Forecasting prices of alternative assets

Samay, Panwar

2021

Forecasting Prices of Alternative Assets

Panwar Samay
School of Physical and Mathematical Sciences

Dr. Pun Chi Seng, Patrick (Asst Prof)
School of Physical and Mathematical Sciences

Abstract - Recently, Cryptocurrencies have been the new alternative asset choice for retail investors around the globe. Due to Cryptocurrencies' characteristic high volatility, Central Banks do not acknowledge them as a store of value and therefore, not as money. While some scholars argue that cryptocurrencies provide a sort of a sink for excess funds, thus keeping inflationary pressures at bay, their high volatility has caused wealth destruction effects at the same time. Cryptos are regarded as unforecastable and the best known strategy regarding crypto so far has been the buy-and-hold strategy which has provided investors returns in excess of 100%.

This paper aims to establish a causal relationship between pairs of cryptocurrencies and utilise the information to forecast returns with the help of a novel OLS-LSTM model proposed within. The practicality of this approach is shown by implementing a simple trading strategy based on the signals generated by the model. We further aim to execute a pairs trading algorithm between different pairs of cryptocurrencies.

Keywords – Granger Causality, LSTM, Cryptocurrency

1 INTRODUCTION

The allude of growing your investment exponentially has been the main driver for the hype around crypto investment. It is also one of the main reasons that they are extremely volatile. Not much research has been devoted into the forecasting or studying the properties of cryptocurrency time series and there is a vacant space in research literature in the same. This paper aims to establish a causal relationship between pairs of cryptocurrencies and utilize the information to forecast returns with the help of a novel RLM-LSTM model proposed within. The practicality of this approach is shown by implementing a simple trading strategy based on the signals generated by the model.

Some of the assumptions made during the study are:there are no trading costs; the study is conducted on a set of 8 crypto currencies and it is assumed that the remaining investment universe of the same has similar behavior and characteristics.

The remainder of this paper is organized as follows: Section 2 introduces the Literature Review and Section 3, the data collection process, along with data cleaning and feature engineering is explained. In Section 4, the methodology followed by the author is introduced as well as the model architecture and pipeline. Section 5 talks about the Experimental results of the models as witnessed empirically with a simple trading strategy and; the paper finally ends with a conclusion in Section 7.

2 LITERATURE REVIEW

There have been attempts to forecast cryptocurrency prices with a variety of machine learning algorithms such as Gated Recurrent Neural Networks [3], LSTMs [7] as well as attempts at regime modelling and bubble detection using Attention based models. Research has also been conducted into using alternative data such as tweets and new headlines as exogenous factors into language models such as BERT [10].

3 DATA COLLECTION

All the data in this paper was collected from https://www.cryptodatadownload.com/data/gemini/

After collection, the data was cleaned by properly indexing and filling NaN values. The data was structured in the form of a multivariate time series and a visualization of each cryptocurrency time series is visible in Figure 1.

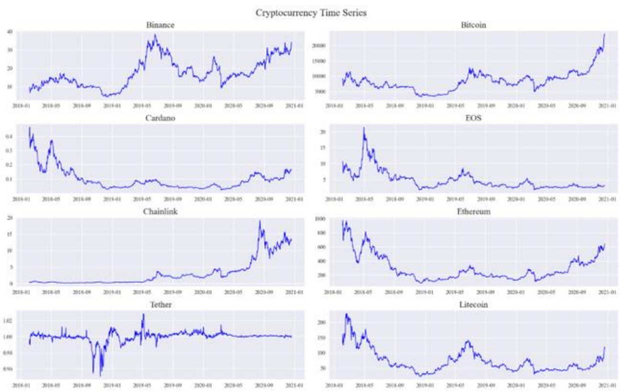

Figure 1: Cryptocurrency Multivariate Time Series

Summary statistics for each time series were collated and presented in the form of graphs, an example of which is visible in Figure 2.

Figure 2: Summary Statistics of Bitcoin

A majority of the time series had positive skews and indicated high level of autocorrelation. The histograms, as well as the boxplots indicated that the cryptocurrencies failed to follow a Gaussian distribution and instead followed some form of gamma or Cauchy distribution. This was again confirmed with the QQ plot, comparing the distribution of our series with that of a normally sampled series.

Fat tails were also indicative from the boxplot and the series exhibited high kurtosis. To statistically confirm non-normality of the data, the Jarque Bera test was utilized with a significance level of 5%. As assumed, the test indicated non-normality of the data. The augmented dickey fuller test was conducted with the same significance level that indicated non-stationarity as well.

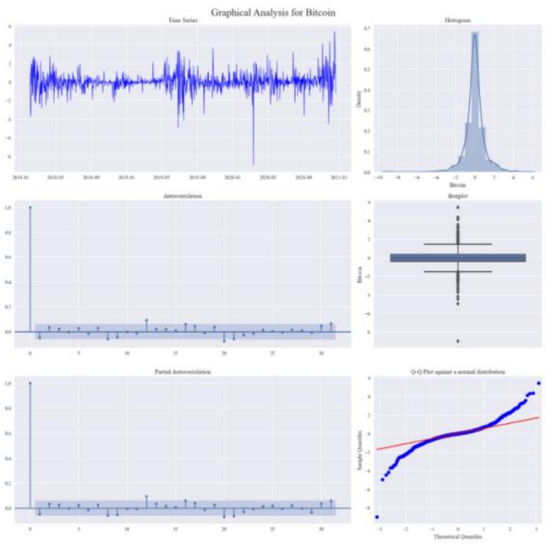

Figure 3: Summary Statistics of Bitcoin After

The multivariate time series was then differenced and standardized by subtracting the $y_{(t-1)}$ observation from y_t observation and then subtracting the mean of the series and dividing by the standard deviation. This made sure that the scale of each series was the same. The summary graphs of the standardized Bitcoin time series is depicted in Figure 3.

While the data looks stationary now, it still is not Gaussian because of the fat tails on both ends. This suggests a t-distribution instead.

4 METHODOLOGY

In our paper, we employ the use of an RLM-LSTM hybrid model that mainly consists of four main components:

1. Finding the most cointegrated granger causal pair.
2. Regressing the predictor on the response series.
3. Taking the residuals of the regression and using them as input to an LSTM model.
4. Combining the RLM and the LSTM forecasts into a series.

4.1 GRANGER CAUSALITY

In this step, we first check for cointegration. Cointegration can be defined as a statistical property of a multivariate time series where two or more time series are individually integrated but a linear combination of them results in a lower order of integration.

Every possible combination of two series is checked for cointegration using the Engle-Granger two step cointegration test and the pair of series that results in the lowest p-value is chosen as the most cointegrated pair.

The pair is then tested for Granger causality using the F-statistic and the lag of the predictor series at which the lowest p-value is achieved is again noted down.

4.2 REGRESSION

After computing the most cointegrated causal pair and the best lag value, the residuals are calculated by regressing the predictor series onto the response variable series.

These residuals are then checked for normality and the other gauss Markov assumptions are ensured. Tests for endogeneity and heteroskedasticity are conducted and we fail to reject the Null Hypothesis in each test. The residuals are then passed through a Fast Fourier transform to remove excessive noise by clipping noisy observations. These modified residuals are then passed onto the LSTM model.

```
-----------------------------------------------------------------
Litecoin is most cointegrated with Bitcoin for the given timeframe
-----------------------------------------------------------------
--------------------------------------------------------------------
We have achieved a F-Statistic of 2.26 and p-value of 0.13272 with a Lag-order of 1
--------------------------------------------------------------------
                    Robust linear Model Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1052
Model:                            RLM   Df Residuals:                     1050
Method:                          IRLS   Df Model:                            1
Norm:                          HuberT
Scale Est.:                       mad
Cov Type:                          H1
Date:                Mon, 01 Mar 2021
Time:                        13:05:34
No. Iterations:                    24
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0093      0.017     -0.545      0.586      -0.043       0.024
x1            -0.0818      0.017     -4.800      0.000      -0.115      -0.048
==============================================================================
```

Figure 4: RLM Regression Summary

4.3 LSTM MODEL

A Long-Short Term memory network is utilised with a custom loss function of taking the mean absolute difference between the true response values and the forecasted response values which is computed by adding the predicted residual by the LSTM model to the predicted response variable from the RLM model.

A window size of 20 is used as input to the model, i.e. the model looks at the past 20 lags of the modified residuals and then predicts the next residual. The model consists of two LSTM layers coupled with Dropout layers and a batch size of 32 observations. Three densely fully connected layers with a ReLu activation function for each. The stochastic gradient descent optimizer is utilised with a decay of 0.05. The model is trained for a total of 100 epochs.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 20, 40)            6720
_________________________________________________________________
dropout (Dropout)            (None, 20, 40)            0
_________________________________________________________________
lstm_1 (LSTM)                (None, 20, 20)            4880
_________________________________________________________________
dropout_1 (Dropout)          (None, 20, 20)            0
_________________________________________________________________
flatten (Flatten)            (None, 400)               0
_________________________________________________________________
dense (Dense)                (None, 32)                12832
_________________________________________________________________
dense_1 (Dense)              (None, 16)                528
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 17
=================================================================
Total params: 24,977
Trainable params: 24,977
Non-trainable params: 0
```

Figure 5: LSTM Model Architecture

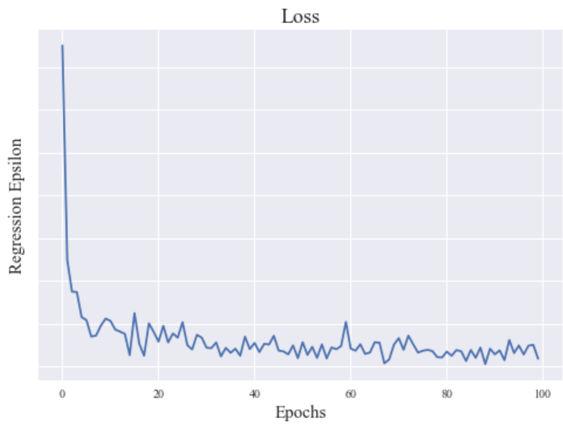

Figure 6: LSTM training loss

4.4 RECOMBINATION

The forecasts from the RLM model and the predicted residuals from the LSTM model are combined together to create a series. The previously calculated mean and standard deviation are used to rescale the series back to the original scale and the previous values are added back to un-difference the series.

This series is then compared with a naïve persistence forecast where we assume that the y_t observation will be the same as y_{t+1} observation. The Mean Absolute Percentage Error is calculated for both the series and the relative MAPE (rMAPE) is then calculated by dividing our model's MAPE by the naïve model MAPE. This ratio displays that our model can accurately forecast better than a naïve baseline model.

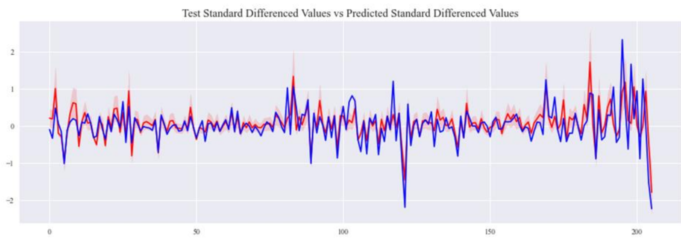

Figure 7: Comparing Predicted vs Actual

5 EXPERIMENTAL RESULTS

The results of the model indicate that our hybrid RLM-LSTM model is able to better forecast the cryptocurrency series as compared to a naïve model and we achieve an rMAPE of 0.684, indicating that our model achieves a lower MAPE as compared to the baseline model.

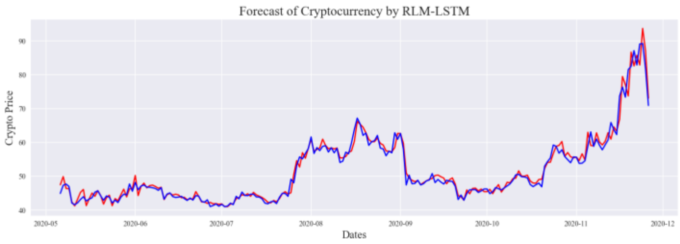

Figure 8: Forecast vs. Actual series

To demonstrate the practicality of the model, we go ahead and simulate a simple trading strategy where we buy a unit of the cryptocurrency at time t if our model forecasts an increase in its price tomorrow, else if our model forecasts a decline in the price of the cryptocurrency as compared to today then we sell a unit of the cryptocurrency.

We start with an initial cash amount of 10,000$. Throughout the strategy, we keep track of our portfolio value and the number of assets we currently hold. Through the period of a year, the backtest yields returns of 15.16% while holding 49 assets of the cryptocurrency. The value of the portfolio is visible throughout time in Figure 9.

It is interesting to note that the strategy suggests that buying the asset is the best strategy for the backtest period which with hindsight makes sense with the bull run of cryptocurrencies throughout the latter half of 2020.

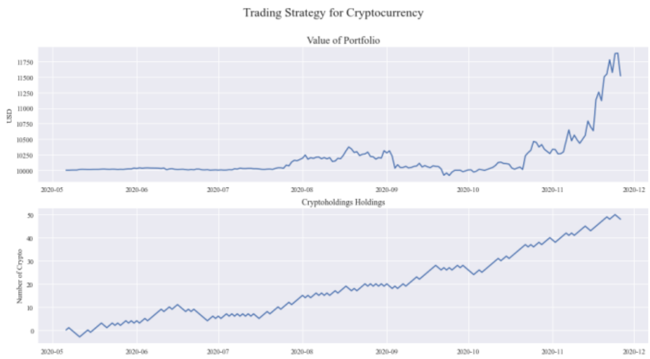

Figure 9: Backtest performance of Strategy

6 CONCLUSION

The results obtained from our hybrid model do in fact beat naïve baseline models in terms of accuracy and the profitability of such a strategy is demonstrated as well through the trading strategy based on the RLM-LSTM forecast. However, as market regimes change, the forecast might not be able to keep up and causal relationships might change through time. Cryptocurrencies are largely sentiment driven and hence, are multiplicative in nature. This causes steep bull as well as bear runs which cause the characteristic high volatility of cryptocurrencies. It is worth noting that further research can be conducted with the use of rolling or expanding models along with ensembles which have shown excellent results in terms of forecast. However such models results come at the cost of interpretability. The research literature in such fields is sparse but through our paper, we aim to add a little light to the unexplored field that is cryptocurrencies.

ACKNOWLEDGMENT

I would like to thank my project supervisor, Dr. Pun Chi Seng for his patience with my project and his continuous support throughout.

I would like to acknowledge the funding support from Nanyang Technological University – URECA Undergraduate Research Programme for this research project.

REFERENCES

[1] A. Cretarola and G. Figà-Talamanca, “Bubble regime identification in an attention-based model for Bitcoin and Ethereum price dynamics,” *Econ. Lett.*, vol. 191, no. 108831, p. 108831, 2020.

[2] C. Liu, W. Hou, and D. Liu, “Foreign exchange rates forecasting with convolutional neural network,” *Neural Process. Lett.*, vol. 46, no. 3, pp. 1095–1119, 2017.

[3] I. E. Livieris, E. Pintelas, S. Stavroyiannis, and P. Pintelas, “Ensemble deep learning models for forecasting cryptocurrency time-series,” *Algorithms*, vol. 13, no. 5, p. 121, 2020.

[4] L. Pichl, International Christian University, Osawa 3-10-2, Mitaka, Tokyo 181-8585 Japan, and T. Kaizoji, “Volatility analysis of bitcoin price time series,” *Quant. Fin. Econ.*, vol. 1, no. 4, pp. 474–485, 2017.

[5] N. Uras, L. Marchesi, M. Marchesi, and R. Tonelli, “Forecasting Bitcoin closing price series using linear regression and neural networks models,” *PeerJ Comput. Sci.*, vol. 6, no. e279, p. e279, 2020.

[6] P. Jay, V. Kalariya, P. Parmar, S. Tanwar, N. Kumar, and M. Alazab, “Stochastic neural networks for cryptocurrency price prediction,” *IEEE Access*, vol. 8, pp. 82804–82818, 2020.

[7] R. Swagat, S. Shruti, S. Sital, and S. Subarna, “Comparison of algorithms in Foreign Exchange Rate Prediction,” in *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*, 2018.

[8] S. Ji, J. Kim, and H. Im, “A comparative study of Bitcoin price prediction using deep

learning," *Mathematics*, vol. 7, no. 10, p. 898, 2019.

[9] T. Shintate and L. Pichl, "Trend prediction classification for high frequency Bitcoin time series with deep learning," *J. Risk Fin. Manag.*, vol. 12, no. 1, p. 17, 2019.

[10] Y. B. Kim *et al.*, "Predicting fluctuations in cryptocurrency transactions based on user comments and replies," *PLoS One*, vol. 11, no. 8, p. e0161197, 2016.