Multifactorial Linear Model for Crptocurrency Prediction: OLS and Lasso

Haoyuan Ma
haoyuanma@ucsb.edu

Financial Mathematics & Statistics Major University of California, Santa Barbara Santa Barbara, California

Abstract—Price forecasting is pretty crucial in the asset management and allocation and quantitative trading industries. With the development of the global economic situation, decentralized finance has gradually entered people's field of vision, and cryptocurrency and cryptocurrency finance have become the research fields of a large number of scientists and practitioners. This article will use two traditional and basic machine learning models to predict the price of cryptocurrencies. Primarily, the tick data and order book information including the price and trading volume of cryptocurrencies will be obtained through OKEX. Subsequently, 16 common technical indicators will be used for stock trading to build OLS and Lasso models. The model is trained on the first part of the data, and finally the model is made to make predictions about future prices and compare it with the second half of the data, which is the true price. According to the evaluations, when the price is in the middle range, the accuracy of the model is relatively high, and it can also follow the market trend well. However, when the price fluctuates, the model predicts the result to be extreme than the reality, i.e., higher than the price above the median price, lower than the price below the median price, and the overall volatility is greater. These results shed light on guiding further exploration of cryptocurrency pricing forecasting.

Keywords: Cryptocurrency; Machine learning; Price forcasting; OLS; Lasso.

1 INTRODUCTION

After the 2008 financial crisis, as a product of rising decentralized finance, cryptocurrencies have gradually occupied an indispensable position in the global capital market [1, 2]. From precious metals to paper money, the importance of money is evident in global trade and commerce. As society enters the digital age, more forms of digital currency are in fierce market competition for rapid dissemination. From this point of view, digital currency trading will be a future financial technology battlefield with great potential. Cryptocurrency price forecasting is the basis for portfolio return optimization. There have been many deep learning and machine learning models for digital currency price prediction. In this article, the price of two weeks of cryptocurrencies obtained on OKEX will be used to construct the OLS and Lasso models, and finally make a price prediction.

For the influencing factors of cryptocurrency prices, 16 commonly used stock technical indicators are selected, which are CR indicator, KDJ indicator, SMA indicator, MACD indicator, BOLL indicator, RSI indicator, WR indicator, CCI indicator, TR, ATR indicator, DMA indicator, DMI, +DI, -DI, DX, ADX, ADXR indicator, TRIX, MATRIX indicator VR, and MAVR indicator. Then, the correlation coefficients between these indicators and the price of the cryptocurrency

FFIT 2022, October 28-30, Shenzhen, People's Republic of China

DOI 10.4108/eai.28-10-2022.2328414

will be calculated in order to filter the indicators that are closely related to the price, and eliminate the indicators that are not related to the price of the cryptocurrency. Subsequently, an OLS model and a Lasso model is constructed for these metrics.

The sample regression model is determined using the principle of minimizing the sum of squares of the estimated residuals, which is called the least squares criterion, also known as the OLS regression model. In this model, CR indicator KDJ indicator SMA indicator MACD indicator BOLL indicator RSI indicator WR indicator, CCI indicator TR, ATR indicator DMA indicator DMI, +DI, -DI, DX, ADX, ADXR indicator, TRIX, MATRIX indicator, VR and MAVR are served as the factors. The conclusion can be obtained by performing OLS regression analysis on the data standardized and satisfying the above statistical assumptions and predicting and comparing the future cryptocurrency prices.

Afterwards, a Lasso regression analysis will be run on the same data. Reasons why not Ridge and Lasso are below. Both Lasso and Ridge are tools for preventing overfitting based on linear regression models. Among them, Lasso uses the L1 penalty to cause the smaller regression coefficient to become 0, and Ridge uses the L2 penalty to reduce all the regression coefficients. Since Lasso will make some regression parameters 0, it is convenient for me to perform feature selection [3-5]. In practice, the industry is also more inclined to use Lasso regression. This article will focus on the two regression models of OLS and Lasso for cryptocurrency currency price prediction. The rest part of the paper is organized as follows. The Sec. II will discuss the methodology that will be applied to conduct this price prediction research. The Sec. III will analysis and discuss the results produced by OLS and Lasso models. Eventually, a summary will be given in Sec. IV.

2 METHODOLOGY

The data is first cleaned, organized, and standardized. The data includes the opening price, closing price, high price, low price and trading volume of cryptocurrencies, etc. The data used in the study consisted of 692,704 seconds of cryptocurrency prices and volumes used to train the machine learning model, plus 188,198 seconds of cryptocurrency prices used to test the model's accuracy. During cleaning, the blanks were filled with the mean before and after, and the errors were deleted. Next, z-score normalization was performed. The processed data has a mean of 0 and a standard deviation of 1. The normalize formula is:

$$x* = \frac{(x - \mu)}{\sigma} \tag{1}$$

This approach is an essential step before proceeding with OLS and Lasso model creation [6].

For the model to be created, CR indicator, KDJ indicator, SMA indicator, MACD indicator BOLL, indicator RSI indicator, WR indicator, CCI indicator TR, ATR indicator DMA indicator DMI, +DI, -DI, DX, ADX, ADXR indicator, TRIX, MATRIX indicator VR, MAVR indicator are chosen as x and the cryptocurrency price is set as y. The value of each x is calculated from the opening price, high price, low price and trading volume in the source data, and then the correlation between these x and y is analyzed. The formula of correlation is

$$r = \frac{\sum[(x_i - x_*)(y_i - y_*)]}{\sqrt{\sum(x_i - x_*)^2 * \sum(y_i - y_*)^2}} \tag{2}$$

Based on the calculation of the correlation value, one sees that there were overfitting and extremely irrelevant values, eliminated them, and retained the more relevant x values for model construction.

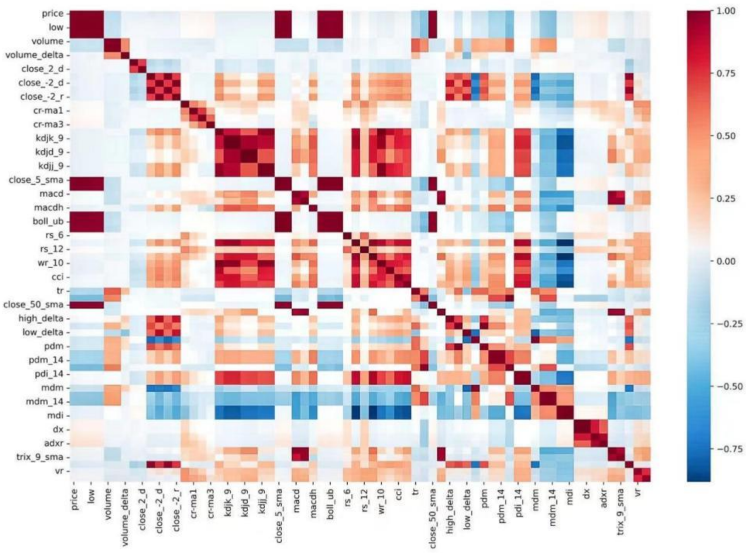

Figure 1. Correlation of features.

After correlation analysis, two most positively correlated factors are chosen and one most negatively correlated factor for plotting. From the calculation of the correlation analysis, I found that the simple moving average and the boll indicator calculated based on 5 time points are the most representative. A simple moving average (SMA) is an arithmetic moving average calculated by adding recent prices and then dividing that figure by the number of time periods in the calculation average.

$$SMA = \frac{A_1 + A_2 + A_3 + \cdots + A_n}{n} \tag{3}$$

where A_n is the price of cryptocurrency at period n and n is the number of total periods. Boll uses statistical principles to find the standard deviation of stock prices and their confidence intervals, so it is also called the Bollinger Band. Its upper and lower limits are not fixed and vary with the rolling stock price. After computing the two most relevant features, all features are fused into the OLS and Lasso models. Finally, a complete algorithm is presented to predict the price. Besides, coefficient of determination (R^2), mean squared error, and mean absolute error are also calculated to test the accuracy of the models. The corresponding formulae are:

$$R^2 = 1 - \frac{SS_{rse}}{SS_{tot}} \tag{4}$$

$$MSE = \frac{\sum_{i=1}^{n}(y_i - y_i^*)^2}{n} \tag{5}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{6}$$

3 RESULT & DISCUSSION

As shown in Figures 2-7, the prediction results of the two models that are most positively correlated with cryptocurrency prices. Subsequently, this research refers to the predictions of the models most negatively correlated with cryptocurrency prices. Finally, the predictions of the entire algorithm are compared with the true values.

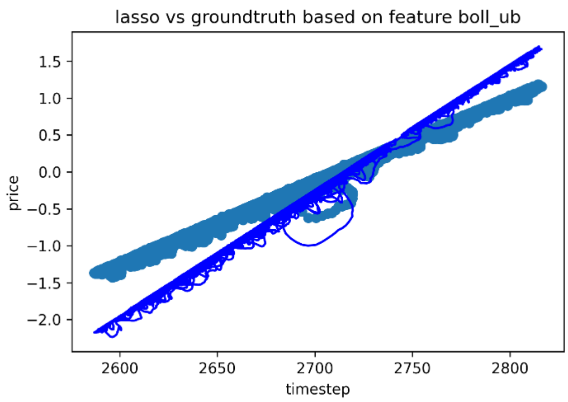

Figure 2. Lasso vs Groundtruth based on feature boll_ub.

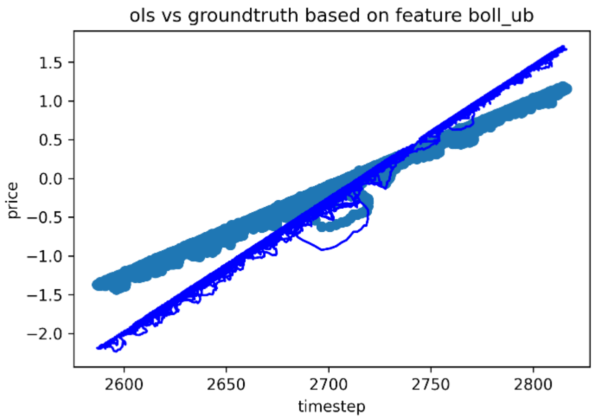

Figure 3. OLS vs groundtruth based on feature boll_ub.

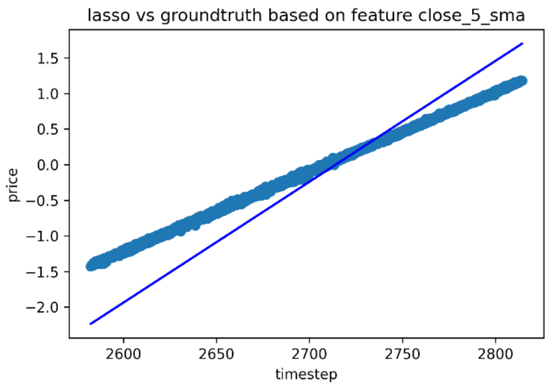

Figure 4. E Lasso vs Groundtruth based on featrue close_5_SMA.

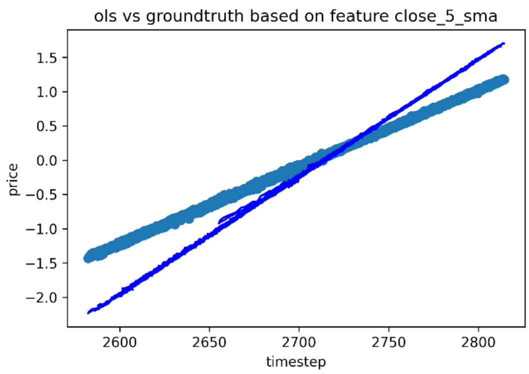

Figure 5. OLS vs Groundtruth based on feature close_5_sma

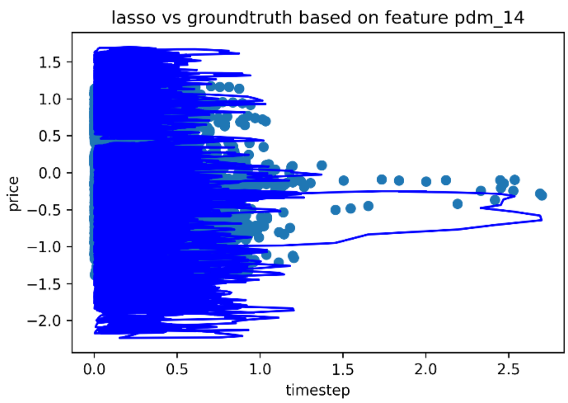

Figure 6. Lasso vs Groundtruth based on feature pdm_14.

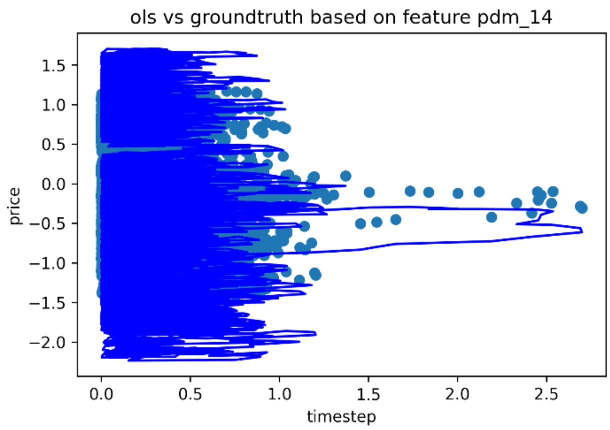

Figure 7. OLS vs groundtruth based on feature pdm_14

First, for the correlation between these 16 indicators and the price of the cryptocurrency, it is not difficult to find that the price changes of stocks and the cryptocurrency have something in common. However, there are also many indicators that are not applicable to the market fluctuations of cryptocurrency currencies. From the perspective of model construction, there is almost no difference between the prediction results of the OLS and Lasso models according to Figures 8 and 9. The model's predictions are very accurate when prices are in the mid-range, and more discrete than the truth when prices are high or low. In other words, when the market is relatively low, the predicted price of the two models will be lower than the real price, and when the market is high, the predicted price of the two models will be higher than the real price. This conclusion holds regardless of whether the single most relevant feature or the entire algorithm is selected. The comparison of the model is given in Table. 1.

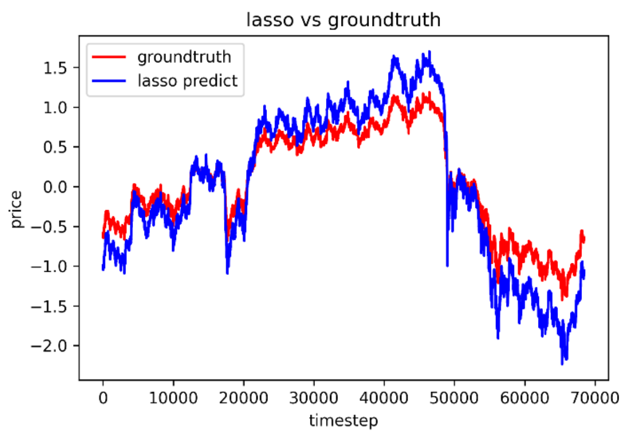

Figure 8. Lasso vs Groundtruth.

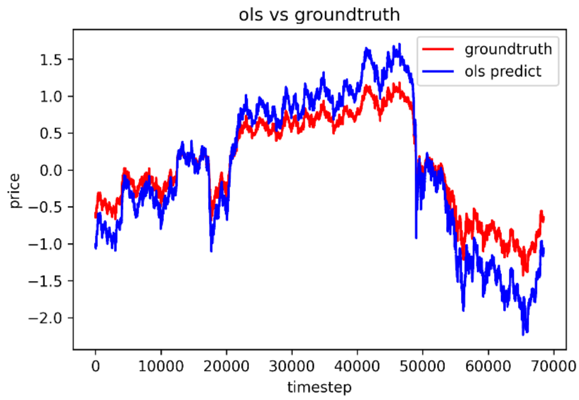

Figure 9. OLS vs Groundtruth.

Although the trend in the forecast results is accurate, and the model works well in the median price range, there are shortcomings in using the model to predict cryptocurrency prices, e.g., lack of macro awareness, too much data, overfitting of certain features. To be specific, a lack of macro awareness can lead to models failing to predict black swan events, e.g., the recent drop in the price of the LUNA virtual currency to near zero. Excessive amount of data also increases the workload for the use of statistical methods and does not significantly improve the accuracy of the model [7-10].

Table 1 10 Model Results.

	OLS	Lasso
Score	0.708472043260	0.724532101
Intercept(w0)	3.00161574e-28	1.54713081e-17
R2	0.708472043260	0.7245321015
MSE	0.12423438847	0.1173904084
RMSE	0.352468989383	0.342622836
MAE	0.303037018158	0.2945051486

4 CONCLUSION

In summary, this paper investigates cryptocurrency price forecasting based on OLS and Lasso models. Regression analysis is a statistical method that determines the quantitative relationship between two or more variables and is widely used. The sample regression model is determined using the principle of minimizing the sum of squares of the estimated residuals, known as the least square criterion. This essay only introduces the least squares criterion under classical assumptions, also known as ordinary least squares estimation, abbreviated as OLS estimation (ordinary least squares estimators). Using them to make predictions about cryptocurrency prices is an effective but relatively narrow means. It is necessary to upgrade in feature selection and model optimization to achieve better results. Cryptocurrency price prediction has far-reaching implications for both asset management and decentralized finance investing. In the future, cryptocurrencies will occupy a larger global financial market. Overall, these results offer a guideline for construction of the basic machine learning model for digital currency price prediction.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system." Decentralized Business Review, vol.1, 2008, 21260.

[2] K. Rudd, "The global financial crisis." Monthly, The Feb 2009, 2009, pp. 20-29.

[3] C. Hans, "Bayesian lasso regression." Biometrika vol. 96.4, 2009, pp. 835-845.

[4] J. Li, et al. "Feature selection: A data perspective." ACM computing surveys (CSUR), vol. 50.6, 2017, pp. 1-45.

[5] J. Ranstam, and J. A. Cook. "LASSO regression." Journal of British Surgery vol. 105.10, 2018, pp. 1348-1348.

[6] C. Dismuke, and R. Lindrooth. "Ordinary least squares." Methods and Designs for Outcomes Research vol. 93, 2006, pp. 93-104.

[7] P. Jay, et al. "Stochastic neural networks for cryptocurrency price prediction." Ieee access vol. 8, 2020, pp. 82804-82818.

[8] M. Glenski, et al. "Improved forecasting of cryptocurrency price using social signals." arXiv preprint arXiv:1907.00558, 2019.

[9] C. H. Wu, et al. "A new forecasting framework for bitcoin price with LSTM." 2018 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 2018.

[10] R. Farell, "An analysis of the cryptocurrency industry." Diss, Repository of UPenv, 2015