Pedro Correia Santos BEZERRA, PhD Candidate  
E-mail: pcsbezerra@gmail.com  
Professor Pedro Henrique Melo ALBUQUERQUE  
E-mail: pedroa@unb.br

VOLATILITY FORECASTING: THE SUPPORT VECTOR REGRESSION CAN BEAT THE RANDOM WALK

Abstract. Financial time series prediction is important, and it is a challenger task in empirical finance due to its chaotic, nonlinear and complex nature. Machine learning techniques that have been employed to forecast financial volatility. In this paper, we implement a standard Support Vector Regression model with Gaussian and Morlet wavelet kernels on daily returns of two stock market indexes - USA (SP&500) and Brazil (IBOVESPA) - over the period 2008-2016. The random walk, GARCH(1,1) and GJR(1,1) on the skewed Student’s t-distribution serve as comparison models by using Mean Squared Error (MSE) and the Diebold-Mariano test. The empirical analysis suggests that the SVR can beat the random walk model in the USA (S&P500) and Brazilian (Ibovespa) markets at one-period ahead forecasting horizon.

Keywords: Financial time series Volatility Forecasting Support Vector Regression Random Walk.

JEL Classification: C45, C58, C53

1. Introduction

The volatility of financial returns is a fundamental metric in finance [5]. The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) is one of the most used volatility forecasting models. However, previous studies showed that non-linear Machine Learning (ML) methods have better forecasting performance than traditional statistical and econometric methods [8,3,9,19]. The Support Vector Regression (SVR) is a ML method that implement the Structural Risk Minimization, is a

1CorrespondingAuthor

DOI: 10.24818/18423264/53.4.19.07
kernel-based methodology and has excellent volatility predictive accuracy compared with the GARCH family and neural networks [15, 25, 28, 10, 20, 4].

Empirical results show that the accuracy of the random walk model (RW) can outperform traditional linear statistical and econometric models in financial time series prediction [13, 1]. Moreover, according to [13], in relation to more sophisticated econometric models, simplest models may present better accuracy in predicting volatility. However, the RW has a linear form and do not capture the nonlinear, complex and noise behaviour of these series. The SVR is very useful in modelling the conditional volatility of stock returns because is a distribution-free approach, a pure-data driven method, allows a flexible structure and can approximate nonlinear characteristics of financial time series [9, 26]. Previous researches showed that the SVR can beat the random walk model for prediction of financial prices [21, 26]. Nevertheless, to the best of our knowledge, this is the first paper to compare the performance of SVR with the RW in the context of volatility forecasting. We develop a SVR algorithm which attempts to improve the one-day ahead volatility forecasts of stock index and also try to beat the RW. The remainder of this paper is organized as follows. The next section describes the Support Vector Machine (SVM) for regression. Section 3 describes the empirical modelling. Section 4 shows the empirical results of the proposed model on daily financial returns of SP&500 and Ibovespa indexes. Section 5 provides the concluding remarks of this paper.

2. Support Vector Regression

Given a set of training data \((x_1, y_1), \ldots, (x_n, y_n)\), where \(x_i \in \mathcal{X} \subseteq \mathbb{R}^p\) is the input vector and \(y_i \in \mathcal{Y} \subseteq \mathbb{R}\) being the output scalar, the goal of SVR is to find a function \(f(x)\) that approximate the output \(y_i\) [27]:

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

where \(w = [w_1, \ldots, w_n]^T\) are the regression coefficients, \(b\) is a constant and \(\phi(.)\) is the nonlinear mapping function, which projects the input vector into a higher dimension feature space \(\mathcal{F}\), where the linear regression is defined.

Vapnik[27] introduced the \(\varepsilon\) -insensitive loss function \(|y_i - f(x)| = \max[0, |y_i - f(x)| - \varepsilon]\) to measure the forecasts errors made by SVR. To denote the errors outside the \(\varepsilon\) -insensitive zone, slack variables \((\xi_i, \xi_i^*)\), \(i = 1, 2, \ldots, n\) are introduced in the SVR primal problem of SVR:

\[
\text{Minimize:} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)
\]
Volatility Forecasting: The Support Vector Regression Can Beat the Random Walk

\[
\begin{align*}
&\text{subject to } \begin{cases} y_i - w^T \phi(x_i) - b \leq \varepsilon + \xi_i^* \\ w^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \text{ for } i = 1, \ldots, n \\
\end{align*}
\]

where \( C \) is the regularization hyperparameter.

The parameters \( C \) and \( \varepsilon \) are the SVR parameters and can be determined by a grid search algorithm with validation (or cross-validation) [18]. We substitute the dot product by a kernel function to overcome the complexity of computing \( \phi(\cdot) \), this is known as the kernel trick approach:

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

where \( 0 \leq \alpha_i, \alpha_i^* \leq C \) (3)

The kernel function \( K(x, x') = \langle \phi(x), \phi(x') \rangle \) is critical to the forecasting performance of the SVR, but until now there is no way to choose an appropriate kernel. For a mathematical function to be admissible as a kernel, it must satisfy the [23] theorem. In this work, the parameters of the Gaussian and Morlet wavelet kernels were determined using a grid-search and hold-out method in the training set [8].

### 2.0.1 Wavelet Kernels

Wavelet analysis is used in a variety of domains, such as: geophysics, engineering, physics, statistics, finance [11]. Wavelets functions can approximate a signal and model the frequency and temporal domain of time series by translations and dilations of a mother wavelet \( \Psi(x) \in L^2(\mathbb{R}^p) \):

\[
\Psi_{k,a}(x) = \frac{1}{\sqrt{a}} \Psi \left( \frac{x - k}{a} \right), \quad x \in \mathbb{R}^p \text{ and } a, k \in \mathbb{R}.
\]

where \( a \) is the dilation factor and \( k \) is the translation factor. With the use of wavelet analysis, Zhang et al. [29] developed admissible wavelet kernels. They proved the existence of two types of wavelets kernels. First, the dot productive kernel:

\[
k(x, x') = \prod_{i=1}^{p} \Psi \left( \frac{x_i - k_i}{a} \right) \Psi \left( \frac{x_i' - k_i}{a} \right)
\]

where \( a, x, x' \in \mathbb{R} \). Second, the translation invariant kernel:

\[
k(x, x') = \prod_{i=1}^{p} \Psi \left( \frac{x_i - x_i'}{a} \right)
\]

Using the Morlet wavelet function \( \Psi(x) = \cos(1.75x)\exp(x^2/2) \), Zhang et al. [29] constructed a translation invariant kernel that satisfies Mercer’s condition:

\[
k(x, x') = \prod_{i=1}^{p} \left( \cos(1.75 \times \frac{(x_i - x_i')}{a}) \exp \left( \frac{-(x_i - x_i')^2}{2a^2} \right) \right)
\] (7)
In the context of volatility forecasting, Li [20] showed that the Morlet wavelet kernel combined with the SVM in estimating the APARCH model has superior predicting ability results than the Gaussian kernel via Monte Carlo simulations. In this paper, we also use the traditional Gaussian kernel with the following form: \( k(x, x') = \exp(-\gamma \| x - x' \|^2) \), where \( \gamma \) is the precision parameter of the kernel function [26].

3. Empirical Modelling

According to [1], the random walk model (RW) is one of the best linear model for financial time series forecasting. As in Dimson and Marsh [13], we use the RW model as a benchmark for judging the other volatility forecasting models. The driftless RW is given by the following equation [13]:

\[
h_t = h_{t-1}
\]  
(8)

where \( h_t \) is the volatility proxy. Although the use of a proxy for daily volatility implies an imperfect estimator of the real conditional variance [24, 2], we use the same proxy as [6, 10]:

\[
\tilde{h}_t = (r_t - \bar{r})^2
\]  
(9)

where \( r_t \) is the daily log-return and \( \bar{r} \) it is the mean of log-returns series, for \( t = 1, ..., T \).

We also apply the GARCH and GJR models on the skewed Student’s t-distribution to model the volatility of financial returns because is a traditional choice in the context of volatility forecasting [17, 22].

3.1 Parametric Volatility Models

We use the log-returns series: \( r_t = \log \left( \frac{P_t}{P_{t-1}} \right) \), where \( P_t \) is the price. The GARCH (1,1) has the following structure [17, 22]:

\[
r_t = u_t + a_t
\]  
(10)

\[
\sqrt{h_t}z_t, \quad z_t \sim i.i.d(0,1)
\]  
(11)

\[
a_t = \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1}
\]  
(12)

where \( \alpha_0 > 0 \) and \( \alpha_1, \beta_1 \geq 0 \).
In order to improve GARCH and capture negative and positive shocks, Glosten et al. [16] introduced the GJR model:

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

(13)

where

$$S_{t-1}^{-} = \begin{cases} 1, & \text{if } a_{t-1} < 0 \\ 0, & \text{otherwise} \end{cases}$$

(14)

where $\alpha_0 > 0$ and $\alpha_1, \beta_1 \geq 0$, $\alpha_1 + \gamma \geq 0$.

In order to model excess of kurtosis and asymmetric effects, we use the skewed Student’s t-distribution [14]:

$$f(x|\nu) = \frac{2}{\nu + 1/4} \left[ g((sx + m)|\nu)l_{(-\infty, 0)}(x + m/s) \right] + \frac{2}{\nu + 1/4} \left[ g((sx + m)/|\nu)l_{(0, +\infty)}(x + m/s) \right],$$

(15)

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

$$m = \frac{\Gamma((\nu+1)/2)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\nu/2)} (t - 1/i),$$

(16)

$$s = \sqrt{(t^2 + 1/i^2 - 1) - m^2}$$

where $t$ is the asymmetric parameter.

### 3.2 SVR Algorithm for volatility forecasting

The SVR used in this work is given by the following structure:

$$\tilde{h}_t = f(\tilde{h}_{t-1})$$

(18)

where $f$ is the decision function estimated by SVR and $\tilde{h}_t$ is the daily volatility proxy. The algorithm steps of SVR for volatility forecasting are as follows:

- **Step 1** Divide the database into three mutually exclusive sets: training, validation and testing. The first 50% composes the training set, the next 20% composes the validation set and the last 30%, are used for testing.
- **Step 2** In the training test, determine the SVR and kernel optimal parameters by the holdout method based on grid-search and sensitivity analysis [8, 7];
- **Step 3** Choose the parameters that has the smallest value of Mean Squared Error (MSE) in the validation set: $\frac{1}{n} \Sigma_{t=1}^{n} \epsilon_t^2$.
Step 4 After the choice of optimal parameters of SVR, make the one-period-ahead volatility forecasts in the test set (i.e. out-of-sample);

Step 5 Evaluate the prediction performance with the Mean Squared Error (MSE) and the Diebold-Mariano test [12].

We use the MSE because is a consistency loss function in the context of volatility forecasting [24, 2].

4 Results

The literature suggests that developed equity markets are more efficient and difficult to predict than emerging markets [19]. Given that, we apply the proposed algorithm in two series of index. These are as follows: (i) USA (daily closing prices of S&P500 from September 12, 2008 to August 23, 2016) and Brazil (daily closing prices of Ibovespa from December 1, 2007 to January 04, 2016).

Table 1: Dataset description

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Source</th>
<th>Period</th>
<th>Training Size</th>
<th>Testing Size</th>
<th>Total Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P500</td>
<td>Yahoo! Finance</td>
<td>2008-09-12 to 2016-08-23</td>
<td>1400</td>
<td>600</td>
<td>2000</td>
</tr>
<tr>
<td>Ibovespa</td>
<td>Yahoo! Finance</td>
<td>2007-12-22 to 2016-01-04</td>
<td>1400</td>
<td>600</td>
<td>2000</td>
</tr>
</tbody>
</table>

The summary statistics of the two series under study are presented in Table 2:

Table 2: Descriptive statistics

<table>
<thead>
<tr>
<th>Statistics</th>
<th>S&amp;P500</th>
<th>Ibovespa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00027</td>
<td>-0.0002</td>
</tr>
<tr>
<td>Median</td>
<td>0.0007</td>
<td>0.0000</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.3448</td>
<td>0.0825</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>10.4765</td>
<td>6.5769</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>0.0137</td>
<td>0.0183</td>
</tr>
</tbody>
</table>
Volatility Forecasting: The Support Vector Regression Can Beat the Random Walk

<table>
<thead>
<tr>
<th>Minimum</th>
<th>-0.0947</th>
<th>-0.1210</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>0.1096</td>
<td>0.1368</td>
</tr>
</tbody>
</table>

The Log likelihood (LL), AIC and BIC values for the GARCH(1,1) and GJR(1,1) are shown in Table 3 and 4. GJR with skewed Student’s t innovation is the best fit to both series.

**Table 3: Goodness of fit for S&P500 returns**

<table>
<thead>
<tr>
<th>Model</th>
<th>LL</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH-Skewed-t</td>
<td>2891</td>
<td>-5.4224</td>
<td>-5.3999</td>
</tr>
<tr>
<td>GJR-Skewed-t</td>
<td>2895</td>
<td>-5.7987</td>
<td>-5.7642</td>
</tr>
</tbody>
</table>

**Table 4: Goodness of fit for Ibovespa returns**

<table>
<thead>
<tr>
<th>Model</th>
<th>LL</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH-Skewed-t</td>
<td>2891</td>
<td>-5.4224</td>
<td>-5.3999</td>
</tr>
<tr>
<td>GJR-Skewed-t</td>
<td>2895</td>
<td>-5.7987</td>
<td>-5.7642</td>
</tr>
</tbody>
</table>

We select the parameters $C$, $\epsilon$ and the kernel parameters via sensitivity analysis and holdout method. In order to save space, Table 5 only reports the optimal parameters of SVR for the Ibovespa series.

**Table 5: Sensitivity analysis of SVR**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
<th>Optimal Value</th>
<th>Smallest MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>[0,10]</td>
<td>5.18400</td>
<td>0.0002154</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>[0,0.1]</td>
<td>0.05929</td>
<td>0.0002146</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>[0,1]</td>
<td>0.98010</td>
<td>0.0002115</td>
</tr>
</tbody>
</table>

Table 6 report the prediction performance for the S&P 500 and Ibovespa indices returns.

**Table 6: Forecasting performance**

<table>
<thead>
<tr>
<th>Model</th>
<th>S&amp;P 500</th>
<th>Ibovespa</th>
</tr>
</thead>
</table>

DOI: 10.24818/18423264/53.4.19.07
For both series, the SVR presented the best prediction results. For S&P 500 series, the SVR has a forecast performance 13% higher than the random walk (RW). For the Ibovespa series, the SVR has a forecast performance 25% higher than the RW. Thus, predictive accuracy is higher in emerging market compared to established financial market, which confirms the findings of Hsu et al. [19].

Besides, to compare the predictive power of two models and investigate the statistical significance of the success of our point forecasts, Sermpinis et al. [26], we use the two-sided Diebold-Mariano test (DM) [12] given by the following structure [10]:

\[ H_0: \frac{1}{600} \sum_{t=1}^{2000} |\tilde{h}_t - \hat{h}_{1,t} - |\tilde{h}_t - \hat{h}_{0,t}| = 0 \]
\[ H_1: \frac{1}{600} \sum_{t=1}^{2000} |\tilde{h}_t - \hat{h}_{1,t} - |\tilde{h}_t - \hat{h}_{0,t}| \neq 0 \]

where \( \tilde{h}_t \) is the volatility proxy, \( \hat{h}_{0,t} \) is the volatility estimated by the random walk model and \( \hat{h}_{1,t} \) is the volatility estimated by a given model. The DM test statistic for the difference of MSE loss function is given by Chen et al. [10]:

\[ DM = \frac{1}{\sqrt{600}} \sqrt{\sum_{t=1400}^{2000} |\tilde{h}_t - \hat{h}_{1,t} - |\tilde{h}_t - \hat{h}_{0,t}| \sim N(0,1) \]  

where \( \sqrt{\hat{\Sigma}} \) is the covariance matrix. Positive values indicate that the random walk has lower predictive ability than other models.

Table 7 and Table 8 report the DM statistics and p-values of the DM test for the difference of MSE loss function for the S&P500 and Ibovespa daily returns, respectively:

**Table 7:** Diebold-Mariano test (Benchmark: Random walk model, one-step-ahead) for S&P500

<table>
<thead>
<tr>
<th>Model</th>
<th>DM Statistics</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RandomWalk</td>
<td>MSE 2.929777×10⁻⁸</td>
<td>MSE 1.620612×10⁻⁷</td>
</tr>
<tr>
<td>SVR Gaussian kernel</td>
<td>MSE 2.541976×10⁻⁸</td>
<td>MSE 1.221150×10⁻⁷</td>
</tr>
<tr>
<td>SVR Morletwavelet</td>
<td>MSE 2.599294×10⁻⁸</td>
<td>MSE 1.225117×10⁻⁷</td>
</tr>
<tr>
<td>GARCH-Skewed-t</td>
<td>MSE 8.647090×10⁻⁵</td>
<td>MSE 8.674276×10⁻⁵</td>
</tr>
<tr>
<td>GJR-Skewed-t</td>
<td>MSE 9.362243×10⁻⁵</td>
<td>MSE 9.417864×10⁻⁵</td>
</tr>
</tbody>
</table>
Volatility Forecasting: The Support Vector Regression Can Beat the Random Walk

<table>
<thead>
<tr>
<th>Model</th>
<th>DM Statistics</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR-Gaussian</td>
<td>5.8767</td>
<td>6.951×10⁻⁹</td>
</tr>
<tr>
<td>SVR-Morlet</td>
<td>7.0304</td>
<td>5.633×10⁻¹²</td>
</tr>
<tr>
<td>GARCH-Skewed-t</td>
<td>-27.156</td>
<td>2.2×10⁻¹⁶</td>
</tr>
<tr>
<td>GJR-Skewed-t</td>
<td>-27.13</td>
<td>2.2×10⁻¹⁶</td>
</tr>
</tbody>
</table>

Table 8: Diebold-Mariano test (Benchmark: Random walk model, one-step-ahead) for Ibovespa

For the S&P500 and Ibovespa index data, the SVR models outperform random walk, GARCH and GJR models on the skewed Student’s t-distribution at any usual confidence level. We reject the null hypothesis of equal forecast accuracy between the mean squared error (MSE) of a random walk to the MSE generated by the point forecasts for all models and series. To best of our knowledge, this is the first study to compare and show that the SVR can beat the random walk in the context of volatility forecasting. Besides, the results of this research show that the SVR algorithm can be exploited with different kernels to improve predictions of volatility.

5 Concluding Remarks

In this paper, we propose a Support Vector Regression (SVR) to forecast daily stock market volatility in United States and Brazil. To evaluate the difference between our point-forecasts and random-walk forecasts, we use the Diebold-Mariano test. The contribution of this paper are twofold. First, we show that the SVR with Gaussian and Morlet wavelet kernel can beat the random walk model in one-period-ahead volatility forecasting. Second, we show that these models outperform the traditional GARCH and GJR with a skewed Student’s t-distribution, which confirms other empirical findings. Despite the limitations of this study, we believe that the results of this work may boost the development of other models that will further improve the predictions of the SVR model in the context of volatility forecasting.

DOI: 10.24818/18423264/53.4.19.07
REFERENCES


Volatility Forecasting: The Support Vector Regression Can Beat the Random Walk