Forecasting High Frequency Volatility: 
A study of the Bitcoin Market using Support Vector Regression

Abstract: This paper provides an evaluation of the predictive performance of Bitcoin volatility using high frequency data. We combined the traditional GARCH model with the Machine Learning approach to volatility estimation. The mean and volatility equations were estimated using Support Vector Regression (SVR). Furthermore, both models’ predictive ability was compared using the Diebold-Mariano test. The analysis was reiterated for both low and high frequency data. Results show that the SVR-GARCH models managed to outperform GARCH models with Normal, Student’s $t$ and Skewed Student’s $t$ distributions. For both time frequencies, the SVR-GARCH model showed statistical significance towards its superiority over GARCH. The findings can aid investors, market practitioners, financial institutions, policy makers and scholars.

Keywords: Machine Learning, Cryptocurrency, Kernel method, Volatility forecasting, Predictive accuracy

1 Introduction

The highest precision on the volatility estimation is reason for controversy among financial economists. Merton (1980) and Nelson (1992) noticed that the volatility forecasting doesn’t need a huge amount of historical data, instead, a short period of observation is enough to do so (Andersen, Bollerslev, Diebold, & Labys, 1999).

It was also noticed that, with an arbitrarily short span of data, it is possible to get an accurate volatility estimation. For this reason, the progress of volatility studies are related to the use of higher frequency data. In this sense, this work provides an evaluation of the predictive performance of Bitcoin volatility, using high frequency data. In order to estimate volatility, researchers most use the GARCH model. However, nowadays the Support Vector Regression (SVR) emerged as a strong and consistent method capable of covering multivariate and dynamic characteristic of the financial series. This method is rooted in the Structural Risk Minimization (SRM) process which aims to estimate the nonlinear data generating process trough a risk minimization and a regularization term to achieve the minimal unknown populational risk.

In this context we proposed the application of this study using the Bitcoin market. This new asset has a new combination of characteristics that makes it so unique in operation and transaction, been unable to relate completely with others markets for several reasons. First, compared to the commodities market, it does not have a great historical background and no future market to be the benchmark, but even thought, we were able to achieve an interesting result in the volatility forecast. Second, the Bitcoin use is different from the traditional currency, even if some countries use it as an official digital currency; the transactions were designed to be done directly between economic agents without the need of an intermediary institution, monetary control or accountability system. Third, the Bitcoin value and distribution is based on a P2P-network, it has no physical representation to handle it, only a string is necessary, which is called wallet, and its password, that is used to send and receive cryptocurrencies. For those reasons it is important to do a specific research in the Bitcoin market, using the traditional and new methodologies to create a model capable of understand the unique characteristics and dynamics provided by this new asset.

This paper is structured as follows: Section 2 of this paper presents the theoretical background of volatility estimating methods in high frequency data and also Bitcoin, Section 3 describes the methodology used to estimate the volatility, Section 4 addresses the empirical analysis of the estimation using the Bitcoin market price (in US dollars) between January 7th 2015 and December
Finally, Section 5 presents conclusions and remarks, showing recommendations and limitations to this approach.

2 Theoretical Background

2.1 Volatility estimation

Within the field of financial study, the learning and analysis of financial time series had risen much interest among researchers until today. Technology advance allowed the expansion of financial market and, consequently, of trading operations, increasing the availability of information, quantity of transactions carried out during the day and, mainly, monitoring real time assets prices.

Literature classifies the financial time series behavior as non-stationary. According to Deboeck (1994), Abu-Mostafa e Atya (1996) and Cao e Tay (2003) these series presents dynamisms and also the distribution shows great variations over time, without exhibiting an apparent and constant pattern in its’ disposal. In the series analysis, one can divide the data according to their frequency over time: (1) monthly frequency are usually classified as low frequency. They present a more extensive analysis on macroeconomic variables and analysis of resource allocation and on investment evaluation (Easley, López de Prado, & O’Hara, 2012); (2) data with appearance in minutes or seconds are commonly classified as high frequency. They are strongly influenced by recent events or the availability of market information, as discussed by Reboredo, Matías, e Garcia-Rubio (2012).

Due to the importance of fresh news, regarding assets price, Andersen e Bollerslev (1998) explain that financial series present an extremely volatility behavior since they incorporate expectations and reactions of economic agents in the face of events. Currently, market asset volatility forecast and estimation are highly relevant in the composition of derivatives prices, in the portfolio risk analysis and in the investment risk analysis itself, so it is of great interest among investors the development of methods that help decision making.

Particularly, the high frequency data analysis has caught the attention of many scholars and financial market agents, given the sharp increasing in worldwide financial transaction flows, which makes high frequency trading a relevant paradigm for nowadays finance, as discussed in Easley et al. (2012). Regarding volatility estimation, the use of high frequency data is seen in Çelik e Ergin (2014) and Barunik e Křehlík (2016).

The standard model used by the academy for volatility estimation is the GARCH model (Bollerslev, 1986), which was derived from the ARCH model (Engle & Bollerslev, 1986). The GARCH model main advantage is the literature of time series behavior connected to the financial sector, such as the volatility change over time, as highlighted by Marcucci et al. (2005). Furthermore, this model is broadly studied and used by financial analysts, for instance Hansen e Lunde (2005) compared 330 ARCH-type models in terms of their ability to describe the conditional variance using Diebold-Mariano (Diebold & Mariano, 1995) predictive accuracy test and found no evidence that a GARCH(1,1) was outperformed by more sophisticated models in their exchange rates analysis, but they concluded that GARCH(1,1) was inferior to models that can accommodate a leverage effect in the stock’s market analysis.

Recently, other techniques to predict volatility have been discussed, a strong and consistent method used is the Support Vector Regression (SVR), which covers the nonlinearity and dynamic characteristic of the financial series. Gavrishchaka e Ganguli (2003), Gavrishchaka e Banerjee (2006) and Chen, Jeong, e Härdle (2008) have already presented empirical results regarding the efficiency and superior predictability of volatility using SVR when compared to the GARCH benchmark and other techniques, such as neural networks and technical analysis (Barunik &
In the last years, there has been a great interest in using traditional methodologies, such as the GARCH model and new ones, like SVR and neural networks, to estimate new assets volatility in the stock or commodities market. The stock market already dispose a great amount of methodologies and models capable of forecasting variables such as price and volatility.

Literature suggests that the most suitable model to forecast this kind of asset is using the future market as a benchmark (French, 1986; Reeve & Vigfusson, 2011; Schwartz & Smith, 2000). In this case, the commodity future price already has in it’s composition the volatility and the market expectations, in a way that it is beating the random walk model.

2.2 Cryptocurrencies and Bitcoin

While a consolidate model is still in discussion, in 2009 a new kind of asset appeared in the market, the Bitcoin, leading the world to analyze this newcomer and try to figure out its’ place and dynamics. Satoshi Nakamoto (Nakamoto, 2008) proposed Bitcoin in order to be an easily way to make transactions over the Internet, which works globally, faster and is independent from an institution to operate it. It is based on a peer-to-peer book of properties that is governed by mathematical restrictions, which only allows real transactions to be written in it. Every transaction and the agents involved are registered, any change over the Bitcoin data nullify its uses in future transactions. Until now the average number of transactions registered are 288,155 per day (BLOCKCHAIN, 2017a; Desk, 2017a), a growth of 40% compared to the same period of 2016, in which registered around 205,246 transactions per day.

This concept can be relatively new to financial theory, as Bitcoin has no association with any authority, has no physical representation, is infinitely divisible and is built using the most sophisticated mathematics and computation techniques. For those reasons, cryptocurrencies can give fear to any specialist that is not used to this kind of ’money’ and is astonishing to the market world widely.

One of the main features of Bitcoin for market practitioners is its ability to maintain value reserve. Although there is still no consensus on this regard, recent academic studies such as Bouri, Gupta, Tiwari, Roubaud, et al. (2016) present evidences that indicate that Bitcoin can be indeed used as a hedge to market specific risk. Dyhrberg (2016a) analyzes the volatility of Bitcoin in comparison to US Dollar and Gold – traditionally regarded as “safe” value reserves – using GARCH (1,1) and EGARCH (1,1). The paper concluded that Bitcoin bears significant similarities to both assets, specially concerning hedging capabilities and volatility reaction to news, suggesting that Bitcoin can be a useful tool for portfolio management, risk analysis and market sentiment analysis. As proof of the recent acknowledgment of the hedge propriety of Bitcoin, economic agents already invested a total of 19 billion dollars in the cryptocurrency until March of 2017, suggesting an increase of the usage, popularization and trust in the Bitcoin (BLOCKCHAIN, 2017b; Desk, 2017b).Dyhrberg (2016a) also points out that the Bitcoin reactions may be quicker than gold and Dollar, thus substantiating the analysis of both high and low data frequencies in this paper. The author replicates the study using T-GARCH(1,1) and find similar conclusions (Dyhrberg, 2016b).

When it comes to comprehending the market and the economic agents involved, Bitcoin uses are still very restricted to some countries and activities. Usually, the one’s that are open to this new asset are technological or innovation centers, which are able to understand the potential and advantages of the cryptomoney. Estonia, United Sates, Denmark, Sweden, South Korea, Netherlands, Finland, Canada, United Kingdom and Australia are ten countries friendly to the uses of Bitcoin.
The asset flexibility in transactions (since there are few regularizations of this new market) and the high level of the cryptography gives the coin enough trust to be used instead of cash, like in Denmark. Even with the importance and its uses, the worldwide acceptance is still far from happening and the impact over the cryptocurrency dynamics in the present is inevitable. The first visible affect is the Bitcoin low volatility, given the aspect of an asset reserve value such as gold or dollar. Yet, there is not enough literature review that discusses or presents efficient methodologies to estimate the new Bitcoin market behavior around the world. Some authors already conducted studies about this new market, (Bouri, Azzi, & Haubo Dyhrberg, 2016; Dyhrberg, 2016a; Xiong, Bao, & Hu, 2014; Yermack, 2013), but they are still restricted to the unknown aspects of the behavior and/or acceptance of Bitcoin.

Analyzing the present techniques and concepts used in the scientific literature for Bitcoin estimation, there are few studies that use the Machine Learning approach in forecasting its volatility, even with well known methodologies such as the GARCH model (X. Li & Wang, 2016). Observing this gap in the literature, the present study represents a ground breaking research since it considers the Bitcoin dynamics close to the behavior of stock and derivatives than commodities (Dyhrberg, 2016a), according to researches developed until now. Since Bitcoin market operates without a supervising organization or entity to ensure it’s value or conduct the transactions, as we see in stock exchanges institutions, the economic agents may found difficult to predict or price this asset, since there is no history or methodology established in the academic or business environment. The construction of a machine, capable to forecast the risk variable, interpreted as volatility, represents an improvement in the studies and business operation using this kind of currency. Furthermore, it may be the first step in the construction of pricing models for the Bitcoin and possibles derivatives of it.

Researches like Urquhart (2016) indicate that Cryptocurrencies such as Bitcoin still present informational inefficiency, even though it has been showing a trend towards efficiency. Thus, separating the volatility analysis of Bitcoin prices considering different time horizons may provide a better understanding regarding this finding. Additionally, studies like Dowd (2014) states that cryptocurrencies are susceptible to speculative bubbles that can mine its fundamental value. Therefore, the application of machine learning methods seems very attractive to capture underlying patterns regarding those issues, thus providing a more comprehensive and accurate view of this new agenda. Bearing in mind those issued related to financial aspects of cryptocurrencies, this paper combined the machine learning approach with volatility forecasting, splitting the analysis into datasets of high and low frequencies.

3 Methodology

3.1 GARCH

Given \( P_t \) the observed price at time \( t \), the GARCH(1,1) (Generalized Auto-regressive Conditional Heteroskedasticity) model can be summarized as follows:

\[
    r_t = \mu_t + \epsilon_t
\]

where \( r_t = \log\left( \frac{P_t}{P_{t-1}} \right) \) is the log return and \( \epsilon_t \) is a stochastic term with zero mean. The mean equation can be defined by:

\[
    \mu_t = \gamma_0 + \gamma_1 r_{t-1}
\]
the volatility equation is given by:

\[ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \]  

such that \( V(\epsilon_t) = h_t \).

Since the actual volatility is not directly obtained, a \textit{ex-post} proxy volatility is needed to estimate the volatility through the SVR. Following Chen, Härdle, e Jeong (2010) and Bezerra e Albuquerque (2016), we defined the proxy volatility as:

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

where \( \tilde{r} \) is the arithmetic mean of the log returns.

The GARCH (1,1) is one of the main models regarding volatility estimation in finance, given its easy estimation, low number of parameters and ability to capture volatility clusters, since its conditional mean is constant, while its conditional variance is not (Hansen & Lunde, 2005). Even when \( \epsilon_t \) is assumed to be normally distributed, GARCH’s presents a fat-tailed behavior in comparison to the Gaussian distribution, even though still not quite incorporating the financial data’s stylized facts (Cont, 2001). Thus, it is common to assume that \( \epsilon_t \) follows non-Gaussian distributions, in order to fit better to the financial data. In this paper, we estimated the GARCH (1,1) assuming 3 distributions for \( \epsilon_t \): Gaussian, Student’s \( t \) and Skewed Student’s \( t \). Although the literature has proposed a wide variety of distributions to be assumed in the GARCH model, authors like Sun e Zhou (2014) argue that the Student’s \( t \) is enough to give a good fit to the financial data’s heavy tail behavior, while innovations concerning GARCH distributions does not seem robust.

### 3.2 Support Vector Regression

Despite its popularity, the GARCH (1,1) still considers linear functional forms in its estimation, which motivates the introduction of nonlinear structural forms. That is one of the main contributions of machine learning and kernel methods, as many studies showed (M. Li & Suohai, 2013; Shen, Chao, & Zhao, 2015) that the introduction of nonlinear interactions can significantly boost the explanatory power and the forecasting ability of many models applied to financial contexts, including volatility estimation (Chen et al., 2010; Premanode & Toumazou, 2013; Santamaría-Bonfil, Frausto-Solís, & Vázquez-Rodarte, 2015). Regarding high frequency forecasting, that issue is also noted (Santos, da Costa, & dos Santos Coelho, 2007).

Therefore, we used Support Vector Regression for the estimation of the GARCH’s mean and volatility equations – described in equations 2 and 3. Instead of using a standard linear regression, we introduced nonlinearities, hoping to provide a better fit to the data. Concerning the high volatility of Bitcoin data – specially in high frequency transactions – this approach seems particularly attractive.

The Support Vector Regression (SVR) is a regression model which aims to (Drucker, Burges, Kaufman, Smola, & Vapnik, 1997; Vapnik, 1995) find a decision function which is the best approximation of a set of observations, bearing in mind the middle ground between a good power of generalization and an overall stable behavior, in order to make good out-of-sample inferences. Associated with these two desirable features, there are two corresponding problems in regression models, constituting the so called “bias-variance dilemma”. To perform the regularization of the decision function, two parameters are added: a band of tolerance \( \varepsilon \), to avoid over-fitting; and a penalty \( C \) to the objective function, for points that lies outside this confidence interval for an amount \( \xi > 0 \).
Therefore, the values $|y_i - f(x_i)| \leq \varepsilon$, where $f(.)$ is the SVR decision function, such that $\mathbb{E}(y)$ are considered to be statistically equal to $y$.

The SVR defined from the addition of these two parameters is known as $\varepsilon$–SVR. The loss function implied in the construction of the $\varepsilon$–SVR is the $\varepsilon$-insensitive loss function (Vapnik, 1995), $L_\varepsilon[y_i, f(x_i)]$, given by:

$$L_\varepsilon[y_i, f(x_i)] = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| > \varepsilon, \\ 0, & |y_i - f(x_i)| \leq \varepsilon. \end{cases} \tag{5}$$

It is worth noting that the $\varepsilon$-insensitive loss function is not the only possible way to define penalties for the SVR; extensions that include different penalty structures includes the $\nu$–SVR (Chang & Lin, 2002). The $\varepsilon$–SVR formulation was chosen for this paper because it is the most commonly form used in the exchange rate forecasting literature, and requires lesser computational time to perform the optimization.

In order to introduce the nonlinear interactions in the regression estimation, a mapping $\varphi$ is applied, such that the objective function to be optimized for the $\varepsilon$–SVR is formulated as follows:

$$\begin{align*}
&\text{Minimize :}\quad \frac{1}{2} w^T w + C \xi^T 1 + C \xi^* 1 \\
&\text{Subject to :}\quad \Phi w + w_0 - y \leq \varepsilon 1 + \xi \\
&\quad \quad \quad \quad \quad \quad \quad y - \Phi w - w_0 \leq \varepsilon 1 + \xi^* \\
&\quad \quad \quad \quad \quad \quad \quad \xi, \xi^* \geq 0
\end{align*} \tag{6}$$

where $\Phi$ is a $n \times q$ matrix created by the Feature Space, i.e., the original explanatory variables $X_{(n \times p)}$ is mapped through the $\varphi(x)$ function, $w$ is a vector of parameters to be estimated, $C$ and $\varepsilon$ are hyper-parameters and $\xi, \xi^*$ are slack variables in the Quadratic Programming Problem.

In other words, $w_{(q \times 1)}$ is the vector of the angular coefficients of the decision hyperplane in $\mathbb{R}^q$; $w_0 \in \mathbb{R}$ is the linear coefficient (intercept) of the decision hyperplane in $\mathbb{R}^q$; $\Phi_{(n \times q)}$ is the augmented matrix of observations, after the original data being transformed by $\varphi$; $y_{(n \times 1)}$ is the vector that provides the dependent variable values of the observed points; $C \in \mathbb{R}$ is the cost of error; $\varepsilon > 0$ is the tolerance band that defines the confidence interval for which there is no penalty; $\xi^*_{(n \times 1)}$ is the vector concerning points above the tolerance band; and $\xi_{(n \times 1)}$ is the vector concerning points below the tolerance band.

After some algebraic manipulations, it can be shown that the decision function of the $\varepsilon$–SVR can be written as:

$$f(x_i) = w^T \varphi(x) - w_0 = \sum_{j=1}^{n} \kappa(x_i, x_j)(\lambda_j^* - \lambda_j) - w_0 \tag{7}$$

where $\kappa(x_i, x_j) = \varphi^T(x_i) \cdot \varphi(x_j) \in \mathbb{R}, i, j = 1, 2, 3..., n$ is the kernel function. Since $\varphi$ transforms the original data to a higher dimension, which can even be infinite, the use of the kernel function prevents the need to explicitly compute the functional form of $\varphi(x)$; instead, $\kappa$ computes the inner product of $\varphi$, a term that appears in SVR’s dual formulation (Drucker et al., 1997), this is known as the kernel trick. In this paper, we used the Gaussian Kernel as $\kappa$, whose expression is given as:

$$\kappa(x_i, x_j) = \exp \left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right), \sigma > 0 \tag{8}$$
The Gaussian Kernel is the most widely used Kernel in the machine learning literature, enjoying huge popularity in various knowledge fields, since this function is able to induce an infinite dimensional feature space while depending on only one scattering parameter $\sigma$.

### 3.3 SVR-GARCH

The SVR-GARCH is the joining result of the GARCH model structure and the nonlinearities introduced by the kernel function via SVR. Santamaría-Bonfil et al. (2015) presented empirical evidences that the SVR-GARCH managed to outperform standard GARCH’s predictions, showing better ability to approximate the nonlinear behavior of financial data and stylized facts, such as heavy tails and volatility clusters. The specification of SVR-GARCH (1,1) is the same of the conventional GARCH (1,1), but the mean and volatility equations were estimated via SVR, such that:

\[
\begin{align*}
    r_t &= f_m(r_{t-1}) + \epsilon_t \\
    h_t &= f_v(h_{t-1}, \epsilon_{t-1}^2)
\end{align*}
\]

where $f_m(.)$ is the SVR decision function for the mean equation 2, and $f_v(.)$ is the SVR decision function for the volatility equation 3.

Depending of which parameters $\epsilon$, $C$ and $\sigma$ are set to the SVR formulation, a different decision function is obtained. In order to decide the “better” combination of parameters, we did a grid search for those three parameters for both mean and volatility equations, and evaluated the Root Mean Square Error (RMSE) of each decision function. The model is first estimated with a subset of the data (known as training dataset) and then the estimated model is used to forecast both mean and volatility in another subset (validation dataset). The combination that minimizes the RMSE for the validation dataset was chosen as the best one. The search intervals for each parameter are displayed in Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Search interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon$</td>
<td>$[0.05, 0.1, \ldots, 0.95, 1]$</td>
</tr>
<tr>
<td>$C$</td>
<td>$[0.5, 1, \ldots, 4.5, 5]$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$[0.05, 0.1, \ldots, 1.95, 2]$</td>
</tr>
</tbody>
</table>

*Table 1: Search intervals used for the parameters’ training*

### 4 Empirical analysis

For the empirical test, we used the Bitcoin market price (in US dollars) between January 7th 2015 and December 31th 2016. The data were collected from Bitcoin Charts (http://bitcoincharts.com/) originally with 388095 observations from 2009 to 2017. We used the daily basis for the low frequency analysis (722 observations) and the hour periodicity for the high frequency estimation (17011 observations).

Both databases were partitioned into three mutually exclusive fractions: training set, validation set and test set. The purpose of this segmentation is to allow the machine learning algorithm to test its predictive performance on data that were not used priorly, in order to better evaluate the real explanatory power of the found decision function when dealing with new data. For this
paper, we choose to allocate 50% of the database for the training set (360 time periods for low frequency database, 8485 time periods for high frequency database), 20% for the validation set (144 time periods for low frequency database, 3410 time periods for high frequency database) and the remaining 30% for the test set (215 time periods for low frequency database, 5096 time periods for high frequency database).

The definition of time periods belonging to each of the subsets was carried out randomly in order to avoid the time trend bias, which might arise due to macroeconomic cycles – i.e., the predictive ability could be hindered or improved as a result of a systemic trend, instead of being influenced by the quality of the model itself. The draft was carried in the low frequency database, then the same days chosen for each of the subsets were maintained for the high frequency database, in order to effectively evaluate the forecasting ability change between the two scenarios.

Firstly, the optimization of the SVR algorithm was applied to the training set, by performing a grid search for each one of the associated parameters for both mean and volatility equations in low and high frequencies. The search ranges for the parameters $\varepsilon$, $C$ and $\sigma$ are listed in Table 1.

Based on each combination of parameters applied to the training set, the accuracy of each optimal obtained decision function was checked for the validation set, by the error metric RMSE, defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (s_t - \hat{s}_t)^2}$$ \hspace{1cm} (11)

Each decision function obtained in the training set was fed with data the validation set to compute the prediction of the dependent variable for these data. This forecast was then confronted with the actual values observed in the validation set, and the RMSE between predicted and observed values was calculated. Repeating the process for every parameter combination, the optimal combination is that which minimizes the RMSE associated with its prediction. After that, the optimal parameters were applied to fit the model for the test set, and then compared with the results generated by GARCH models. For this step, we considered the error metrics RMSE (square root of the mean square error), defined in equation 11; and MAE (mean absolute error), whose expression is given by:

$$MAE = \frac{1}{N} \sum_{t=1}^{N} |s_t - \hat{s}_t|$$ \hspace{1cm} (12)

Finally, the Diebold-Mariano Diebold e Mariano (1995) predictive accuracy test was applied for the three GARCH models in both low and high frequencies, using SVR-GARCH as benchmark using the test dataset.

4.1 Results and discussion

The optimal parameters for the SVR-GARCH training and validation steps are displayed below:
The optimal parameters were used to fit the SVR-GARCH (1,1), for both low and high frequency data, alongside the GARCH (1,1) with three different distributions for the error term $\epsilon_t$. The volatility predictions’ RMSE and MAE for each model are displayed in Table 4.

<table>
<thead>
<tr>
<th>Data</th>
<th>Model</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low frequency</td>
<td>Normal GARCH (1,1)</td>
<td>0.03369</td>
<td>0.02799</td>
</tr>
<tr>
<td></td>
<td>Student’s $t$ GARCH (1,1)</td>
<td>0.03469</td>
<td>0.02917</td>
</tr>
<tr>
<td></td>
<td>Skewed Student’s $t$ GARCH (1,1)</td>
<td>0.03466</td>
<td>0.02894</td>
</tr>
<tr>
<td></td>
<td>SVR-GARCH (1,1)</td>
<td>0.01818</td>
<td>0.01518</td>
</tr>
<tr>
<td>High frequency</td>
<td>Normal GARCH (1,1)</td>
<td>0.01089</td>
<td>0.00585</td>
</tr>
<tr>
<td></td>
<td>Student’s $t$ GARCH (1,1)</td>
<td>0.01070</td>
<td>0.00533</td>
</tr>
<tr>
<td></td>
<td>Skewed Student’s $t$ GARCH (1,1)</td>
<td>0.01067</td>
<td>0.00532</td>
</tr>
<tr>
<td></td>
<td>SVR-GARCH (1,1)</td>
<td>0.00457</td>
<td>0.00182</td>
</tr>
</tbody>
</table>

Table 4: Forecasting performance for low and high frequency test set data

The results show that the SVR models presented a significantly lower value for both error metrics RMSE and MAE in comparison to all three GARCH models. For the low frequency dataset, the volatility levels and the error metrics for all models were higher than the high frequency one. This result is consistent with the finding in the literature: as seen in Xie e Li (2010), the RMSE for the volatility forecasting tend to decrease as the frequency increases.

However, while the error metrics RMSE and MAE do give a good overview of the performance of the analyzed models, they fail to provide stronger statistical evidences about the
predictive superiority of SVR-GARCH over conventional GARCH models. Bearing that in mind, we applied the Diebold-Mariano test, which compares the predictive ability of a given model against a benchmark. The null hypothesis of this test is defined as:

\[ H_0 : \frac{1}{N} |\tilde{h}_t - \hat{h}_{i,t}| - |\tilde{h}_t - \hat{h}_{\text{benchmark},t}| = 0 \]  

(13)

where \( N \) is the number of observations in the test set. Therefore, this test evaluates whether the predictive ability of the \( i \)-th model is statistically equal to the benchmark model. The rejection of the null hypothesis of the Diebold-Mariano test indicates predictive superiority in favor of one of the models, while its non-rejection implies that the two models are statistically equivalent.

In this paper, the benchmark is the SVR-GARCH (1,1), tested against the GARCH (1,1) with Normal, Student’s \( t \) and Skewed Student’s \( t \) distributions. The test was carried for both low and high frequencies in its two-tailed version. The test statistics and p-values are displayed in Table 5.

<table>
<thead>
<tr>
<th>Data</th>
<th>Model</th>
<th>Test statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low frequency</td>
<td>Normal GARCH (1,1)</td>
<td>3.84050</td>
<td>0.00013</td>
</tr>
<tr>
<td></td>
<td>Student’s ( t ) GARCH (1,1)</td>
<td>3.50970</td>
<td>0.00046</td>
</tr>
<tr>
<td></td>
<td>Skewed Student’s ( t ) GARCH (1,1)</td>
<td>3.51560</td>
<td>0.00047</td>
</tr>
<tr>
<td>High frequency</td>
<td>Normal GARCH (1,1)</td>
<td>3.61930</td>
<td>0.00029</td>
</tr>
<tr>
<td></td>
<td>Student’s ( t ) GARCH (1,1)</td>
<td>2.62970</td>
<td>0.00855</td>
</tr>
<tr>
<td></td>
<td>Skewed Student’s ( t ) GARCH (1,1)</td>
<td>2.18750</td>
<td>0.02872</td>
</tr>
</tbody>
</table>

*Table 5: Forecasting performance for low and high frequency test set data*

The results of the Diebold-Mariano test show that the SVR-GARCH (1,1) managed to outperform the conventional GARCH models, given the small p-values for both low and high frequency data, although the p-values were slightly higher for the high frequency case, in which the statistical difference between GARCH and SVR-GARCH is smaller. At the usual 95% confidence level, all null hypothesis would be rejected, favoring the predictive superiority of SVR models. Particularly, the Skewed Student’s \( t \) distribution, which performed slightly better then Normal and Unskewed \( t \) GARCHs (as seen in Table 4, had a much higher p-value for the Diebold-Mariano test in the high frequency case, even though still strongly suggesting the superiority of the SVR-GARCH model.

Comparing the obtained results for the two data frequencies, we observed that both error metrics and Diebold-Mariano test p-values were lower in the low frequency dataset, indicating that the SVR-GARCH has indeed predictive superiority over GARCH models, but the statistical difference between those models are slightly lesser concerning high frequency data, albeit still rejecting the Diebold-Mariano test’s null hypothesis quite emphatically.

Based on the results and the explanations set up above, it is of great remark to the financial field and to other researches to better understand the unique characteristics and dynamics of the Bitcoin.

5 Conclusion and remarks

This paper evaluated the predictive performance for Bitcoin volatility combining the traditional GARCH model with the Machine Learning approach, by estimating the mean and volatility
equations using Support Vector Regression. Furthermore, we compared both models’ predictive ability with the Diebold-Mariano test, and reiterated the analysis for both low and high frequency data. The results show that the SVR-GARCH models managed to outperform all three GARCHs – with Normal, Student’s t and Skewed Student’s t distributions, as seen by the value of error metrics RMSE and MAE and the Diebold-Mariano test p-value.

The findings of this paper have the potential to aid scholars and market practitioners two main contributions. First, insights regarding financial time series’ forecasting and provide a overview of the Bitcoin market features; second, the performance of the SVR is proved to be better than the actual benchmark, based on GARCH (1,1), for volatility forecast. The outcome of this research is a tool capable of estimate the risk for the Bitcoin in the future and can be used as a risk management for portfolios, as proposed by Dyhrberg (2016a). Future researches are encouraged to replicate this study for other financial assets’ volatility estimation, as well as to consider other distributions for the GARCH models’ error term – such as Generalized Pareto Distribution (McNeil & Frey, 2000) – and other well known extension models for GARCH, such as EGARCH (Nelson, 1991) and GJR-GARCH (Glosten, Jagannathan, & Runkle, 1993). As the fact that those models were not included in this paper constitutes a limitation, verifying whether the inclusion of SVR estimation can contribute to better volatility predictions is an attractive and relevant issue in the finance literature.

Bearing in mind the huge popularity and prominence of Machine Learning methods many scientific fields, finance included, testing for different extensions of SVR – like Chang e Lin (2002)’s ν–SVR – is also quite desirable. Finally, replications with different time periods and frequencies (e.g.: even higher frequency data, by minutes or even seconds) can contribute further for this research agenda.

References


