Pedro Correia Santos BEZERRA, PhD Candidate
E-mail: pcsbezerra@gmail.com
Professor Pedro Henrique Melo ALBUQUERQUE
E-mail: pedroa@unb.br[1]

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

[1]CorrespondingAuthor

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 $(x_1, y_1), \cdots, (x_n, y_n)$, where $x_i \in \mathcal{X} \subseteq \mathbb{R}^p$ 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 y_i [27]:

$$f(x) = w^T \phi(x) + b, \qquad with \qquad \phi: \mathbb{R}^p \to \mathcal{F}, w \in \mathcal{F} \tag{1}$$

where $w = [w_1, \cdots, w_n]^T$ 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 ε -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 ε-insensitive zone, slack variables (ξ_i, ξ_i^*), $i = 1,2, \cdots, n$ are introduced in the SVR primal problem of SVR:

$$Minimize: \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{2}$$

DOI: 10.24818/18423264/53.4.19.07

$$subject\ \ to \begin{cases} y_i - w^T\phi(x_i) - b \leq \varepsilon + \xi_i, \\ w^T\phi(x_i) + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0 \qquad for i = 1, \dots, n \end{cases}$$

where C is the regularization hyperparameter.

The parameters C and ε 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: $f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)K(x_i, x) + b, \qquad where \qquad 0 \leq \alpha_i, \alpha_i^* \leq C$ (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 $\Psi(x) \in L^2(\mathbb{R}^p)$:

$$\Psi_{k,a}(x) = \frac{1}{\sqrt{a}}\Psi\left(\frac{x-k}{a}\right), \qquad x \in \mathbb{R}^p \quad and \quad a, k \in \mathbb{R}. \tag{4}$$

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:

$$k(x, x') = \prod_{i=1}^{p} \Psi(\frac{x_i - k_i}{a})\Psi(\frac{x\prime_i - k\prime_i}{a}) \tag{5}$$

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

$$k(x, x') = \prod_{i=1}^{p} \Psi(\frac{x_i - x_i'}{a}) \tag{6}$$

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:

$$k(x, x') = \prod_{i=1}^{p} (cos(1.75 \times \frac{(x_i - x_{i\prime})}{a})\exp(\frac{-(x_i - x_{i\prime})^2}{2a^2})) \tag{7}$$

DOI: 10.24818/18423264/53.4.19.07

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: $k(x,x') = \exp(-\gamma \parallel x - x' \parallel^2)$, where γ 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]:

$$h_t = h_{t-1} \tag{8}$$

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]:

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

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

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]:

$$r_t = u_t + a_t \tag{10}$$

$$a_t = \sqrt{h_t} z_t, \qquad z_t \sim i.i.d(0,1) \tag{11}$$

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1} \tag{12}$$

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

DOI: 10.24818/18423264/53.4.19.07

In order to improve GARCH and capture negative and positive shocks, Glostenet al.[16] introduced the GJR model:

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

where

$$S_{t-1}^- = \begin{cases} 1, & if a_{t-1} < 0 \\ 0, & otherwise \end{cases} \qquad (14)$$

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]:

$$f(x|\iota, \nu) = \frac{2}{\iota + 1/\iota}[g(\iota(sx+m)|\nu) I_{(-\infty,0)}(x + m/s)] + \frac{2}{\iota+1/\iota}[g((sx + m)/\iota|\nu) I_{(0,+\infty)}(x + m/s)], \qquad (15)$$

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

$$m = \frac{\Gamma((\nu+1)/2)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\nu/2)}(\iota - 1/\iota), \qquad (16)$$

$$s = \sqrt{(\iota^2 + 1/\iota^2 - 1) - m^2} \qquad (17)$$

where ι is the asymmetric parameter.

3.2 SVR Algorithm for volatility forecasting

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

$$\tilde{h}_t = f(\tilde{h}_{t-1}) \qquad (18)$$

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: $\frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2$;

DOI: 10.24818/18423264/53.4.19.07

• 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

Dataset	Source	Period	Training Size	Testing Size	Total Size
S&P500	Yahoo! Finance	2008-09-12 to 2016-08-23	1400	600	2000
Ibovespa	Yahoo! Finance	2007-12-22 to 2016-01-04	1400	600	2000

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

Table 2: Descriptive statistics

	S&P500	Ibovespa
Statistics	Value	Value
Observations	2000	2000
Mean	0.00027	-0.0002
Median	0.0007	0.0000
Skewness	-0.3448	0.0825
Kurtosis	10.4765	6.5769
Std.Dev.	0.0137	0.0183

DOI: 10.24818/18423264/53.4.19.07

Minimum	-0.0947	-0.1210
Maximum	0.1096	0.1368

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

Model	LL	AIC	BIC
GARCH-Skewed-t	2891	-5.4224	-5.3999
GJR-Skewed-t	2895	-5.7987	-5.7642

Table 4: Goodness of fit for Ibovespa returns

Model	LL	AIC	BIC
GARCH-Skewed-t	2891	-5.4224	-5.3999
GJR-Skewed-t	2895	-5.7987	-5.7642

We select the parameters C, ε 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

Parameter	Range	OptimalValue	Smallest MSE
C	[0,10]	5.18400	0.0002154
ε	[0,0.1]	0.05929	0.0002146
γ	[0,1]	0.98010	0.0002115

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

Table 6: Forecasting performance

Model	S&P 500	Ibovespa

DOI: 10.24818/18423264/53.4.19.07

	MSE	MSE
RandomWalk	2.929977×10^{-8}	1.620612×10^{-7}
SVR Gaussian kernel	2.541976×10^{-8}	1.221150×10^{-7}
SVR Morletwavelet	2.599294×10^{-8}	1.225117×10^{-7}
GARCH-Skewed-t	8.647090×10^{-5}	8.674276×10^{-5}
GJR-Skewed-t	9.362243×10^{-5}	9.417864×10^{-5}

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]:

$$H_0: \frac{1}{600}\left|\tilde{h}_t - \hat{h}_{1,t}\right| - \left|\tilde{h}_t - \hat{h}_{0,t}\right| = 0$$

$$H_1: \frac{1}{600}\left|\tilde{h}_t - \hat{h}_{1,t}\right| - \left|\tilde{h}_t - \hat{h}_{0,t}\right| \neq 0$$

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]:

$$DM = \frac{1}{\sqrt{600}}\frac{1}{\sqrt{\hat{S}^2}}\sum_{t=1400}^{2000} \left|\tilde{h}_t - \hat{h}_{1,t}\right| - \left|\tilde{h}_t - \hat{h}_{0,t}\right| \sim N(0,1) \tag{19}$$

where $\sqrt{\hat{S}^2}$ 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

Model	DM Statistics	p-value

DOI: 10.24818/18423264/53.4.19.07

SVR-Gaussian	5.8767	6.951×10^{-9}
SVR-Morlet	7.0304	5.633×10^{-12}
GARCH-Skewed-t	-27.156	2.2×10^{-16}
GJR-Skewed-t	-27.13	2.2×10^{-16}

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

Model	DM Statistics	p-value
SVR-Gaussian	11.539	2.2×10^{-16}
SVR-Morlet	10.161	2.2×10^{-16}
GARCH-Skewed-t	-51.669	2.2×10^{-16}
GJR-Skewed-t	-43.635	2.2×10^{-16}

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

[1] **Adhikari, R., Agrawal, R. K. (2014**), ***A Combination of Artificial Neural Network and Random Walk Models for Financial Time Series Forecasting***. Neural Computing and Applications 24 (6), 1441–1449;
[2] **Amendola, A., Candila, V. (May 2016**), ***Evaluation of Volatility Predictions in a VaR Framework***. Quantitative Finance 16 (5), 695–709;
[3] **Bahrammirzaee, A. (2010), *A Comparative Survey of Artificial Intelligence Applications in Finance: Artificial Neural Networks, Expert System and Hybrid Intelligent Systems***. Neural Computing and Applications 19 (8), 1165–1195;
[4] **Bezerra, P. C. S., Albuquerque, P. H. M. (2017**), ***Volatility Forecasting via SVR–GARCH with Mixture of Gaussian Kernels***. Computational Management Science 14 (2), 179–196;
[5] **Bollerslev, T. (1987**), ***A Conditionally Hetroskedastic Time Series Model for Speculative Prices and Rates of Return.***
[6] **Brooks, C., Persand, G. (2003), *Volatility Forecasting for Risk Management.*** Journal of Forecasting 22 (1), 1–22;
[7] **Cao, L., Tay, F. (2003**), ***Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting.*** IEEE Transactions on Neural Networks 14 (6), 1506–1518;
[8] **Cao, L., Tay, F. E. (2001**), ***Financial Forecasting Using Support Vector Machines.***Neural Computing &Applications 10 (2), 184–192;
[9] **Cavalcante, R. C., Brasileiro, R. C., Souza, V. L., Nobrega, J. P., Oliveira, A. L. (2016**), ***Computational Intelligence and Financial Markets: A Survey and Future Directions.*** Expert Systems with Applications 55, 194–211;
[10] **Chen, S., Hardle, W. K., Jeong, K.(2010), *Forecasting Volatility with Support Vector Machine-Based GARCH Model.*** Journal of Forecasting 433 (29), 406–433;
[11] **Daubechies, I. (1992**), ***Ten Lectures of Wavelets***. *Springer-Verlag*;
[12] **Diebold, F. X., Mariano, R. S. (1995**) ***Comparing Predictive Accuracy.*** Journal of Business & Economic Statistics13 (3), 253–263;
[13] **Dimson, E., Marsh, P. (1990**), ***Volatility Forecasting without Data-snooping***. Journal of Banking and Finance 14 (2-3), 399–421;

DOI: 10.24818/18423264/53.4.19.07

[14] **Fernandez, C., Steel, M. F. J. (1998**), ***On Bayesian Modeling of Fat Tails and Skewness***. Journal of the American Statistical Association 93 (441), 359;
[15] **Fernando, P.-C., Afonso-Rodrıguez, J. A., Giner, J. (2003**), ***Estimating GARCH Models Using Support Vector Machines.*** Quantitative Finance 3, 1–10;
[16] **Glosten, L. R., Jagannthan, R., Runkle, D. E. (1993**), ***On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks***. The Journal of Finance 48 (5), 1779–1801;
[17] **Hansen, P. R., Lunde, A. (2005**), ***A Forecast Comparison of Volatility Models: Does anything Beat a GARCH(1,1)?*** Journal of Applied Econometrics 20 (February), 873–889;
[18] **Haykin, S. (1999*), Neural Networks-A Comprehensive Foundation***, 2nd Edition;
[19] **Hsu, M.-W., Lessmann, S., Sung, M.-C., Ma, T., Johnson**, **J. E**. **(Nov 2016), *Bridging the Divide in Financial Market Forecasting: Machine Learners vs. Financial Economists***. Expert Systems with Applications 61, 215–234;
[20] **Li, Y. (2014)*, Estimating and Forecasting APARCH-Skew- t Model by Wavelet Support Vector Machines.*** Journal of Forecasting 269 (March), 259–269;
[21] **Lu, C.-J., Lee, T.-S., Chiu, C.-C. (2009**), ***Financial Time Series Forecasting Using Independent Component Analysis and Support Vector Regression***. Decision Support Systems 47 (2), 115–125;
URLhttp://linkinghub.elsevier.com/retrieve/pii/S0167923609000323;
[22]. **Marcucci, J., 2005**. ***Forecasting Stock Market Volatility with Regime-Switching GARCH models.*** Studies in Nonlinear Dynamics & Econometrics 9 (4).
[23] **Mercer, J. (1909**), ***Functions of Positive and Negative Type and their Connection with the Theory of Integral Equations***. Philosophical Transactions of the Royal Society of London 209 (A), 415–446;
[24] **Patton, A. J. (2011)*, Volatility Forecast Comparison Using Imperfect Volatility Proxies***. Journal of Econometrics 160 (1), 246–256;
[25] **Santamarıa-Bonfil, G., Frausto-Solıs, J., Vazquez-Rodarte, I. (2015)*, Volatility Forecasting Using Support Vector Regression and a Hybrid Genetic Algorithm***. Computational Economics 45, 111–133;
[26] **Sermpinis, G., Stasinakis, C., Rosillo, R., de la Fuente, D.(2016)*, European Exchange Trading Funds Trading with Locally Weighted Support Vector Regression***. European Journal of Operational Research.
URL http://linkinghub.elsevier.com/retrieve/pii/S0377221716307354;
[27] **Vapnik, V. N. (1995),*The Nature of Statistical Learning Theory***. *Springer Science+Business Media*;

DOI: 10.24818/18423264/53.4.19.07

[28] **Wang, B., Huang, H., Wang, X. (2011),** ***A Support Vector Machine Based MSM Model for Financial Short-Term Volatility Forecasting***. Neural Computing and Applications 22 (1), 21–28;
[29] **Zhang, L., Zhou, W., Jiao, L. (2004),** ***Wavelet Support Vector Machine***. IEEE Transactions on Systems, Man, and Cybernetics, Part B 34 (1), 34–39.

DOI: 10.24818/18423264/53.4.19.07