Examining Predictive Pricing Indicators for Cryptocurrency Assets

Tiffany Valdecantos, Terrill Toe, Adetayo Adegoke

https://github.com/adegade/crytoanalysisproject

2022-05-04

Research Question and Motivation

According to the U.S. government's Internal Revenue Service (IRS), a virtual currency is a digital representation of value, other than a representation of the U.S. dollar or a foreign currency ("real currency"), that functions as a unit of account, a store of value, and a medium of exchange. The IRS defines cryptocurrencies as a virtual currency that uses cryptography to secure digitally recorded transactions on a distributed ledger, such as a blockchain [1].

Cryptocurrencies are mainstream assets electronically traded by many investors globally, similar to stocks and other equity types. There are many different cryptocurrencies, with several more being introduced to the public every few months for the opportunity to make transactions based on these assets. There is value to examining factors that drive cryptocurrency asset prices since the market size of these currencies is currently over 1 trillion USD in traditional fiat USD and over 32 trillion USD by 2027 [2].

Cryptocurrencies significantly influence the global economy because of their growing market size. Researchers examining cryptocurrency pricing predictors can inform government, business, and retail investors in making more profitable trading strategies.

This research paper examines the following question as it pertains to cryptocurrencies:

Can the technical analysis method of analyzing moving averages in Bitcoin prices be effective in predicting future upward price movement?

Hypothesis

We explore the method of analyzing moving averages used by technical analysts in the capital markets industry to make recommendations for future stock price movements. In doing this, technical analysts search for momentum in price movement caused by the disposition effect, which Bodie defined as "the tendency of investors to hold on to losing investments" (Bodie 384) [3]. To take advantage of this pattern, the investor must sell their holdings once the signal of the momentum of falling prices is determined. Prices would then continue falling until they bottom out and head back up again. At this point, the investor will watch for the signal of the rising momentum of stock prices. Once this signal is determined, that will be the investor's chance to buy the stock and ride the prices up. Through this strategy, the investor can cut losses when prices are falling and buy stocks cheaper when prices are increasing further, which will result in profits for the investor if all goes according to plan. This paper will only focus on buy strategies associated with upward price movement.

We hypothesize that investors can use a similar approach of analyzing moving averages to predict future movements in Bitcoin prices. Cryptocurrencies are emerging asset classes that researchers have not extensively studied. As such, unseen opportunities may exist. We suspect that the cryptocurrency market has a larger share of retail investors versus institutional investors relative to the traditional stock market. Retail investors do not have the same sophisticated tools that institutional investors have access to, so they may be more predisposed to the disposition effect. This may mean that there are more opportunities to execute successful trading strategies.

Data Exploration and Methodology

We collected data for this experiment from the following source:

(1) Cryptocurrency Historical Prices [4]

The following features were evaluated for this analysis:

1. Bit_returns - the simple daily returns using the closing price at time t and t-1
2. Bit_log_returns - the daily log returns utilizing the closing price at time t and t-1
3. Bit_days_range - the range in price movements for the given day defined as the highest price - the lowest price of that day
4. Bit_log_wtd_ret - the week to date returns using the closing price at time t and t-7
5. Bit_log_biwtd_ret - the bi-weekly to date returns using the closing price at time t and t-14
6. Bit_log_mtd_ret - the month to date returns using the closing price at time t and t-30
7. ema_log_ret_7 - the exponential moving average of the week to date returns defined as the weighted average using values in the range of time t and t-7
8. ema_log_ret_14 - the exponential moving average of the bi-weekly to date returns defined as the weighted average using values in the range of time t and t-14
9. ema_log_ret_30 - the exponential moving average of the month to date returns defined as the weighted average using values in the field of time t and t-30
10. ema_log_biret_7 - the exponential moving average using bi-weekly returns between time t and t-7
11. ema_log_biret_14 - the exponential moving average using bi-weekly returns between time t and t-14
12. ema_log_biret_30 - the exponential moving average using bi-weekly returns between time t and t-30
13. moving_avg_10 - the 10-day simple moving average of the closing price
14. moving_avg_20 - the 20-day simple moving average of the closing price
15. moving_avg_50 - the 50-day simple moving average of the closing price
16. moving_avg_50_diff - the difference between the 50-day moving average at t and t-1
17. moving_avg_50_diff_7 - the difference between the 50-day moving average at t and t-7
18. Bit_rsi_3 - Momentum indicator using 30 days of prior prices at time t
19. Bit_rsi_60 - Momentum indicator using 60 days of prior prices at time t
20. Bit_rsi_90 - Momentum indicator using 90 days of prior prices at time t

Different windows of simple moving averages and exponential moving averages were analyzed. Exponential moving averages are more sensitive to more recent data [5], so they may prove to be better predictors for Bitcoin price movements. The RSI or Relative Strength Index is another technical analysis indicator used to determine momentum in price movements. The RSI has values between 0 to 100, with lower values indicating that the stock is oversold and higher values indicating the stock is overbought [6].

There are specific properties with time-series data that require extra consideration. Patterns observed in the data, such as trends and seasonality, may result in temporal dependencies between data points. Trends may lead to means that change over time and seasonality may result in variances that change over time. Therefore, the usual regression model assumption of errors being independent normal random variables with a mean of zero and constant variance will not hold.

We run the Augmented Dickey-Fuller (ADF) [7] test on our continuous variables to check for stationary properties. A stationary time series is one whose properties do not depend on the time at which the series is observed and will therefore have more stable means and variances [8].

```
1] "################# Testing for the Close #################"

            Augmented Dickey-Fuller Test

data:  btc_data$Close
Dickey-Fuller = -1.5558, Lag order = 14, p-value = 0.7664
alternative hypothesis: stationary

[1] "################# Testing for the Marketcap #################"

            Augmented Dickey-Fuller Test

data:  btc_data$Marketcap
Dickey-Fuller = -1.4905, Lag order = 14, p-value = 0.794
alternative hypothesis: stationary

[1] "################# Testing for 20 day moving average #################"

            Augmented Dickey-Fuller Test

data:  btc_data$moving_avg_20[21:n]
Dickey-Fuller = -1.7186, Lag order = 14, p-value = 0.6975
alternative hypothesis: stationary

[1] "################# Testing for 10 day moving average #################"

            Augmented Dickey-Fuller Test

data:  btc_data$moving_avg_10[11:n]
Dickey-Fuller = -1.4954, Lag order = 14, p-value = 0.792
alternative hypothesis: stationary
```

The ADF test shows that the 10-day moving average, 20-day moving average, the Closing Price, and Market Cap have unit roots were non-stationary. We account for the non-stationary features in statistical models for time series data sets by creating features based on calculating the differences between consecutive observations in the data [9]. This is the motivation behind creating the variables moving_avg_50_diff and moving_avg_50_diff_7.

Now that we understand what is stationary and non-stationary and have accounted for non-stationary features let's discuss the target variables. Observe the graphs below:
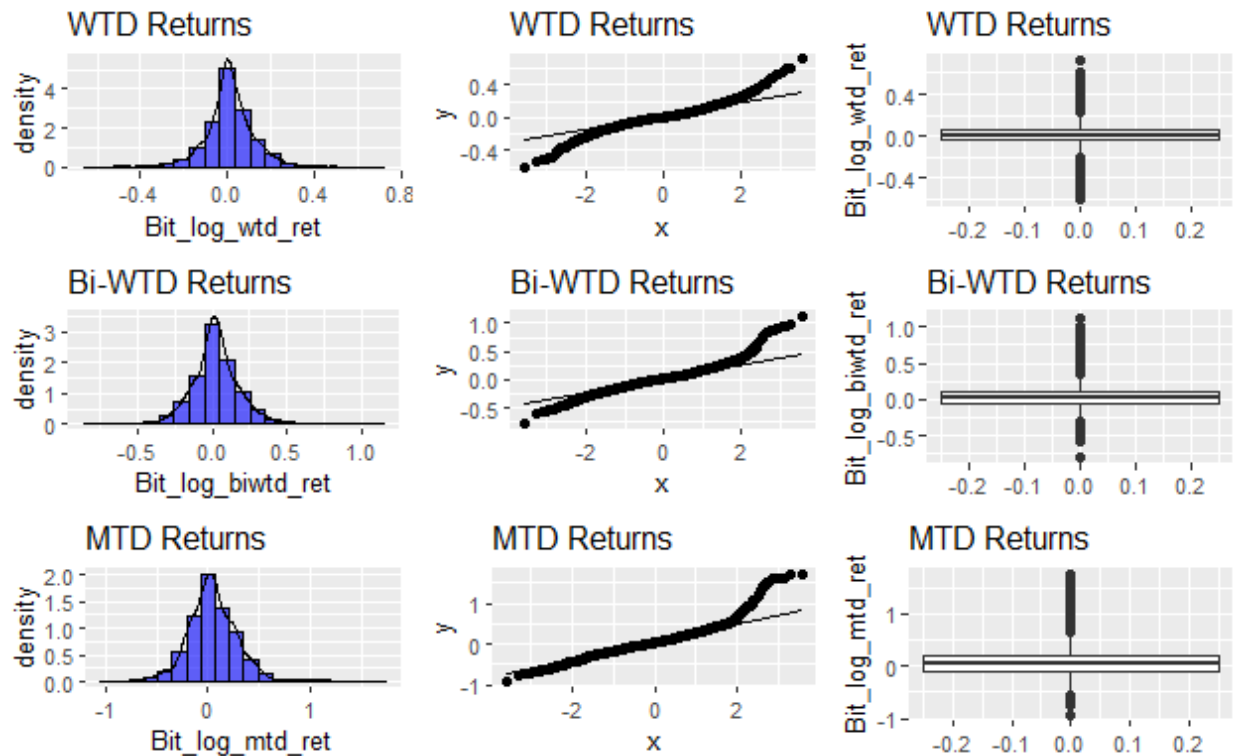
Since the log weighted returns appear closer to normal, we use this variable to label our indicator target variables. Even though there is some variation in the tails, the overall shape is still normal. Our purpose for this analysis is to search for the momentum of Bitcoin upward price movements since investors can create different buy strategies if they can predict these movements. For this purpose, we use the following thresholds to label our momentum indicator target variables. If the log of average returns over the future month period is greater than 20%, it is marked "1" as a potential buy opportunity. Otherwise, the target variable is marked "0" to indicate a window of price movement due to random variation or downward price movement, which is outside the scope of our analysis.

The choice of the 20% threshold is arbitrary so further analysis may be required to determine the optimum threshold. Certain considerations may need to be made in deciding this threshold, such as the level of an investor's conservativeness and transaction expenses. The price movement needs to be big enough for the investor to pay transaction costs to initiate their strategy, so they do not want to be too sensitive and react to events that end up being immaterial. Our team decided that 20% up is a large enough price movement to warrant action.

The green points in the plot below show the target variables as potential buy opportunities:

**Buy Indicators between Jun 2013 - Jan 2015 (Green)**
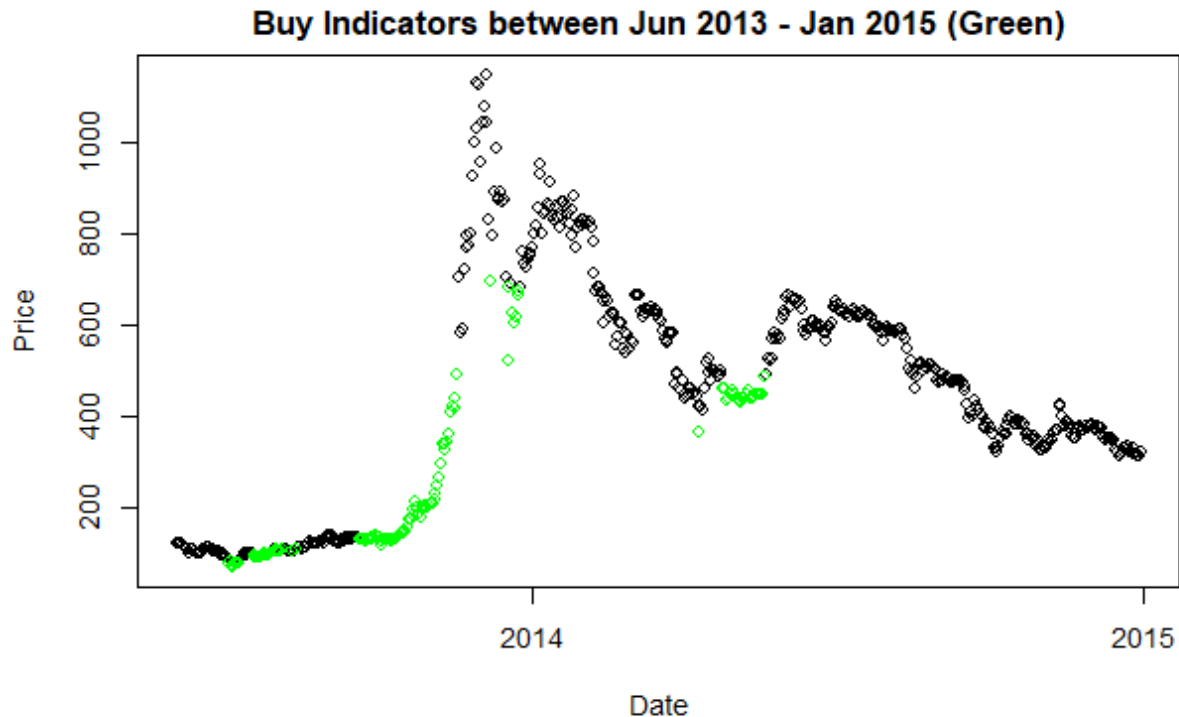
Since the form of the question we are trying to answer is categorical - Is this a good time to buy? – we decided to look at classification models instead of regression models. Logistic regression was ruled out due to the assumptions being violated with time-series data as explained above. Additionally, the predictor variables are highly correlated since they are moving averages over overlapping periods and linear models do not perform well with high correlation between predictors. From among the remaining tools available for classification problems, we chose Extreme Gradient Boosting for this analysis.

Results and Conclusions

Due to the unique nature of time series data, special consideration was needed to create the training and test data sets. The data points are time-dependent, and we needed to keep the temporal structure of the data, so the usual way of randomly splitting the data into training and test sets would not work. Instead, we kept the older 70% of the data as the training set and left the more recent 30% of the data as test. This way, we retain the order of the data points in both the training and test data sets.

From among the features we examined, only the features below were found to be important in our Extreme Gradient Boosting model:

Variable Importance
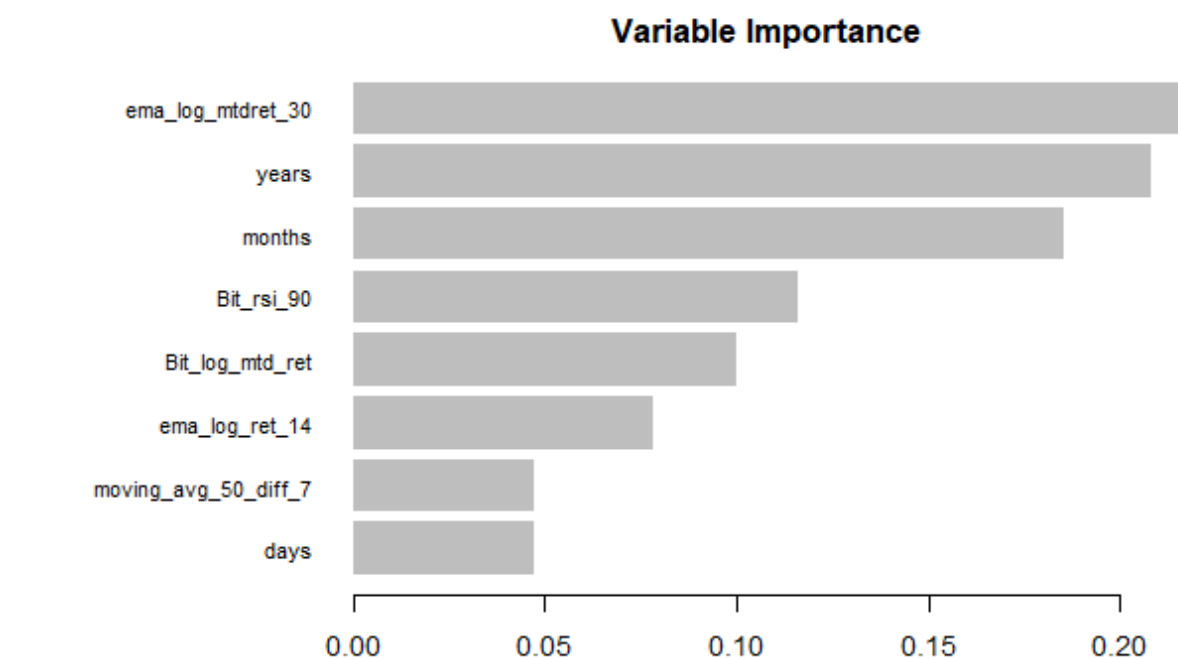
Our final model includes only the most essential features to predict momentum in Bitcoin price movement. Below are the results of our final model:

```
[1] "Results on training data:"
Confusion Matrix and Statistics

         Reference
Prediction   0    1
         0 1588    0
         1    0  422

             Accuracy : 1
               95% CI : (0.9982, 1)
  No Information Rate : 0.79
  P-Value [Acc > NIR] : < 2.2e-16

                Kappa : 1

 Mcnemar's Test P-Value : NA

          Sensitivity : 1.00
          Specificity : 1.00
       Pos Pred Value : 1.00
       Neg Pred Value : 1.00
           Prevalence : 0.21
       Detection Rate : 0.21
 Detection Prevalence : 0.21
    Balanced Accuracy : 1.00

     'Positive' Class : 1

[1] "Results on testing data:"
Confusion Matrix and Statistics

         Reference
```

```
Prediction  0  1
        0 570 282
        1  9  0

              Accuracy : 0.662
                95% CI : (0.6293, 0.6936)
    No Information Rate : 0.6725
    P-Value [Acc > NIR] : 0.7555

                  Kappa : -0.0207

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.00000
            Specificity : 0.98446
         Pos Pred Value : 0.00000
         Neg Pred Value : 0.66901
             Prevalence : 0.32753
         Detection Rate : 0.00000
   Detection Prevalence : 0.01045
      Balanced Accuracy : 0.49223

       'Positive' Class : 1
```

Although we have an accuracy of 100% on the training data, the accuracy dropped down to 66% on the test data, which points to the possibility of over-fitting. The model, therefore, may not be as effective in predicting momentum in Bitcoin price movement when faced with unseen data. Additionally, although the accuracy of 100% was higher than the 79% baseline rate on the training data, this failed to translate over when the model was applied to the test data, with the accuracy of 66% being below the baseline rate of 67%. Sadly, these results do not point to an adequate model, and more analysis will be required.

The ROC curves also indicate possible over-fitting, with the curve on the training data almost touching the upper left corner with an AUC of 100%, while the curve on the test data is closer to the diagonal of the graph with an AUC of 52%:

ROC Curves: Bitcoin Price Momentum
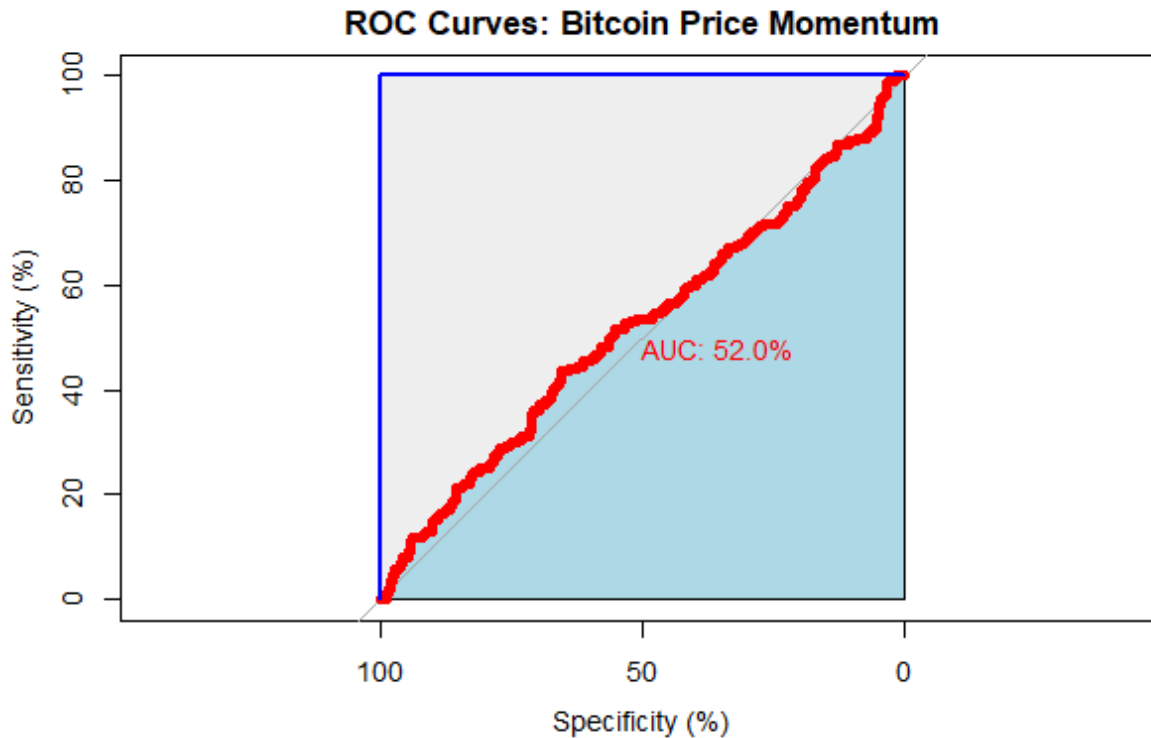
Challenges and Limitations

This analysis examined different windows of moving averages as predictors of future Bitcoin price movement. An inherent limitation of this approach is that a certain period would need to elapse before a moving average over the period can be calculated. Any model built using moving averages as predictors therefore will not react as quickly as other models. Depending on the size of the investment, this flaw could lead to substantial losses for the investor, especially in the highly volatile cryptocurrency markets. However, the conservative nature of this approach could also be seen as an advantage depending on the investor's risk appetite. More conservative investors would not want to react too quickly to random price movements as there are substantial costs to initiate any buy strategy - transaction expenses and tax implications, to name a few.

An additional limitation is the use of daily data for this analysis. Cryptocurrency markets operate 24/7, and data is readily available to the fraction of a second. However, processing data of this magnitude would be no easy task, not to mention the storage and computing power required. Although there are advantages to processing more timely data as it might overcome the limitations of using moving averages as explained above, we are forced to rely on daily data due to the challenges mentioned above.

Further Discussion and Research

This analysis used the threshold of 20% to label the target variables for Bitcoin price momentum upwards. The threshold needed to be big enough to not incorrectly flag the random noise in price

movement as momentum events. The investor would not want to be too reactive to random price movement as each buy transaction requires payment of fees, and there may be tax implications as well. The 20% threshold, therefore, made the most intuitive sense, but further study will be needed to determine the optimum threshold.

We focused on predicting Bitcoin price movement upward but did not examine price movements down as this was deemed outside the scope of our initial research question. Further investigation would need to be done to see if an effective model can be created to predict downward price movements.

Our analysis resulted in an overfit model. It is possible that increasing the minimum loss reduction parameter will lead to a better performing model. Further research is required to determine if this is true.

Additionally, this analysis examined many different windows of moving averages and exponential moving averages for features - 7 days, 14 days, 30 days, etc. - and average log returns over the future month period for the target variable. Different windows of moving averages may be more predictive, but further research is required.

One more interesting topic to discuss is the Efficient Market Hypothesis and the implications this may have for this analysis. According to Bodie, the Efficient Market Hypothesis is "the notion that stocks already reflect all available information" (Bodie 335). If the Efficient Market Hypothesis is accurate, stock prices immediately adjust to any new information made public so that historical trading data - such as what was used in this analysis - cannot be used to predict future price movement to execute successful trading strategies. The Efficient Market Hypothesis refers to traditional stock prices and it is unknown whether it also holds for cryptocurrency. Further research is needed, but if the Efficient Market Hypothesis does hold, it might partially explain the results from this analysis that predictors using moving average were not effective in predicting future Bitcoin price movements.

References

[1] "Frequently Asked Questions on Virtual Currency Transactions | Internal Revenue Service." *Irs.gov.* U.S. Internal Revenue Service. N.d. Web. 04 May 2022. https://www.irs.gov/individuals/international-taxpayers/frequently-asked-questions-on-virtual-currency-transactions

[2] "Global Cryptocurrency Market Report 2022-2027 - Industry to Cross a Staggering $32.4 Trillion by 2027, Exploding with a CAGR of 58.4%." *yahoo.com.* yahoo.com Research and Markets. N.d. Web. 04 May 2022. https://finance.yahoo.com/news/global-cryptocurrency-market-report-2022-120800380.html

[3] "Investments." *Investments, 11th edition, Bodie, Kane, and Marcus.* Vk.com. N.d. Web. 04 May 2022. https://vk.com/doc399904795_482800477?hash=dsrEOFu8Ngw46PuoF8SCx47HjNUti1QlTC96UzdSEWL

[4] "Cryptocurrency Historical Prices | Kaggle." *Kaggle.com.* Kaggle bitcoin dataset. N.d. Web. 04 May 2022. https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory?resource=download

[5] "Exponential Moving Average (EMA) Definition." *Investopedia.com.* Technical Analysis Basic Education. N.d. Web. 04 May 2022. https://www.investopedia.com/terms/e/ema.asp

[6] "Relative Strength Index (RSI) Definition." *Investopedia.com.* Technical Analysis Basic Education. N.d. Web. 04 May 2022. https://www.investopedia.com/terms/r/rsi.asp

[7] "Augmented Dickey-Fuller Test in R (With Example)." *Statology.org.* Statology. N.d. Web. 04 May 2022. https://www.statology.org/dickey-fuller-test-in-r/

[8] "6.4.4.2 Stationarity." *Nist.gov.* Engineering Statistics Handbook. N.d. Web. 04 May 2022. https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc442.htm

[9] "How to Remove Non-Stationarity in Time Series Forecasting | by Bex T. | Towards Data Science." *Towardsdatascience.com.* How to Remove Non-Stationarity in Time Series Forecasting. N.d. Web. 04 May 2022. https://towardsdatascience.com/how-to-remove-non-stationarity-in-time-series-forecasting-563c05c4bfc7