for Solvency II, Basel II and III. VaR Methodology for Non-Gaussian Finance. London: ISTE/Wiley. https://doi.org/10.1002/9781118733691.ch1

Costanzino N., Curran M., 2018. A Simple Traffic Light Approach to Backtesting Expected Shortfall. Risks 6(1), 2. https://doi.org/10.3390/risks6010002

Duda J., 2017. Rapid parametric density estimation. arXiv preprint arXiv:1702.02144.

Duda J., 2018. Hierarchical correlation reconstruction with missing data, for example for biology-inspired neuron. arXiv preprint arXiv:1804.06218.

Duda J., Snarska M., 2018. Modeling joint probability distribution of yield curve parameters. arXiv preprint arXiv:1807.11743.

Duda J., 2019. Exploiting statistical dependencies of time series with hierarchical correlation reconstruction. arXiv preprint arXiv:1807.04119.

Duda J., Szulc A., 2020. Social Benefits Versus Monetary and Multidimensional Poverty in Poland: Imputed Income Exercise. In: Tsounis N., Vlachvei A. (eds) Advances in Cross-Section Data Methods in Applied Economic Research. ICOAE 2019. Springer Proceedings in Business and Economics. Springer, Cham. https://doi.org/10.1007/978-3-030-38253-7_6

Embrechts P., Kluppelberg C., Mikosch T., 1997. Modelling extremal events for Insurance and Finance. Springer-Verlag, 295-305. http://doi.org/10.1007/978-3-642-33483-2

Embrechts P., Resnick S., Samorodnitsky G., 1999. Extreme value theory as a risk management tool. North American Actuarial Journal 3(2). https://doi.org/10.1080/10920277.1999.10595797

Engle R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50(4), 987–1007. https://doi.org/10.2307/1912773

Engle R., Manganelli S., 1999. CAViaR: Conditional Value at Risk by Quantile Regression. NBER, Working Paper 7341.

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

Epanechnikov A., 1969. Non-parametric estimation of a multivariate probability density. Theory of Probability and Its Applications,14: 153–158. https://doi.org/10.1137/1114019

Ergun A. T., Jun J., 2010. Time-varying higher-order conditional moments and forecasting intraday VaR and expected shortfall. The Quarterly Review of Economics and Finance, 50(3), 264-272. https://doi.org/10.1016/j.qref.2010.03.003

Gerlach R., Wang C., 2020. Semi-parametric dynamic asymmetric Laplace models for tail risk forecasting, incorporating realized measures. International Journal of Forecasting, 36(2), 489-506. https://doi.org/10.1016/j.ijforecast.2019.07.003

Hansen B. E., 1994. Autoregressive conditional density estimation. International Economic Review, 705-730. https://doi.org/10.2307/2527081

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

Jones M., Marron J., Sheather S., 1996. A Brief Survey of Bandwidth Selection for Density Estimation. Journal of the American Statistical Association 91, 401–407. https://doi.org/10.1080/01621459.1996.10476701

Jorion P., 2001. Value at Risk: The New Benchmark for Managing Financial Risk. McGraw-Hill, United States of America

Kaszyński D., Kamiński B., Pankratz B., 2020. Assessment of the size of VaR backtests for small samples. Przegląd Statystyczny, 67 (2): 114-151. https://doi.org/10.5604/01.3001.0014.5726

Khan A.I., 2011. Modelling daily value-at-risk using realized volatility, non-linear support vector machine and ARCH type models. J Econ Int Finance 3(5), 305–321 https://doi.org/10.5897/JEIF.9000078

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

Li Z., Tran M. N., Wang C., Gerlach R., Gao J., 2020. A Bayesian Long Short-Term Memory Model for Value at Risk and Expected Shortfall Joint Forecasting. arXiv preprint arXiv:2001.08374.

Lux M., Härdle W., Lessmann S., 2020. Data driven value-at-risk forecasting using a SVR-GARCH-KDE hybrid. Computational Statistics, 35, 947–981. https://doi.org/10.1007/s00180-019-00934-7

Morgan J. P., 1996. Riskmetrics technical document, 4th ed. J.P. Morgan, New York

Mostafa F., Dillon T., Chang E., 2017. Value at Risk. Computational Intelligence Applications to Option Pricing, Volatility Forecasting and Value at Risk. Studies in Computational Intelligence, 697. Springer, Cham. https://doi.org/10.1007/978-3-319-51668-4_8

Nguyen N., Tran M. N., Gunawan D., Kohn R., 2019. A long short-term memory stochastic volatility model. arXiv preprint arXiv:1906.02884.

Patton A. J., Ziegel J. F., Chen R., 2019. Dynamic semiparametric models for expected shortfall (and value-at-risk). Journal of econometrics, 211(2), 388-413. https://doi.org/10.1016/j.jeconom.2018.10.008

Polanski A., Stoja E., 2010. Incorporating higher moments into value-at-risk forecasting. Journal of Forecasting, 29(6), 523-535. https://doi.org/10.1002/for.1155

PricewaterhouseCoopers, 2016. Basel IV: Revised internal models approach for market risk.

Radovi´c O., Stankovi´c J., Stankovi´c J., 2015. Tail risk assessment using support vector machine. J Eng Sci Technol Rev 8(1), 61–64 https://doi.org/10.25103/jestr.081.11

Risk.net, 2020. Market risk. Risk glossary. https://www.risk.net/definition/market-risk

Rockafellar R., Uryasev S., 2002. Conditional value-at-risk for general loss distributions. Journal of banking & finance, 26(7), 1443-1471. https://doi.org/10.1016/S0378-4266(02)00271-6

Sarykalin S., Serraino G., Uryasev S., 2008. Value-at-risk vs. conditional value-at-risk in risk management and optimization. 270–294. https://doi.org/10.1287/educ.1080.0052

Silverman B., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Scott D., 1992. Multivariate Density Estimation. Theory, Practice and Visualization, New York.

Sheather S., Jones M., 1991. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society series B, 53: 683–690. https://doi.org/10.1111/j.2517-6161.1991.tb01857.x

Szubzda F., Chlebus M., 2019. Comparison of Block Maxima and Peaks Over Threshold Value-at-Risk models for market risk in various economic conditions. Central European Economic Journal, 6(53), 70-85. https://doi.org/10.2478/ceej-2019-0005

Taylor J. W., 2019. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. Journal of Business & Economic Statistics, 37(1), 121-133. https://doi.org/10.1080/07350015.2017.1281815

Vapnik V. N., 1995. The nature of statistical learning theory. Springer, New York.

Wang C.S., Zhao Z., 2016. Conditional Value-at-Risk: Semiparametric estimation and inference. Journal of Econometrics, 195(1), 86-103. https://doi.org/10.1016/j.jeconom.2016.07.002

Xu Q., Jiang C., He Y., 2016. An exponentially weighted quantile regression via SVM with application to estimating multiperiod VaR. Stat Methods Appl 25(2), 285–320. https://doi.org/10.1007/s10260-015- 0332-9

Yang L., Hamori S., 2020. Forecasts of Value-at-Risk and Expected Shortfall in the Crude Oil Market: A Wavelet-Based Semiparametric Approach. Energies, 13(14), 3700. https://doi.org/10.3390/en13143700

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