

Volatility forecasting via SVR–GARCH with mixture of Gaussian kernels

Pedro Correia S. Bezerra¹ · Pedro Henrique M. Albuquerque¹

Received: 20 April 2016 / Accepted: 1 November 2016 / Published online: 16 November 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract The support vector regression (SVR) is a supervised machine learning technique that has been successfully employed to forecast financial volatility. As the SVR is a kernel-based technique, the choice of the kernel has a great impact on its forecasting accuracy. Empirical results show that SVRs with hybrid kernels tend to beat single-kernel models in terms of forecasting accuracy. Nevertheless, no application of hybrid kernel SVR to financial volatility forecasting has been performed in previous researches. Given that the empirical evidence shows that the stock market oscillates between several possible regimes, in which the overall distribution of returns it is a mixture of normals, we attempt to find the optimal number of mixture of Gaussian kernels that improve the one-period-ahead volatility forecasting of SVR based on GARCH(1,1). The forecast performance of a mixture of one, two, three and four Gaussian kernels are evaluated on the daily returns of Nikkei and Ibovespa indexes and compared with SVR–GARCH with Morlet wavelet kernel, standard GARCH, Glosten–Jagannathan–Runkle (GJR) and nonlinear EGARCH models with normal, student-t, skew-student-t and generalized error distribution (GED) innovations by using mean absolute error (MAE), root mean squared error (RMSE) and robust Diebold–Mariano test. The results of the out-of-sample forecasts suggest that the SVR–GARCH with a mixture of Gaussian kernels can improve the volatility forecasts and capture the regime-switching behavior.

✉ Pedro Henrique M. Albuquerque
pedroa@unb.br

Pedro Correia S. Bezerra
pcsbezerra@gmail.com

¹ Departamento de Management, Universidade de Brasília, Brasília, Brazil

Keywords Volatility forecasting · Statistical learning theory · Support vector regression · Mixture of Kernels · Gaussian mixtures · Market regimes

1 Introduction

Financial time series prediction is important and challenging task in empirical finance. The data generation process of these series are complex because of its chaotic, noisy, non-stationary and nonlinear nature (Cao and Tay 2001). Thus, the use of support vector regression (SVR) in financial forecasting has been proposed in the literature, because it do not establish hypotheses about the distribution of data, is a pure data-driven technique, is very flexible, has excellent forecasting accuracy, and show theoretical and empirical superior results than artificial neural networks and traditional statistical methods (Sapankevych and Sankar 2009; Cavalcante et al. 2016). Moreover, the SVR is a machine learning technique based on the statistical learning theory that implement the Structural Risk Minimization Principle, which results in better generalization performance (Cao and Tay 2003).

Volatility is a measure of the degree of fluctuation of financial return and is a proxy for risk, thus is a key variable in risk management, asset pricing and portfolio selection (Brownlees and Gallo 2009). Linear and nonlinear parametric generalized autoregressive conditional heteroscedasticity (GARCH) models make assumptions about the functional form of the data generating process and the error distribution. Besides, empirical studies provide evidence that GARCH has low forecasting performance (Jorion 1995; Brailsford and Faff 1996; Mcmillan and Speight 2000; Choudhry and Wu 2008). Therefore, modifications have been proposed to improve its forecasts accuracy such as: changes in specification and model estimation, the use of different proxies for volatility, changes in evaluation metrics of forecast (Chen et al. 2010). To overcome these limitations, volatility forecasting models based on SVR have been proposed in the literature, because it able to capture non-linear characteristics of financial time series such as volatility clustering, leptokurtosis and leverage effect, without any assumptions about the data distribution properties. As shown by Fernando et al. (2003), Chen et al. (2010), Li (2014) and Santamaría-Bonfil et al. (2015), SVR shows superior results on volatility forecasting compared with GARCH models, due to its ability to capture the dynamic and nonlinear behavior of financial time series.

The volatility of financial asset returns change over time due to the capital market behavior. Empirical evidence shows that there are oscillation between several regimes in the financial market, in which the overall distribution of returns is a mixture of normals (Levy and Kaplanski 2015). In general, researches report the existence of two regimes (one with high and the other with low volatility) for the distribution of stock returns in the equity market. However, the markets can have more than two states (Bae et al. 2014). Then, it is necessary a mixture of more than two normal distributions to model the regime-switching behavior (Guidolin 2011). As the SVR is a kernel-based methodology, its forecasting performance is greatly dependent upon the selection of kernel function. To improve the SVR learning and generalization ability and take advantage of different kernel functions, it is possible to construct hybrid kernels via linear or non-linear combination of kernels (Huang et al. 2014). Empirical evidence

shows that the hybrid kernel has superior empirical results on forecasting accuracy than the SVR with a single-kernel (Huang et al. 2014). However, to the best of our knowledge, no research on volatility forecasting via SVR used a hybrid kernel. In this context, the main purpose of this article is to use a mixture of Gaussian kernels in the SVR based on GARCH (1,1) (heretofore SVR–GARCH) in order to improve the prediction accuracy and model: (i) market regimes and (ii) financial returns stylized characteristics such as high kurtosis, heavy tails and volatility clustering.

The forecasting accuracy of SVR–GARCH with a linear combination of one, two, three and four Gaussian kernels in one-period-ahead volatility forecasting is compared with the SVR–GARCH with Morlet wavelet kernel, GARCH, EGARCH, GJR. For each GARCH models four different distributions for the innovations are considered: the Normal, the Student's t, the skewed Student's t and the GED. in terms of two evaluation metrics of mean absolute error (MAE) and root mean squared error (RMSE).

The remainder of this paper is organized as follows. Section 2 provides a brief explanation of the Support Vector Machine (SVM) for regression. Section 3 explains the use of mixture of Gaussian kernels in the SVR–GARCH. Section 4 describes the empirical modeling. Section 5 shows the empirical results of the proposed model on daily financial returns of Nikkei 225 and Ibovespa indexes. Section 6 provides the concluding remarks of this paper.

2 Support vector regression

The support vector machine (SVM) is a machine learning algorithm based on the statistical learning theory developed by Vapnik and Chervonenkis (1974). SVM for linear and non-linear regressions is called support vector regression (SVR) (Smola and Schölkopf 2004). The non-linear SVR can be written as follows: given a set of training data $(x_1, y_1), \dots, (x_n, y_n)$, where $x_i \in \mathcal{X} \subseteq \mathbb{R}$ is the input vector and $y_i \in Y \subseteq \mathbb{R}$ being the output scalar. The goal of SVR is to find a function $f(x)$ that approximate the output scalar y_i less than a forecast error. To achieve this goal, the SVR nonlinearly maps the input vector space (\mathbb{R}^n) into a higher dimension feature space (\mathcal{F}), where the non-linear relations of the input space are approximated by a linear regression in the feature space Vapnik (1995):

$$f(x) = w^T \phi(x) + b, \quad \text{with } \phi : \mathbb{R}^n \rightarrow \mathcal{F}, w \in \mathcal{F} \quad (1)$$

where w and b are the regression parameter vectors and $\phi(\cdot)$ is the nonlinear mapping function, which projects the input vector into a higher dimension feature space, where the linear regression is defined. Vapnik (1995) introduced the ϵ -insensitive loss function (L_ϵ) to measure the difference between the actual and predicted values. The goal of ϵ -SVR is to find a function that has at least ϵ deviation from y_i . The vector w and constant b can be found by minimizing the following regularized function $R(C)$ (Vapnik 1995):

$$\text{Minimize : } R(C) = \frac{1}{2} \|w\|^2 + \frac{C}{n} \sum_{i=1}^n (L_\epsilon(f(y_i), f(x_i))); \quad (2)$$

where:

$$L_\epsilon(y, f(x)) = \begin{cases} |y_i - f(x_i)| - \epsilon, & \text{if } |y_i - f(x_i)| > \epsilon \\ 0, & \text{otherwise} \end{cases}, \epsilon \geq 0 \quad (3)$$

is the ϵ -insensitive loss function (L_ϵ). Only the observations on or outside the ϵ -insensitive zone will serve as the support vectors to construct the decision function ($f(\mathbf{x})$). To indicate the errors outside the ϵ -insensitive zone, slack variables (ξ_i, ξ_i^*), $i = 1, 2, \dots, n$) are introduced (Smola and Schölkopf 2004). Then, the primal problem of SVR is given by:

$$\begin{aligned} \text{Minimize : } & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \\ \text{subject to } & \begin{cases} y - w^T \phi(x) - b \leq \epsilon + \xi_i, \\ w^T \phi(x) + b - y \leq \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (4)$$

The convex quadratic programming and the linear restrictions of the primal problem above assure that the SVR will always achieve the optimal global solution. The term $\frac{1}{2} \|w\|^2$ characterizes the model complexity. The parameter C denotes the trade-off between the function complexity and training error $\sum_{i=1}^n (\xi_i + \xi_i^*)$ (Sermpinis et al. 2014). If the value C is large, the algorithm will overfit the data and will have lower generalization ability. The parameter ϵ controls the width of the ϵ -insensitive zone: the higher ϵ is, less support vectors are selected (Cherkassky and Ma 2004). The parameters C and ϵ are the SVR meta-parameters and, in general, are determined by cross-validation (Haykin 1998). In order to solve equation 4, it is possible to use the Lagrangian multipliers (α_i e α_i^*) and the Karush–Kuhn–Tucker conditions (Karush (1939); Kuhn and Tucker (1951)), transforming the problem in its dual form (Vapnik 1995):

$$\begin{aligned} \text{Maximize : } \mathcal{L} = & -\frac{1}{2} \sum_{i=1}^n (\alpha_i - \alpha_i^*)(\alpha_j^* - \alpha_j) \langle \phi(x_i), \phi(x) \rangle \\ & + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) \\ \text{subject to } & \begin{cases} \sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0, \\ 0 \leq \alpha_i \leq C, i = 1, \dots, n \\ 0 \leq \alpha_i^* \leq C, i = 1, \dots, n \end{cases} \end{aligned} \quad (5)$$

The Lagrange multipliers are calculated, then we find the support vector in expansion: $w = \sum_{i=1}^n (\alpha_i^* - \alpha_i) \phi(x_i)$. From the solution of the dual problem, the ϵ -SVR function can be written as (Vapnik 1995):

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle \phi(x_i), \phi(x) \rangle + b \quad (6)$$

where $\langle \phi(x_i), \phi(x) \rangle$ is the dot product in the feature space \mathcal{F} . To avoid the complexity of computing $\phi(\cdot)$, we can substitute the dot product by a kernel function:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b^* \quad (7)$$

The kernel function $K(x, x') = \langle \phi(x), \phi(x') \rangle$ is critical to the forecasting performance of the SVR. Any function that satisfies the (Mercer 1909) theorem is an admissible kernel. So far there is no analytical method for choosing the most appropriate kernel for a given problem (Sangeetha and Kalpana 2010). Before estimating w and b with SVR, it is necessary to choose the regularized parameter C , loss function parameter ϵ and the parameters of the chosen kernel (Sermpinis et al. 2014).

3 SVR–GARCH with a mixture of Gaussian kernels

In empirical finance, the normal distribution is the most convenient distributional assumption of asset returns (Wirjanto and Xu 2009). Nevertheless, empirical studies show that the distribution of returns depart from the normal distribution shape due to their substantial leptokurtosis (fat tails) and skewness (assymetry) (Wang and Taaffe 2015). Besides, the stock market oscillates between regimes (or states) (Tu 2010). Then, to explain these facts, it is more appropriate to use a mixture of two or more normal distributions (Wang and Taaffe 2015), because they are more flexible and can model complex phenomena (Marron and Wand 1992; McLachlan and Peel 2004).

Conditional volatility models such as the Autoregressive Conditional Heteroscedasticity (ARCH) (Engle 1982) and Generalized ARCH (GARCH) (Bollerslev 1986) can capture volatility clustering and time-varying volatility. However, empirical evidence shows that these models with Gaussian or heavier distribution innovations can not model the full extent of skewness and kurtosis (Wirjanto and Xu 2009). Since any continuous distribution can be well approximated by a finite mixture of normal distributions, the use of mixtures of normals in GARCH innovations have been proposed by Bai et al. (2003), Haas et al. (2004), Marcucci (2005), Alexander and Lazar (2006), Wirjanto and Xu (2009).

Given that financial returns are subject to regime-switching behavior between k states, even if the distribution of financial returns of each regime is normal, the overall distribution, given the probability of each state, is not normal. In fact, it is a mixture of k normal distributions (Levy and Kaplanski 2015). One way to accommodate this situation in the context of volatility forecasting via SVR is to use a mixture of Gaussian kernels. The mixture of normal distributions can capture extreme events, high curtosis, heavy tails of financial returns and approximate arbitraly any continuous probability distribution (McLachlan and Peel 2004; Wang and Taaffe 2015). In this context, we attempt to find the optimal number of mixtures of Gaussian kernels for the SVR–GARCH model. We will test a linear combination of one, two, three and four Gaussian

Table 1 Kernel choice in volatility forecasting

References	Journal	Model name	Kernel
Fernando et al. (2003)	Quantitative finance	GARCH–SVR linear	Not applicable
Gavrishchaka and Ganguli (2003)	Neurocomputing	SVR–GARCH	Gaussian
Gavrishchaka and Banerjee (2006)	Computational management science	SVR–GARCH	Gaussian
Tang et al. (2009b)	Expert systems with applications	WSVM	Gaussian and wavelet
Tang et al. (2009a)	Neural Computing and applications	SWSVM	Gaussian and <i>Spline</i> wavelet
Chen et al. (2010)	Journal of forecasting	SVM–GARCH	Linear, polynomial e Gaussian
Wang et al. (2011)	Neural Computing and Applications	SVM–MSM	Gaussian
Ou and Wang (2013)	International Arab journal of information technology	SVRCGA	Gaussian
Li (2014)	Journal of forecasting	SVM for APARCH	Gaussian and wavelet
Santamaría-Bonfil et al. (2015)	Computational economics	SVR–GARCH	Linear, polynomial and Gaussian

kernels in SVR based on GARCH(1,1), which can capture market regimes and perhaps show better out-of-sample forecasting results than the SVR–GARCH with a single-kernel.

To verify which are the most used kernels in volatility forecasting via SVR, we conduct a search from 2000 to 2016 through each of the following publishers online search engines: Elsevier, Wiley Online Library, IEEE Xplore Digital Library, SCOPUS, ISI Web of Knowledge, Scencedirect, Google Scholar and ProQuest Journals. We use the same query for every search engine:

(“support vector machine” OR ”support vector regression”) AND (“financial time series forecasting” OR “volatility forecasting” OR “volatility”)

Then, we select papers about volatility forecasting via SVR published in peer-reviewed journals with an impact factor. We found that the ten selected research articles used only a single-kernel and that the Gaussian is the most widely used kernel function (Table 1):

The kernel function maps non-linear observations of input data into a higher dimensional feature space, in which the data is linearly separable (Vapnik 1995). In this paper, we use a linear combination of $k = 1, 2, 3, 4$ Gaussian kernels:

$$K_{mix}(x, x') = \sum_{k=1}^K \rho_k \times K_k(x, x'), \quad \rho_k \geq 0 \quad \text{and} \quad \sum_{k=1}^K \rho_k = 1 \quad (8)$$

where ρ is the weighting coefficient and $K(x, x')_k = \exp(-\gamma \|x - x'\|^2)$. Following Huang et al. (2014) the optimal ρ is obtained by a grid search with the search step length 0.1.

Empirical research show that wavelet kernels have superior volatility forecasting results than the Gaussian kernel (Tang et al. 2009b, a; Li 2014). Then, we also use the Morlet wavelet kernel in the SVR–GARCH (Zhang et al. 2004):

$$k(x, x') = \prod_{i=1}^N \left(\cos\left(1.75 \times \left(\frac{x_i - x'_i}{a}\right)\right) \right) \exp\left(\frac{-\|x_i - x'_i\|^2}{2a^2}\right), \quad x, x' \in \mathbb{R}^N \quad (9)$$

4 Empirical modelling

4.1 Parametric volatility models

Let P_t be asset price at time t . Then, the return of asset in time t is given by:

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right) \quad (10)$$

The GARCH can capture volatility clustering and persistence. The GARCH(1,1) is specified as follows:

$$r_t = u_t + a_t \quad (11)$$

$$a_t = \sqrt{h_t} z_t, \quad z_t \sim i.i.d(0, 1) \quad (12)$$

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1} \quad (13)$$

where $\alpha_0 > 0$ and $\alpha_1, \beta_1 \geq 0$. One of the drawbacks of the standard GARCH is that negative and positive shocks have the same impact in the volatility forecasts (Franses and van Dijk 1996). Glosten et al. (1993) introduced the GJR model to capture the asymmetric response of volatility to shocks. The GJR(1,1) is defined as:

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1} + \gamma S_{t-1}^- a_{t-1}^2, \quad (14)$$

where

$$S_{t-1}^- = \begin{cases} 1, & \text{se } a_{t-1} < 0 \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

where $\alpha_0 > 0$ and $\alpha_1, \beta_1 \geq 0, \alpha_1 + \gamma \geq 0$. The exponential generalized autoregressive conditional heteroscedasticity (EGARCH) (Nelson 1991) can model the the skewness of financial returns and ensure that the variance is always postive. The EGARCH(1,1) is written as the following:

$$\ln(h_t) = \alpha_0 + \beta_1 \ln(h_{t-1}) + \alpha_1 \left[\frac{|a_{t-1}|}{\sqrt{h_{t-1}}} - \frac{2}{\sqrt{\pi}} \right] - \gamma \left(\frac{|a_{t-1}|}{\sqrt{h_{t-1}}} \right) \tag{16}$$

where γ is the asymmetric response parameter. If this parameter is positive, a negative return generates more volatility than positive returns. In order to model the fat-tails of the empirical distribution of financial returns, the errors z_t can follow a Student's t, skewed Student's t or Generalized Error Distribution (GED) distributions (Marcucci 2005):

1. A random variable X that follows a Student's t distribution has the following probability density function (pdf) (Casella and Berger 2001):

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu} \right)^{-\frac{\nu+1}{2}} \tag{17}$$

where ν is the degree of freedom parameter and $\Gamma(\cdot)$ is the Gamma function.

2. Generalized error distribution (GED): a random variable X that follows a GED distribution with zero mean and unit variance has the following pdf Tsay (2010):

$$f(x) = \frac{\text{vexp}[-(\frac{1}{2})|(x/\lambda)|^\nu]}{\lambda 2^{(\nu+1/\nu)} \Gamma(1/\nu)}, \quad 0 < \nu \leq \infty \tag{18}$$

where:

$$\lambda = \left[\frac{2^{-(2/\nu)} \Gamma(1/\nu)}{\Gamma(3/\nu)} \right]^{1/2} \tag{19}$$

where the parameter ν denotes the thickness-of-tail. When $0 < \nu < 2$, GED has thicker tails than the normal distribution.

3. The skewed Student's t-distribution can model the asymmetric effects and excess of kurtosis, the pdf takes the following form Fernandez and Steel (1998):

$$f(x|\iota, \nu) = \frac{2}{\iota + 1/\iota} [g(\iota(sx + m)|\nu) I_{(-\infty, 0)}(x + m/s)] \tag{20}$$

$$+ \frac{2}{\iota + 1/\iota} [g((sx + m)/\iota|\nu) I_{(0, +\infty)}(x + m/s)], \tag{21}$$

where $g(\cdot/\nu)$ is a Student's t-distribution with ν degrees of freedom,

$$m = \frac{\Gamma((\nu + 1)/2) \sqrt{\nu - 2}}{\sqrt{\pi} \Gamma(\nu/2)} (\iota - 1/\iota), \tag{22}$$

$$s = \sqrt{(\iota^2 + 1/\iota^2 - 1) - m^2} \tag{23}$$

where ι is the asymmetric parameter.

4.2 SVR based on GARCH

In order to forecast volatility, we have to define the inputs and outputs of the SVR decision function. Previous studies showed that the GARCH(1,1) is sufficient to model financial volatility (Poon and Granger 2003; Hansen and Lunde 2005). Thus, in this work the conditional variance is modeled by a GARCH (1,1), while the conditional mean is modeled by an AR (1) process (Franses and van Dijk 1996). Then, to forecast volatility, we use a SVR based on GARCH (1,1) (heretofore SVR–GARCH). The output variable is h_t and the input vector is: $x_t = [a_{t-1}^2, h_{t-1}]$. The SVR–GARCH is given by the following structure:

$$r_t = f(r_{t-1}) + a_t \quad (24)$$

where f is the decision function estimated by SVR for the mean equation. We get the squared residuals from the conditional mean estimation of the SVR–GARCH, then we estimate the conditional variance equation given by:

$$\tilde{h}_t = g(\tilde{h}_{t-1}, a_{t-1}^2) \quad (25)$$

where g is the decision function estimated by SVR, a_t^2 it is the squared residuals and \tilde{h} is the volatility proxy. In the mean equation, we use a single Gaussian kernel ($K(x, x') = \exp(-\gamma \|x - x'\|^2)$), because it is a common choice in financial time series forecasting via SVR (Sapankevych and Sankar 2009). In the volatility equation of SVR–GARCH, we use a linear combination of one, two, three and four Gaussian kernels given by Eq. 8.

As volatility is not directly observable, is necessary the use of proxy. As in Brooks (2001), Brooks and Persaud (2003), Chen et al. (2010) we use the following proxy:

$$\tilde{h}_t = (r_t - \bar{r})^2 \quad (26)$$

where r_t are the financial returns and \bar{r} it is the mean of returns. Any volatility proxy is an imperfect estimator of the true conditional variance (Patton 2011). Perhaps the use of another proxy may alter the results presented here. However, this issue is beyond the scope of this paper.

Before applying the SVR–GARCH for volatility prediction, we use the validation procedure (also known as holdout method) based on grid search and sensitivity analysis to select the kernel parameter γ (for Gaussian kernel), the regularized parameter C and loss function parameter ϵ (Stone 1974; Kohavi 1995; Arlot and Celisse 2010). We divide the database into three mutually exclusive sets: training, validation and testing (Shalev-shwartz and Ben-david 2014). The training set is used to estimate model parameters, then the performance of various values of the parameters are evaluated in the validation set. Following Cao and Tay (2001) and Chen et al. (2010), we make a sensitivity analysis to assess the effects of variation of parameters C , ϵ , γ in the MAE of volatility forecasting in the validation set. Therefore, we make a grid search for each parameter, keeping the others fixed. For the variation of each of the parameters, we make the forecasting in the validation set and then calculate the MAE. We choose the

parameters that has the smallest value of MAE. Finally, we evaluate the SVR–GARCH generalization performance in the test set. In this paper, the whole data are divided into three subsets: the first 50% composes the training set, the next 20% composes the validation set and the last, 30%, is reserved for the test set.

We use the MAE and the RMSE to evaluate the prediction performance. The RMSE is given by:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n \epsilon_t^2} \quad (27)$$

MAE measures the magnitude of overall error and is given by the following equation (Hyndman and Koehler 2006):

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (28)$$

where y_t denote the observation at time t and \hat{y}_t denote the forecast of y_t . The model which produces the smallest values of MAE and RMSE is judged to be the best model. MAE is a random variable and we have to use a statistical procedure to determine if one model shows superior predictive performance over another model. We use the two-sided (Diebold and Mariano 1995) (DM) test to compare the forecast performance of two competing models. Then, the DM test statistic is based on the difference of MAE loss function and it has the following null and alternative hypothesis:

$$H_0 : MAE_1 - MAE_0 = 0 \quad \text{versus} \quad H_1 : MAE_1 - MAE_0 \neq 0$$

where MAE_0 is the MAE of the competing model and MAE_1 is the MAE of the proposed model. Thus, if the null hypothesis is rejected, there is evidence that some model is superior to the other. Moreover, according to Chen et al. (2010), the Diebold–Mariano (DM) statistic in a robust form for a time series with volatility σ_t is given by:

$$DM = \frac{1}{\sqrt{n}} \frac{1}{\sqrt{\hat{S}^2}} \sum_{t=T_1}^{T-1} (|\sigma_{t+1}^2 - \hat{\sigma}_{1,t+1}^2| - |\sigma_{t+1}^2 - \hat{\sigma}_{0,t+1}^2|) \sim N(0, 1) \quad (29)$$

where $\hat{\sigma}_{0,t+1}^2$ is the volatility estimated by the competing model, $\hat{\sigma}_{1,t+1}^2$ is the volatility estimated by the proposed model and \hat{S}^2 denotes the co(variance) matrix estimated by the Newey and West (1987) procedure. Negative (positive) values of DM statistic indicates that the proposed model performs better (worse) than the competing model.

5 Empirical results

In this section, we apply the SVR–GARCH with a linear combination of one, two, three and four Gaussian kernels to volatility forecast and compare its performance to three

Table 2 Descriptive statistics for daily returns

Statistics	Nikkei return	Ibovespa return
Observations	1422	2000
Mean	0.00032	−0.00021
Median	0.0006	0.0000
Skewness	−0.5757	0.0825
Kurtosis	4.9158	6.5769
Std. dev.	0.0139	0.0239
Minimum	−0.1115	−0.1210
Maximum	0.0743	0.1368

Table 3 Goodness of fit for Nikkei returns

Model	LL	AIC	BIC
GARCH-N	2878	−5.7709	−5.7512
GARCH-t	2889	−5.7915	−5.7669
GARCH-skewed-t	2891	−5.4224	−5.3999
GARCH–GED	2886	−5.7859	−5.7613
EGARCH-N	2880	−5.7732	−5.7486
EGARCH-t	2894	−5.7997	−5.7702
EGARCH-skewed-t	2891	−5.4224	−5.3999
EGARCH–GED	2890	−5.7908	−5.7612
GJR-N	2881	−5.7749	−5.7503
GJR-t	2893	−5.7973	−5.7678
GJR-skewed-t	2895	−5.7987	−5.7642
GJR–GED	2890	−5.7903	−5.7608

other parametric volatility models, specifically, the GARCH, EGARCH and GJR. The first dataset consists of the Nikkei 225 daily closing price from May 1, 2010 to January 28, 2016 for a total of 1422 observations all obtained from Yahoo Finance and then transformed into log return as in 10. The second dataset consists of the daily closing price of the Bovespa index for the period December 22, 2007 to January 04, 2016. The first half of the whole data are used as the training data, 20% are reserved for the validation set and the remaining data, 30%, as test set. Table 2 shows the summary of the descriptive statistics for the Nikkei 225 and Ibovespa along the whole sample period.

The returns are characterized by excess kurtosis and deviate from normal distribution. Table 3 and Table 4 show the parameter estimates for the GARCH (1,1), EGARCH (1,1) and GJR (1,1) models for the Nikkei and Ibovespa returns. For each model four different distributions for the innovations are considered: the Normal, the Student's t, the skewed Student's t and the GED (Generalized Error Distribution). The Nikkei and Ibovespa series best fit to the GJR with skewed Student's t innovation, according to highest value of Log likelihood (LL) and smallest value of AIC and BIC.

Then, we select the parameters C , ϵ and the kernel parameters for the conditional mean and volatility equation via cross-validation. For the mean equation, we use a

Table 4 Goodness of fit for Ibovespa returns

Model	LL	AIC	BIC
GARCH-N	3787	-5.4037	-5.3887
GARCH-t	3800	-5.421	-5.4023
GARCH-skewed-t	3802	-5.4224	-5.3999
GARCH-GED	3800	-5.4213	-5.4026
EGARCH-N	3808	-5.4323	-5.4135
EGARCH-t	3816	-5.4427	-5.4202
EGARCH-skewed-t	3819	-5.4454	-5.4192
EGARCH-GED	3816	-5.4425	-5.4201
GJR-N	3813	-5.4401	-5.4213
GJR-t	3819	-5.4478	-5.4253
GJR-skewed-t	3822	-5.4503	-5.4261
GJR-GED	3820	-5.4483	-5.4258

Gaussian kernel and for the volatility equation we use a linear combination of one, two three and four Gaussian kernels. The first 711 observations of Nikkei returns series are used for training, from 712 to 996 for validation and from 997 to 1422 for the test set. We use the training set to estimate the function f of the mean equation and g of the volatility equation of SVR-GARCH. In this section, we only report the parameter selection for the SVR-GARCH with a linear combination of two Gaussian kernels. The parameter selection for the SVR-GARCH with one, two, three Gaussian kernels and Morlet kernel is similar and not reported here to save space. For the same reason, we do not report the results for the Ibovespa returns.

First, we estimate the conditional mean equation in the training set:

$$r_t = f(r_{t-1}) \quad \text{for } i \in (2, \dots, 711) \quad (30)$$

For the selection of optimal parameters, we use a grid search for each parameter, while keeping the others fixed. For the variation of each parameter, we make a forecast in the validation set in order to minimize the following expression:

$$MAE = \frac{1}{284} \sum_{t=712}^{996} |r_t - f(r_{t-1})| \quad (31)$$

For the sensitive analysis of C , we fix $\epsilon = 0,0001$, $\gamma = 1,25$ and parameter C takes values in the range $[0, 10]$. The value of $C = 0.004$ leads to the best validation performance. Epsilon varies in the range $[0, 5]$, with $\gamma = 1,25$, $C = 0,025$. The validating MAE attains the minima when $\epsilon = 0.2205$. Parameter γ takes value in the range $[0, 10]$, with $C = 0.004$ and $\epsilon = 0.2205$. The value of $\gamma = 0.9$ results in the best validation performance. Thus, the best parameters of SVR-GARCH for the conditional mean returns are: $C = 0.004$, $\epsilon = 0.2205$ and $\gamma = 0.9$ (Table 5):

Therefore, we estimate the conditional mean equation by using the SVR-GARCH with the best parameters for the conditional mean until the 996 observation to obtain

Table 5 Sensitivity analysis of SVR in conditional mean estimation

Parameter	Range	Optimal value	Smallest MAE
C	[0,10]	0.004	0.01217182
ϵ	[0,5]	0.2205	0.01216
γ	[0,10]	0.9	0.01216106

Table 6 Sensitivity analysis of SVR in conditional variance estimation

Parameter	Range	Optimal value	Smallest MAE
C	[0,10]	5.184	0.0002154
ϵ	[0,0.1]	0.05929	0.0002146
γ_1	[0,1]	0.9801	0.0002115
γ_2	[0,1]	0.01	0.0002102
ρ	[0,1]	0.37	0.0002067

the residuals a_t in the following way:

$$a_t = r_t - f(r_t) \quad \text{for } i \in (2, \dots, 996) \tag{32}$$

Then we estimate the volatility equation of SVR–GARCH(1,1):

$$\tilde{h}_t = g(\tilde{h}_{t-1}, a_{t-1}^2) \quad \text{for } i \in (2, \dots, 711) \tag{33}$$

where a_t^2 is the squared residuals. The volatility proxy \tilde{h}_t is calculated until the 996 observation and the parameter selection is made in order to minimize the following expression:

$$MAE = \frac{1}{284} \sum_{t=712}^{996} |\tilde{h}_t - g(\tilde{h}_{t-1}, a_{t-1}^2)| \tag{34}$$

For the sensitive analysis of C , we fix $\epsilon = 0.0001$, $\gamma_1 = 0.01$, $\gamma_2 = 0.07$, $\rho = 0.25$ and parameter C takes values in the range [0, 10]. The validating MAE attains the minima when $C = 0, 625$. As in the mean equation, we do the same procedure for the others parameters (Table 6).

Thus, the appropriate parameters of SVR–GARCH for the conditional variance are $C = 5.184$, $\epsilon = 0.05929$, $\gamma_1 = 0.9801$, $\gamma_2 = 0.01$ and $\rho = 0.37$.

5.1 Volatility forecasting evaluation

With the SVR–GARCH optimal parameters (C , ϵ and kernel parameters), we make the one-period-ahead volatility forecasts in the test set (i.e. out-of-sample). After each forecast, we calculate the forecast errors and repeat the forecasting process for the next period. Table 7 report the values of MAE and RMSE obtained from different models for the Nikkei and Ibovespa returns.

Table 7 Out-of-sample evaluation of one-period-ahead volatility forecasts

Model	NIKKEI 225		IBOVESPA	
	MAE	RMSE	MAE	RMSE
SVR–GARCH ¹	0.000205	0.000450	0.0003958	0.000666
SVR–GARCH ²	0.0002071	0.0004746	0.0003620	0.0005665
SVR–GARCH ³	0.0001691	0.0004114	0.0003594	0.0005587
SVR–GARCH ⁴	0.0001717	0.0004100	0.0003988	0.0006786
SVR–GARCH–Morlet	0.000184	0.0004059	0.0003901	0.0006536
GARCH–N	0.01264	0.01316	0.01465	0.01499
GARCH–t	0.01255	0.01295	0.01463	0.01495
GARCH–skewed–t	0.01251	0.01292	0.0146	0.01493
GARCH–GED	0.01261	0.01307	0.01463	0.01496
EGARCH–N	0.01232	0.01281	0.01468	0.0152
EGARCH–t	0.01232	0.01274	0.01467	0.0152
EGARCH–skewed–t	0.01229	0.0127	0.01469	0.01523
EGARCH–GED	0.01234	0.01278	0.03379	0.0152
GJR–N	0.0127	0.01315	0.01417	0.01463
GJR–t	0.01267	0.01303	0.01417	0.01461
GJR–Skewed–t	0.01263	0.01299	0.01418	0.01464
GJR–GED	0.01271	0.01311	0.01418	0.01463

Note: SVR–GARCH 1, 2, 3, 4, denote Support Vector Regression based on GARCH with one, two, three and four Gaussian Kernels, respectively

For the Nikkei 225 series, the SVR–GARCH with a mixture of three Gaussian kernels achieve smallest value of MAE. But, the SVR–GARCH with Morlet wavelet kernel achieve the smallest value of RMSE. According to MAE and RMSE measures in Table 7, the SVR–GARCH with a linear combination of four Gaussian kernels is the best one for the Ibovespa series. To compare the predictive power of two models we use the two-sided Diebold–Mariano test given by the following null and alternative hypotheses for the Nikkei returns:

$$\begin{aligned}
 H_0 &: \frac{1}{426} \left| \tilde{h}_t - \hat{h}_{1,t} \right| - \left| \tilde{h}_t - \hat{h}_{0,t} \right| = 0 \quad \text{versus} \\
 H_1 &: \frac{1}{426} \left| \tilde{h}_t - \hat{h}_{1,t} \right| - \left| \tilde{h}_t - \hat{h}_{0,t} \right| \neq 0,
 \end{aligned}
 \tag{35}$$

where \tilde{h}_t is the volatility proxy, $\hat{h}_{0,t}$ is the volatility estimated by the proposed model and $\hat{h}_{1,t}$ is the volatility estimated by the competing model. Moreover, the DM test statistic is given by [Chen et al. \(2010\)](#):

$$DM = \frac{1}{\sqrt{426}} \frac{1}{\sqrt{\hat{S}^2}} \sum_{t=996}^{1422} \left| \tilde{h}_t - \hat{h}_{1,t} \right| - \left| \tilde{h}_t - \hat{h}_{0,t} \right| \sim N(0, 1)
 \tag{36}$$

Table 8 Diebold–Mariano test (benchmark: SVR–GARCH⁴, one-step-ahead)

Model	DM Statistics
SVR–GARCH ¹	−0.98†
SVR–GARCH ²	−1.5†
SVR–GARCH ³	−2.3*
SVR–GARCH–Morlet	0.18†
GARCH–N	−30 **
GARCH–t	−35 **
GARCH–skewed–t	−34 **
GARCH–GED	−32 **
EGARCH–N	−33 **
EGARCH–t	−35 **
EGARCH–skewed–t	−35 **
EGARCH–GED	−34 **
GJR–N	−33 **
GJR–t	−37 **
GJR–skewed–t	−37 **
GJR–GED	−35 **

Note: † indicate the DM statistics for which one fail to reject the null hypothesis of equal predictive accuracy. * and ** represent the DM statistics for which one can reject the null hypothesis at 5 and 1% respectively

Table 9 Diebold–Mariano test (benchmark: SVR–GARCH³, one-step-ahead)

Model	DM Statistics
SVR–GARCH ¹	−5.2**
SVR–GARCH ²	−5.5**
SVR–GARCH ⁴	−5**
SVR–GARCH–Morlet	−4.8**
GARCH–N	−51**
GARCH–t	−52**
GARCH–skewed–t	−52**
GARCH–GED	−52**
EGARCH–N	−42**
EGARCH–t	−42**
EGARCH–skewed–t	−42**
EGARCH–GED	−42**
GJR–N	−44**
GJR–t	−44**
GJR–skewed–t	−44**
GJR–GED	−44**

Note: † indicate the DM statistics for which one fail to reject the null hypothesis of equal predictive accuracy. * and ** represent the DM statistics for which one can reject the null hypothesis at 5 and 1% respectively

Tables 8 and 9 report the DM statistics and p-values of the Diebold–Mariano test for the difference of MAE loss function for the Nikkei 225 and Ibovespa daily returns, respectively:

For the Nikkei 225 and Ibovespa series, it is evident that all SVR–GARCH models significantly outperform every GARCH model at any usual confidence level. For the Nikkei 225 series, except for SVR–GARCH with Morlet wavelet kernel, the sign of the DM statistic is always negative, implying that the benchmark's loss is lower than the loss implied by the competing models. However, we cannot reject the null hypothesis for the SVR–GARCH with Morlet wavelet kernel and with one and two Gaussian kernels, which means that these models have equal forecasting ability. For the other models we always reject the null hypothesis of equal forecast accuracy at any usual confidence level. For the Ibovespa series, the sign of the DM statistic for the SVR–GARCH with a linear combination of three kernels is always negative and we always reject the null hypothesis of equal forecast accuracy.

6 Concluding remarks

The main contributions of this paper is to use a mixture of one, two, three and four Gaussian kernels in the SVR based on GARCH(1,1) to take into account the existence of market regimes. We compare these models with SVR–GARCH with Morlet wavelet kernel, GARCH, EGARCH and GJR models in terms of their ability to forecast volatility by using MAE, RMSE and Diebold–Mariano test. All GARCH models are estimated assuming Gaussian, Student's t, skewed Student's t and GED innovations. To determine the SVR optimal parameters we use used the validation technique (hold-out method) based on grid-search and sensitivity analysis. Nikkei 225 and Ibovespa daily returns were used as the dataset. The empirical results indicate that the mixture of Gaussian kernels can improve the SVR–GARCH one-period-ahead volatility forecasts. In sum, the mixture of normal distributions can model the overall distribution of financial returns when markets display regime behaviour and also better approximate nonlinear characteristics of financial returns such as heavy tails, volatility clustering and time-varying skewness.

References

- Alexander C, Lazar E (2006) Normal mixture GARCH(1,1): applications to exchange rate modelling. *J Appl Econom* 21(3):307–336
- Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. *Stat Surv* 4:40–79
- Bae GI, Kim WC, Mulvey JM (2014) Dynamic asset allocation for varied financial markets under regime switching framework. *Eur J Oper Res* 234(2):450–458
- Bai X, Russell JR, Tiao GC (2003) Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *J Econom* 114(2):349–360
- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econom* 31:307–327
- Brailsford TJ, Faff RW (1996) An evaluation of volatility forecasting techniques. *J Bank Finance* 20:419–438
- Brooks C (2001) A Double-threshold GARCH Model for the French Franc/Deutschmark exchange rate. *J Forecast* 20(2):135–143
- Brooks C, Persaud G (2003) Volatility forecasting for risk management. *J Forecast* 22(1):1–22
- Brownlees CT, Gallo GM (2009) Comparison of volatility measures: a risk management perspective. *J Financial Econom* 8(1):29–56
- Cao L, Tay F (2003) Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Trans Neural Netw* 14(6):1506–1518

- Cao L, Tay FE (2001) Financial forecasting using support vector machines. *Neural Comput Appl* 10(2):184–192
- Casella G, Berger RL (2001) *Statistical inference*, 2nd edn. Duxbury Press, California
- Cavalcante RC, Brasileiro RC, Souza VL, Nobrega JP, Oliveira AL (2016) Computational intelligence and financial markets: a survey and future directions. *Expert Syst Appl* 55:194–211
- Chen S, Härdle WK, Jeong K (2010) Forecasting volatility with support vector machine-based GARCH model. *J Forecast* 43(29):406–433
- Cherkassky V, Ma Y (2004) Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Netw* 17(1):113–126
- Choudhry T, Wu HAO (2008) Forecasting ability of GARCH vs Kalman filter method: evidence from Daily UK Time-Varying Beta. *J Forecast* 689:670–689
- Diebold FX, Mariano RS (1995) Comparing predictive accuracy. *J Bus Econ Stat* 13(3):253–263
- Engle RF (1982) Autoregressive conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4):987–1007
- Fernandez C, Steel MFJ (1998) On Bayesian modeling of fat tails and skewness. *J Am Stat Assoc* 93(441):359
- Fernando P-C, Afonso-Rodríguez JA, Giner J (2003) Estimating GARCH models using support vector machines. *Quant Finance* 3:1–10
- Franses PH, van Dijk D (1996) Forecasting stock market volatility using (non-linear) Garch models. *J Forecast* 15(3):229–235
- Gavrishchaka VV, Banerjee S (2006) Support vector machine as an efficient framework for stock market volatility forecasting. *Comput Manag Sci* 3(2):147–160
- Gavrishchaka VV, Ganguli SB (2003) Volatility forecasting from multiscale and high-dimensional market data. *Neurocomputing* 55(1–2):285–305
- Glosten LR, Jagannathan R, Runkle DE (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *J Finance* 48(5):1779–1801
- Guidolin M (2011) Markov switching models in empirical finance. In: Drukker DM (ed) *Missing data methods: time-series methods and applications (advances in econometrics)*, vol 27. Emerald Group Publishing Limited, UK, pp 1–86
- Haas M, Mittnik S, Paoletta MS (2004) Mixed normal conditional heteroskedasticity. *J Financial Econom* 2(2):211–250
- Hansen PR, Lunde A (2005) A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *J Appl Econom* 20(7):873–889
- Haykin S (1998) *Neural networks: a comprehensive foundation*, 2nd edn. Prentice Hall, New York (ISBN-10: 0132733501, ISBN-13: 978-0132733502)
- Huang, C., Gao, F., Jiang, H., 2014. Combination of biorthogonal wavelet hybrid kernel OCSVM with feature weighted approach based on EVA and GRA in financial distress prediction. In: *Mathematical problems in Engineering 2014*
- Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. *Int J Forecast* 22(4):679–688
- Jorion P (1995) Predicting volatility in the foreign exchange market. *J Finance* 50(2):507–528
- Karush W (1939) *Minima of functions of several variables with inequalities as side constraints*. Ph.D. thesis, University of Chicago
- Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th international joint conference on artificial intelligence*, vol 14. Morgan Kaufmann Publishers Inc., Monreal, pp 1137–1143
- Kuhn HW, Tucker A (1951) *Nonlinear programming*. University of California Press, California
- Levy M, Kaplanski G (2015) Portfolio selection in a two-regime world. *Eur J Oper Res* 242(2):514–524
- Li Y (2014) Estimating and forecasting APARCH-Skew- t model by wavelet support vector machines. *J Forecast* 269(March):259–269
- Marcucci J (2005) Forecasting stock market volatility with regime-switching GARCH models. *Stud Non-linear Dyn Econom* 9(4):1–55
- Marron JS, Wand M (1992) Exact mean integrated squared error. *Ann Stat* 20(2):712–736
- McLachlan G, Peel D (2004) *Finite mixture models*. Wiley, Canada
- Mcmillan DG, Speight A (2000) Forecasting UK stock market volatility. *Appl Financial Econ* 10:435–448
- Mercer J (1909) Functions of positive and negative type and their connection with the theory of integral equations. *Philos Trans R Soc Lond* 209(A):415–446

- Nelson DB (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59(2):347–370
- Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708
- Ou P, Wang H (2013) Volatility modelling and prediction by hybrid support vector regression with chaotic genetic algorithms. *Int Arab J Inf Technol* 11(3):287–292
- Patton AJ (2011) Volatility forecast comparison using imperfect volatility proxies. *J Econ* 160(1):246–256
- Poon SH, Granger CW (2003) Forecasting volatility in financial markets: a review. *J Econ Lit* 41(2):478–539
- Sangeetha R, Kalpana B (2010) A comparative study and choice of an appropriate kernel for support vector machines. In: *Information and communication technologies*, pp 549–553
- Santamaría-Bonfil G, Frausto-Solís J, Vázquez-Rodarte I (2015) Volatility forecasting using support vector regression and a hybrid genetic algorithm. *Comput Econ* 45:111–133
- Sapankevych NI, Sankar R (2009) Time series prediction using support vector machines: a survey. *IEEE Comput Intell Mag* 4(2):24–38
- Sermpinis G, Stasinakis C, Theofilatos K, Karathanasopoulos A (2014) Inflation and unemployment forecasting with genetic support vector regression. *J Forecast* 33(6):471–487
- Shalev-shwartz S, Ben-david S (2014) *Understanding machine learning: from theory to algorithms*. Cambridge University Press, New York
- Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14:199–222
- Stone M (1974) Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc* 36(2):111–147
- Tang L-B, Sheng H-Y, Tang L-X (2009a) GARCH prediction using spline wavelet support vector machine. *Neural Comput Appl* 18(8):913–917
- Tang L-B, Tang L-X, Sheng H-Y (2009b) Forecasting volatility based on wavelet support vector machine. *Expert Syst Appl* 36(2):2901–2909
- Tsay RS (2010) *Analysis of financial time series*, 3rd edn, vol 48. Wiley, Newyork
- Tu J (2010) Is regime switching in stock returns important in portfolio decisions? *Manag Sci* 56(7):1198–1215
- Vapnik VN (1995) *The nature of statistical learning theory*. Springer, Berlin
- Vapnik VN, Chervonenkis AY (1974) *Theory of pattern recognition: statistical problems of learning*. Nauka, Moscow
- Wang B, Huang H, Wang X (2011) A support vector machine based MSM model for financial short-term volatility forecasting. *Neural Comput Appl* 22(1):21–28
- Wang J, Taaffe MR (2015) Multivariate mixtures of normal distributions: properties, random vector generation, fitting, and as models of market daily changes. *INFORMS J Comput* 27(2):193–203
- Wirjanto TS, Xu D (2009) The applications of mixtures of normal distributions in empirical finance: a selected survey, Working paper 0904. University of Waterloo, Department of Economics. <http://economics.uwaterloo.ca/documents/mn-review-paper-CES.pdf>
- Zhang L, Zhou W, Jiao L (2004) Wavelet support vector machine. *IEEE Trans Syst Man Cybern Part B* 34(1):34–39