Bitcoin Price Prediction Using EMD-SVR (EEN-351: Course Project)

Nitin Aggarwal
Department of Electrical Engineering
Indian Institute of Technology, Roorkee
Email: naggarwal@ee.iitr.ac.in

Simrat Singh
Department of Electrical Engineering
Indian Institute of Technology, Roorkee
Email: ssingh6@ee.iitr.ac.in

***Abstract*—This report presents an ensemble method composed of Emprirical Mode Distribution (EMD) algorithm and Support Vector Regression (SVR) for Bitcoin price prediction by analyzing the time series data of Bitcoin process. Bitcoin price prediction is one of the most challenging tasks of time series forecasting due to the inherent non-linearity and nonstationary characteristics of the bitcoin market and financial time series. It has recently attracted considerable attention in the fields of economics, cryptography and computer science due to its inherent nature of combining encryption technology and monetary units. Here, we attempt to predict the Bitcoin price accurately taking into consideration various parameters that affect the Bitcoin value. Our data set consists of various blockchain parameters relating to the Bitcoin price and payment network over the course of five years recorded daily. Along with this, we also used the data set of 5 other cryptocurrencies to observe the relation of Bitcoin price to the price of these cryptocurrencies. The proposed EMD-SVR model is evaluated with recent one month price data set.**

I. Introduction

Bitcoin is a successful cipher currency introduced into the financial market based on its unique protocol and Nakamoto's systematic structural specification [1]. Unlike existing fiat currencies with central banks, Bitcoin aims to achieve complete decentralization. Participants in the Bitcoin market build trust relationships through the formation of Blockchain based on cryptography techniques using hash functions. Inherent characteristics of Bitcoin derived from Blockchain technologies have led to diverse research interests not only in the field of economics but also in cryptography and machine learning.

Bitcoin is a crypto currency which is used worldwide for digital payment or simply for investment purposes. Bitcoin is decentralized i.e. it is not owned by anyone. Transactions made by Bitcoins are easy as they are not tied to any country. Investment can be done through various marketplaces known as "Bitcoin exchanges". These allow people to sell/buy Bitcoins using different currencies. The largest Bitcoin exchange is Mt Gox. Bitcoins are stored in a digital wallet which is basically like a virtual bank account. The record of all the transactions, the timestamp data is stored in a place called Blockchain. Each record in a blockchain is called a block. Each block contains a pointer to a previous block of data. The data on blockchain is encrypted. During transactions the user's name is not revealed, but only their wallet ID is made public.

Formation of Blockchain, a core technology of Bitcoin, distinguishes Bitcoin from other fiat currencies and is directly related to Bitcoin's supply and demand. To the best of our knowledge, in addition to macroeconomic variables and Blockchain information, direct use of other cryptocurrencies prices has not been investigated to describe the process of Bitcoin price. To fill this gap, the current study systematically evaluates and characterizes the process of Bitcoin price by modeling and predicting Bitcoin prices using Blockchain information and other cryptocurrencies prices in the financial market. We also try to account for the remarkable recent fluctuation, which is shown in Figure 1.

The rest of this report is structured as follows: Section II describes the previous work in this field of Bitcoin and Blockchain technique. Section III briefly reviews the EMD algorithm and SVR employed to model the process of Bitcoin prices. Section IV presents the experimental design and data specifications. Section V outlines empirical results. Section VI concludes the report.

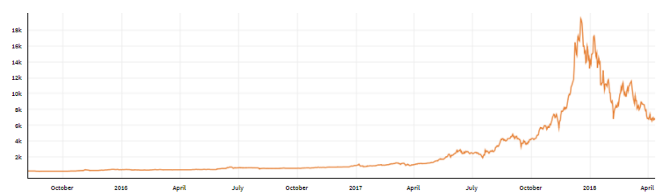

Fig. 1. Bitcoin daily price (USD) from Sep 2016 to April 2018

II. Literature Survey

Numerous studies have been conducted recently on modeling the time series of Bitcoin prices as a new market variable with specific technical rules. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) volatility analysis is performed to explore the time series of Bitcoin price [2], [3]. Various studies on statistical or economical properties and characterizations of Bitcoin prices refer to its capabilities as a financial asset; these research focus on statistical properties [4], [5], inefficiency of Bitcoin according to efficient market hypothesis [6], [7], hedging capability [8], [9], speculative bubbles in Bitcoin [10], the relationship between Bitcoin and search information, such as Google Trends and Wikipedia [11], and wavelet analysis of Bitcoin [12].

Relatively few studies have thus far been conducted on estimation or prediction of Bitcoin prices. Reference [13] evaluates Bitcoin price formation based on a linear model by considering related information that is categorized into several factors of market forces, attractiveness for investors, and global macro-financial factors. They assume that the first and second factors mentioned above significantly influence Bitcoin prices but with variation over time. The same researchers limit the number of regressors to facilitate linear model analysis. Reference [14] predicts the Bitcoin pricing process using machine learning techniques, such as recurrent neural networks (RNNs) and long short-term memory (LSTM), and compare results with those obtained using autoregressive integrated moving average (ARIMA) models. A machine trained only with Bitcoin price index and transformed prices exhibits poor predictive performance. Reference [15] conducts practical analysis on modeling and predicting of the Bitcoin process by employing a Bayesian neural network (BNN), which can naturally deal with increasing number of relevant features in the evaluation. A BNN includes a regularization term into the objective function to prevent the overfitting problem that can be crucial to this framework. When the machine considers a lot of input variables, a trained machine can be complex and suffer from the overfitting problem. BNN models showed their effect to the financial derivative securities analysis [16]. Reference [17] compares the accuracy of predicting Bitcoin price through binomial logistic regression, support vector machine, and random forest.

There are few practical and systematic empirical studies on the analysis of the time series of Bitcoin. In this study, we conduct practical analysis on modeling and prediction of the Bitcoin prices by employing an ensemble method composed of Emprirical Mode Distribution (EMD) algorithm and Support Vector Regression (SVR). First of all, the historical bitcoin price time series were decomposed into several intrinsic mode functions (IMFs). Then each IMF was modeled by a SVR model to generate the corresponding forecasting IMF value. Finally, the prediction results of all IMFs were combined to formulate an aggregated output for bitcoin price. The bitcoin price dataset of last one month is used to test the effectiveness of the proposed EMD-SVR method.

III. Methodology

This section reviews the Emprirical Mode Distribution (EMD) algorithm used for the bitcoin price prediction in the proposed ensemble method.

A. Empirical Mode Distribution

EMD, also known as Hilbert-Huang transform (HHT), is a method to decompose a signal into several intrinsic mode functions (IMF) along with a residue (R_N) which stands for the trend. EMD is an empirical approach to obtain instantaneous frequency data from non-stationary and nonlinear data sets. Analyzing the IMFs and residue separately is often easier than analyzing the original TS directly. EMD is based on local characteristic time scale and the operation is also adaptive and efficient. The shifting process which EMD uses to decompose the signal into IMFs is described as follows [18], [19].

1) Identify all local maxima and local minima in the TS x(t) and interpolate all local maxima with an interpolation method such as cubic spline to form an upper envelope U(t) and use the similar interpolation method for all local minima to form a lower envelope L(t).

2) Calculate the mean of upper and lower envelopes M(t) = (U(t) + L(t))/2 and subtract it from the original TS to obtain a local detail H(t) = x(t) − M(t).

3) Repeat Steps 1 and 2 on H(t) until: 1) M(t) approaches zero and 2) the number of local extrema and the number of zero crossings differs at most by one.

This process is known as shifting. The first IMF $IMF_1(t)$ equals H(t) and the residue $R_1(t)$ equals X(t) − H(t).

4) Repeat Steps 1–3 on $R_i(t)$ to obtain the subsequent IMF $IMF_{i+1}(t)$ and the second residue $R_{i+1}(t)$ until $R_{i+1}(t)$ do not have more than two local extrema, where i = 1,2 . . . N − 1. Finally, the original TS is decomposed as

$$x(t) = \sum_{i=1}^{N} IMF_i(t) + R_N(t).$$

IV. Proposed Work

A. Data Description

The first step towards Bitcoin price prediction is database collection. For our project we have collected database from the following sources:

1. Quandl

Quandl holds databases related to financial, economic and social background from over 500 publishers. The fact that bitcoin markets do not possess purchasing power nor interest rate parity suggests the need for completely new determinants of Bitcoin price: the Blockchain information that includes relevant features as main determinants for pricing Bitcoin.We were able to procure datasets for 25 parameters of Blockchain for up to 5 years of timestamp data including specifications such as –Bitcoin Total Transaction Fees, Market Capitalization, Difficulty, Hash Rate, Average Block Size, My Wallet Number of Users, USD Exchange Trade Volume, Estimated Transaction Volume USD, Total Output Volume etc. Figure 2 shows the Sample Partial Autocorrelation of Bitcoin Price with Price lag. It is clear that the bitcoin price on a particular day is highly correlated with the price on the previous day, i.e. price on day t is highly correlated with price on day (t-1). Table 1 presents some of the Blockchain data and variables that are used in predicting the evolution of Bitcoin prices.

TABLE I
Bitcoin Features and their Equations

S.no	Features	Equations/Definitions
1.	Block Size	Average block size (MB)
2.	Hash Rate	Processing power of Bitcoin network
3.	Total Bitcoins	Total number of bitcoins mined
4.	Difficulty	Difficulty of finding hash below a target
5.	Number of transactions	No. of unique Bitcoin transactions per day
6.	Trade Volume	USD trade volumes from top exchanges
7.	Day high, day low	Highest & lowest values of different days

2. CoinMarketCap

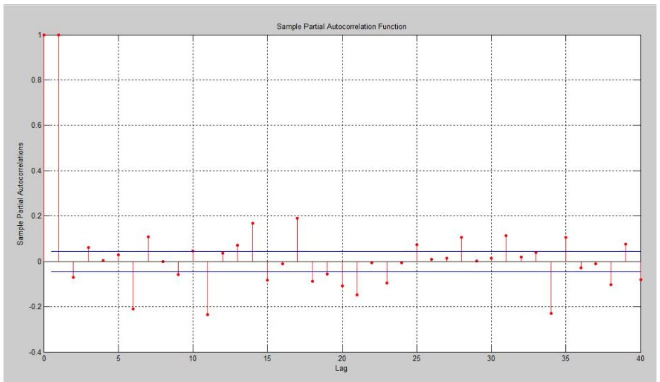

Fig. 2. Sample Partial Autocorrelation of Bitcoin Price with Price lag

CoinMarketCap keeps a track of all the cryptocurrencies available in the market. They keep a record of all the transactions by recording the amount of coins in circulation and the volume of coins traded in the last 24-hours. CoinMarketCap provides with historical data for Bitcoin price changes. From CoinMarketCap, we took the dataset of 5 cryptocurrencies other than bitcoin, namely, Litecoin, Ripple, Ethereum, Dash, Stellar to observe the relation of Bitcoin price to the price of these cryptocurrencies. Among these 5 cryptocurrencies, Ethereum was introduced recently so we fetched the dataset from 8 Aug. 2015 to 12 April 2018 having specifications such as –Date, Price open, High, Low, Close, Volume, Market Capitalization.

The next step is database normalization. We basically perform this step to achieve consistency i.e. reduce or eliminate duplicate data, insignificant points and other redundancies. For normalizing our data, we have used Z score normalization. This method uses a technique similar to standard deviation method by considering the mean value. Figure 3 shows the plot of bitcoin price after applying Z-score normalization technique.

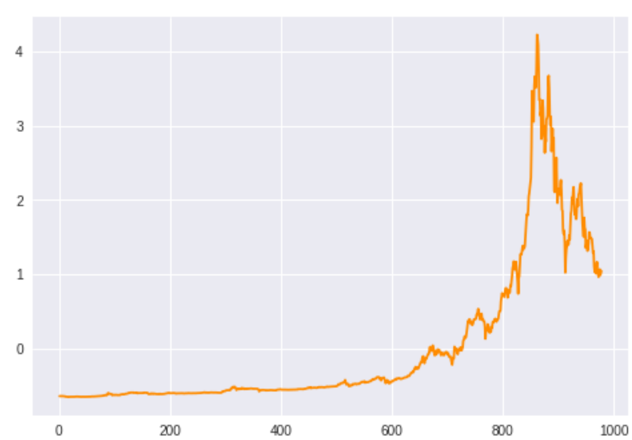

Fig. 3. Graph of Bitcoin price after Z-score normalization technique

B. EMD-SVR Model

"Divide and conquer" is an ensemble method which works by decomposing the original TS into a series of sub-datasets until they are simple enough to be analyzed. For proposed EMD-SVR approach, as mentioned above, the bitcoin price data is decomposed into several IMFs and one residue by EMD method. Then a SVR model is trained for each IMF including the residue. The final prediction results are given by combining the outputs from all sub-series using another SVR model. Figure 4 is the schematic diagram of this proposed ensemble method and the procedure can be stated as:

1. Use EMD to decompose the original TS into several IMFs and one residue.

2. Construct the training matrix as the input of each SVR for each IMF and residue.

3. Train SVR models to obtain the prediction results for each of the extracted IMF and residue.

4. Combine all the prediction results by another SVR model to formulate an ensemble output for TS forecasting.

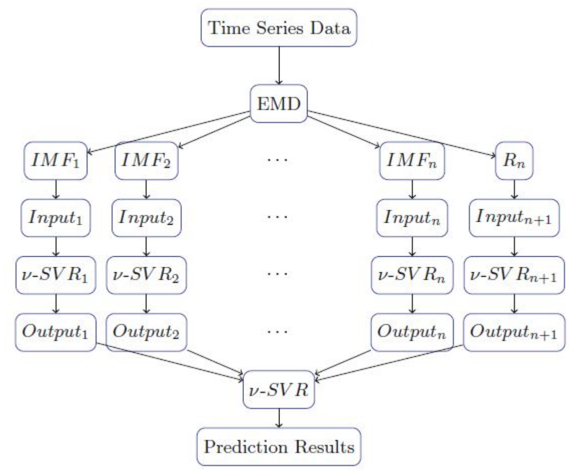

Fig. 4. Schematic diagram of the proposed EMD-SVR approach [20]

C. Experimental Design

For each dataset, we used 80% of the data points for training and the remaining 20% for testing. In this project, the bitcoin price dataset of last one month has been used for evaluating the performance of our proposed ensemble method. Daily closing bitcoin prices were used for our analysis.

To examine the accuracy of the prediction model, Root Mean Square Error (RMSE) has been used as evaluation metric.

RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i' - y_i)^2}$

where y_i' is the predicted value of corresponding y_i and n is the number of data points in the testing time series.

V. Results

Our ensemble method comprising of Emprirical Mode Distribution (EMD) algorithm and Support Vector Regression

(SVR) achieved a root mean square error (RMSE) of 0.02989 on the test data and 0.0184 on the training data. Figure 5 is the Regression Plot for the training data and Figure 6 shows the actual and predicted values of Bitcoin Price for a time span of 930 days starting from 8^{th} Aug. 2015. Orange curve depicts the actual values and Blue curve depicts the predicted values of bitcoin price.

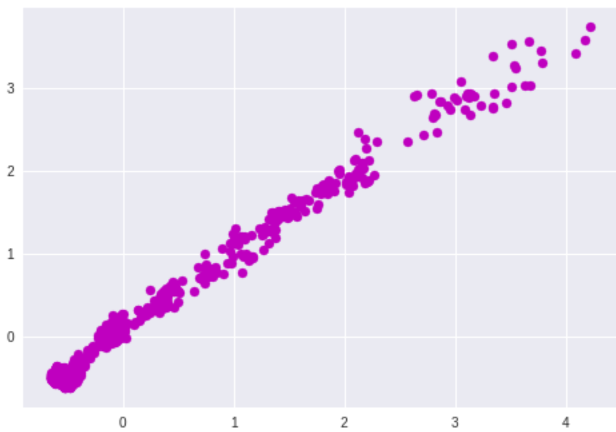

Fig. 5. Regression Plot for Training Data

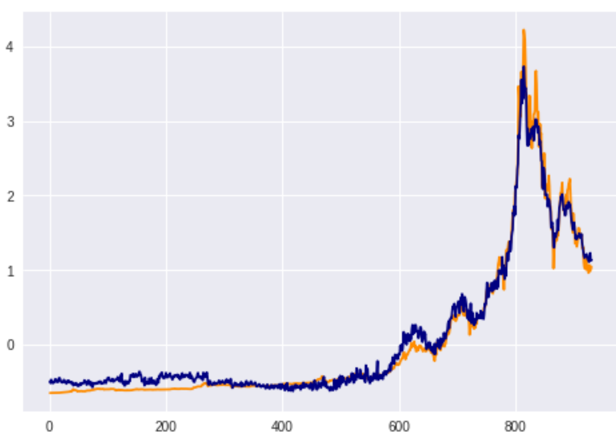

Fig. 6. Actual and Predicted values of Bitcoin Price

VI. Conclusions

Bitcoin is a successful cryptocurrency, and it has been extensively studied in fields of economics and computer science. In this report, we proposed an ensemble learning approach for short-term bitcoin price forecasting composed of EMD and SVR. Given the data of the entire time range, experimental results show that the EMD-SVR model learned with the selected features effectively describes the process of Bitcoin price.

Investigating nonlinear relationships between input functions based on network analysis can explain analysis of Bitcoin price time series. Variability of Bitcoin must be modeled and predicted more appropriately. This goal can be achieved by adopting other extended machine learning methods or considering new input capabilities related to the variability of Bitcoin. Such study will contribute to rich Bitcoin time series analysis in addition to existing Bitcoin studies.

Acknowledgement

We would like to thank our project guide Mr. G.N. Pillai for encouraging us to work on this project.

References

[1] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” Tech. Rep., 2008.

[2] A. H. Dyhrberg, “Bitcoin, gold and the dollar—A GARCH volatility analysis,” Finance Res. Lett., vol. 16, pp. 85–92, Feb. 2016.

[3] P. Katsiampa, “Volatility estimation for Bitcoin: A comparison of GARCH models,” Econ. Lett., vol. 158, pp. 3–6, Sep. 2017.

[4] A. F. Bariviera, M. J. Basgall, and W. Hasperué, and M. Naiouf, “Some stylized facts of the Bitcoin market,” Phys. A, Stat. Mech. Appl., vol. 484, pp. 82–90, Oct. 2017.

[5] J. Chu, S. Nadarajah, and S. Chan, “Statistical analysis of the exchange rate of Bitcoin,” PLoS ONE, vol. 10, no. 7, p. e0133678, 2015.

[6] A. Urquhart, “The inefficiency of Bitcoin,” Econ. Lett., vol. 148, pp. 80–82, Nov. 2016.

[7] S. Nadarajah and J. Chu, “On the inefficiency of Bitcoin,” Econ. Lett., vol. 150, pp. 6–9, Jan. 2017.

[8] A. H. Dyhrberg, “Hedging capabilities of Bitcoin. Is it the virtual gold?” Finance Res. Lett., vol. 16, pp. 139–144, Feb. 2016.

[9] E. Bouri, P. Molnár, G. Azzi, D. Roubaud, and L. I. Hagfors, “On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?” Finance Res. Lett., vol. 20, pp. 192–198, Feb. 2017.

[10] E.-T. Cheah and J. Fry, “Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin,” Econ. Lett., vol. 130, pp. 32–36, May 2015.

[11] L. Kristoufek, “Bitcoin meets Google trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era,” Sci. Rep., vol. 3, p. 3415, Dec. 2013.

[12] L. Kristoufek, “What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis,” PLoS ONE, vol. 10, no. 4, p. e0123923, 2015.

[13] P. Ciaian, M. Rajcaniova, and D. Kancs, “The economics of Bitcoin price formation,” Appl. Econ., vol. 48, no. 19, pp. 1799–1815, 2016.

[14] S. McNally, “Predicting the price of Bitcoin using machine learning,” Ph.D. dissertation, School Comput., Nat. College Ireland, Dublin, Ireland, 2016.

[15] Huisu Jang, Jaewook Lee, “An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information” Dec. 2017

[16] R. Gencay and M. Qi, “Pricing and hedging derivative securities with neural networks: Bayesian regularization, early

stopping, and bagging," IEEE Trans. Neural Netw., vol. 12, no. 4, pp. 726–734, Jul. 2001.

[17] I. Madan, S. Saluja, and A. Zhao, "Automated Bitcoin trading via machine learning algorithms," Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep., 2015.

[18] N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. Roy. Soc. London A, vol. 454, no. 1971, pp. 903–995, 1998.

[19] Z.-W. Zheng, Y.-Y. Chen, X.-W. Zhou, M.-M. Huo, B. Zhao, and M.-Y. Guo, "Short-term wind power forecasting using empirical mode decomposition and RBFNN," Int. J. Smart Grid Clean Energy, vol. 2, no. 2, pp. 192–199, May 2013.

[20] Xueheng Qiu, Huilin Zhu, P.N. Suganthan, and Gehan A.J. Amaratunga, "Stock Price Forecasting with Empirical Mode Decomposition Based Ensemble Support Vector Regression Model", CICBA 2017, Part I, CCIS 775, pp. 22–34, 2017.

[21] Siddhi Velankar, Sakshi Valecha, Shreya Maji, "Bitcoin Price Prediction using Machine Learning", ICACT 2018.

[22] J. Yao and C. L. Tan, "Time dependent directional prot model for nancial time series forecasting," in Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw. (IJCNN), vol. 5. Jul. 2000, pp. 291296.

[23] P. Wang, "Pricing currency options with support vector regression and stochastic volatility model with jumps," Expert Syst. Appl., vol. 38, no. 1, pp. 17, 2011.

[24] A. Kazem, E. Shari, F. K. Hussain, M. Saberi, and O. K. Hussain, "Support vector regression with chaos-based rey algorithm for stock market price forecasting," Appl. Soft Comput., vol. 13, no. 2, pp. 947958, 2013.

[25] F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac, "Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines," Int. J. Electr. Power Energy Syst., vol. 67, pp. 431438, May 2015.