Artificial Neural Networks for Realized Volatility Prediction in Cryptocurrency Time Series

Ryotaro Miura, Lukáš Pichl(✉), and Taisei Kaizoji

International Christian University, Osawa 3-10-2, Mitaka, Tokyo 181-8585, Japan
lukas@icu.ac.jp
http://www.icu.ac.jp/

Abstract. Realized volatility (RV) is defined as the sum of the squares of logarithmic returns on high-frequency sampling grid and aggregated over a certain time interval, typically a trading day in finance. It is not a priori clear what the aggregation period should be in case of continuously traded cryptocurrencies at online exchanges. In this work, we aggregate RV values using minute-sampled Bitcoin returns over 3-h intervals. Next, using the RV time series, we predict the future values based on the past samples using a plethora of machine learning methods, ANN (MLP, GRU, LSTM), SVM, and Ridge Regression, which are compared to the Heterogeneous Auto-Regressive Realized Volatility (HARRV) model with optimized lag parameters. It is shown that Ridge Regression performs the best, which supports the auto-regressive dynamics postulated by HARRV model. Mean Squared Error values by the neural-network based methods closely follow, whereas the SVM shows the worst performance. The present benchmarks can be used for dynamic risk hedging in algorithmic trading at cryptocurrency markets.

Keywords: ANN · MLP · LSTM · GRU · CNN · SVM · HARRV · Ridge regression · Realized volatility

1 Introduction

Uncertainty modeling in financial markets in econometrics has traditionally been based on the notion of volatility inferred indirectly from daily time series data, an approach which has extended to high frequency data since the pioneering work on realized volatility (RV) by Andersen et al. [1]. The quantity is defined on aggregation period between t and $t + T$ sampled N times as

$$RV(t) = \sum_{i=1}^{N} R_i^2, \quad R_i = \log\left(P_i/P_{i-1}\right) \quad P_i \equiv P(t + iT/N). \tag{1}$$

There has been a recent surge in the the number of work related to deep learning (DL) algorithm for market prediction [3,4] including lately cryptocurrency markets [5–7], however, the applications of machine learning (ML) to RV

H. Lu et al. (Eds.): ISNN 2019, LNCS 11554, pp. 165–172, 2019.
https://doi.org/10.1007/978-3-030-22796-8_18

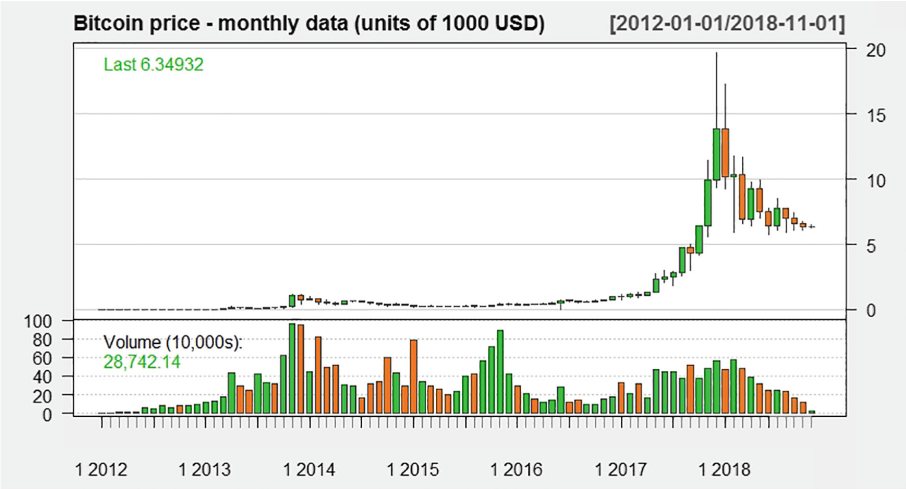

Fig. 1. Candlestick diagram for monthly Bitcoin prices in USD (aggregated from minute data at the Bitstamp exchange). The body of each box ranges between open and close prices, with low and high values indicated by vertical lines. Green color codes price increase and red color price decrease. The volume series in the bottom are counted in units of 10 thousand Bitcoin. (Color figure online)

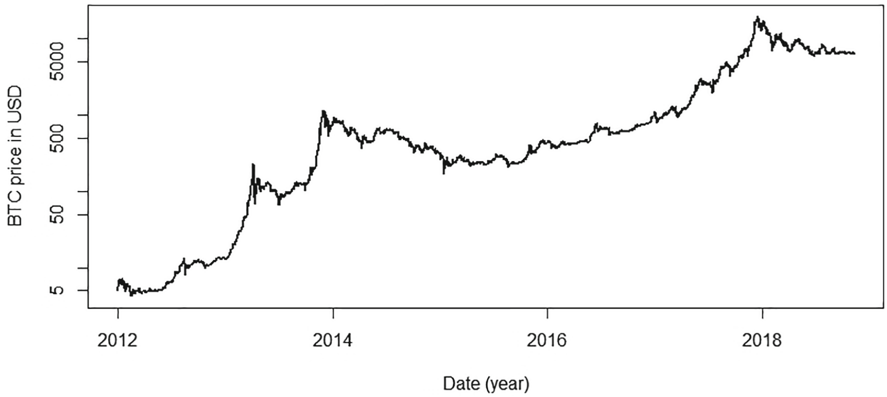

Fig. 2. BTCUSD daily-aggregated close price in logarithmic scale. Time period ranges from 2012-01-01 to 2018-11-11 contains 2507 days. Source data from [2].

of digital currencies have not appeared yet. It is the purpose of this work to explore Bitcoin time series data depicted in Figs. 1 and 2 (source: [2]) and provide a benchmark study of modern ML algorithms to RV prediction. The paper is organized as follows. The next section sums up empirical properties of data and lists the ML/DL models applied. Results are given and discussed in Sect. 3, followed by a brief conclusion.

2 Data and Models

Figures 3 and 4 show the distribution of the log returns, from which volatility clustering and a fat-tail distribution may be observed.

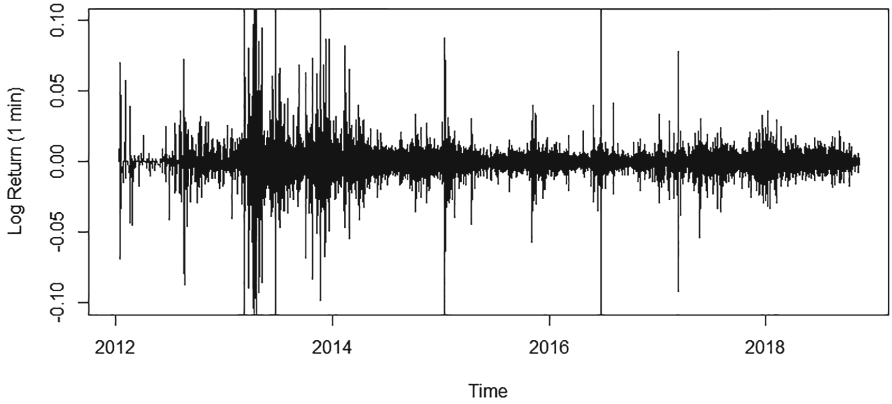

Fig. 3. BTCUSD minute log returns. After removal of missing prices and non-adjacent returns, there are 1,804,479 values. Source data fom [2].

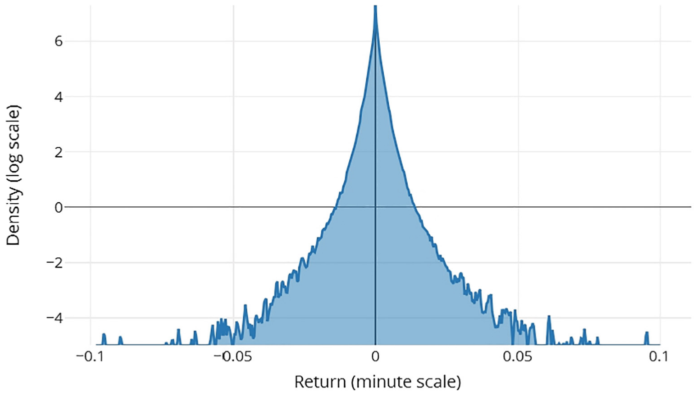

Fig. 4. Distribution of log returns of BTCUSD price on minute-sampling scale.

The distribution in Fig. 4 is practically symmetric, and the absolute values of the log return can be fitted with power law as shown in Fig. 5. One-lag correlations are shown in Fig. 6. Realized volatility distribution is shown in Fig. 7(a), along with a log-value differencing transform in Fig. 7(b). Figure 8 shows the autocorrelation function (ACF) values for the first 20 lags. Correlations between RV and

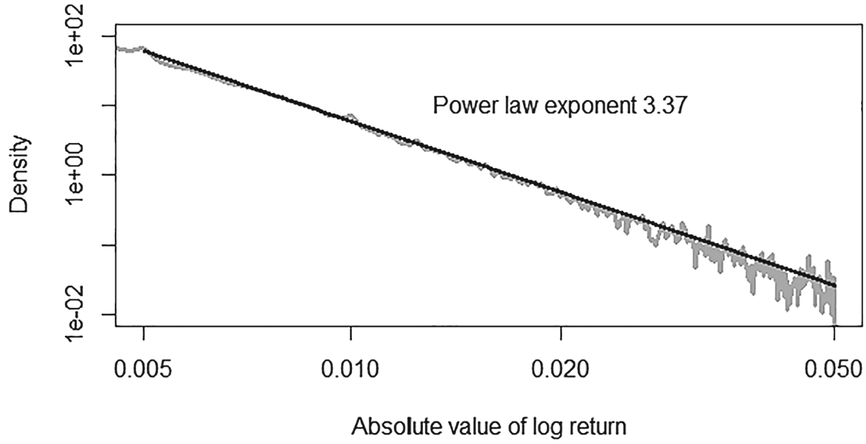

Fig. 5. Power-law distirbution of the fat-tail of BTCUSD minute log returns and the fitted exponent.

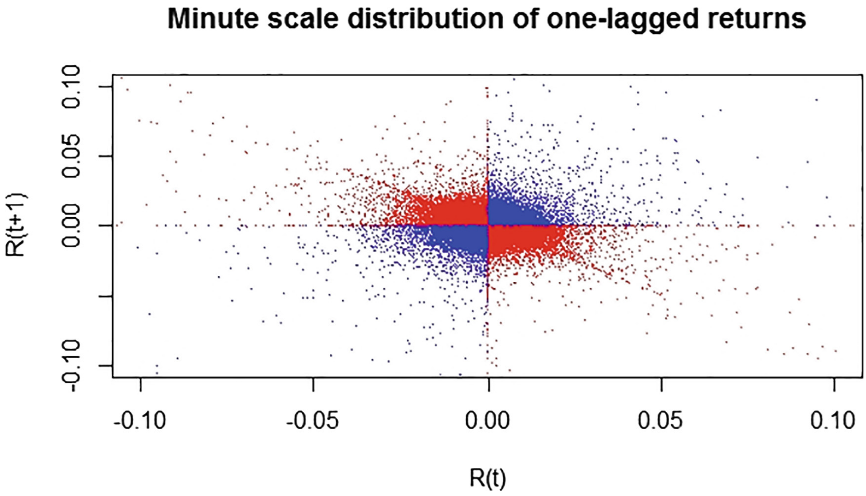

Fig. 6. Distirbution patterns of adjacent minute returns. Note the line structure on the reverse diagonal of II and IV quadrant, which corresponds to negative autocorrelation for lag 1.

intraday variance, defined as $2(H-L)/(H+L)$, where H stands for the highest and L for the lowest price within each 3-h bin interval, are shown in Fig. 9(a). Approximate relation between RV and aggregated trading volume is depicted in Fig. 9(b). The methods of analysis are the heterogenous auto-regressive model of realized volatility (HARRV) [1] with 3 lags, multi-layer perceptron (MLP, dense layer neural network) [8], convolutional neural network (CNN) [9], long short-term memory (LSTM) [10] (cf. Fig. 10), gated recurrent unit (GRU) [11],

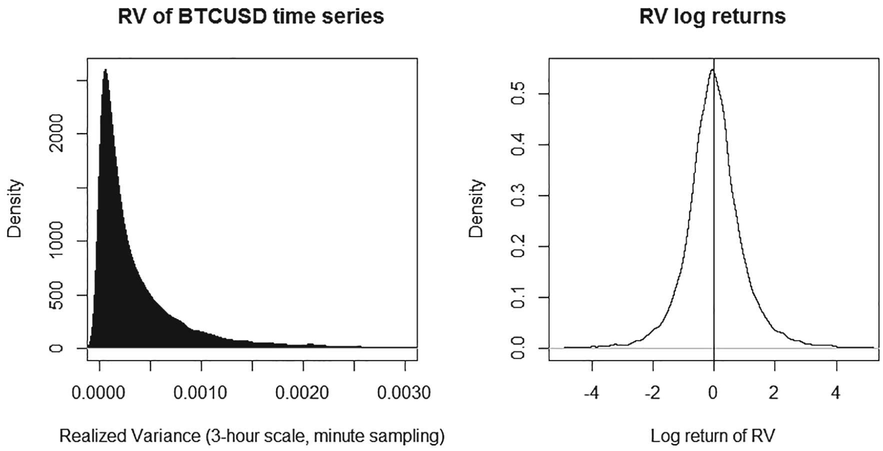

Fig. 7. Realized Volatility distributions: (a) Sum of the squared log returns on minute scale aggregated over 180 min, (b) log returns taken from adjacent values of RV in (a).

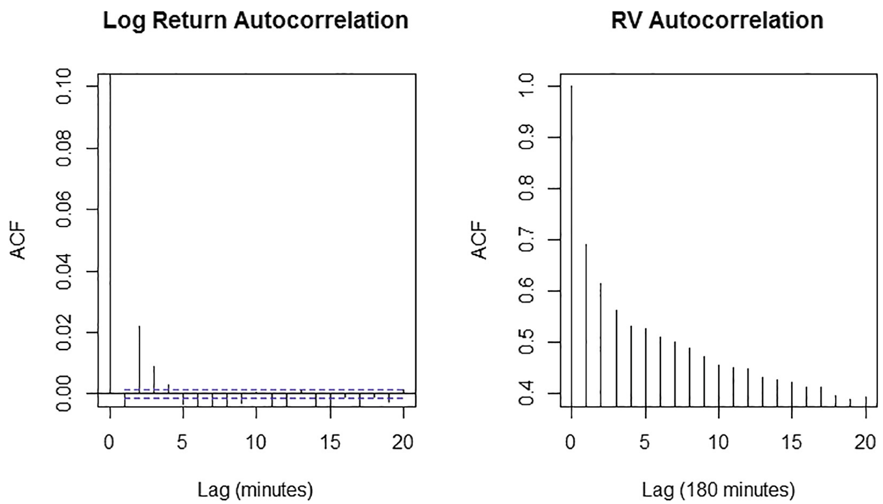

Fig. 8. Autocorrelation function (a) Left: correlation of log returns: notice the negative value for lag 1, and the significant yet small values for lags 2 and 3; (b) Right: persistent correlations of RV time series.

support vector machine (SVM) [12] and ridge regression [13]. In addition, we implemented Dropout mechanism to LSTM, GRU, and Batch Normalisation (BN) to MLP and CNN.

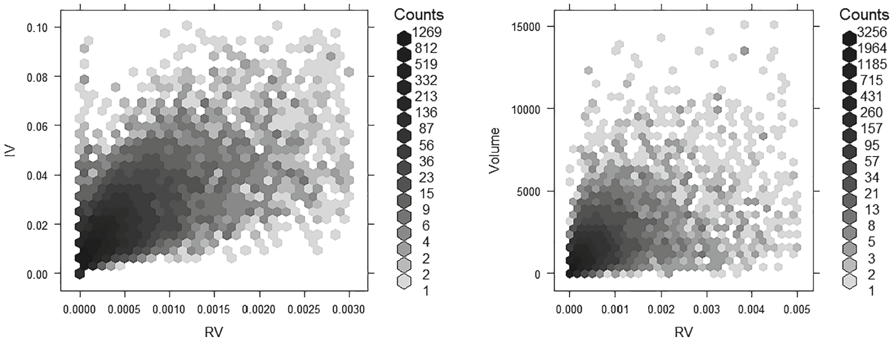

Fig. 9. 2D Histograms in the form of heatmap diagram: (a) Correlation of RV and intraday variance (see text for definition) (b) Right: correlation of RV and trading volume (time scale of 3 h).

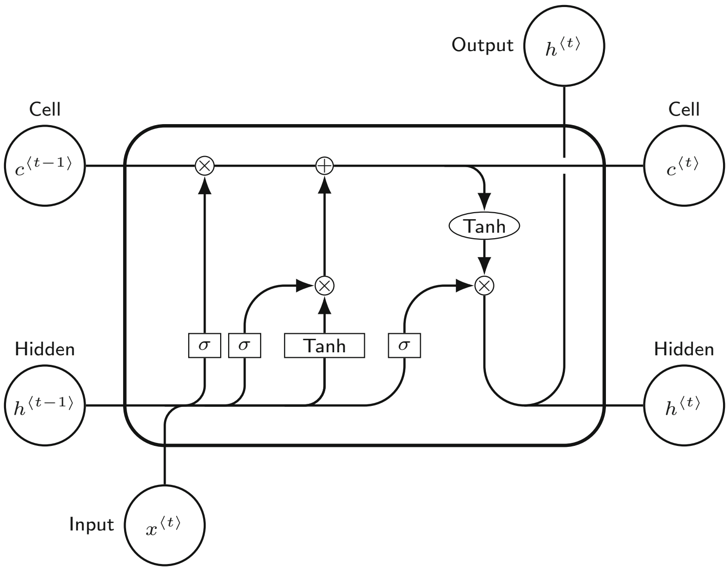

Fig. 10. Long short-term memory unit operation following notation in Ref. [10].

3 Results

Table 1 shows the results for several models referenced above, with model configuration listed in the first column. The models were validated with 10-fold cross validation (CV) for time series using 100 runs, and benchmark on the test set, using 30 random weight initializations. Optimized parameters (length of sequence used for prediction) are shown in column 2, and the statistical characteristics of the mean squared error (MSE) and the rooted MSE (RMSE) are listed. Result variances are also displayed where applicable. The best performing method on the validation data set is the GRU 1 layer + 2 dense layer result; however, it does not generalize well on the test data set, where the ridge regression method benchmarks as the most superior method with penalty parameter 0.6951.

Table 1. Resuts of machine learning models trained with cross validation and benchmarked on the test set.

Model	Sequence length	CV MSE	CV RMSE	Test MSE	Test RMSE
HARRV	1, 6, 16	3.1617e-04	1.7781e-02	4.8442e-06	2.2010e-03
MLP 4 layers with dropout	10	2.9613e-04	1.7209e-02	4.8704e-06 (var 1.3311e-14)	2.2067e-03 (var 6.8202e-10)
MLP 4 layers with BN	10	3.0151e-04	1.7364e-02	4.8212e-06 (var 4.9997e-14)	2.1952e-03 (var 2.5325e-09)
LSTM 2 layers + 1 Dense	12	2.9845e-04	1.7276e-02	4.8233e-06 (var 2.0082e-14)	2.1960e-03 (var 1.0293e-09)
GRU 1 layer + 2 Dense	5	**2.9607e-04**	**1.7207e-02**	4.7433e-06 (var 8.8148e-15)	2.1778e-03 (var 4.5781e-10)
CNN 2 layers + 1 Dense	6	3.0608e-04	1.7495e-02	4.7605e-06 (var 1.2904e-14)	2.1817e-03 (var 6.7603e-10)
SVM	7	3.2120e-04	1.7922e-02	4.3463e-05	6.5293e-03
Ridge Regression	6	3.0615e-04	1.7497e-02	**4.6667e-06**	**2.1603e-03**

4 Conclusion

We have analyzed high-frequency Bitcoin time series sampled on minute scale using statistical method and machine learning algorithms. First, the autocovariance function of the minute-based log return values not only shows a negative significant value at lag 1, but also small positive values distinct from zero at lags 2, 3 and 4. Since cryptocurrency exchanges continue trading with customers all over the globe, 24 h a day, there is no a priori reason to sample the realized volatility at daily scale. Given the length of the data set (about 6 years) and the data-savvy machine learning algorithms, we decided to aggregate the RV values using 3-h long intervals. The RV values show a weak correlation with relative values of the 3-h interval based high-low price extent. This work has focused solely on the heterogeneous autoregressive dynamics. We have found that albeit at the validation data set level neural network algorithms provide good fits of the RV dynamics, this does not carry over to the test set benchmarks. In particular, the best performing method is the ridge regression, in which past RV values are used as predictors. The optimized lags were 1, 6, and 16. We remark, nevertheless, that the assumptions of heterogenous time scales for the auto-regressive process is not necessary valid at crypto-currency exchanges, which have their specific dynamics. For instance, it may be possible to use differenced log values of realized volatility to model the increment process. Future work along these lines will include predicting the RV values from a broader set of indicators, in additiion to the past RV data, especially the minute-scale time series of logarithmic returns, and transaction volume series.

References

1. Andersen, T.G., Bollerslev, T., Diebold, F., Labys, P.: Modeling and forecasting realized volatility. Econometrica **71**, 579–625 (2003)
2. Kaggle, Bitcoin historical data. https://www.kaggle.com/mczielinski/bitcoin-historical-data. Accessed 1 Dec 2018. Released under CC BY-SA 4.0 license
3. Moews, B., Herrmann, J.M., Ibikunle, G.: Lagged correlation-based deep learning for directional trend change prediction in financial time series. Expert Syst. Appl. **120**, 197–206 (2019)
4. Cao, J., Li, Z., Li, J.: Financial time series forecasting model based on CEEMDAN and LSTM. Phy. A: Stat. Mech. Appl. **519**, 127–139 (2019)
5. Mallqui, D.C.A., Fernandes, R.A.S.: Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. Appl. Soft Comput. **75**, 596–606 (2019)
6. Lahmiri, S., Bekiros, S.: Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos, Solitons Fractals **118**, 35–40 (2019)
7. Nakano, M., Takahashi, A., Takahashi, S.: Bitcoin technical trading with artificial neural network. Phys. A: Stat. Mech. Appl. **510**, 587–609 (2018)
8. Rosenblatt, F.: Principles of Neurodynamics Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington (1961)
9. LeCun, Y., et al.: Back-propagation applied to handwritten zip code recognition. Neural Compu. **1**(4), 541–551 (1989)
10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
11. Cho, K., van Merrienboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: encoder-decoder approaches. In: 8th Workshop on Syntax. Semantics and Structure in Statistical Translation, pp. 102–111. Association for Computational Linguistics, Doha (2014)
12. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, Heidelberg (1995). https://doi.org/10.1007/978-1-4757-3264-1
13. Hoerl, A.E., Kennard, R.W.: Ridge regression: biased estimation for nonorthogonal problems. Technometrics **12**, 55–67 (1970)