

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

Yaohao Peng  
Mariana Rosa Montenegro  
Ana Julia Akaishi Padula  
Jader Martins Camboim de Sá  
Pedro Henrique Melo Albuquerque

*University of Brasilia, Campus Darcy Ribeiro, Brasilia, 70910-900, Brazil*

---

## Abstract

This paper provides an evaluation of the predictive performance of Bitcoin

*Keywords:* Machine Learning, Cryptocurrency, Kernel method, Cross validation, SVR-GARCH, Volatility forecasting, Data Mining, Predictive accuracy

---

## 1. Introduction

2 The highest precision on the volatility estimation is reason for controversy  
3 among financial economists. [1] and [2] noticed that the volatility forecasting  
4 doesn't need a huge amount of historical data, instead, a short period of  
5 observation is enough to do so[3].

6 It was also noticed that, with an arbitrarily short span of data, it is pos-  
7 sible to get an accurate volatility estimation. For this reason, the progress  
8 of volatility studies are related to the use of higher frequency data. In this  
9 sense, this work provides an evaluation of the predictive performance of Bit-  
10 coin volatility, using high frequency data. In order to estimate volatility,  
11 researchers most use the GARCH model. However, nowadays the Support  
12 Vector Regression (SVR) emerged as a strong and consistent method capa-  
13 ble of covering multivariate and dynamic characteristic of the financial series.  
14 This method is rooted in the Structural Risk Minimization (SRM) process  
15 which aims to estimate the nonlinear data generating process through 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 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 31th 2016. 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 the 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 [4], [5] and [6] these series presents dynamisms and also the dis-

52 tribution shows great variations over time, without exhibiting an apparent  
53 and constant pattern in its' disposal. In the series analysis, one can divide  
54 the data according to their frequency over time: (1) monthly frequency are  
55 usually classified as low frequency. They present a more extensive analy-  
56 sis on macroeconomic variables and analysis of resource allocation and on  
57 investment evaluation [7]; (2) data with appearance in minutes or seconds  
58 are commonly classified as high frequency. They are strongly influenced by  
59 recent events or the availability of market information, as discussed by [8].

60 Due to the importance of fresh news, regarding the price of assets, [9]  
61 explain that financial series present an extremely volatility behavior since  
62 they incorporate expectations and reactions of economic agents in the face of  
63 events. Currently, market asset volatility forecast and estimation are highly  
64 relevant in the composition of derivatives prices, in the portfolio risk analysis  
65 and in the investment risk analysis itself, so it is of great interest among  
66 investors the development of methods that help decision making.

67 Particularly, the high frequency data analysis has caught the attention  
68 of many scholars and financial market agents, given the sharp increasing in  
69 worldwide financial transaction flows, which makes high frequency trading  
70 a relevant paradigm for nowadays finance, as discussed in [7]. Regarding  
71 volatility estimation, the use of high frequency data is seen in [10] and [11].

72 The standard model used by the academy for volatility estimation is the  
73 GARCH model, created by [12], which was derived from the [13] ARCH  
74 model. The GARCH model main advantage is the literature of the time se-  
75 ries behavior connected to the financial sector, such as the volatility change  
76 over time, as highlighted by [14]. Furthermore, this model is broadly studied  
77 and used by financial analysts, for instance [15] compared 330 ARCH-type  
78 models in terms of their ability to describe the conditional variance using  
79 Diebold-Mariano [16] predictive accuracy test and found no evidence that a  
80 GARCH(1,1) was outperformed by more sophisticated models in their ex-  
81 change rates analysis, but they concluded that GARCH(1,1) was inferior to  
82 models that can accommodate a leverage effect in the stock's market analysis.

83 Recently, other techniques to predict volatility have been discussed, a  
84 strong and consistent method used is the Support Vector Regression (SVR),  
85 which covers the nonlinearity and dynamic characteristic of the financial se-  
86 ries. [17], [18] and [19] have already presented empirical results regarding the  
87 efficiency and superior predictability of volatility using SVR when compared  
88 to the GARCH benchmark and other techniques, such as neural networks  
89 and technical analysis [11].

90 In the last years, there has been a great interest in using traditional  
91 methodologies, such as the GARCH model and new ones, like SVR and  
92 neural networks, to estimate new assets volatility in the stock or commodities  
93 market. The stock market already dispose a great amount of methodologies  
94 and models capable of forecasting variables such as price and volatility.

95 Literature suggests [20], [21], [22] that the most suitable model to forecast  
96 this kind of asset is using the future market as a benchmark. In this case,  
97 the commodity future price already has in it's composition the volatility and  
98 the market expectations, in a way that it is beating the random walk model.

## 99 2.2. *Cryptocurrencies and Bitcoin*

100 While a consolidate model is still in discussion, in 2009 a new kind of  
101 asset appeared in the market, the Bitcoin, leading the world to analyze this  
102 newcomer and try to figure out its' place and dynamics. Satoshi Nakamoto  
103 [23] proposed Bitcoin in order to be an easily way to make transactions  
104 over the Internet, which works globally, faster and is independent from an  
105 institution to operate it. It is based on a peer-to-peer book of properties that  
106 is governed by mathematical restrictions, which only allows real transactions  
107 to be written in it. Every transaction and the agents involved are registered,  
108 any change over the Bitcoin data nullify its uses in future transactions, until  
109 now the average number of transactions registered are 288,155 per day [24]  
110 [25], a growth of 40% compared to the same period of 2016 where were  
111 registered around 205,246 transactions per day.

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

118 One of the main features of Bitcoin for market practitioners is its abil-  
119 ity to maintain value reserve. Although there is still no consensus on this  
120 regard, recent academic studies such as Bouri et al. [26] present evidences  
121 that indicate that Bitcoin can be indeed used as a hedge to market specific  
122 risk. [27] analyzes the volatility of Bitcoin in comparison with the US Dollar  
123 and Gold – traditionally regarded as “safe” value reserves – using GARCH  
124 (1,1) and EGARCH (1,1), the paper concluded that Bitcoin bears signifi-  
125 cant similarities to both assets, specially concerning hedging capabilities and  
126 volatility reaction to news, suggesting that Bitcoin can be a useful tool for

127 portfolio management, risk analysis and market sentiment analysis. As proof  
128 of the recent acknowledgment of the hedge propriety of the Bitcoin, economic  
129 agents already invested a total of 19 billion dollars in the cryptocurrency un-  
130 til March of 2017, suggesting an increase of the usage, popularization and  
131 trust in the Bitcoin [28] [29]. [27] also points out that the Bitcoin reactions  
132 may be quicker than gold and Dollar, thus substantiating the analysis of both  
133 high and low data frequencies in this paper. The author replicates the study  
134 using T-GARCH(1,1) and find similar conclusions [30].

135 When it comes to comprehending the market and the economic agents  
136 involved, Bitcoin uses are still very restricted to some countries and activi-  
137 ties. Usually, the one's that are open to this new asset are technological or  
138 innovation centers, which are able to understand the potential and advan-  
139 tages of the cryptomoney. Estonia, United Sates, Denmark, Sweden, South  
140 Korea, Netherlands, Finland, Canada, United Kingdom and Australia are  
141 ten countries friendly to the uses of Bitcoin.

142 The asset flexibility in transactions (since there are few regularizations of  
143 this new market) and the high level of the cryptography gives the coin enough  
144 trust to be used instead of cash, like in Denmark. Even with the importance  
145 and its uses, the worldwide acceptance is still far from happening and the  
146 impact over the cryptocurrency dynamics in the present is inevitable. The  
147 first visible affect is the Bitcoin low volatility, given the aspect of an asset  
148 reserve value such as gold or dollar. Yet, there is not enough literature review  
149 that discusses or presents efficient methodologies to estimate the new Bitcoin  
150 market behavior around the world. Some authors already conducted studies  
151 about this new market, [31], [32], [27] and [33] , but they are still restricted  
152 to the unknown aspects of the behavior and/or acceptance of Bitcoin.

153 Analyzing the present techniques and concepts used in the scientific lit-  
154 erature for Bitcoin estimation, there are few studies that use the Machine  
155 Learning approach in forecasting its volatility, even with well known method-  
156 ologies such as the GARCH model [34]. Observing this gap in the literature,  
157 the present study represents a ground breaking research since it considers  
158 the Bitcoin dynamics close to the behavior of stock and derivatives than  
159 commodities [27], according to researches developed until now. Since Bitcoin  
160 market operates without a supervising organization or entity to ensure it's  
161 value or conduct the transactions, as we see in stock exchanges institutions,  
162 the economic agents may found difficult to predict or price this asset, since  
163 there is no history or methodology established in the academic or business  
164 environment. The construction of a machine, capable to forecast the risk

variable, interpreted as volatility, represent 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 possible derivatives of it.

Researches like [35] 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 [36] 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 issues 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:

$$r_t = \mu_t + \epsilon_t \quad (1)$$

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

$$\mu_t = \gamma_0 + \gamma_1 r_{t-1} \quad (2)$$

the volatility equation is given by:

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \quad (3)$$

such that  $\mathbb{V}(\epsilon_t) = h_t$ .

190 Since the actual volatility is not directly obtained, a *ex-post* proxy volatil-  
 191 ity is needed to estimate the volatility through the SVR. Following [37] and  
 192 [38], we defined the proxy volatility as:

$$\tilde{h}_t = (r_t - \bar{r})^2 \quad (4)$$

193 where  $\bar{r}$  is the arithmetic mean of the log returns.

194 The GARCH (1,1) is one of the main models regarding volatility esti-  
 195 mation in finance, given its easy estimation, low number of parameters and  
 196 ability to capture volatility clusters, since its conditional mean is constant,  
 197 while its conditional variance is not [15]. Even when  $\epsilon_t$  is assumed to be  
 198 normally distributed, GARCH's presents a fat-tailed behavior in comparison  
 199 to the Gaussian distribution, even though still not quite incorporating the  
 200 financial data's stylized facts [39]. Thus, it is common to assume that  $\epsilon_t$  fol-  
 201 lows non-Gaussian distributions, in order to fit better to the financial data.  
 202 In this paper, we estimated the GARCH (1,1) assuming 3 distributions for  
 203  $\epsilon_t$ : Gaussian, Student's  $t$  and Skewed Student's  $t$ . Although the literature  
 204 has proposed a wide variety of distributions to be assumed in the GARCH  
 205 model, authors like [40] argue that the Student's  $t$  is enough to give a good  
 206 fit to the financial data's heavy tail behavior, while innovations concerning  
 207 GARCH distributions does not seem robust.

### 208 3.2. Support Vector Regression

209 Despite its popularity, the GARCH (1,1) still considers linear functional  
 210 forms in its estimation, which motivates the introduction of nonlinear struc-  
 211 tural forms. That is one of the main contributions of machine learning and  
 212 kernel methods, as many studies showed [41, 42] that the introduction of  
 213 nonlinear interactions can significantly boost the explanatory power and the  
 214 forecasting ability of many models applied to financial contexts, including  
 215 volatility estimation [37, 43, 44]. Regarding high frequency forecasting, that  
 216 issue is also noted [45].

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

223 The Support Vector Regression (SVR) is a regression model which aims  
 224 to [46, 47] 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  $|y_i - f(\mathbf{x}_i)| \leq \varepsilon$ , where  $f(\cdot)$  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 [46],  $L_\varepsilon[y_i, f(\mathbf{x}_i)]$ , given by:

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

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 [48]. 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:

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

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

In other words,  $\mathbf{w}_{(q \times 1)}$  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



250 decision hyperplane in  $\mathbb{R}^q$ ;  $\Phi_{(n \times q)}$  is the augmented matrix of observations,  
 251 after the original data being transformed by  $\varphi$ ;  $\mathbf{y}_{(n \times 1)}$  is the vector that  
 252 provides the dependent variable values of the observed points;  $C \in \mathbb{R}$  is the  
 253 cost of error;  $\varepsilon > 0$  is the tolerance band that defines the confidence interval  
 254 for which there is no penalty;  $\boldsymbol{\xi}_{(n \times 1)}^*$  is the vector concerning points above  
 255 the tolerance band; and  $\boldsymbol{\xi}_{(n \times 1)}$  is the vector concerning points below the  
 256 tolerance band.

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

$$f(\mathbf{x}_i) = \mathbf{w}^T \varphi(\mathbf{x}) - w_0 = \sum_{j=1}^n \kappa(\mathbf{x}_i, \mathbf{x}_j) (\lambda_j^* - \lambda_j) - w_0 \quad (7)$$

257 where  $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \varphi^T(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) \in \mathbb{R}, i, j = 1, 2, 3 \dots, n$  is the kernel function.  
 258 Since  $\varphi$  transforms the original data to a higher dimension, which can even  
 259 be infinite, the use of the kernel function prevents the need to explicitly  
 260 compute the functional form of  $\varphi(\mathbf{x})$ ; instead,  $\kappa$  computes the inner product  
 261 of  $\varphi$ , a term that appears in SVR's dual formulation [47], this is known as  
 262 the *kernel trick*. In this paper, we used the Gaussian Kernel as  $\kappa$ , whose  
 263 expression is given as:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \sigma > 0 \quad (8)$$

264 The Gaussian Kernel is the most widely used Kernel in the machine  
 265 learning literature, enjoying huge popularity in various knowledge fields, since  
 266 this function is able to induce an infinite dimensional feature space while  
 267 depending on only one scattering parameter  $\sigma$ .

### 268 3.3. SVR-GARCH

269 The SVR-GARCH is the joining result of the GARCH model structure  
 270 and the nonlinearities introduced by the kernel function via SVR. [44] pre-  
 271 sented empirical evidences that the SVR-GARCH managed to outperform  
 272 standard GARCH's predictions, showing better ability to approximate the  
 273 nonlinear behavior of financial data and stylized facts, such as heavy tails  
 274 and volatility clusters. The specification of SVR-GARCH (1,1) is the same of  
 275 the conventional GARCH (1,1), but the mean and volatility equations were  
 276 estimated via SVR, such that:

$$r_t = f_m(r_{t-1}) + \epsilon_t \quad (9)$$

$$h_t = f_v(h_{t-1}, \epsilon_{t-1}^2) \quad (10)$$

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

Depending of which parameters  $\varepsilon$ ,  $C$  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**

Parameter	Search interval
$\varepsilon$	$[0.05, 0.1, \dots, 0.95, 1]$
$C$	$[0.5, 1, \dots, 4.5, 5]$
$\sigma$	$[0.05, 0.1, \dots, 1.95, 2]$

Table 1: Search intervals used for the parameters’ training

289

## 290 4. Empirical analysis

291 For the empirical test, we used the Bitcoin market price (in US dollars)  
 292 between January 7th 2015 and December 31th 2017. The data were collected  
 293 from Altcoin Charts (<http://alt19.com/>) originally with 388095 observations  
 294 from 2009 to 2017

295 We used the daily basis for the low frequency analysis (722 observations)  
 296 and the hour periodicity for the high frequency estimation (17011 observa-  
 297 tions).

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:

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

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:

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

Finally, the Diebold-Mariano [16] 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:

Equation	Parameter	Optimal value	RMSE
Mean	$\varepsilon$	0.85	0.03074
	$C$	0.5	
	$\sigma$	1.2	
Volatility	$\varepsilon$	0.1	0.01454
	$C$	5	
	$\sigma$	1.5	

Table 2: SVR-GARCH optimal parameters for low frequency data mean and volatility equation.

Equation	Parameter	Optimal value	RMSE
Mean	$\varepsilon$	1.0	0.00880
	$C$	0.5	
	$\sigma$	1.85	
Volatility	$\varepsilon$	0.2	0.00086
	$C$	5	
	$\sigma$	1.7	

Table 3: SVR-GARCH optimal parameters for high frequency data mean and volatility equation.

343 The optimal parameters were used to fit the SVR-GARCH (1,1), for  
 344 both low and high frequency data, alongside the GARCH (1,1) with three  
 345 different distributions for the error term  $\epsilon_t$ . The volatility predictions' RMSE  
 346 and MAE for each model are displayed in table 4.

Data	Model	RMSE	MAE
Low frequency	Normal GARCH (1,1)	0.03369	0.02799
	Student's t	0.03469	0.02917
	Skewed Student's t	0.03466	0.02894
	SVR-GARCH (1,1)	0.01818	0.01518
High frequency	Normal GARCH (1,1)	0.01089	0.00585
	Student's t	0.01070	0.00533
	Skewed Student's t	0.01067	0.00532
	SVR-GARCH (1,1)	0.00457	0.00182

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

348 The results show that the SVR models presented a significantly lower  
 349 value for both error metrics RMSE and MAE in comparison to all three  
 350 GARCH models. For the low frequency dataset, the error metrics for all  
 351 models were higher than the high frequency one. This result is consistent  
 352 with the finding in the literature: as seen in [49], the RMSE for the volatility  
 353 forecasting tend to decrease as the frequency increases.

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

$$H_0 : \frac{1}{N} |\tilde{h}_t - \hat{h}_{i,t}| - |\tilde{h}_t - \hat{h}_{benchmark,t}| = 0 \quad (13)$$

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.

Data	Model	Test statistic	P-value
Low frequency	Normal GARCH (1,1)	3.84050	0.00013
	Student's $t$	3.50970	0.00046
	Skewed Student's $t$	3.51560	0.00047
High frequency	Normal GARCH (1,1)	3.61930	0.00029
	Student's $t$	2.62970	0.00855
	Skewed Student's $t$	2.18750	0.02872

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

370

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 than 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 with insights regarding financial time series' forecasting and provide a overview of the Bitcoin market features. 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 [50] – and other well known extension models for GARCH, such as EGARCH [51] and GJR-GARCH [52]. 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 [48]'s  $\nu$ -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

- [1] R. C. Merton, On estimating the expected return on the market: An exploratory investigation, *Journal of financial economics* 8 (1980) 323–361.

- 421 [2] D. B. Nelson, Filtering and forecasting with misspecified arch models i:  
422 Getting the right variance with the wrong model, *Journal of Economet-*  
423 *rics* 52 (1992) 61–90.
- 424 [3] T. G. Andersen, T. Bollerslev, F. X. Diebold, P. Labys, Realized volatil-  
425 ity and correlation (1999).
- 426 [4] G. Deboeck, Trading on the edge: neural, genetic, and fuzzy systems  
427 for chaotic financial markets, volume 39, John Wiley & Sons, 1994.
- 428 [5] Y. S. Abu-Mostafa, A. F. Atiya, Introduction to financial forecasting,  
429 *Applied Intelligence* 6 (1996) 205–213.
- 430 [6] L.-J. Cao, F. E. H. Tay, Support vector machine with adaptive param-  
431 eters in financial time series forecasting, *IEEE Transactions on neural*  
432 *networks* 14 (2003) 1506–1518.
- 433 [7] D. Easley, M. M. López de Prado, M. OHara, The volume clock: Insights  
434 into the high-frequency paradigm, *The Journal of Portfolio Management*  
435 39 (2012) 19–29.
- 436 [8] J. C. Reboredo, J. M. Matías, R. Garcia-Rubio, Nonlinearity in fore-  
437 casting of high-frequency stock returns, *Computational Economics* 40  
438 (2012) 245–264.
- 439 [9] T. G. Andersen, T. Bollerslev, Deutsche mark–dollar volatility: intra-  
440 day activity patterns, macroeconomic announcements, and longer run  
441 dependencies, *the Journal of Finance* 53 (1998) 219–265.
- 442 [10] S. Çelik, H. Ergin, Volatility forecasting using high frequency data:  
443 Evidence from stock markets, *Economic modelling* 36 (2014) 176–190.
- 444 [11] J. Baruník, T. Křehlík, Combining high frequency data with non-linear  
445 models for forecasting energy market volatility, *Expert Systems With*  
446 *Applications* 55 (2016) 222–242.
- 447 [12] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity,  
448 *Journal of econometrics* 31 (1986) 307–327.
- 449 [13] R. F. Engle, T. Bollerslev, Modelling the persistence of conditional  
450 variances, *Econometric reviews* 5 (1986) 1–50.



- 451 [14] J. Marcucci, et al., Forecasting stock market volatility with regime-  
452 switching garch models, *Studies in Nonlinear dynamics and Economet-*  
453 *rics* 9 (2005) 1–53.
- 454 [15] P. R. Hansen, A. Lunde, A forecast comparison of volatility models:  
455 does anything beat a garch (1, 1)?, *Journal of applied econometrics* 20  
456 (2005) 873–889.
- 457 [16] F. X. Diebold, R. S. Mariano, Comparing predictive accuracy, *Journal*  
458 *of Business & Economic Statistics* 13 (1995) 253–263.
- 459 [17] V. V. Gavrishchaka, S. B. Ganguli, Volatility forecasting from multiscale  
460 and high-dimensional market data, *Neurocomputing* 55 (2003) 285–305.
- 461 [18] V. V. Gavrishchaka, S. Banerjee, Support vector machine as an effi-  
462 cient framework for stock market volatility forecasting, *Computational*  
463 *Management Science* 3 (2006) 147–160.
- 464 [19] S. Chen, K. Jeong, W. K. Härdle, Support vector regression based  
465 garch model with application to forecasting volatility of financial returns  
466 (2008).
- 467 [20] K. R. French, Detecting spot price forecasts in futures prices, *Journal*  
468 *of Business* (1986) S39–S54.
- 469 [21] E. Schwartz, J. E. Smith, Short-term variations and long-term dynamics  
470 in commodity prices, *Management Science* 46 (2000) 893–911.
- 471 [22] T. A. Reeve, R. Vigfusson, Evaluating the forecasting performance of  
472 commodity futures prices (2011).
- 473 [23] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system (2008).
- 474 [24] C. Desk, Bitcoin - Daily Number of Transactions, [http://www.  
475 coindesk.com/data/bitcoin-daily-transactions/](http://www.coindesk.com/data/bitcoin-daily-transactions/), 2017. [Online;  
476 accessed 16-March-2017].
- 477 [25] BLOCKCHAIN, Bitcoin - Confirmed Transactions Per Day, [https:  
478 //blockchain.info/charts/n-transactions](https://blockchain.info/charts/n-transactions), 2017. [Online; accessed  
479 16-March-2017].

- 480 [26] E. Bouri, R. Gupta, A. K. Tiwari, D. Roubaud, et al., Does Bitcoin  
481 Hedge Global Uncertainty? Evidence from Wavelet-Based Quantile-in-  
482 Quantile Regressions, Technical Report, 2016.
- 483 [27] A. H. Dyhrberg, Bitcoin, gold and the dollar—a garch volatility analysis,  
484 Finance Research Letters 16 (2016) 85–92.
- 485 [28] C. Desk, Bitcoin - Market Capitalization, [http://www.coindesk.com/  
486 data/bitcoin-market-capitalization/](http://www.coindesk.com/data/bitcoin-market-capitalization/), 2017. [Online; accessed 16-  
487 March-2017].
- 488 [29] BLOCKCHAIN, Bitcoin - Market Capitalization, [https:  
489 //blockchain.info/en/charts/market-cap](https://blockchain.info/en/charts/market-cap), 2017. [Online; accessed  
490 16-March-2017].
- 491 [30] A. H. Dyhrberg, Hedging capabilities of bitcoin. is it the virtual gold?,  
492 Finance Research Letters 16 (2016) 139–144.
- 493 [31] D. Yermack, Is Bitcoin a real currency? An economic appraisal, Tech-  
494 nical Report, National Bureau of Economic Research, 2013.
- 495 [32] T. Xiong, Y. Bao, Z. Hu, Multiple-output support vector regression  
496 with a firefly algorithm for interval-valued stock price index forecasting,  
497 Knowledge-Based Systems 55 (2014) 87–100.
- 498 [33] E. Bouri, G. Azzi, A. Haubo Dyhrberg, On the return-volatility rela-  
499 tionship in the Bitcoin market around the price crash of 2013, Technical  
500 Report, Economics Discussion Papers, 2016.
- 501 [34] X. Li, C. A. Wang, The technology and economic determinants of cryp-  
502 tocurrency exchange rates: The case of bitcoin, Decision Support Sys-  
503 tems (2016).
- 504 [35] A. Urquhart, The inefficiency of bitcoin, Economics Letters 148 (2016)  
505 80–82.
- 506 [36] K. Dowd, New private monies: A bit-part player? (2014).
- 507 [37] S. Chen, W. K. Härdle, K. Jeong, Forecasting volatility with support  
508 vector machine-based garch model, Journal of Forecasting 29 (2010)  
509 406–433.

- [38] P. C. S. Bezerra, P. H. M. Albuquerque, Volatility forecasting via svr-garch with mixture of gaussian kernels, *Computational Management Science* (2016) 1–18.
- [39] R. Cont, Empirical properties of asset returns: stylized facts and statistical issues (2001).
- [40] P. Sun, C. Zhou, Diagnosing the distribution of garch innovations, *Journal of Empirical Finance* 29 (2014) 287–303.
- [41] M. Li, F. Suohai, Forex Prediction Based on SVR Optimized by Artificial Fish Swarm Algorithm, *Fourth Global Congress on Intelligent Systems* (2013) 47–52.
- [42] F. Shen, J. Chao, J. Zhao, Forecasting exchange rate using deep belief networks and conjugate gradient method, *Neurocomputing* 167 (2015) 243–253.
- [43] B. Premanode, C. Toumazou, Improving prediction of exchange rates using Differential EMD, *Expert Systems with Applications* 40 (2013) 377–384.
- [44] G. Santamaría-Bonfil, J. Frausto-Solís, I. Vázquez-Rodarte, Volatility forecasting using support vector regression and a hybrid genetic algorithm, *Computational Economics* 45 (2015) 111–133.
- [45] A. A. P. Santos, N. C. A. da Costa, L. dos Santos Coelho, Computational intelligence approaches and linear models in case studies of forecasting exchange rates, *Expert Systems with Applications* 33 (2007) 816–823.
- [46] V. N. Vapnik, *The nature of statical learning theory*, Springer Verlag, 1995.
- [47] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, V. N. Vapnik, Support vector regression machines, in: *Advances in neural information processing systems*, Morgan Kaufmann Publishers, 1997, pp. 155–161.
- [48] C.-C. Chang, C.-J. Lin, Training  $\nu$ -support vector regression: Theory and algorithms, *Neural Computation* 14 (2002) 1959–1977.
- [49] H. Xie, J. Li, Intraday volatility analysis on s&p 500 stock index future, *International Journal of Economics and Finance* 2 (2010) 26.

- 541 [50] A. J. McNeil, R. Frey, Estimation of tail-related risk measures for het-  
542 eroscedastic financial time series: an extreme value approach, *Journal*  
543 *of empirical finance* 7 (2000) 271–300.
- 544 [51] D. B. Nelson, Conditional heteroskedasticity in asset returns: A new  
545 approach, *Econometrica: Journal of the Econometric Society* (1991)  
546 347–370.
- 547 [52] L. R. Glosten, R. Jagannathan, D. E. Runkle, On the relation between  
548 the expected value and the volatility of the nominal excess return on  
549 stocks, *The journal of finance* 48 (1993) 1779–1801.