# An experimental study of Bitcoin fluctuation using machine learning methods

Tian Guo
ETH Zurich, COSS
Zurich, Switzerland
tian.guo@gess.ethz.ch

Nino Antulov-Fantulin
ETH Zurich, COSS
Zurich, Switzerland
anino@gess.ethz.ch



## ABSTRACT

In this paper, we study the ability to make the short-term prediction of the exchange price fluctuations (measured with volatility) towards the United States dollar for the Bitcoin market. We use the data of realized volatility collected from one of the largest Bitcoin digital trading offices in 2016 and 2017 as well as order information. Experiments are performed to evaluate a variety of statistical and machine learning approaches.

**ACM Reference Format:**
Tian Guo and Nino Antulov-Fantulin. 2018. An experimental study of Bitcoin fluctuation using machine learning methods. In *Proceedings of ACM Conference (Conference'18).* ACM, New York, NY, USA, Article 4, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

Bitcoin (BTC) [32] is a novel digital currency system which functions without central governing authority. Instead, payments are processed by a peer-to-peer network of users connected through the Internet. Bitcoin users announce new transactions on this network, which are verified by network nodes and recorded in a public distributed ledger called the blockchain. Bitcoin is the largest of its kind in terms of total market capitalization value. They are created as a reward in a competition in which users offer their computing power to verify and record transactions into the blockchain. Bitcoins can also be exchanged for other currencies, products, and services. The exchange of the Bitcoins with other currencies is done on the exchange office, where "buy" or "sell" orders are stored on the order book. "Buy" or "bid" offers represent an intention to buy certain amount of Bitcoins at some price while "sell" or "ask" offers represent an intention to sell certain amount of Bitcoins at some price. The exchange is done by matching orders by price from order book into a valid trade transaction between buyer and seller.

Volatility [2, 16] as a measure of price fluctuations has a significant impact on trade strategies and investment decisions [12] as well as on option pricing [3, 10] and measures of systemic risk [8, 19, 28, 34]. Bitcoin, as a pioneer in the blockchain financial renaissance [7, 32] plays a dominant role in a whole cryptocurrency market capitalization ecosystem. Therefore, it is of great interest of data mining and machine learning community to be able to predict Bitcoin price fluctuations


*Conference'18, ,*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 2 RELATED WORK

Different studies have tried to explain various aspects of the Bitcoin such as its price formation, price fluctuations, systems dynamics and economic value.

From general complex systems perspective, different studies quantify either the evolution of transaction network [25] or the evolution of whole cryptocurrency ecosystem [11]. From economy perspective, the main studies [4, 18, 26] are focused around the fundamental and speculative value of Bitcoin.

Prediction of price was done with the following studies: (i) Garcia et. al. used autoregression techniques and identified two positive feedback loops (word of mouth, and new Bitcoin adopters) that lead to price bubbles [15], (ii) Amjad et. al. used the historical time series price data for price prediction and trading strategy [1] and (iii) Garcia et. al. also showed that the increases in opinion polarization and exchange volume precede rising of Bitcoin prices [14].

The price fluctuations were studied from different data sources: (a) Kondor et. al. used the Principal Component Analysis of the blockchain transaction networks data to find correlations between principal variables and changes in the exchange price, (b) Kim et. al. used cryptocurrency web communities data to extract sentiment and predict price fluctuations [23] and (c) Donier et. al. used order book data and found that the lack of buyers stoked the panic [9] before April 10$^{\text{th}}$ 2013 Bitcoin price crash.

Separately from cryptocurrency markets, a huge amount of volatility models [2, 16, 22, 38] and order book models [21, 31, 33] for standard financial markets exist. However, in this paper, we focus on the short-term prediction of volatility from machine learning perspective.

## 3 DATA AND PROCESSING

In this paper we use the actual historical hourly volatility data of Bitcoin market, which refers to the standard deviation of minute returns over one hour time range [16]. The return is defined as the relative change in consecutive prices of BTC. Our dataset contains time series of hourly volatility spanning more than one year, which outlines the fluctuation of Bitcoin market over time.

In addition, we have the order book data from the OKCoin, which is a digital asset trading platform providing trading services between fiat currencies and cryptocurrencies. In the last 2 years, trading volume of BTC at the OKCoin exchange office was approximately 39 % of the total traded BTC volume, which implies that our data source (OkCoin) can be used as a good proxy for BTC trading. Order book was collected through the exchange API with the granularity of one minute, with negligible missing values due to the API downtime or communication errors. Each order contains two attributes, price and amount.

Intuitively, our idea is to first transform order book data into features over time, referred as feature series and then to develop prediction models to consume volatility and feature series simultaneously. Concretely, from each minute snapshot of order book we extract the features related to:

- Price spread: The spread is the difference between the highest price that a buyer is willing to pay for a BTC (bid) and the lowest price that a seller is willing to accept (ask).
- Weighted spread: Weighted spread is the difference between cumulative price over 10 % of bid depth and the cumulative price over 10 % of ask depth.
- ask/bid depth and their difference: Depth is the number of orders on the bid or ask side.
- ask/bid volume and their difference: Volume is the number of BTCs on the bid or ask side.
- ask/bid slope: Slope is estimated as the volume until $\delta$ price offset from the current traded price, where $\delta$ is estimated by the bid price at the order that has at least 10 % of orders with the higher bid price.

We formulate the problem to resolve in this paper as follows. Given an hourly series of volatility $\{v_0, \ldots, v_H\}$, the $h$-th observation is denoted by $v_h \in \mathbb{R}^+$. The features of order book at each minute are denoted by a vector $\mathbf{x}_m \in \mathbb{R}^n$, where $n$ is the dimension of the feature vector. We define an index mapping function $i(\cdot)$ to map an hour index to the starting minute index of that hour (e.g. i(0)=1, i(1)=61), such that order book features associated with volatility observation $v_h$ is denoted by a matrix

$$\mathbf{X}_{[i(h),-l_b]} = (\mathbf{x}_{i(h)-1}, \ldots, \mathbf{x}_{i(h)-l_b}) \in R^{n \times l_b}$$

,where $l_b$ is the look-back time horizon. Likewise, a set of historical volatility observations w.r.t. $v_h$ is denoted by $\mathbf{v}_{[h,-l_v]} = (v_{h-1}, \ldots, v_{h-l_v}) \in \mathbb{R}^{l_v}$. Given historical volatility observations $\mathbf{v}_{[h,-l_v]}$ and order book features $\mathbf{X}_{[i(h),-l_b]}$, we aim to predict the one-step ahead volatility $v_h$.

# 4 EXPERIMENTS

In this section we present a comprehensive evaluation and reasoning behind the results of the approaches. We first introduce the dataset, baselines and set-up details. Then, prediction performance is reported.

## 4.1 Dataset

We have collected volatility and order book data ranging from September 2015 to April 2017. It consists of 13730 hourly volatility observations and 701892 order book snapshots. Each order book snapshot contains several hundreds of ask and bid orders. The maximum number of ask and bid orders at each minute are 1021 and 965. For the volatility time series, Augmented Dickey-Fuller (AD-Fuller) test is rejected at 1% significance level and therefore there is no need for differencing [24]. Seasonal patterns of volatility series is examined via periodogram and the result shows no existence of strong seasonality [37]. The dataset is split into training 70%, validation 10% and testing sets 20%.

## 4.2 Approaches

The **first category of statistics baselines** are only trained either on volatility or return time series. The included methods are as follows:

**EWMA** represents the exponential weighted moving average approach, which simply predicts volatility by performing moving average over historical ones [20].

**GARCH** refers to generalized autoregressive conditional heteroskedasticity model. It is a widely used approach to estimate volatility of returns and prices [16]. The basic idea of such methods is to model the variance of the error term of a time series as a function of the the previous error terms.

**BEGARCH** represents the Beta-t-EGARCH model [17]. It extends upon GARCH models by letting conditional log-transformed volatility dependent on past values of a t-distribution score.

**STR** is the structural time series model [35, 36]. It is formulated in terms of unobserved components via the state space method and used to capture local trend variation in time series.

Under this category, we also have plain **ARIMA** model.

The **second category of machine learning baselines** learns volatility and order book data simultaneously.

**RF** refers to random forests. Random forests are an ensemble learning method consisting several decision trees for classification, regression and other tasks [27, 30].

**GBT** is the gradient boosted tree, which is the application of boosting methods to regression trees [13]. GBT trains a sequence of simple regression trees and then adds together the prediction of individual trees to provide final prediction.

**XGT** refers to the extreme gradient boosting [6]. It was developed to improve the training efficiency and model performance of GBT by efficiently making use of computing resources and a more regularized model formalization.

**ENET** represents elastic-net, which is a regularized regression method combining both L1 and L2 penalties of the lasso and ridge methods [29].

**GP** stands for the Gaussian process based regression [5, 38], which has been successfully applied to time series. It is a supervised learning method which provides a Bayesian nonparametric approach to smoothing and interpolation.

**STRX** is the **STR** method augmented by adding regression terms on external features, similar to the way of **ARIMAX**.

For RF, GBT, XGT, ENET, and GP methods, input features are built by concatenating historical volatility and order book features.

## 4.3 Evaluation set-up

In GARCH and BEGARCH, the orders of autoregressive and moving average terms for the variance are both set to one [16]. Smoothing parameter in EWMA is chosen from $\{0.01, 0.1, 0.2, \ldots, 0.9\}$. In ARIMA and ARIMAX, the orders of auto-regression and moving-average terms are set via the correlogram and partial autocorrelation. For decision tree based approaches including RF, GBT, XGT,

**Table 1: Test errors (RMSE)**

| Model | RMSE | MAE |
|---|---|---|
| EWMA | 0.082 | 0.047 |
| GARCH | 0.136 | 0.099 |
| BEGARCH | 0.134 | 0.096 |
| STR | 0.111 | 0.065 |
| ARIMA | 0.109 | 0.059 |
| ARIMAX | 0.126 | 0.083 |
| STRX | 0.159 | 0.103 |
| GBT | 0.078 | 0.049 |
| RF | 0.082 | 0.052 |
| XGT | 0.076 | 0.047 |
| ENET | 0.080 | 0.052 |
| GP | 0.082 | 0.050 |

hyper-parameter tree depth and the number of iterations are chosen from range [3, 10] and [3, 200] via grid search. For XGT, L2 regularization is added by searching within $\{0.0001, 0.001, 0.01, 0.1, 1, 10\}$. As for ENET, the coefficients for L2 and L1 penalty terms are selected from the set of values $\{0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$. In GP, we adopt the radial-basis function (RBF) kernel, white noise, and periodic kernels[5]. The hyper-parameters in GP are optimized via maximum likelihood estimation.

We use root mean squared error (RMSE) and mean absolute error (MAE) as evaluation metrics [16], which are defined as follows: $RMSE = \sqrt{\sum_i (v_i - \hat{v}_i)^2 / n}$ and $MAE = 1/n \sum_i |v_i - \hat{v}_i|$.

## 4.4 Prediction performance

In this part, we report on the prediction performance. The time horizon of order book features, i.e. parameter $l_b$, is set to 30, that is, order book features within the 30 minutes prior to the prediction hour are fed into models. Then, the sensitivity of prediction performance w.r.t. $l_b$ is also presented.

Tab. 1 shows prediction errors of each model. The results are reported in two groups of approaches. The simple EWMA can beat all others in some intervals. Particularly, in the second group of Tab. 1, ARIMAX and STRX using both volatility and order book data fail to outperform their counterparts, i.e. ARIMA and STR in the top group. It suggests that simply adding features from order book does not necessarily improve the performance. Ensemble and regularized regression perform better, e.g., XGT and ENET perform the best in most of cases, within this group.

Tab. 2 demonstrates the effect of time horizon (i.e. $l_b$) of order book features on prediction performance of models using order book features. The results are obtained by evaluating each model with increasing size of $l_b$. It exhibits that short term order book features is sufficient for most of the models and further more data, e.g. 40 and 50 minutes of order book features, lead to no improvement of the performance. In particular, models like ARIMAX and STRX are prone to overfit by redundant data of long horizon, while ensemble method XGT, and ENET are relatively robust to the horizon.

**Table 2: Test error sensitivity to time horizon of order book features (RMSE)**

| Time horizon | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| ARIMAX | 0.112 | 0.125 | 0.126 | 0.126 | 0.135 |
| STRX | 0.138 | 0.142 | 0.159 | 0.157 | 0.161 |
| GBT | 0.081 | 0.081 | 0.083 | 0.084 | 0.085 |
| RF | 0.081 | 0.082 | 0.082 | 0.083 | 0.083 |
| XGT | 0.077 | 0.077 | 0.076 | 0.077 | 0.076 |
| ENET | 0.080 | 0.080 | 0.080 | 0.080 | 0.081 |
| GP | 0.095 | 0.081 | 0.082 | 0.084 | 0.085 |

## 5 CONCLUSION

In this paper, we study the short-term fluctuation of Bitcoin market by using real data from a major Bitoin exchange office. The dataset comprises realized volatility observations and order book snapshots of different time scales and data types. By reformatting the oder book data into feature series, we formulate the volatility prediction problem as learning predictive models over both volatility and order book feature series. Through experiments, we found out that ensemble method XGT and regularized regression ENET outperform other methods in most of the cases. Meanwhile, XGT and ENET are robust to the time horizon of order features.

## REFERENCES

[1] Muhammad Amjad and Devavrat Shah. 2017. Trading Bitcoin and Online Time Series Prediction. In *NIPS 2016 Time Series Workshop*. 1–15.
[2] Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2003. Modeling and Forecasting Realized Volatility. *Econometrica* 71, 2 (2003), 579–625.
[3] Fischer Black and Myron Scholes. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81, 3 (1973), 637–654.
[4] Wilko Bolt. 2016. On the Value of Virtual Currencies. *SSRN Electronic Journal* (2016).
[5] Sofiane Brahim-Belhouari and Amine Bermak. 2004. Gaussian process for non-stationary time series prediction. *Computational Statistics & Data Analysis* 47, 4 (2004), 705–712.
[6] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *SIGKDD*. ACM, 785–794.
[7] D.L.K. Chuen. 2015. *Handbook of Digital Currency: Bitcoin, Innovation, Financial Instruments, and Big Data*. Academic Press.
[8] Jamil Civitarese. 2016. Volatility and correlation-based systemic risk measures in the US market. *Physica A: Statistical Mechanics and its Applications* 459 (2016), 55–67.
[9] Jonathan Donier and Jean-Philippe Bouchaud. 2015. Why Do Markets Crash? Bitcoin Data Offers Unprecedented Insights. *PLOS ONE* 10, 10 (2015), 1–11.
[10] Huu Nhan Duong, Petko S. Kalev, and Chandrasekhar Krishnamurti. 2009. Order aggressiveness of institutional and individual investors. *Pacific-Basin Finance Journal* 17, 5 (2009), 533–546.
[11] Abeer ElBahrawy, Laura Alessandretti, Anne Kandler, Romualdo Pastor-Satorras, and Andrea Baronchelli. 2017. Evolutionary dynamics of the cryptocurrency market. *Royal Society Open Science* 4, 11 (2017), 170623.
[12] Jeff Fleming, Chris Kirby, and Barbara Ostdiek. 2003. The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics* 67, 3 (2003), 473–509.
[13] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.
[14] David Garcia and Frank Schweitzer. 2015. Social signals and algorithmic trading of Bitcoin. *Royal Society Open Science* 2, 9 (2015), 150288.
[15] D. Garcia, C. J. Tessone, P. Mavrodiev, and N. Perony. 2014. The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. *Journal of The Royal Society Interface* 11, 99 (2014), 20140623–20140623.
[16] Peter R. Hansen and Asger Lunde. 2005. A forecast comparison of volatility models: does anything beat a GARCH(1, 1)? *Journal of Applied Econometrics* 20, 7 (2005), 873–889.
[17] Andrew C Harvey and Tirthankar Chakravarty. 2008. Beta-t-(e) garch. (2008).

[18] Adam Hayes. 2015. Cryptocurrency Value Formation: An Empirical Analysis Leading to a Cost of Production Model for Valuing Bitcoin. *SSRN Electronic Journal* (2015).

[19] Dirk Helbing. 2013. Globally networked risks and how to respond. *Nature* 497, 7447 (2013), 51–59.

[20] Rob J Hyndman and George Athanasopoulos. 2014. *Forecasting: principles and practice*. OTexts.

[21] Pankaj K. Jain, Pawan Jain, and Thomas H. McInish. 2011. The Predictive Power of Limit Order Book for Future Volatility, Trade Price, and Speed of Trading. *SSRN Electronic Journal* (2011).

[22] Paraskevi Katsiampa. 2017. Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters* 158 (2017), 3–6.

[23] Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, and Chang Hun Kim. 2016. Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies. *PLOS ONE* 11, 8 (2016), e0161197.

[24] Gebhard Kirchgässner and Jürgen Wolters. 2007. *Introduction to modern time series analysis*. Springer Science & Business Media.

[25] Daniel Kondor, Marton Posfai, Istvan Csabai, and Gabor Vattay. 2014. Do the Rich Get Richer? An Empirical Analysis of the Bitcoin Transaction Network. *PLOS ONE* 9, 2 (2014), 1–10.

[26] Ladislav Kristoufek. 2015. What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. *PLOS ONE* 10, 4 (2015), e0123923.

[27] Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by randomForest. *R news* 2, 3 (2002), 18–22.

[28] Jianxu Liu, Songsak Sriboonchitta, Panisara Phochanachan, and Jiechen Tang. 2015. Volatility and Dependence for Systemic Risk Measurement of the International Financial System. In *Lecture Notes in Computer Science*. Springer International Publishing, 403–414.

[29] Yan Liu, Alexandru Niculescu-Mizil, Aurelie C Lozano, and Yong Lu. 2010. Learning temporal causal graphs for relational time-series analysis. In *ICML*. 687–694.

[30] Christopher Meek, David Maxwell Chickering, and David Heckerman. 2002. Autoregressive tree models for time-series analysis. In *SDM*. SIAM, 229–244.

[31] Randi Næs and Johannes A. Skjeltorp. 2006. Order book characteristics and the volume–volatility relation: Empirical evidence from a limit order market. *Journal of Financial Markets* 9, 4 (2006), 408–432.

[32] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. (2008). http://bitcoin.org/bitcoin.pdf

[33] Roberto Pascual and David Veredas. 2008. Does the Open Limit Order Book Matter in Explaining Informational Volatility? *SSRN Electronic Journal* (2008).

[34] Matija Piškorec, Nino Antulov-Fantulin, Petra Kralj Novak, Igor Mozetič, Miha Grčar, Irena Vodenska, and Tomislav Šmuc. 2014. Cohesiveness in Financial News and its Relation to Market Volatility. *Scientific Reports* 4, 1 (2014).

[35] Kira Radinsky, Krysta Svore, Susan Dumais, Jaime Teevan, Alex Bocharov, and Eric Horvitz. 2012. Modeling and predicting behavioral dynamics on the web. In *WWW*. ACM, 599–608.

[36] Steven L Scott and Hal R Varian. 2014. Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation* 5, 1-2 (2014), 4–23.

[37] Xiaozhe Wang, Kate Smith, and Rob Hyndman. 2006. Characteristic-based clustering for time series data. *Data mining and knowledge Discovery* 13, 3 (2006), 335–364.

[38] Yue Wu, José Miguel Hernández-Lobato, and Zoubin Ghahramani. 2014. Gaussian process volatility model. In *NIPS*. 1044–1052.