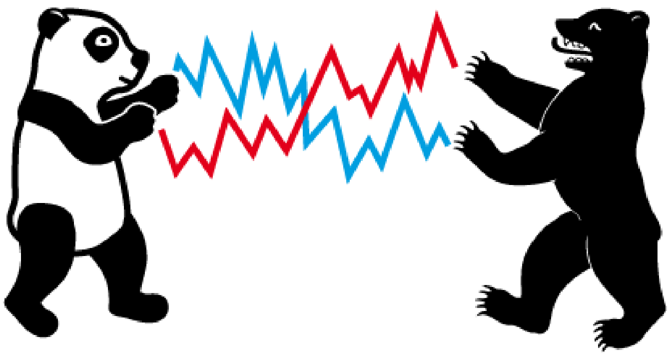
IRTG 1792 Discussion Paper 2018-001

Data Driven Value-at-Risk Forecasting using a SVR-GARCH-KDE Hybrid

Marius Lux *
Wolfgang Karl Härdle *
Stefan Lessmann *



* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Data Driven Value-at-Risk Forecasting using a SVR-GARCH-KDE Hybrid

Marius Lux[a,*], Wolfgang Karl Härdle[a,b], Stefan Lessmann[a]

[a]*School of Business and Economics, Humboldt-University Berlin, Unter den Linden 6, D-10099 Berlin, Germany*

[b]*SKBI School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899*

Abstract

Appropriate risk management is crucial to ensure the competitiveness of financial institutions and the stability of the economy. One widely used financial risk measure is Value-at-Risk (VaR). VaR estimates based on linear and parametric models can lead to biased results or even underestimation of risk due to time varying volatility, skewness and leptokurtosis of financial return series. The paper proposes a nonlinear and nonparametric framework to forecast VaR. Mean and volatility are modeled via support vector regression (SVR) where the volatility model is motivated by the standard generalized autoregressive conditional heteroscedasticity (GARCH) formulation. Based on this, VaR is derived by applying kernel density estimation (KDE). This approach allows for flexible tail shapes of the profit and loss distribution and adapts for a wide class of tail events.

The SVR-GARCH-KDE hybrid is compared to standard, exponential and threshold GARCH models coupled with different error distributions. To examine the performance in different markets, one-day-ahead forecasts are produced for different financial indices. Model evaluation using a likelihood ratio based test framework for interval forecasts indicates that the SVR-GARCH-KDE hybrid performs competitive to benchmark models. Especially models that are coupled with a normal distribution are systematically outperformed.

Keywords: Value-at-Risk, Support Vector Regression, Kernel Density Estimation, GARCH

1. Introduction

Events like the 2008 financial crisis or the outcome of the 2016 referendum in the UK came unexpected for many people. Yet, as these examples illustrate, unlikely events occur at times and they might have far reaching consequences. Risk management is the practice to

*Corresponding author

Email addresses: lux.marius@gmail.com (Marius Lux), haerdle@hu-berlin.de (Wolfgang Karl Härdle), stefan.lessmann@hu-berlin.de (Stefan Lessmann)

analyze the macro-environment of an organization, identify possible adverse developments, and design suitable countermeasures.

For financial institutions and systemically important institutions in particular, a key risk management responsibility is to sustain solvency under adverse economic conditions (e.g., Silva et al., 2017; Kraus and Czado, 2017). One of the most popular measures of uncertainty in financial markets is VaR (e.g., Alexander (2008)). VaR is based on the quantiles of a portfolio's profit and loss ($P\&L$) distribution and can be interpreted as an upper bound on the potential loss that will not be exceeded with a given level of confidence. Its use is appealing because it summarizes the downside risk of an institution in one easily interpretable figure (e.g., Chen et al., 2012). Regulatory frameworks for the banking and insurance industry such as Basel III or Solvency II also rely on VaR for determining capital requirements. Compared to expected shortfall, an alternative risk measure with some superior mathematical properties (e.g., Kim and Lee, 2016), an advantage of VaR may be seen in the fact that its estimation is more robust due to putting less weight on tail events and large losses, which may deteriorate the quality of statistical estimation routines (Sarykalin et al., 2008).

Several approaches have been proposed to estimate VaR including parametric statistical models and data-driven machine learning algorithms such as neural networks (NN) and SVR. In a seminal study, Kuester et al. (2006) review several statistical methods and compare these in a forecasting benchmark. Using more than 30 years of historical returns data, they find standard GARCH models to forecast VaR with the highest accuracy on average.

GARCH models are also employed by Chen et al. (2012) to estimate VaR for four daily series of stock market indices. More specifically, Chen et al. (2012) rely on an asymmetric Laplace distribution and model volatility using a GJR-GARCH model to introduce leverage effects. They then develop a time-varying model to allow for dynamic higher moments. These extensions allow for wider application of the model beyond forecasting.

Unlike the parametric approach of Chen et al. (2012), Franke and Diagne (2006) estimate VaR for the German stock index through fitting the mean and volatility of the return series using NNs. More specifically, they model the mean and volatility as an autoregressive (AR) and autoregressive conditionally heteroscedastic (ARCH) process, respectively. To derive VaR and expected shortfall, Franke and Diagne (2006) use the predicted mean and variance with the normal distribution. This model outperforms a standard GARCH model in terms of VaR exceedances and proofs capable of quickly adjusting volatility in case of shocks with only short impact. Dunis et al. (2010) also propose a NN-based approach towards forecasting VaR and expected shortfall.

Khan (2011) develops a VaR-model that forecasts realized volatilities using a combination of a heterogeneous AR model and SVR. VaR is then computed based on the normal, t- and skewed t-distribution. Applying this model to 5- and 15-minutes return data, Khan (2011) is able to confirm the suitability of the SVR component. O. Radović et al. (2015) provide further evidence that SVR is a useful method for VaR forecasting. Likewise, Xu et al. (2016) introduce a multi-period VaR model using SVR in a quantile regression framework and show this approach to outperform GARCH models.

The findings of Xu et al. (2016) seem to disagree with prior results of Kuester et al. (2006) where GARCH models predict VaR with highest average accuracy and more accurately than quantile regression approaches in particular. Implementations of the quantile regression using a data-driven SVR model might explain the results of Xu et al. (2016). More specifically, the linear and parametric structure of standard GARCH models might be a limiting factor in VaR forecasting. Moreover, the parameters of GARCH-type models are usually estimated via maximum likelihood estimation (Bollerslev, 1986). This necessitates distributional assumptions, which might be problematic since the distribution of financial returns is skewed and exhibits fat tails (Bali et al., 2008; Harvey and Siddique, 2000).

Noting the possible limitation of the parametric framework, Schaumburg (2012) combines extreme value theory with nonparametric VaR estimation to forecast return distributions of four financial stock indices. A parametric conditional autoregressive value at risk (CAViaR) model serves as benchmark. The benchmark and the proposed model both circumvent the estimation of the mean and variance of the $P\&L$ distribution through predicting a quantile directly. In this regard, the approach of Schaumburg (2012) can be characterized as a nonparametric CAViaR model.

VaR forecasts based on CAViaR frameworks have also been considered in the benchmarking study of Kuester et al. (2006). In fact, the authors also introduce a novel CAViaR model in the paper and test it alongside various other VaR models. However, GARCH models and models relying on the t-distribution in particular emerge as most suitable for VaR modeling. Therefore, a novel VaR model that grounds on the GARCH framework but estimates the individual components in a purely nonparametric manner is proposed here. More specifically, we estimate the mean and variance of the $P\&L$ distribution using SVR and employ KDE to model the density of the standardized residuals (e.g., Härdle et al., 2004). We then integrate these components to derive a VaR forecast. In other words, we propose to start from the most effective parametric modeling approach of Kuester et al. (2006) and develop models that estimate its components in a purely data-driven manner. Although KDE is not a new approach in VaR forecasting (e.g., Chen et al., 2016; Schaumburg, 2012; Malec and Schienle, 2014), the particular combination of data-driven VaR estimation using SVR and nonparametric density estimation, which we propose in this paper, has, to the best of our knowledge, not been considered in prior work.

We assess the performance of the proposed model in comparison to GARCH-type models with different error distributions also including skewed and fat-tailed distributions. Empirical experiments using data from three major financial indices, namely the Euro STOXX 50, Nikkei 225 and Standard & Poor's 500 (S&P 500), suggest that the SVR-GARCH-KDE hybrid typically outperforms models that are coupled with a normal distribution and performs competitive to other benchmark models.

The remainder of the paper is organized as follows. In Section 2 VaR is defined and the methods underlying the proposed VaR modeling framework are presented. Specifically, the standard GARCH approach, nonparametric density estimation via KDE and SVR are introduced. The proposed SVR-GARCH-KDE hybrid is then developed based on these building blocks. After outlining the theoretical background, the SVR-GARCH-KDE hybrid is compared to other models on different datasets in Section 3. Concluding remarks and suggestions for future research are provided in the last section.

2. Methodology

2.1. Defining Value-at-Risk

In general, VaR can be derived from the portfolio's $P\&L$ distribution. However, since today's portfolio value is usually known, it suffices to model the return distribution. For a formal description of VaR, let the portfolio returns r_t in period t have the cumulative distribution function (CDF) F_t. Then, the VaR in d trading days for a confidence level $1-\alpha$ is defined as

$$VaR^{\alpha}_{t+d} = -F^{-1}_{t+d}(\alpha) = -\inf\{x \in \mathbb{R} : F_{t+d}(x) \geq \alpha\} \quad \text{with} \quad \alpha \in (0, 1). \tag{1}$$

In the rest of the paper, VaR refers to the negative α-quantile of the next period's portfolio return distribution.

2.2. Estimating VaR Using Location-Scale Models

The proposed VaR modeling framework is based on the location-scale approach. Models of this class estimate the entire distribution of asset returns and derive VaR as a quantile of that distribution (e.g., Kuester et al., 2006). Such approaches assume the return process is described as

$$r_t = \mu_t + u_t = \mu_t + \sigma_t z_t, \qquad z_t \sim (0, 1) \text{ i.i.d.} \tag{2}$$

In (2), μ_t is the location and $\sigma_t > 0$ the scale parameter. Given r_t belongs to the location-scale family and F_z is the CDF of z, we can compute VaR as

$$VaR^{\alpha}_t = -\left\{\mu_t + \sigma_t F^{-1}_z(\alpha)\right\}. \tag{3}$$

Autoregressive moving average (ARMA) processes and GARCH-type models are commonly used to estimate μ_t and σ_t in (2).

2.3. Modeling Volatility Using GARCH Models

Bollerslev (1986) introduces GARCH models by generalizing the volatility modeling approach of Engle (1982). In deriving the GARCH regression model Bollerslev (1986) starts by assuming conditional normality of the return process r_t:

$$r_t | \mathcal{F}_{t-1} \sim N(\beta^{\top} x_t, \sigma^2_t). \tag{4}$$

where x_t is a vector of lagged endogenous as well as exogenous variables, β an unknown parameter vector and $\mathcal{F}_{t-1}$ the information set available at $t-1$. Rewriting (4) as linear model with conditionally heteroscedastic and normally distributed disturbances gives:

$$r_t = \beta^{\top} x_t + u_t, \qquad u_t | \mathcal{F}_{t-1} \sim N(0, \sigma^2_t). \tag{5}$$

Then, the GARCH(p,q) representation of the variance σ^2_t is

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \delta_i u_{t-i}^2 + \sum_{j=1}^{p} \theta_j \sigma_{t-j}^2. \tag{6}$$

Bollerslev (1986) notes that (6) has an ARMA representation. To see this let $\nu_t = u_t^2 - \sigma_t^2$ and substitute σ_t^2 in (6) with $u_t^2 - \nu_t$ to obtain

$$u_t^2 - \nu_t = \omega + \sum_{i=1}^{q} \delta_i u_{t-i}^2 + \sum_{j=1}^{p} \theta_j (u_{t-j}^2 - \nu_{t-j}). \tag{7}$$

Rearranging (7) yields an ARMA representation for u_t^2:

$$u_t^2 = \omega + \sum_{i=1}^{q} \delta_i u_{t-i}^2 + \sum_{j=1}^{p} \theta_j (u_{t-j}^2 - \nu_{t-j}) + \nu_t \tag{8}$$

$$= \omega + \sum_{i=1}^{\max(p,q)} (\delta_i + \theta_i) u_{t-i}^2 - \sum_{j=1}^{p} \theta_j \nu_{t-j} + \nu_t. \tag{9}$$

Based on (9) nonlinear and nonparametric volatility models can be introduced. Hence, the volatility model in the SVR-GARCH-KDE hybrid is motivated by the ARMA representation of σ^2.

2.4. Nonparametric Density Estimation

The volatility of stock returns varies over time and a similar behavior has been observed for the third and fourth moment of the return distribution. For example, Bali et al. (2008) show that VaR forecasts can be improved by using past estimates of skewness and kurtosis. Given the evidence for the leptokurtic nature of stock returns (Franke et al., 2015), parametric distributional models might lack the flexibility to capture such distributional characteristics, which motivates the use of nonparametric methods such as KDE (e.g., Härdle et al., 2004).

Let X be a random variable with an absolutely continuous distribution function F. Further, denote the corresponding density function as f and let $\{x_1, \ldots, x_n\}$ be a sample of i.i.d. realizations of X. Then, the kernel density estimator $\hat{f}_h(x)$ of $f(x)$ is defined as

$$\hat{f}_h(x) = \frac{1}{hn} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) \tag{10}$$

where h is a bandwidth parameter with $h > 0$ and K is a so-called kernel function. Usually, a kernel function is assumed to be a symmetric density function, i.e.

$$\int_{-\infty}^{\infty} K(u)du = 1 \quad \text{with } K(u) \geq 0 \tag{11}$$

and

$$\int_{-\infty}^{\infty} uK(u)du = 0. \tag{12}$$

Conveniently, (11) implies that $\hat{f}_h(x)$ is also a density. Note that $\hat{f}_h(x)$ inherits all properties of K regarding continuity and differentiability.

The KDE based quantile estimator to forecast VaR can be derived as follows. First, the estimator for $F(x)$ that is based on KDE needs to be derived. Denote $\widehat{F}_h(x)$ as the KDE based estimate of $F(x)$. Then, $\widehat{F}_h(x)$ can be derived as follows:

$$\widehat{F}_h(x) = \int_{-\infty}^{x} \hat{f}_h(z)dz \tag{13}$$

$$= \int_{-\infty}^{x} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{z-x_i}{h}\right) dz \tag{14}$$

$$= \frac{1}{nh}\sum_{i=1}^{n} \int_{-\infty}^{x} K\left(\frac{z-x_i}{h}\right) dz. \tag{15}$$

Since the given kernel function K is a density, let Γ denote the corresponding CDF. Moreover, using the substitution $u = (z - x_i)/h$ one obtains

$$\widehat{F}_h(x) = \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\frac{x-x_i}{h}} K(u)\, du \tag{16}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \Gamma\left(\frac{x-x_i}{h}\right). \tag{17}$$

Thus, $\widehat{F}_h(x)$ is the mean of the CDF corresponding to K evaluated at $(x - x_i)/h$ for $i = 1, \ldots, n$. Then, for $\alpha \in (0, 1)$ the KDE based quantile function $\widehat{Q}_h$ is obtained as

$$\widehat{Q}_h(\alpha) = \widehat{F}_h^{-1}(x). \tag{18}$$

2.5. Support Vector Regression

SVR can be understood as a learning method to solve nonlinear regression tasks (e.g., Smola and Schölkopf, 2004). It shares some similarities with a three-layer feed-forward NN and is able to approximate arbitrarily complex functions (Chen et al., 2010). To describe the SVR model, let $\{(y_i, x_i)|i = 1, \ldots, n;\ n \in \mathbb{N}\}$ with $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$ denote the training data set. Suppose f is a linear function such that

$$f(x) = \omega^\top x + b \tag{19}$$

where $\omega \in \mathbb{R}^p$ and $b \in \mathbb{R}$. Then, SVR aims to find an approximation of f that deviates at most by ϵ from the observed target y while being as flat as possible (i.e., in the sense that weights in ω are small). This translates into the following convex optimization problem:

$$
\begin{aligned}
&\text{minimize} \quad \frac{1}{2}\|\omega\|^2 \\
&\text{subject to} \quad \begin{cases} y_i - \omega^\top x_i - b \le \epsilon \\ \omega^\top x_i + b - y_i \le \epsilon. \end{cases}
\end{aligned} \tag{20}
$$

In view that (20) might lack a feasible solution, Vapnik (1995) introduces an ϵ-insensitive loss function:

$$
L_\epsilon \{y - f(x)\} = \begin{cases} 0 & \text{if } |y - f(x)| \le \epsilon \\ |y - f(x)| - \epsilon & \text{otherwise.} \end{cases} \tag{21}
$$

To measure empirical loss (and thus model fit) using (21) Vapnik (1995) reformulates (20) using slack variables ζ and ζ^* that capture losses above and below the ϵ-tube around $f(x)$, respectively. Figure 1 depicts this approach. Only points outside the gray shaded ϵ-tube contribute linearly to the loss function.

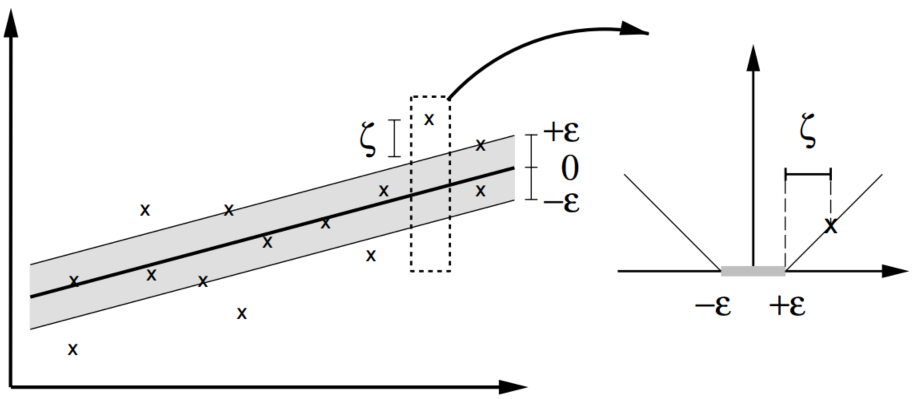


Figure 1: The ϵ-insensitive loss function of the SVR algorithm. Slack variable ζ captures the loss above the ϵ-tube. Points within the grey shaded area have no impact on the loss. In contrast, all other observations contribute linearly to the loss. Source: Smola and Schölkopf (2004).

Integrating the slack variables ζ and ζ^* into (20), the task to estimate a SVR model is equivalent to solving:

$$
\begin{aligned}
&\text{minimize} \quad \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) \\
&\text{subject to} \quad \begin{cases} y_i - \omega^\top x_i - b \le \epsilon + \zeta_i \\ \omega^\top x_i + b - y_i \le \epsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \ge 0, \end{cases}
\end{aligned} \tag{22}
$$

where $C > 0$ is a regularization parameter to balance between model fit and complexity (e.g., Hastie et al., 2009). Larger (smaller) values of C put more (less) weight on maximizing model fit during SVR learning.

To capture nonlinear relationships between covariates and the response variable, SVR maps the input data into a higher dimensional feature space. The linear regression is then constructed in the transformed space, which corresponds to a nonlinear regression in the input space. The transformation is feasible from a computational point of view because SVR calculates the mapping by implicitly using a kernel function $k(x_i^\top x) = \Theta(x^\top)\Theta(x)$. To implement this approach, it is common practice to estimate a SVR model through solving the dual of (22), which is given as (e.g., Smola and Schölkopf, 2004):

$$\begin{aligned} &\text{maximize} -\frac{1}{2}\sum_{i,j=1}^{n}(\rho_i-\rho_i^*)(\rho_j-\rho_j^*)x_i^\top x_j - \epsilon\sum_{i=1}^{n}(\rho_i+\rho_i^*) + \sum_{i=1}^{n} y_i(\rho_i-\rho_i^*) \\ &\text{subject to} \begin{cases} \sum_{i=1}^{n}(\rho_i-\rho_i^*) = 0 \\ \rho_i, \rho_i^* \in [0, C]. \end{cases} \end{aligned} \tag{23}$$

The dual program (23) includes the input data only in the form of scalar products $x_i^\top x_j$. Replacing the scalar product by means of a kernel function is thus straightforward and does not affect the solver. In this work, we employ the Gaussian radial basis function (RBF) kernel (24) which is defined as

$$k(x_i^\top x) = \exp\left(\frac{\|x - x_i\|}{2\gamma^2}\right) \tag{24}$$

where the meta-parameter $\gamma > 0$ governs the width of the Gaussian function and needs to be set by the modeler. The RBF kernel is used because it includes other kernels as special cases, possesses numerical advantages compared to alternatives, and often performs well in practical applications (Keerthi and Lin, 2003).

In order to construct the regression function (19), the weight vector ω is represented as a linear combination of observations in the training set:

$$\omega = \sum_{i=1}^{n}(\rho_i - \rho_i^*)x_i. \tag{25}$$

More specifically, for observations x_i where $f(x_i)$ is within the ϵ-tube holds that $\rho_i = \rho_i^* = 0$. Consequently, $f(x)$ depends only on the observations outside the ϵ-tube. These x_i are called support vectors. Accordingly, (25) is also called the support vector expansion of ω. Rewriting the regression function in terms of the support vector expansion gives the SVR forecasting model:

$$f(x) = \sum_{i=1}^{n}(\rho_i - \rho_i^*)x_i^\top x + b. \tag{26}$$

In the nonlinear case, the scalar product in (26) is once again replaced by a kernel function.

2.6. SVR-GARCH-KDE Hybrid

In the following section, we introduce a nonlinear GARCH hybrid to forecast VaR based on a combination of SVR and KDE. Subsequently, we elaborate on the estimation of the corresponding forecasting model.

We assume the distribution of the return series r_t to belong to the location-scale class, such that:

$$r_t = \mu_t + u_t = \mu_t + \sigma_t z_t, \qquad z_t \sim (0,1) \text{ i.i.d.} \tag{27}$$

Consider an ARMA structure for the mean model where the only assumption about the error distribution is a zero mean and a finite variance. In addition, recall equation (9), which shows that GARCH processes can also be given an ARMA representation. This leads to the following mean and variance model:

$$r_t = c + \sum_{i=1}^{s} \alpha_i r_{t-i} + \sum_{j=1}^{d} \phi_j u_{t-j} + u_t, \quad u_t \sim (0, \sigma_t^2) \tag{28}$$

$$u_t^2 = \omega + \sum_{i=1}^{\max(p,q)} (\delta_i + \beta_i) u_{t-i}^2 - \sum_{j=1}^{p} \beta_j \nu_{t-j} + \nu_t. \tag{29}$$

Let $e = \max(p,q)$, $\mathbf{r}_{t,s} = (r_{t-1}, r_{t-2}, \ldots, r_{t-s})$, $\mathbf{u}_{t,k} = (u_{t-1}, u_{t-2}, \ldots, u_{t-k})$ and $\boldsymbol{\nu}_{t,p} = (\nu_{t-1}, \nu_{t-2}, \ldots, \nu_{t-p})$. Then, following Chen et al. (2010), we introduce the nonlinear and nonparametric functions h and g such that the conditional mean and variance models of r_t are

$$r_t = h(\mathbf{r_{t,s}}, \mathbf{u_{t,d}}) + u_t \qquad u_t \sim (0, \sigma_t^2) \tag{30}$$

$$u_t^2 = g(\mathbf{u_{t,e}}, \boldsymbol{\nu_{t,p}}) + \nu_t \qquad \nu_t \sim WN(0, a_t^2) \tag{31}$$

where $WN(0, a_t^2)$ denotes white noise with expectation zero and variance a_t^2. We propose to estimate $h(\cdot)$ and $g(\cdot)$ using SVR. The estimates for μ_t and σ_t in (27) are then obtained as:

$$\widehat{\mu}_t = h(\mathbf{r_{t,s}}, \mathbf{u_{t,d}}) \tag{32}$$

$$\widehat{\sigma}_t = \sqrt{g(\mathbf{u_{t,e}}, \boldsymbol{\nu_{t,p}})}. \tag{33}$$

By defining the estimated residuals as $\widehat{u}_t = r_t - \widehat{\mu}_t$, estimates of z_t are obtained as

$$\widehat{z}_t = \frac{\widehat{u}_t}{\widehat{\sigma}_t}. \tag{34}$$

Then, for $\widehat{Q}_{\widehat{z}}(\alpha)$ being the estimated quantile function of z, the VaR estimate for r_t is:

$$\widehat{VaR}_t^{\alpha} = -\left\{ h(\mathbf{r_{t,s}}, \mathbf{u_{t,d}}) + \sqrt{g(\mathbf{u_{t,e}}, \boldsymbol{\nu_{t,p}})}\widehat{Q}_{\widehat{z}}(\alpha) \right\}. \tag{35}$$

whereby we estimate $\widehat{Q}_{\widehat{z}}(\alpha)$ using KDE.

We now present a procedure to estimate VaR as in (35) and describe it in the context of producing one-day-ahead VaR forecasts. A step-by-step overview is given in Algorithm 1. Let $\{r_t\}_{t=1}^T$ be the training set consisting of the daily returns from a portfolio where r_T is the most recent observation. In the first step, we model the mean process (30). To do this, we estimate an AR(s) model using SVR to obtain the estimated returns $\{\widehat{r}_t\}_{t=1+s}^T$. The set of estimated residuals $\{\widehat{u}_t\}_{t=1+s}^T$ is derived as $\widehat{u}_t = r_t - \widehat{r}_t$. Then, a moving average (MA) part can be introduced to the model such that r_t can be modeled as an ARMA(s,d) process by running SVR and including $\widehat{u}_t$. The sets of estimated returns and residuals from the ARMA(s,d) model are denoted as $\{\widehat{r}_t^*\}_{t=1+s+d}^T$ and $\{\widehat{u}_t^*\}_{t=1+s+d}^T$, respectively.

We also estimate the variance process in a two step approach and start by fitting the squared mean model residuals $\{\widehat{u}_t^{*2}\}_{t=1+s+d}^T$ in the way of an AR(e) process with SVR. Based on this, fitted variances $\{\widehat{\sigma}_t^2\}_{t=1+s+d+e}^T$ are obtained. Then, an ARMA model for (31) is obtained in the same way as for the mean process by using the estimated model residuals $\{\widehat{\nu}_t\}_{t=1+s+d+e}^T$ where $\widehat{\nu}_t = \widehat{u}_t^{*2} - \widehat{\sigma}_t^2$. Consequently, the final set of fitted variances is denoted as $\{\widehat{\sigma}_t^{*2}\}_{t=1+s+d+e+p}^T$. No assumptions are made about the starting values of the residuals. Hence, the final estimation of (31) is done using data for $T - s - d - e - p$ time points. Since SVR is applied without introducing further restrictions, it is not ensured that $\widehat{\sigma}_t^2$ and $\widehat{\sigma}_t^{*2}$ are positive. Therefore, if the SVR estimate is $\widehat{\sigma}_t^{(*)2} \leq 0$, it will be replaced by the last positive estimated variance. In case the first fitted variance is negative, it will be replaced by the first squared residual from the final mean model.

The set of estimated standardized residuals $\{\widehat{z}_t\}_{t=1+s+d+e+p}^T$ can be computed by applying (34). However, $\widehat{z}_t$ does not necessarily have zero mean and unit variance. Hence, we perform the quantile estimation using scaled standardized residuals $\widehat{z}_t^*$:

$$\widehat{z}_t^* = \frac{\widehat{z}_t - \overline{\widehat{z}_t}}{\sqrt{\frac{1}{T-1}\sum_{i=1}^T (\widehat{z}_t - \overline{\widehat{z}_t})^2}} \tag{36}$$

where $\overline{\widehat{z}_t}$ denotes the empirical mean of $\widehat{z}_t$. The forecasted mean $\widehat{\mu}_{T+1}$ and standard deviation $\widehat{\sigma}_{T+1}$ are obtained from the mean and variance model. Finally, we use KDE to estimate the α-quantile of $\{\widehat{z}_t^*\}_{t=1+s+d+e+p}^T$. Then, the one-day-ahead VaR forecast is:

$$\widehat{VaR}_{T+1}^{\alpha} = -[\widehat{\mu}_{T+1} + \widehat{\sigma}_{T+1}\widehat{Q}_{\widehat{z}^*}(\alpha)]. \tag{37}$$

Algorithm 1 SVR-GRACH-KDE Estimation Algorithm for Forecasting VaR

1: AR(s) model for $\{r_t\}_{t=1}^{T}$ using SVR
2: Get errors from Step 1 $\{\widehat{u}\}_{t=1+s}^{T}$
3: ARMA(s,d) model for $\{r_t\}_{t=1+s}^{T}$ with results from Step 2 using SVR
4: Get errors from Step 3 $\{\widehat{u}^{*}\}_{t=1+s+d}^{T}$
5: AR(e) model for $\{\widehat{u}_t^{*2}\}_{t=1+s+d}^{T}$ using SVR
6: Get errors from Step 5 $\{\widehat{\nu}_t\}_{t=1+s+d+e}^{T}$
7: ARMA(e,p) model for $\{\widehat{u}_t^{*2}\}_{t=1+s+d+e}^{T}$ with results from Step 6 using SVR
8: Obtain volatility estimates $\{\widehat{\sigma}^{*}\}_{t=1+s+d+e+p}^{T}$ from Step 7
9: Get standardized residuals $\widehat{z}_t = \widehat{u}_t^{*}/\widehat{\sigma}_t$ for $t = 1+s+d+e+p, \ldots, T$
10: Scale $\{\widehat{z}_t\}_{t=1+s+d+e+p}^{T}$ to zero mean and unit variance and obtain $\{\widehat{z}_t^{*}\}_{t=1+s+d+e+p}^{T}$
11: Estimate the α-quantile $\widehat{Q}_{\widehat{z}^{*}}(\alpha)$ with KDE
12: Obtain $\widehat{r}_{T+1}$ and $\widehat{\sigma}_{T+1}$ by using the models from Step 3 and 7
13: VaR forecast: $\widehat{VaR}_{T+1}^{\alpha} = -[\widehat{\mu}_{T+1} + \widehat{\sigma}_{T+1}\widehat{Q}_{\widehat{z}^{*}}(\alpha)]$

3. Empirical Study

3.1. General Setting

The SVR-GARCH-KDE hybrid is tested using stock indices to evaluate the performance for different regions. We consider three indices, namely the Euro STOXX 50, S&P 500 and Nikkei 225 which represent the Euro zone, the USA and Japan, respectively. The analysis is based on the log-returns of the adjusted index closing prices P_t:

$$r_t = \log(P_t) - \log(P_{t-1}). \tag{38}$$

We forecast VaR for the quantiles $\alpha \in \{0.01, 0.025, 0.05\}$, always considering a forecast horizon of one trading day. In empirical applications, the quality of SVR depends on the kernel and parameter values which need to be set manually. The prevailing approach to determine parameter settings is grid search (e.g., Lessmann and Voß, 2017), which we also apply in this study. For the density estimation via KDE, the Gaussian kernel function in combination with Silverman's rule of thumb are used to reduce computational cost.

The evaluation of the models is based on Christoffersen (1998) who proposes a likelihood ratio (LR) test framework, which assesses the unconditional and conditional coverage as well as the independence of VaR exceedances. Moreover, Christoffersen (1998) shows that the test statistic for conditional coverage can be derived as the sum of the test statistics of the test for unconditional coverage and independence of VaR exceedances. Hence, it is possible to test whether the performance of a VaR model in terms of conditional coverage is determined by its ability to achieve correct unconditional coverage or adjust for changing volatility. This is useful in situations where the model has relatively bad conditional coverage but only one of the test statistics for unconditional coverage or independence is small. As criterion for selecting a model from grid search and evaluate the performance of the SVR-GARCH-KDE hybrid with respect to benchmark models, the p-value of the test for conditional coverage is used. Since the null hypothesis corresponds to correct conditional coverage which is the desired property, the one with the highest p-value is considered to be the best.

We perform all analyses using the statistical software R. The data has been downloaded from Yahoo Finance using the quantmod package. For SVR, we use the package e1071, which is the R implementation of the LIBSVM library of Chang and Lin (2011). To reduce computational time, we employ the doParallel package for parallelization of computations. The benchmark methods introduced below are implemented by using the rugarch package. All codes are available on `www.quantlet.de`. For details we refer to Borke and Härdle (2016) and Borke and Härdle (2017).

3.2. Benchmark Methods

To test the SVR-GARCH-KDE hybrid, we compare its performance to the standard GARCH model and two of its variations. In particular, Franke et al. (2015) state that the most important variations are the EGARCH and TGARCH model. Hence, they serve as benchmarks in the empirical comparison. The EGARCH and TGARCH models are introduced by Nelson (1991) and Zakoian (1994), respectively. In contrast to standard GARCH models, both can account for asymmetric behavior with respect to past positive or negative returns. The two main differences between EGARCH and TGARCH models are that the former has a multiplicative and the latter an additive model structure. Moreover, TGARCH models allow for different coefficients depending on the lags whereas EGARCH models capture the asymmetric behavior for all lags with one coefficient. The GARCH-type models that Kuester et al. (2006) analyze are coupled with different error distributions, i.e. the normal distribution, t-distribution and skewed t-distribution. We adopt this approach, which implies that we compare the SVR-GARCH-KDE hybrid to nine benchmarks.

3.3. Results

3.3.1. Model Setting and Tuning

We assume the mean process of r_t in (32) is zero. Moreover, we assume the variance process to have one AR and one MA part. These assumptions are imposed on both the SVR-GARCH-KDE hybrid and the benchmark models. We then forecast VaR for every index trading day from 2011-07-01 until 2016-06-30. The data is scaled to zero mean and unit variance in the SVR step; as suggested in the documentation of the e1071 package. Model training is done in a moving window approach using 251 return observations, which corresponds to approximately one trading year. Tuning is done using the same moving window approach for forecasting VaR from 2006-07-01 until 2011-06-30. The considered parameter values in the grid for SVR are

- $C \in \{10^{-4}, 10^{-3}, \ldots, 10^{4}\}$
- $\psi \in \{0, 0.1, \ldots, 0.9\}$ where $\epsilon = Q_{u^2_{scale}}(\psi)$
- $\gamma \in \{10^{-4}, 10^{-3}, \ldots, 10^{4}\}$.

Note that the second point indicates that tuning is not done over fixed values of ϵ. Instead, in every step of the estimation, ϵ is determined based on the ψ-quantile of the squared scaled disturbances of the mean model. This corresponds to the squared scaled returns because we assume zero mean of r_t. The motivation behind this is the tendency of returns to form volatility clusters. Hence, a fixed ϵ can lead to good results in one volatility regime but might have a poor performance after a regime change. For instance, in the case of a financial crisis, the right tail of the distribution of past volatilities gets thicker. Hence, an ϵ that depends on the quantile of the distribution will increase such that large volatilities have automatically a higher influence on the estimated parameters from the SVR optimization. By using squared values it is ensured that only positive values are obtained for ϵ. However, notice that the distribution of the scaled squared disturbances, which are used in the SVR training, is shifted to the left of the distribution of the squared scaled disturbances. Hence, it is possible that for high values of ψ no observations are outside the ϵ-tube such that the model cannot be estimated.

The parameter settings that resulted in the best models for the SVR-GARCH-KDE hybrid during the tuning period are shown in Table 2. It can be seen that especially for ψ only a certain range appears among the best models. Moreover, the optimal ψ tends to be higher for lower quantiles. Based on the obtained parameters, VaR forecasts are produced from 2011-07-01 until 2016-06-30.

3.3.2. Model Comparison

The results of the simulations are presented for each quantile separately in the Tables 3, 4 and 5 at the end of the section. To clearly identify the best performing models, every table is sorted in descending order for every index according to the p-value of the conditional coverage test. The abbreviations NORM, STD and SSTD indicate the normal, t- and skewed t-distribution, respectively. Additionally, the column headers UC, ID and CC refer to the corresponding p-value of the LR test for unconditional coverage, independence of violations and conditional coverage.

Model Evaluation for $\boldsymbol{\alpha = 0.01}$. The SVR-GARCH-KDE hybrid is the best model for the Euro STOXX 50 for $\alpha = 0.01$. A visualization of its performance is given in Figure 2. It is compared to the EGARCH-NORM model, which ranked worst. Here, the SVR-GARCH-KDE hybrid estimates in general higher values for VaR than the EGARCH-NORM model and exhibits more variability. For the S&P 500 and Nikkei 225 the SVR-GARCH-KDE hybrid outperforms all models that are coupled with a normal distribution. However, all models using a skewed t-distribution perform better. Especially for the Nikkei 225 this is caused by having low unconditional coverage due to risk overestimation. In general, the models coupled with a normal distribution perform poorly for all indices. This comes as no surprise since the distribution of asset returns is usually leptokurtic.

Model Evaluation for $\boldsymbol{\alpha = 0.025}$. The performance of the SVR-GARCH-KDE hybrid at the 2.5% level is not as good as for $\alpha = 0.01$ for the Euro STOXX 50. We observe the lowest p-value in the test for independence of violations but the third best regarding the test for unconditional coverage. Interestingly, although the TGARCH and EGARCH model account for asymmetries in volatility, the former is the best and the latter the worst variance model. A relatively high risk overestimation for the S&P 500 and Nikkei 225 causes the SVR-GARCH-KDE hybrid to be on the sixth and ninth rank, respectively.

Model Evaluation for $\boldsymbol{\alpha = 0.05}$. The best performance of the SVR-GARCH-KDE hybrid for $\alpha = 0.05$ is rank two for the Nikkei 225. Here, it is only beaten by the EGARCH-SSTD model. For the other indices the SVR-GARCH-KDE hybrid ranks on place five. This is caused by having relatively low p-values for the ID test. In terms of UC, the SVR-GARCH-KDE hybrid is the second and third best model for the S&P 500 and Nikkei 225, respectively. In comparison to the results for $\alpha = 0.01$, the models with a normal distribution show a better performance for $\alpha \in \{0.025, 0.05\}$. However, using the skewed t-distribution leads also for $\alpha \in \{0.025, 0.05\}$ to the best rankings.

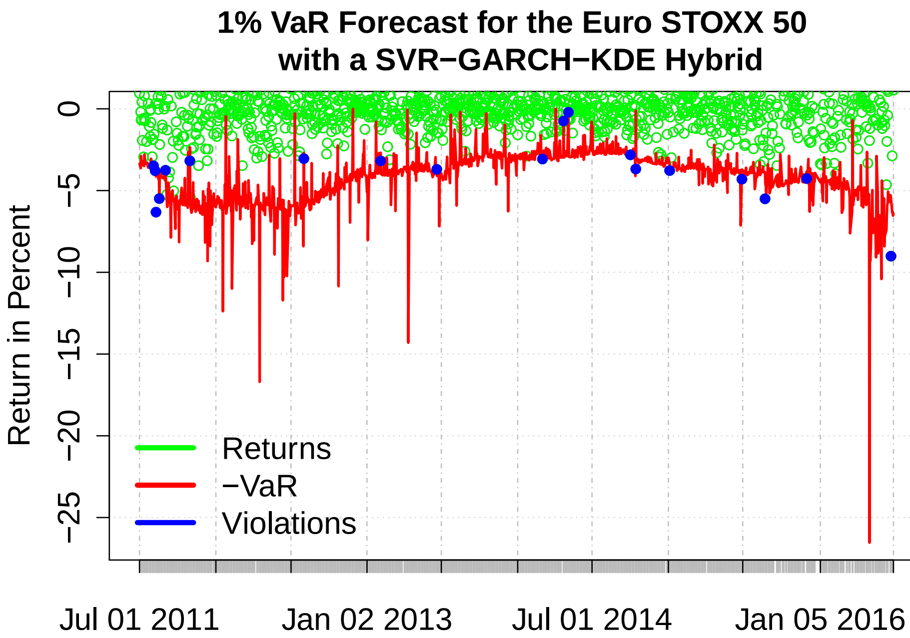
Evaluation Summary. Summarizing the results observed across all indices and quantiles, we conclude that the SVR-GARCH-KDE hybrid displays a competitive performance. This can be seen in Table 1 where the mean ranks per index and quantile are presented for each model. The SVR-GARCH-KDE hybrid is the third best model. Benchmark models coupled with a normal distribution are usually outperformed by the SVR-GARCH-KDE hybrid. Additionally, there is no setting where the models using a normal distribution perform best. This provides further evidence that using the normal distribution to measure market risk is inappropriate. However, models with a skewed t-distribution show in many cases the best performance. For instance, the TGARCH model coupled with the skewed t-distribution is always among the top three. This confirms the results of previous research showing usually skewed return distributions. Unlike the benchmarks, the SVR-GARCH-KDE hybrid tends to overestimate market risk. This might come from the choice of time interval for SVR parameter tuning. In particular, the tuning period covers the financial crisis of 2008 where market risk was extremely high. However, in the context of risk management, risk underestimation is more critical than risk overestimation because it can lead to bankruptcy in the short term. For instance, assume a hypothetical situation with the goal to forecast the 5% VaR, where the SVR-GARCH-KDE hybrid and a benchmark model have the same p-value regarding the independence test. Additionally, assume the p-value of the benchmark in the test of conditional coverage is higher because it has an unconditional coverage of 5.5% whereas that of the SVR-GARCH-KDE hybrid is 4%. The 1% overestimation of the SVR-GARCH-KDE hybrid works like a buffer for model risk since all estimation techniques exhibit statistical uncertainty. Hence, the use of the SVR-GARCH-KDE hybrid may be still more appealing than the use of benchmark models that tend to underestimate risk.

Model	**Mean Rank**
TGARCH-SSTD	2.1
GARCH-SSTD	2.1
SVR-GARCH-KDE	4.9
EGARCH-SSTD	5.0
TGARCH-STD	5.3
TGARCH-NORM	5.9
GARCH-STD	6.0
GARCH-NORM	6.0
EGARCH-NORM	8.6
EGARCH-STD	8.7

Table 1: The models are presented with their mean rank. The mean rank was computed per index and quantile using the ties method *min* in the *frankv* function of the R package data.table. The lower the mean rank, the better the model.

C	ψ	γ	**Index**	**Quantile**	**Violations**	**UC**	**ID**	**CC**
10	0.7	0.1	S&P500	1.0	0.95	86.62	63.08	87.84
10	0.7	0.01	S&P500	2.5	2.46	93.15	96.55	96.20
10	0.6	0.001	S&P500	5.0	5.00	99.48	99.57	99.57
100	0.8	0.01	Nikkei225	1.0	0.82	50.63	68.48	73.85
100	0.7	0.01	Nikkei225	2.5	2.70	66.43	57.80	52.61
100	0.7	0.10	Nikkei225	5.0	4.49	40.84	94.38	67.06
100	0.7	0.01	EuroStoxx50	1.0	1.02	93.28	60.41	87.11
0.1	0.6	0.001	EuroStoxx50	2.5	2.52	96.42	97.76	97.66
10000	0.6	10	EuroStoxx50	5.0	4.96	94.86	99.71	99.50

Table 2: The best models in the tuning period according to the p-value of the test for conditional coverage. UC, ID and CC indicate the p-value of the corresponding LR test. All values in the columns Quantile, Violations, UC, ID and CC are given in percent.



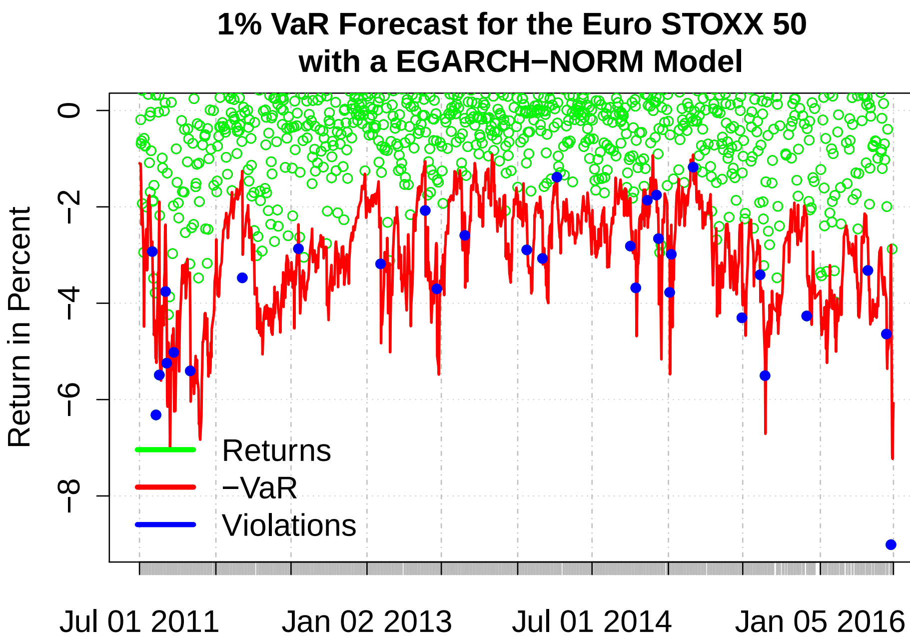


Figure 2: VaR forecast model comparison for the Euro STOXX 50 at $\alpha = 0.01$ in the period from July 1, 2011 to June 30, 2016. The upper panel shows VaR forecasts and violations for the SVR-GARCH-KDE hybrid and the lower panel for the EGARCH-NORM model.

Model	Index	Violations	UC	ID	CC
EGARCH-SSTD	S&P500	1.11	69.27	57.46	79.01
TGARCH-SSTD	S&P500	0.72	28.52	71.87	52.94
GARCH-SSTD	S&P500	1.27	35.24	43.34	28.13
TGARCH-STD	S&P500	1.43	14.91	46.97	27.20
SVR-GARCH-KDE	S&P500	0.79	44.83	18.28	13.71
GARCH-STD	S&P500	1.67	2.95	65.77	6.15
EGARCH-STD	S&P500	1.75	1.58	37.62	3.67
TGARCH-NORM	S&P500	2.15	0.04	27.64	0.10
EGARCH-NORM	S&P500	2.23	0.02	25.88	0.04
GARCH-NORM	S&P500	2.38	0.00	94.85	0.01
TGARCH-SSTD	Nikkei225	1.21	47.61	54.48	64.58
GARCH-SSTD	Nikkei225	1.21	47.61	54.48	64.58
EGARCH-SSTD	Nikkei225	1.37	21.61	49.21	36.76
GARCH-STD	Nikkei225	1.69	2.60	66.33	5.56
TGARCH-STD	Nikkei225	1.77	1.37	37.30	3.23
SVR-GARCH-KDE	Nikkei225	0.24	0.13	90.41	0.55
EGARCH-STD	Nikkei225	2.09	0.07	29.17	0.19
GARCH-NORM	Nikkei225	2.17	0.03	88.16	0.14
TGARCH-NORM	Nikkei225	2.33	0.01	93.19	0.03
EGARCH-NORM	Nikkei225	2.42	0.00	95.21	0.01
SVR-GARCH-KDE	EuroStoxx50	1.53	8.17	44.23	16.36
GARCH-SSTD	EuroStoxx50	1.69	2.60	39.54	5.84
TGARCH-SSTD	EuroStoxx50	1.69	2.60	66.33	5.56
TGARCH-STD	EuroStoxx50	1.77	1.37	70.53	3.39
GARCH-STD	EuroStoxx50	1.85	0.70	35.15	1.70
TGARCH-NORM	EuroStoxx50	2.09	0.07	85.17	0.28
GARCH-NORM	EuroStoxx50	2.17	0.03	27.33	0.08
EGARCH-STD	EuroStoxx50	2.33	0.01	93.19	0.03
EGARCH-SSTD	EuroStoxx50	2.42	0.00	95.21	0.01
EGARCH-NORM	EuroStoxx50	2.50	0.00	96.87	0.00

Table 3: Results for the VaR forecasts from July 1, 2011 to June 30, 2016 for $\alpha = 0.01$. UC, ID and CC indicate the p-value of the corresponding LR test. The results are in descending order with respect to CC for each index. All values in the columns Quantile, Violations, UC, ID and CC are given in percent.

Model	Index	Violations	UC	ID	CC
GARCH-SSTD	S&P500	2.62	78.12	98.97	95.22
EGARCH-SSTD	S&P500	3.02	25.18	98.94	51.31
TGARCH-SSTD	S&P500	2.38	79.19	22.60	46.40
GARCH-STD	S&P500	3.50	3.24	93.45	9.47
GARCH-NORM	S&P500	3.82	0.55	99.2	2.10
SVR-GARCH-KDE	S&P500	1.43	0.83	52.45	1.62
TGARCH-STD	S&P500	3.90	0.33	75.52	1.02
TGARCH-NORM	S&P500	4.05	0.12	69.37	0.36
EGARCH-STD	S&P500	4.37	0.01	96.11	0.06
EGARCH-NORM	S&P500	4.53	0.00	92.48	0.02
GARCH-SSTD	Nikkei225	2.90	38.01	99.90	67.96
TGARCH-SSTD	Nikkei225	2.42	84.78	22.29	46.72
TGARCH-STD	Nikkei225	3.14	16.44	97.67	37.15
TGARCH-NORM	Nikkei225	3.22	11.91	96.33	28.60
EGARCH-NORM	Nikkei225	3.22	11.91	96.33	28.60
GARCH-NORM	Nikkei225	3.38	5.87	92.85	15.55
GARCH-STD	Nikkei225	3.38	5.87	89.33	14.96
EGARCH-SSTD	Nikkei225	3.06	22.21	12.14	14.30
SVR-GARCH-KDE	Nikkei225	1.69	5.26	14.27	2.18
EGARCH-STD	Nikkei225	3.86	0.43	99.42	1.68
TGARCH-SSTD	EuroStoxx50	2.98	29.36	72.9	42.00
TGARCH-NORM	EuroStoxx50	3.22	11.91	83.46	24.78
TGARCH-STD	EuroStoxx50	3.38	5.87	89.33	14.96
GARCH-SSTD	EuroStoxx50	3.46	3.99	91.84	11.13
SVR-GARCH-KDE	EuroStoxx50	3.14	16.44	11.43	4.35
GARCH-STD	EuroStoxx50	3.70	1.11	97.42	3.88
GARCH-NORM	EuroStoxx50	3.95	0.26	75.92	0.81
EGARCH-STD	EuroStoxx50	4.35	0.02	91.29	0.07
EGARCH-SSTD	EuroStoxx50	4.51	0.00	95.50	0.02
EGARCH-NORM	EuroStoxx50	4.59	0.00	97.11	0.01

Table 4: Results for the VaR forecasts from July 1, 2011 to June 30, 2016 for $\alpha = 0.025$. UC, ID and CC indicate the p-value of the corresponding LR test. The results are in descending order with respect to CC for each index. All values in the columns Quantile, Violations, UC, ID and CC are given in percent.

Model	**Index**	**Violations**	**UC**	**ID**	**CC**
GARCH-SSTD	S&P500	5.01	98.97	89.07	89.06
GARCH-NORM	S&P500	5.72	24.94	99.79	51.40
TGARCH-SSTD	S&P500	5.41	51.47	60.38	48.83
GARCH-STD	S&P500	5.80	20.21	99.23	43.99
SVR-GARCH-KDE	S&P500	4.37	29.68	30.25	17.55
EGARCH-SSTD	S&P500	6.44	2.47	83.92	6.73
TGARCH-NORM	S&P500	6.60	1.30	77.88	3.56
EGARCH-NORM	S&P500	6.76	0.65	94.35	2.33
TGARCH-STD	S&P500	6.76	0.65	71.36	1.77
EGARCH-STD	S&P500	7.00	0.21	87.46	0.78
EGARCH-SSTD	Nikkei225	5.23	70.78	97.26	90.66
SVR-GARCH-KDE	Nikkei225	4.59	50.1	97.11	77.43
TGARCH-SSTD	Nikkei225	4.67	58.95	89.41	77.30
GARCH-SSTD	Nikkei225	5.07	90.69	75.29	74.78
TGARCH-NORM	Nikkei225	5.23	70.78	68.85	64.18
TGARCH-STD	Nikkei225	5.56	37.71	89.62	60.67
GARCH-NORM	Nikkei225	5.88	16.68	78.41	30.16
EGARCH-NORM	Nikkei225	5.72	25.68	49.18	25.85
GARCH-STD	Nikkei225	6.04	10.34	71.98	19.11
EGARCH-STD	Nikkei225	6.04	10.34	71.98	19.11
TGARCH-SSTD	EuroStoxx50	5.15	80.55	65.27	63.32
GARCH-SSTD	EuroStoxx50	5.96	13.21	95.83	30.84
GARCH-NORM	EuroStoxx50	5.96	13.21	95.83	30.84
TGARCH-NORM	EuroStoxx50	6.04	10.34	78.10	20.73
SVR-GARCH-KDE	EuroStoxx50	5.48	44.90	26.86	20.17
EGARCH-SSTD	EuroStoxx50	6.12	7.99	68.65	14.82
TGARCH-STD	EuroStoxx50	6.20	6.11	59.29	10.26
EGARCH-NORM	EuroStoxx50	6.52	1.85	99.10	6.18
GARCH-STD	EuroStoxx50	6.60	1.33	96.58	4.52
EGARCH-STD	EuroStoxx50	6.84	0.46	69.51	1.27

Table 5: Results for the VaR forecasts from July 1, 2011 to June 30, 2016 for $\alpha = 0.05$. UC, ID and CC indicate the p-value of the corresponding LR test. The results are in descending order with respect to CC for each index. All values in the columns Quantile, Violations, UC, ID and CC are given in percent.

4. Conclusion

In a large-scale empirical comparison Kuester et al. (2006) find VaR models belonging to the location-scale class superior to alternative approaches. However, the location-scale models considered in their study are parametric and based on distributional assumptions. Motivated by the potential shortcomings of a parametric approach, the paper introduces a nonparametric and nonlinear VaR forecasting framework based on the location-scale class. The mean and volatility model are modeled with SVR in an ARMA and GARCH like fashion, respectively. In addition, the VaR forecast is obtained by estimating the distribution function of the standardized residuals via KDE.

To evaluate the performance of the SVR-GARCH-KDE hybrid, VaR is forecasted for three indices: Euro STOXX 50, Nikkei 225 and S&P 500, considering the different quantiles of $\alpha \in \{0.01, 0.025, 0.05\}$. GARCH, EGARCH and TGARCH models coupled with the normal, t- and skewed t-distribution serve as benchmarks and are compared to the proposed model using a LR testing framework for interval forecasts (Christoffersen, 1998). Grid search with the goal to maximize the p-value of the LR test for conditional coverage is used to set the SVR parameters. The SVR-GARCH-KDE hybrid delivers competitive results. For instance, it is the best model for the Euro STOXX 50 for $\alpha = 0.01$ and the third best model overall. The TGARCH and GARCH model in combination with the skewed t-distribution show the best results. In contrast to the benchmark models, which usually underestimate risk, the SVR-GARCH-KDE tends to overestimate risk. This can lead to situations where the SVR-GARCH-KDE hybrid has an average performance regarding the used model evaluation criterion. However, with respect to risk management, the use of the SVR-GARCH-KDE might be still favorable over using the benchmarks since all approaches exhibit statistical uncertainty. The tendency to overestimate risk can, therefore, serve as a model risk buffer.

In general, the competitive results indicate that the proposed SVR-GARCH-KDE hybrid is a promising alternative. This conclusion follows from considering the applied tuning routine. There exist several ways that can lead to an improved performance. First, tuning could be done for more parameters. For instance, in the KDE part of the estimation procedure, the kernel function and bandwidth estimator are set without tuning. Hence, considering different kernel functions and more flexible bandwidth estimators are potential ways to improve the performance further. Moreover, the kernel in the SVR part is also fixed and could be varied. Second, more recent information could be used in the parameter selection by re-tuning the model. Here, tuning is done for a block of five years of data. Then, based on the optimal parameters found for this data block, one-day-ahead forecasts for five years are made and the parameters are held fixed. Thus, annual or even shorter re-tuning periods could result in parameters that are more appropriate for the existing market risk. Additionally, refining the grid can also result in better parameter choices.

In addition to modifying the tuning routine, the SVR-GARCH-KDE hybrid could be improved by changing the model specification. Overall, the TGARCH model with the skewed t-distribution achieves very good results. Hence, the SVR-GARCH-KDE hybrid could be modified such that it also accounts for asymmetric reactions of the volatility to past returns in a TGARCH like manner. Moreover, the proposed procedure does not ensure that the estimated variances are positive. Here, this problem is handled by replacing non-positive estimates with the last positive. However, positive variance estimates could be ensured by modeling the logarithm of the squared mean model residuals instead.

All above mentioned adjustments are potential starting points for future research to further improve the proposed framework. However, although the suggested modifications of the tuning procedure are reasonable approaches to improve the model performance, they also increase the computational complexity. After all, this is a slight drawback of the SVR-GARCH-KDE hybrid in comparison to standard models. It is, however offset by potential performance gains.

Alexander, C., 2008. Value-at-Risk Models. Vol. 4 of Market Risk Analysis. Wiley, Chichester.

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.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31 (3), 307–327.

Borke, L., Härdle, W. K., 2016. Q3-D3-LSA (November 17, 2016). SFB 649 Discussion Paper (2016-049). URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2871111

Borke, L., Härdle, W. K., 2017. GitHub API based QuantNet mining infrastructure in R (March 6, 2017). SFB 649 Discussion Paper (2017-008).
URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2927901

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2 (3), 27:1–27:27.

Chen, C. Y.-H., Chiang, T. C., Härdle, W. K., 2016. Downside risk and stock returns. Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, Submitted to Journal of Banking & Finance.

Chen, Q., Gerlach, R., Lu, Z., 2012. Bayesian value-at-risk and expected shortfall forecasting via the asymmetric Laplace distribution. Computational Statistics & Data Analysis 56 (11), 3498 – 3516, 1st issue of the Annals of Computational and Financial Econometrics Sixth Special Issue on Computational Econometrics.

Chen, S., Härdle, W. K., Jeong, K., 2010. Forecasting volatility with support vector machine-based GARCH model. Journal of Forecasting 29, 406–433.

Christoffersen, P. F., 1998. Evaluating interval forecasts. International Economic Review 39 (4), 841–862.

Dunis, C. L., Laws, J., Sermpinis, G., 2010. Modelling commodity value at risk with higher order neural networks. Applied Financial Economics 20 (7), 585–600.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 987–1007.

Franke, J., Diagne, M., 2006. Estimating market risk with neural networks. Statistics & Decisions 24 (2), 233–253.

Franke, J., Härdle, W. K., Hafner, C. M., 2015. Statistics of Financial Markets: An Introduction, 4th Edition. Universitext. Springer, Berlin, Heidelberg.
URL http://dx.doi.org/10.1007/978-3-642-16521-4

Härdle, W., Müller, M., Sperlich, S., Werwatz, A., 2004. Nonparametric and Semiparametric Models. Springer Series in Statistics. Springer, Berlin, Heidelberg.

Harvey, C. R., Siddique, A., 2000. Conditional skewness in asset pricing tests. The Journal of Finance 55 (3), 1263–1295.

Hastie, T., Tibshirani, R., Friedman, J. H., 2009. The elements of statistical learning: Data mining, inference, and prediction, 2nd Edition. Springer Series in Statistics. Springer, New York NY.

Keerthi, S. S., Lin, C.-J., 2003. Asymptotic behaviors of support vector machines with gaussian kernel. Neural Computation 15 (7), 1667–1689.

Khan, A. I., 2011. Modelling daily value-at-risk using realized volatility, non-linear support vector machine and ARCH type models. Journal for Economics and International Finance 3 (5), 305–321.

Kim, M., Lee, S., 2016. Nonlinear expectile regression with application to value-at-risk and expected shortfall estimation. Computational Statistics & Data Analysis 94 (Supplement C), 1–19.
URL http://www.sciencedirect.com/science/article/pii/S0167947315001681

Kraus, D., Czado, C., 2017. D-vine copula based quantile regression. Computational Statistics & Data Analysis 110 (Supplement C), 1–18.
URL http://www.sciencedirect.com/science/article/pii/S0167947316303073

Kuester, K., Mittnik, S., Paolella, M. S., 2006. Value-at-risk prediction: A comparison of alternative strategies. Journal of Financial Econometrics 4 (1), 53–89.

Lessmann, S., Voß, S., 2017. Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy. International Journal of Forecasting 33 (4), 864877.
URL https://doi.org/10.1016/j.ijforecast.2017.04.003

Malec, P., Schienle, M., 2014. Nonparametric kernel density estimation near the boundary. Computational Statistics & Data Analysis 72 (Supplement C), 57–76.
URL http://www.sciencedirect.com/science/article/pii/S0167947313003770

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59 (2), 347–370.

O. Radović, J. Stanković, J. Stanković, 2015. Tail risk assessment using support vector machine. Journal of Engineering Science and Technology Review 8 (1), 61–64.

Sarykalin, S., Serraino, G., Uryasev, S., 2008. Value-at-Risk vs. Conditional Value-at-Risk in Risk Management and Optimization. pp. 270–294.
URL http://pubsonline.informs.org/doi/abs/10.1287/educ.1080.0052

Schaumburg, J., 2012. Predicting extreme value at risk: Nonparametric quantile regression with refinements from extreme value theory. Computational Statistics & Data Analysis 56 (12), 4081 – 4096.

Silva, W., Kimura, H., Sobreiro, V. A., 2017. An analysis of the literature on systemic financial risk: A survey. Journal of Financial Stability 28 (Supplement C), 91–114.
URL http://www.sciencedirect.com/science/article/pii/S1572308916302285

Smola, A. J., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and Computing 14 (3), 199–222.

Vapnik, V. N., 1995. The Nature of Statistical Learning Theory. Springer, New York.
URL http://dx.doi.org/10.1007/978-1-4757-2440-0

Xu, Q., Jiang, C., He, Y., 2016. An exponentially weighted quantile regression via svm with application to estimating multiperiod VaR. Statistical Methods & Applications 25 (2), 285–320.
URL http://link.springer.com/content/pdf/10.1007%2Fs10260-015-0332-9.pdf

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

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

001 "Data Driven Value-at-Risk Forecasting using a SVR-GARCH-KDE Hybrid" by Marius Lux, Wolfgang Karl Härdle and Stefan Lessmann, January 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.