Cryptocurrencies short-term forecast: application of ARIMA, GARCH and SVR models

Dovilė Kuizinienė
Department of Applied Informatics
Vytautas Magnus University
Kaunas, Lithuania
dovile.kuiziniene@vdu.lt

Aušra Varoneckienė
Department of Applied Informatics
Vytautas Magnus University
Kaunas, Lithuania
ausra.varoneckiene@vdu.lt

Tomas Krilavičius
Department of Applied Informatics
Vytautas Magnus University
Kaunas, Lithuania
Baltic Institute of Advanced Technology
tomas.krilavicius@vdu.lt

***Abstract*— Cryptocurrency are difficult to forecast due to it's globality and availability to everyone and every time. There is no Friday or Holidays effect, seasonality, market news and other aspects, which influence the course direction. It is the phenomena of the market and it is useful to spread forecast methods research to find out the best fitting model for this phenomenon. In this paper is presented short-term forecast of five different cryptocurrencies (Bitcoin, BitcoinCash, Ethereum, Litecoin, Ripple). Forecast methods split in two groups: 1) real value (ARIMA and SVR models) 2) volatility (GARCH and SVR models). The model's suitability is evaluated by RMSE and MAE. The best results for real value forecast were achieved using ARIMA, for volatility forecast - SVR. In further research it would be useful to analyze methods variety of Artificial Neural Networks and others connected models' modifications.**

Keywords — Cryptocurrency, Bitcoin, forecast, ARIMA, GARCH, SVR.

I. Introduction

Around a ago decade cryptocurrency was presented as a new phenomenon on global financial markets. By providing an alternative money and investment opportunity, they function outside centralized financial institutions [1]. The basic idea of cryptocurrency sustains of electronic payment system based on cryptographic proof instead of trusted third party [2]. Distributed or decentralized cryptocurrency system allows its users transfers make faster, cheaper and secure, without any intermediate. Scientists are interested in this new cryptocurrency phenomena as well. Holub & Johnsonkuri (2018) analyzed 13.5 thousand research papers from 20 different databases (EBSCOhost, Elsevier, JSTOR, SSRN and other) on Bitcoin and other cryptocurrency topics. This fast-growing topic is across many disciplines: technical fields (29.9%), economics (24.9%), regulation (17.1%), finance (8.3%) and others [3]. Miau & Yang (2017) estimated that the number of literatures on Blockchain, Bitcoin and other cryptocurrencies topics is still increasing over the year in different disciplines [4]. Looking from financial perspective, cryptocurrency literature review can be split in to several main themes: factors valuation, analysis on returns, forecasting, market speculation, market efficiency. It is complicated to make forecast models for cryptocurrencies. Cryptocurrencies are traded all over the world, 24 hours a day, in more than 200 market places. Comparing the same cryptocurrency at the same time in different market places, the price may vary by a few or even several tens of dollars. The main aim of this paper is to compere different forecast models for short-run cryptocurrency: Bitcoin (BTC), Bitcoin-Cash (BCH), Ethereum (ECH), Litecoin (LTC), Ripple (XRP).

II. Literature review

Forecast technics are being developed and applied in order to make better decisions. In this article we discuss time series forecast whose observations depends on the time. We offer to split Time series forecast models in three deferent groups, from each group shortly presented one model used in further research:

1. _Classic forecast methods_ (moving average (MA), autoregressive models (AR), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) and others). These methods can be split in *data-driven forecasting methods* (averaging or smoothing), where there is no difference between a predictor and a target, and *sophisticated model-driven forecasting methods* (ARIMA) [5]. An ARIMA(p, d, q) is a model in which the time series has been differenced d times, and the resulting values are predicted from the previous p actual values and q previous errors [6]. The p, d, q values identifies ACF and PACF plots.
2. _Volatility forecast methods_ (autoregressive conditional heteroskedasticity (ARCH), generalized autoregressive conditional heteroskedasticity (GARCH) and others). These methods are different from classic because it fits for non-stationary time series, where mean and variance changes over time. GARCH statistical model which includes not only the variation of error, but also the variance itself. It's ARMA variation of the model for the variance variable.
3. _Advanced forecast methods_ (Support Vector Machines for Regression (SVR), Artificial Neural Networks (ANNs)). These artificial intelligence methods are easier to use, do not include assumption constraints such as linearity, normal distribution and a specific observation number and produce more accurate forecasts [7]. The main idea behind SVR is to minimize error in hyperplane, knowing error tolerance rate and by using support vectors. Forecast results depend on the proper selection of the hyper-parameters (kernel, degree of deviations (C) and support vector used (ε)) [8]. Thera are many different flavours of ANNs, and there are very interesting, but in this stage we investigate only very basic models, like in [9]. However we plan to investigate performance of ANNs in future.

III. DATA AND RESEARCH METODOLOGY

We have chosen five cryptocurrencies with highest trading volume: Bitcoin (BTC), Bitcoin-Cash (BCH), Ethereum (ECH), Litecoin (LTC), Ripple (XRP), for period from 1st of June to 30th of September 2018. Period was selected due to availability download data from coindesk website [10]. The periods of cryptocurrency and dollar course is 5 minutes; hence we have 35136 different values for one cryptocurrency. During the period 68 records of data were missing (excluding Bitcoin), for 5 – 10 min. period. The missing values were replaced by the previous one. Table 1 and fig. 1 shows cryptocurrencies rate statistics of the period.

TABLE I. CRYPTOCURRENCY DESCRIPTIVE STATISTIC

	BTC	BCH	ETH	LTC	XRP
Min.:	5785	413	170.3	47.67	0.25
1st Qu.:	6387	532.4	277.8	58.02	0.33
Median:	6594	699.6	421.8	76.55	0.45
Mean:	6792	691.4	384.5	75.94	0.4364
3st Qu.:	7222	810.9	473.6	85.08	0.51
Max.:	8479	1206.3	624.5	127.37	0.77

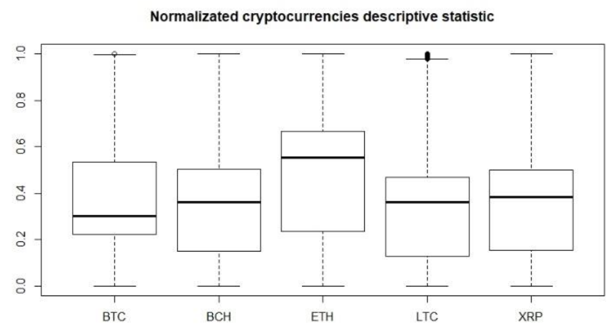

Fig. 1. Cryptocuriencies normalizated descriptive statistic

Forecast research was separated in two groups: 1) real value and 2) volatility. ARIMA, SVR methods and data normalization were used for the *first forecast group*. Data normalization is used to compere data results between different cryptocurrencies, as in:

$$x_{new} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x_{new} – new value, x_i – real value, x_{min} – minimum value of x sample, x_{max} – maximum value of x sample.

GARCH and SVR methods were used for the *second forecast group*. Variable of interest is counted, as in:

$$y_t = \frac{x_t - x_{t-1}}{x_{t-1}} \quad (2)$$

where y – interest, t – time, x – real value [11].

Research results are evaluated by RMSE and MAE coefficients, as in:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(a_t - f_t)^2} \quad (3)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|a_t - f_t| \quad (4)$$

where f_t – forecast expected value, a_t – real value [6]. These measures of error were chosen due to possibility to compare results with other authors research. RMSE and MAE coefficients are calculated for test sample and 3 forecast samples: 1 step forward, 1 hour forward (12 steps) and 1 day forward (288 steps).

Limitations:

The basic version of model is analyzed. The chosen models' coefficients are used to all cryptocurrencies the same. Only in GARCH model ARFIMA hypothesis for high-frequency components coefficients vary for different cryptocurrencies.

IV. RESULTS

A. Real value forecast (ARIMA and SVR)

ARIMA. Assumption of stationarity can be evaluated with visual inspection by autocorrelation function (ACF) and partial autocorrelation (PACF) plots [13] and by using Augmented Dickey-Fuller (ADF) test [6]. We rejected hypothesis of stationarity, because of trend in fig. 1 and ADF test results (p-value >0.5 (from all cryptocurrencies)). So, differencing order is equal 1. The parameters of p (stands of AR model) and q (stands of MA model) depends ACF and PACF plots. ACF and PACF plots are the same for ARIMA models: 110, 011 or 111. These models were checked by Akaike Information Criteria (AIC) [6] and the best results showed, that the best model for most cryptocurrencies – ARIMA(1,1,1). The ARIMA and SVR results are presented in table 3.

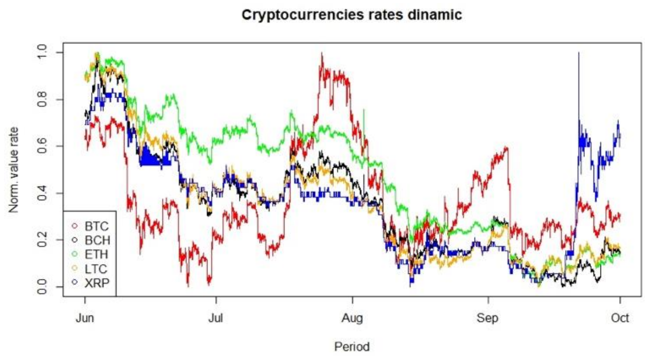

Fig. 2. Cryptocuriencies normalizated value rate dinamic

SVR. Only basic linear kernel was used due to high calculation requirements of other kernels (kernel = 2, C = 3 and ε=0.1). SVR model on Bitcoin normalized prices presented in fig. 3.

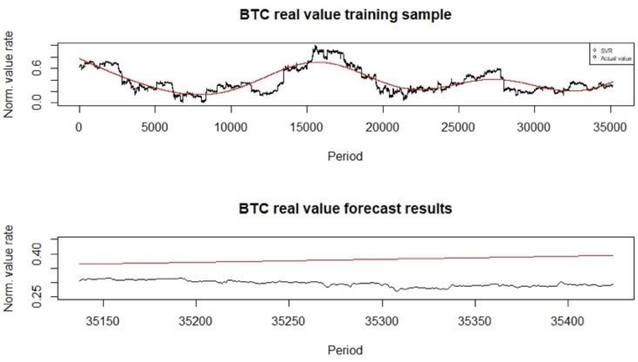

Fig. 3. Real value forecast BTC results

B. Volatility foracast (GARCH and SVR)

GARCH. In research is used GARCH (1,1) model, also known as basic GARCH. For each cryptocurrency we found out the best ARFIMA model by AIC and BIC criteria's results and Ljung-Box test confirmation (table 2). The ARFIMA model is compound GARCH method part for mean model calculations.

TABLE II. ARFIMA MODEL SELECTION FOR GARCH MODEL

	BTC	**BCH**	**ETH**	**LTC**	**XRP**
ARFIMA	(0,0,0)	(1,0,2)	(0,0,2)	(1,0,3)	(1,0,1)

SVR. Due to limited speed resources was used the same basic linear regression based SVR model (kernel = 2, C = 3 and ε=0.1). SVR model on Bitcoin volatility presented in fig. 3. The GARCH and SVR forecast results are presented in table 4.

TABLE III. ARIMA AND SVR MODELS RESULTS

$*10^{-2}$		**TRAINING SAMPLE**					**FORECAST (1 step forward - 5 minutes)**				
		BTC	BCH	ETH	LTC	XRP	BTC	BCH	ETH	LTC	XRP
RMSE	ARIMA	**0.438**	**0.220**	**0.168**	**0.249**	**0.747**	**0.135**	**0.268**	**0.110**	**0.020**	**0.019**
	SVR	11.585	5.492	4.78	4.674	7.335	5.951	6.002	2.760	3.974	9.570
MAE	ARIMA	**0.240**	**0.134**	**0.093**	**0.160**	**0.270**	**0.135**	**0.268**	**0.110**	**0.020**	**0.019**
	SVR	8.896	4.213	3.846	3.726	4.673	5.951	6.002	2.760	3.974	9.570
$*10^{-2}$		**FORECAST (12 step forward – 1 hour)**					**FORECAST (288 step forward – 1 day)**				
		BTC	BCH	ETH	LTC	XRP	BTC	BCH	ETH	LTC	XRP
RMSE	ARIMA	**0.796**	**0.690**	**0.269**	**0.490**	**2.586**	**1.217**	**0.832**	**0.538**	**0.926**	**3.255**
	SVR	5.385	5.685	2.651	3.629	7.559	8.421	7.126	3.673	5.487	14.52
MAE	ARIMA	**0.762**	**0.667**	**0.251**	**0.405**	**2.230**	**1.000**	**0.718**	**0.464**	**0.772**	**2.359**
	SVR	5.381	5.683	2.649	3.620	7.451	8.222	7.031	3.629	5.36	13.79

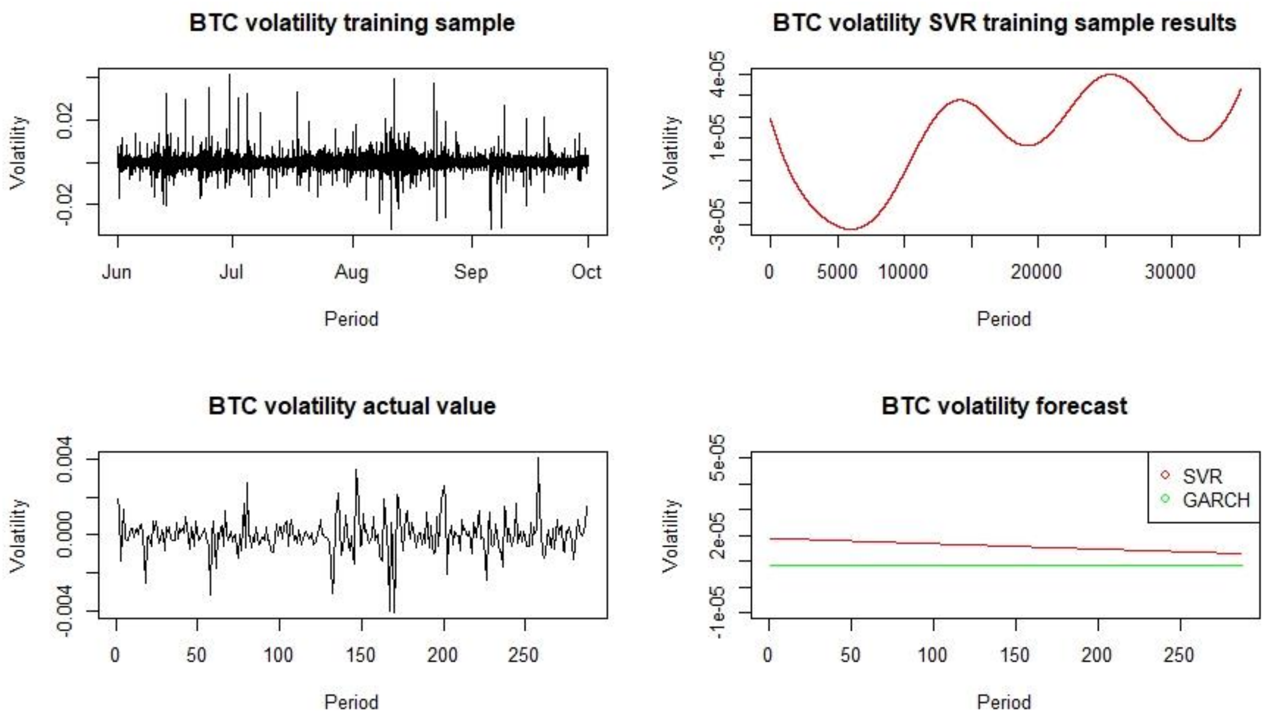

Fig. 4. Volatility forecast BTC results

TABLE IV. GARCH AND SVR MODELS RESULTS

*10^{-3}		TRAINING SAMPLE					FORECAST (1 step forward - 5 minutes)				
		BTC	BCH	ETH	LTC	XRP	BTC	BCH	ETH	LTC	XRP
RMSE	GARCH	**0.008**	**0.191**	**0.400**	**0.313**	**2.648**	1.876	4.680	0.750	0.266	**0.117**
	SVR	1.764	2.540	2.100	2.758	9.056	**1.866**	**4.277**	**0.525**	**0.257**	0.723
MAE	GARCH	**0.008**	**0.566**	**0.223**	**0.201**	**1.047**	1.876	4.680	0.750	0.266	0.117
	SVR	0.955	1.574	1.178	1.765	3.056	**1.866**	**4.277**	**0.525**	**0.257**	**0.723**

*10^{-3}		FORECAST (12 step forward – 1 hour)					FORECAST (288 step forward – 1 day)				
		BTC	BCH	ETH	LTC	XRP	BTC	BCH	ETH	LTC	XRP
RMSE	GARCH	0.912	2.759	0.990	1.741	9.785	**0.940**	2.755	**1.350**	**1.129**	**8.380**
	SVR	**0.909**	**2.701**	**0.965**	**1.720**	**9.601**	0.941	**2.753**	1.351	1.136	8.407
MAE	GARCH	0.673	2.183	0.857	1.299	**5.663**	**0.6318**	**2.114**	1.041	**0.871**	**3.778**
	SVR	**0.671**	**2.161**	**0.812**	**1.291**	6.011	0.6321	2.115	**1.038**	0.878	4.320

V. CONCLUSIONS

Globalization and market availability let us to trade in cryptocurrencies every day at any time. Cryptocurrencies is difficult to forecast due to complexity of the involved parties, low regulation and other reasons. The main aim of this research was to try different forecast models and find out one, most suitable for cryptocurrency. Five cryptocurrencies (BTC, BCH, ECH, LTC, XRP) were selected due to their market capitalization and spread availability to buy or sell in different market places. These cryptocurrencies were analyzed for period from the 1st of June to the 30th of September 2018. The strategy was to start from the simple models. Analyses was performed on (1) real values (ARIMA and SVR) and (2) volatility (GARCH and SVR). Performance was evaluated using RMSE and MAE for test sample and 3 forecast samples: 1 step forward, 1 hour forward (12 steps) and 1 day forward (288 steps).

The main results are:

1) The best results in real value forecast were achieved using ARIMA for ETH.

2) The best results for volatility forecast were achieved using SVR for BTC.

3) If we compare this analysis with Peng and others (2017) research, this analysis GARCH (1,1) results are better due to the intensity of the data (5 minutes not 1 day) and test sample length (1 day not 1 months). But Peng and others analyzed 9 different GARCH methods and the best results received by combined the traditional GARCH model with Support Vector Regression (SVR-GARCH) [13].

In future research we are planning to investigate performance of SVR-Garch, and different ANN (especially Deep Learning, e.g. LSTM) models.

REFERENCES

[1] P. Ciaian, M. Rajcaniova and d. Kancs, "Virtual relationships: Short- and long-run evidence from BitCoin and altcoin markets," *Journal of International Financial Markets Institutions and Money,*, vol. 52(C), p. 173–195, 2017.

[2] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. [Online].

[3] M. J. J. Holub, "Bitcoin research across disciplines," *Information Society,* Vols. 34:2, 1, pp. 14-126,, 2018.

[4] S. Y. J. Miau, "Bibliometrics-based evaluation of the Blockchain research trend: 2008 – March 2017.," *Technology Analysis & Strategic Management,* vol. 30:9, pp. 1029-1045, 2017.

[5] V. Katu and B. Deshpande, "Chapter 10 - Time Series Forecasting," in *Predictive Analytics and Data Mining*, Elsevier Inc. , 2015, pp. 305-327.

[6] J. Mount and N. Zumel, Exploring Data Science, Manning Publications, 2016.

[7] A. C.H. and E. Eğrioğlu, Advanced Time Series Forecasting Methods, ProQuest Ebook Central, 2012.

[8] B. Ojemakinde, "Support Vector Regression for Non-Stationary Time Series," Master's Thesis, University of Tennessee, 2006.

[9] N. Ahmed, A. Atiya, N. E. Gayar and H. El-Shishiny, "AN EMPIRICAL COMPARISON OF MACHINE LEARNING," *Econometric Reviews,* vol. 29, no. 0747-4938, p. 29(5–6):594–621, 2010.

[10] coindesk, 2018. [Online]. Available: https://www.coindesk.com/. [Accessed 30 09 2018].

[11] "ARCH/GARCH Models," The Pennsylvania State University, 2018. [Online].

[12] N. R., Fuqua School of Business Duke University, 2018. [Online].

[13] Y. Peng, P. H. M. Albuquerque, A. J. A. Padula and M. R. Montenegro, "The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression.," Expert Systems with Applications, vol. 97, p. 177–192, 2017.

[14] Capizzi, G., Sciuto, G. L., Napoli, C., & Tramontana, E. (2017). Advanced and adaptive dispatch for smart grids by means of predictive models. IEEE Transactions on Smart Grid, 9(6), p. 6684-6691.

[15] Napoli, C., Pappalardo, G., Tina, G. M., & Tramontana, E. (2015). Cooperative strategy for optimal management of smart grids by wavelet rnns and cloud computing. IEEE transactions on neural networks and learning systems, 27(8), 1672-1685.