Forecasting Cryptocurrency Prices Using ARIMA and LSTM

Xinruo Hu
Columbia University
xh2464@columbia.edu

Nthabi Rapoeea
Columbia University
njr2143@columbia.edu

Daniel Schmidle
Columbia University
dps2150@columbia.edu

Junfeng Yan
Columbia University
jy3118@columbia.edu

Abstract

This paper aims to predict future prices of Bitcoin using daily sample price data between January 02, 2019 and March 31, 2021. The volatility index (VIX) is used to divide the analysis period into normal and abnormal market conditions. This paper analyzes the performance of both an Autoregressive Integrated Moving Average (ARIMA) and Long Term-Short Memory (LSTM) networks. The LSTM is commonly used in financial time series and is known for its ability to learn relevant dependency order, making it a powerful tool for trend recognition. The ARIMA model is utilized as the baseline for comparison against the "more-advanced" LSTM model. This paper finds that both ARIMA and LSTM models are suitable candidates for a long-term, risk-averse strategy. However, the ARIMA model currently seems to be the best choice as a predictive base for trading Bitcoin.

1 Introduction

1.1 Cryptocurrency

Bitcoin is the oldest decentralized cryptocurrency and has the biggest market capitalization of approximately $896bn, as of December 19, 2021. Bitcoin and other cryptocurrencies trade 24/7. Bitcoin is a digital asset that uses public-key cryptography to record, sign and send transactions over the Bitcoin blockchain technology. Bitcoin transactions are processed electronically and don't require intermediaries to be facilitated. There is a supply limit of 21 million Bitcoins as set by its inventor(s). Currently, there are over 18,7 million Bitcoins in circulation. Despite the misconception about its anonymity, Bitcoin is pseudo-anonymous. All Bitcoin transactions are recorded and stored on a public ledger. The ledger has balances and Bitcoin addresses, and it can be accessed by anyone on the network.

Bitcoin's nascent stage and lack of formal regulations make it susceptible to arbitrage opportunities often attributed to its volatility induced by speculation, inefficiencies, and other unknown influencing factors. There are also opposing views about how its value is derived and whether it can be recognized as a financial asset. Notwithstanding all these elements, Bitcoin has become increasingly popular over the years, and with the arrival of the cryptocurrency market, the economy may be standing at the precipice of a monetary revolution.

1.2 Related Works

The topic of machine learning has risen to prevalence in recent years, across multiple fields. For instance, machine learning has become the bedrock of high frequency trading (HFT). Although, there may still be some skepticism regarding complete control of algorithmic trading over a human-decision based market, both concepts are concurrently occupying the same space. There is also a fair amount of academic literature that encourages the use of machine learning in finance. The use of deep learning in financial time series has gained some momentum and various methods such as the LSTM networks have shown positive results in their predictive power. The work that supports the use of the LSTM networks in financial time series include that of [1]Fischer and Krauss (2018), where they used the LTSM to predict future prices of the S&P 500 Index constituents between 1992 and 2015. Their results showed that the LSTM outperformed memory-free classification models such as RandomForest and Logistic Regression Classifier. [4]Adil Moghar et al.(2020) found that the LSTM network has more predictive power on financial time series when compared to the ARIMA model. Their work deployed the RNN based LSTM to predict future prices of GOOGL and NKE using adjusted daily closing prices. Furthermore, [6]Sima Siami Namin et al.(2018) compared the predictive power between the ARIMA and LSTM on financial time series, their results showed that on average the LSTM produced an accuracy rate of 85%. Some of these results resonate with the idea that the use of machine learning on financial data can help enhance returns.

(For full reference info please see the Bibliography section)

1.3 Data

We have sourced Bitcoin daily price data from Bloomberg. The daily data begins on 01/01/2019 and ends on 03/31/21. We split the assessment periods into two categories, the presence and absence of market shocks, to assess the model performance under different market conditions. We define the presence of market shocks as periods of unprecedented events that adversely impacted the economy such as the COVID-19 pandemic, the 2008 Global Financial Crisis, etc. These events often lead to unprecedented action within markets that typically involve elevated levels of volatility. We measure volatility using the volatility index (VIX), and where the level of the VIX for five consecutive trading days indicates a market shock, we categorize it as different market conditions for the purpose of our study. The concept behind the VIX is founded on how option prices behave in response to different market conditions. Period 1 indicates absence of market shocks while period 2 indicates the presence of market shocks.

Dataset split using daily closing prices
Period 1

- Training – January 02, 2019 to June 28, 2019
- Validation – July 01, 2019 to October 31, 2019
- Testing – November 01, 2019 to February 21, 2020

Period 2

- Training – February 24, 2020 to August 31, 2020
- Validation – September 01, 2020 to December 31, 2020
- Testing – January 04, 2021 to March 31, 2021

Additionally, lagged values of the daily closing prices are calculated and added to the data as a feature for future use.

2 Methodology

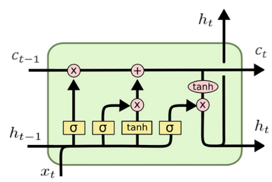
We fit two models to the Bitcoin daily spot data, one “base” ARIMA model and one “advanced” LSTM model. The base ARIMA model will serve as an industry standard for outcome comparison against the more advanced LSTM model. To establish a basis for model comparison, we also define a trading methodology to assess model performance.

2.1 ARIMA

The acronym ARIMA stands for Auto-Regressive Integrated Moving Average. Lags of the stationarized series in the forecasting equation are called "autoregressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series. These three components of the model are controlled by the hyper-parameters p,q,d, and we perform various diagnostic tests and grid searches to find values for these hyper-parameters. To find the optimal p,d,q for the ARIMA model, we searched various ARIMA model's Akaike's Information Criterion (AIC) as an estimator of prediction error. Letting k be the number of estimated parameters in the model and L be the maximum value of the likelihood function for the model then, the AIC value of the model is $AIC = 2k - 2ln(L)$. The data is split into train and test sets from each Period for model training and evaluation.

2.2 LSTM

Long short-term memory networks are an extension of recurrent neural networks (RNN). Compared to vanilla RNN, LSTMs are better at learning long-term dependencies that might present in the data. This is because LSTM are made up of memory cells, which are much like the memory of a computer and can read, write, and delete information from its memory. A memory cell has three gates: input, forget and output gates.



LSTM
(Long-Short Term Memory)
Image via Wikimedia Commons

The information in LSTMs flows through its gates: 1) forget gate: a gate that decides what information should be thrown away or kept, 2) input gate: a gate to update the cell state, 3) cell state: a gate to update the cell state to new values that the network finds relevant, 4) output gate: a gate to decide what the next hidden state should be. [3]Mathematically, a forward pass of an LSTM cell can be formulated as such:

$$
\begin{aligned}
f_t &= \sigma_g \left(W_f x_t + U_f h_{t-1} + b_f\right) \\
i_t &= \sigma_g \left(W_i x_t + U_i h_{t-1} + b_i\right) \\
o_t &= \sigma_g \left(W_o x_t + U_o h_{t-1} + b_o\right) \\
\tilde{c}_t &= \sigma_c \left(W_c x_t + U_c h_{t-1} + b_c\right) \\
c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \\
h_t &= o_t \circ \sigma_h \left(c_t\right)
\end{aligned}
$$

Where x_t is the input vector, f_t, i_t, and o_t are the forget gate, input gate and output gate, $\tilde{c}_t$ is the cell input vector, c_t is the cell state, and h_t is the hidden state vector and output vector of the LSTM cell.

We initialize our model using the Sequential module from Keras which allows us to create a Neural Network object with sequential layers. We chose Adam as our optimizer with a small learning rate of 0.0008 to accommodate the small sample size. Then, we add LSTM layers to predict the future price of Bitcoin.We used ReLu as our activation functions instead of tanh. Given that predicting the price of a financial product is quite a complex task, our aim is to have an expressive model that can capture upward and downward trends in the bitcoin price. Therefore, we decide to stack LSTM layers and set the hidden dimension size to be relatively large. We experimented with different numbers of LSTM layers and hidden state dimensions for each period. For the first period, we chose an architecture with two LSTM layers of hidden state dimension of 128 and 64 respectively. For the second period, our architecture contains a single LSTM layer with a hidden state dimension of 64. Dropout layers are added to prevent overfitting. Finally, we add a dense layer to produce a prediction.

The data is split into training and validation sets from each Period for model training and evaluation. Our LSTM models will be trained to make a single prediction for the next day based on five days of historical data. In other words, the training example for time step t would be a tuple of historical data from t-5 to t-1, which is the input, and the historical data for t, which is the ground truth. We use the TimeseriesGenerator function from keras to create a training data generator with a batch size of 5 and a validation data generator. We trained both period 1 and period 2 models with a maximum epoch of 500 and used early stopping with a patience of 40. Once training is terminated, the best model so far based on validation MAE will be returned.

2.3 Trading Strategy and Evaluation

Standard model evaluation techniques offer a reasonable gauge for model selection however, with the intention of creating a trading strategy on a predictive model, such model assessments can blindly fall into the trap of bias vs variance. For example, a model with a lower MAE might seem to fit the data better but, a given trading strategy based on that model might result in inferior returns. The closer a model fits the data, i.e., lower MAE, the more sensitive it might be to slight movements within the underlying asset. From this, consistent losses would occur as the model would perpetually lag the slight movements and might consistently predict incorrect market direction. A low MAE may result in a model that has overfit the data, so an increase in bias would also increase performance in this case. Thus, we propose the following trading strategy as the basic, main measurement for model performance. We first calculate the daily returns of the underlying daily prices.

Daily Returns:

$$R_t = \frac{P_t}{P_{t-1}} - 1$$

P_t is the price at time t

R_t is the rate of return at time t

Next, model forecasts of predicted next-day prices are then evaluated to create BUY signals based on this methodology:

$$\text{If } Pred_{t+1} > Pred_t : Buy$$
$$Else : Hold$$

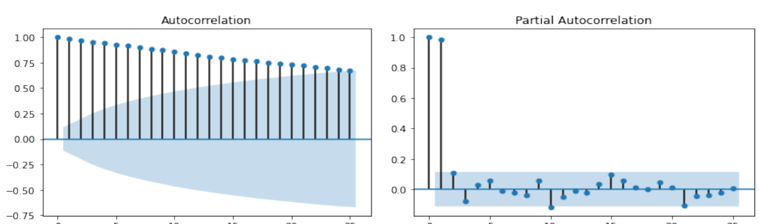
Where $Pred_{t+1}$ is the the model's predicted price at day $t+1$

Essentially, if the model predicts the price of tomorrow to be higher than today, then a simple BUY signal is generated. Models are then evaluated on Daily Returns based on concurrent BUY signals. We also compare a plain Buy and Hold strategy to illustrate model performance over the movement in the Bitcoin spot rate.

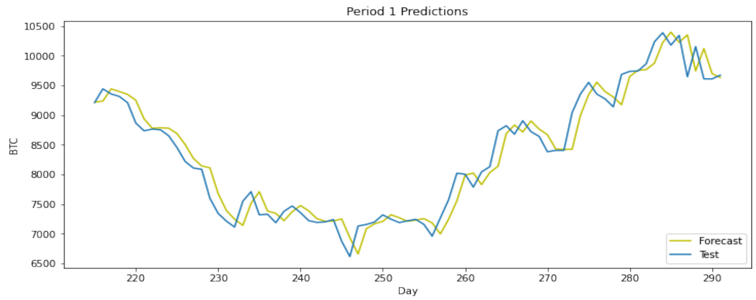
3 Results

3.1 ARIMA

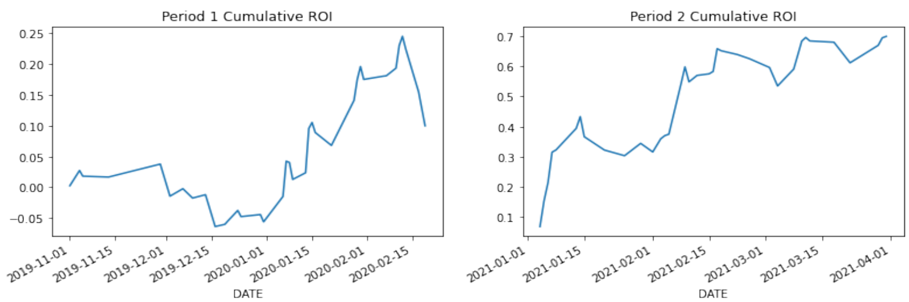
The ACF plot of a time series describes how well the present value of the series is related with its past values.



Our ACF plot suggests that the autocorrelations are significant for many lags, but perhaps the autocorrelations at later lags are merely due to the propagation of the autocorrelation at lag 1 and lag 2. This is confirmed by the PACF plot below. The partial correlation plot is a plot of the partial correlation coefficients between the series and lags of itself. The PACF plot spikes at lag 1 and lag 2, meaning that all the higher-order autocorrelations are effectively explained by the lag1 and lag 2 autocorrelation. This suggests that the data possess an AR(2) component. Using this as a general guideline for parameter span, we then searched for optimality using the min AIC. From this we obtain a (1, 1 ,0) model for Period 1 and a (0,1,1) model for Period 2.



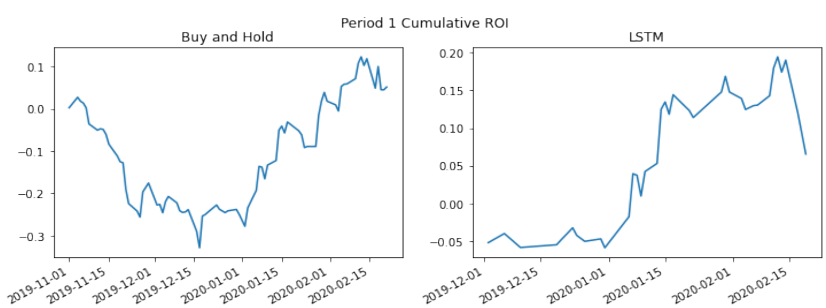
Visualizing the predictions against the validation set we can estimate the model fits well to the data. With our trading strategy and evaluated BUY signals we can see each model's cumulative return on investment (ROI) over the validation data. Below we can see the ARIMA model performs well over both time periods.



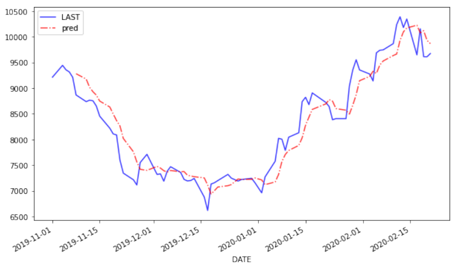
By the returns seen above, the ARIMA model seems to be a strong choice for model performance. So, let us now consider the LSTM model while also providing a market basis of a Buy and Hold strategy to compare the performance of both ARIMA and LSTM models.

3.2 LSTM

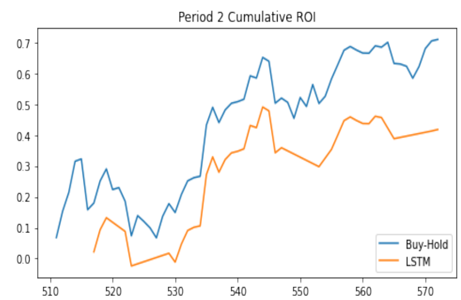
As seen in the charts below, The LSTM model produces similar results to the ARIMA model for Period 1.



The results indicated here for Period 1 show a significant return over a basic Buy and Hold strategy. However, when compared to the ARIMA model, the performance of the LSTM is lackluster, performing slightly worse than the ARIMA's ROI seen above. We can see in the plot below the visual representation of the LSTM predicted values over the Period 1 test data.



The LSTM model introduces bias over the ARIMA model predictions seen above and is less affected by slight movements in the market; however, this strategy does not prove to be a successful option over the higher-variance ARIMA above, at least for Period 1 test data. So, moving forward, let us assess the predictive ability of LSTM for Period 2.



For Period 2 the LSTM performs worse than the simple Buy and Hold strategy and worse than the ARIMA model seen above. The ARIMA model's performance is comparable to the Buy and Hold movement in the market. In the presence of market shocks and extreme volatility, the expectation of a model would not be to outperform the market but produce a reasonable return with the added benefit of risk aversion. In the event of extreme directional movement, the price of an asset cannot rise or fall forever and will eventually return to "normal" oscillating patterns. Thus, a model with long-term performance over the market, while also being resistant to extremes, would prove to be an advantageous strategy. As seen here, both ARIMA and LSTM models are suitable candidates for a long-term, risk-averse strategy. However, the ARIMA model currently seems to be the best choice for a predictive base when trading Bitcoin with this selected trading strategy.

4 Discussion

In this paper, the results of the ARIMA and LSTM models are presented as potential underlying strategies for trading the cryptocurrency markets. As the primary cryptocurrency, Bitcoin is utilized as the representative currency to illustrate a model's effectiveness in the markets. And from the results, we can see that there is great potential for a model-based trading strategy to ensure profits while also safeguarding against high volatility in the emerging crypto markets. Going forward, applying this project's methodology to other cryptocurrencies should be considered as a potential modeling and trading strategy to diversify risk and realize opportunities. Although intentionally outside the scope of this paper, adding complexity to the trading strategy, such as hedging with the use of derivatives, should be examined for additional gains. Moreover, including additional complexities to the LSTM model might further advance the performance over the ARIMA. For example, in the paper [5]"Deep LSTM with Reinforcement Learning Layer for Financial Trend Prediction in FX High Frequency Trading Systems," the author Francesco Rundo adds a reinforcement learning layer after the deep learning block to classify short- and long-term trends within FOREX currency pairs. By taking more of an "ensemble" way of thinking to the methodology of the model, potential efficiencies may be realized.

Bibliography

[1] Fischer, T., & Krauss, C. (2018) Deep learning with long short-term memory networks for financial market predictions.*European Journal of Operational Research, 270*(2), 654-669. doi:10.1016/j.ejor.2017.11.054

[2] Fischer, T., Krauss, C., & Deinert, A. (2019). Statistical Arbitrage in Cryptocurrency Markets. *Journal of Risk and Financial Management, 12*(1), 31. doi:10.3390/jrfm12010031

[3]Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. *Neural computation.* 9. 1735-80. 10.1162/neco.1997.9.8.1735.

[4]Moghar, A.,& Hamiche, M. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Computer Science*, 170, 1168-1173. doi:10.1016/j.procs.2020.03.049

[5] Rundo, F., (2019) Deep LSTM with Reinforcement Learning Layer for Financial Trend Prediction in FX High Frequency Trading Systems.*Applied Sciences, 9*(20) 4460. doi:10.3390/app9204460

[6]Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series.*2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. doi:10.1109/icmla.2018.00227

[7] Zhang, X., Liang, X. Zhiyuli, A., Zhang, S., Xu, R., & Wu, B. (2019). AT-LSTM: An Attention-based LSTM Model for Financial Time Series Prediction. *IOP Conference Series: Materials Science and Engineering, 569*(5), 052037. doi:10.1088/1757-899x/569/5/052037