April 1st, 2022

# Timeseries Analysis of Bitcoin in the Context of Cryptocurrencies


Niklas Gutheil

Niklas.s.gutheil@gmail.com

https://www.linkedin.com/in/niklas-gutheil/

https://github.com/NikGutheil/Capstone


## Introduction.

Cryptocurrencies have been on the rise both in popularity and in market value since Bitcoin's inception in November 2009. Cryptocurrencies are digital assets that have some form of utility; either as a digital currency, a store of value or as a utility token that provides functionality in a larger ecosystem called a dApp. What these coins/tokens have in common is that they are publicly traded assets. Anyone can buy or sell these digital assets at any time of the year. Unlike traditional finance markets, there are no open or closing times and no holidays. Additionally, there are many unique features such as reward halving's and price metrics only found in Cryptocurrency markets. **The goal of this project is to identify what unique features are present in Bitcoin; and Crypto markets as a whole, that can be transferred to the creation of timeseries forecasts.** Currently there are many timeseries analyses on Bitcoin, but many of them fail to include either features or concepts that are unique to Bitcoin that can be used in modelling. This report will summarize many of the findings made throughout 5 Python JupyterNotebooks, available on GitHub, that can guide other timeseries analysts on what to look for when working with Bitcoin or other Cryptocurrencies.

## Data Acquisition and Processing.

As with any data science project, the first step after formalizing the project idea is to acquire and clean relevant data. Bitfinex is a centralized exchange for cryptocurrencies that offers a free API for downloading historical price data. The price data for Bitcoin was downloaded in 5-minute intervals from April 15th, 2013 (the first data point Bitfinex has) to January 1st, 2022. This price data was cleaned up of missing or NULL entries and resampled so data for 5-minute, 1-hour and 1-day intervals could be saved for the future. The data was also cut to start at July 20th, 2014 to account for the previous 463 days being needed for calculations during Feature engineering. The resulting data has the following features:

| Date | Open ($) | Close ($) | High ($) | Low ($) | Volume (BTC) |
|---|---|---|---|---|---|
| 2013-05-19 20:35:00 | 116.11 | 116.00 | 116.11 | 116.00 | 18.000000 |
| 2013-05-19 20:40:00 | 116.46 | 116.46 | 116.46 | 116.46 | 10.000000 |

## Feature Engineering.

A significant part of this project was to acquire several metrics that could later be used as additional information for modelling. Traditional metrics were calculated such as RSI, Exponential Moving Averages and On-balance Volume. 3 unique metrics were also created; **The Puell Multiple, The Fear and Greed Index,** and **The Stock-to-Flow model.** What these metrics are is covered in the notebooks in detail, but it is important to mention that these are very popular in the Cryptocurrency trading communities. Custom functions were written to calculate the features, resulting in a dataset that is quite informative and unique, with little to no comparison to publicly available data. The final dataframe can be seen to the right for the 1-day time intervals. This dataset was also used in the upcoming EDA and Modelling sections.

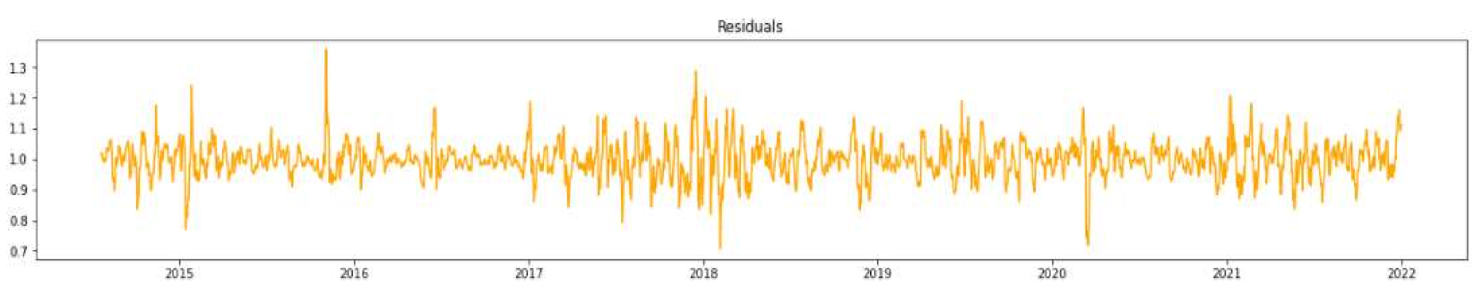
| ***Date*** | **2021-12-31** |
|---|---|
| *Open* | $47,109.52 |
| *Close* | $47,122.20 |
| *High* | $47,154,21 |
| *Low* | $47,06494 |
| *Volume* | 9.08 |
| *Stock* | 12,713,200 BTC |
| *S2F_463MA* | $60,193.44 |
| *Puell* | 0.995 |
| *Fg_index* | 22 |
| *OBV* | -11.10 |
| *RSI* | 0.357 |
| *EMA_7* | $48,662.38 |
| *EMA_14* | $48,973.19 |
| *EMA_21* | $49,491.46 |
| *EMA_28* | $50,179.90 |
| *EMA_50* | $51,928.66 |
| *EMA_100* | $52,383.62 |
| *EMA_250* | $48,143.59 |

## Exploratory Data Analysis

During exploratory data analysis the unique metrics to Bitcoin were analyzed for their relationship to the price of Bitcoin. One of the most popular and interesting metrics is the Fear and Greed Index, which looks similar to the price chart of Bitcoin.


The chart clearly reveals that the two metrics are complimentary in some fashion. The Fear and Greed index is meant to be a predictive metric compared to a lagging indicator, meaning a usefulness for future predictions could be present. Some key dates were also explored such as Bitcoins halving cycles and Bull/Bear markets, where halving's were an indicator of a Bull-run to start within 4 months.

The next step of EDA was to perform seasonal decomposition on the time series to identify seasonal components and an overall trend line. Decomposition proved difficult and there was little evidence that Bitcoin's price chart is stationary. The best result came from multi-seasonal decomposition where yearly, quarterly and monthly seasonality's were taken into account.



The residuals component of the multi-seasonal analysis shows that there is still significant variance at the beginning, middle and end of the timeseries. Accounting for a 3-year seasonality, or Bitcoin halving's, as suggested by the residuals, did not improve the results.

## Modelling and Evaluation

This is where the most interesting parts of the project come into play. The final 1-day time interval data acquired from the feature engineering step was used for modelling. To start, an ARIMA model was built to get base accuracies, where only the target variable was used to evaluate the model. The S component of SARIMAX was left alone as SARIMAX can only account for a single seasonality, something that is not present in the timeseries. The forecast range was also set to 2 weeks as SARIMAX does not handle long time forecasts well without a seasonal component. Two weeks in the context of cryptocurrencies is enough. ACF and PACF plots were drawn to determine the starting points for some of the hyperparameters. From there a grid search was used to find all optimal hyperparameters.

The next step was to re-build and optimize a model where only one of the additional features created is added at a time, making it an ARIMAX model. After running all the models, they were scored using the Mean Absolute Percentage Error. The results for all models are as follows, where only improvements on the base model are listed:

| ***Model Name*** | **Test MAPE** |
|---:|---|
| *ARIMA Base Model* | 5.11% |
| *ARIMAX w/ Volume* | 4.82% |
| *ARIMAX w/ OBV* | 4.67% |
| *ARIMAX w/ S2F_463MA* | 4.58% |
| *ARIMAX w/ fg_index* | 4.16% |
| *ARIMAX w/ Puell* | 2.43% |
| *ARIMAX w/ RSI* | 1.32% |

The results tell us that the Bitcoin specific features that were created had a positive affect on the prediction. The second part of modelling was done with Facebooks Prophet, where multi-seasonality was explored. Here the forecast was 6 months into the future, which asks Prophet to account for the rise in prices shortly after the train data ends. An optimized model was able to predict this upward trend in price, as indicated by the darker blue line.


Through various experimentation it was found that adding seasonal components did not have a positive affect, meaning it can be concluded that accounting for Bitcoin's seasonality's is either not possible, or far more complex methods need to be used.

## Conclusion & Discussion

Many things were learned over the course of this project about timeseries analysis in the context of Bitcoin and other Cryptocurrencies. Some key takeaways were that Bitcoin has a unique seasonality in the form of halving's that seemed to indicate a Bull-run was within 4 months away. It was also found that the created features unique to Bitcoin; **Puell Multiple, Fear and Greed Index and Stock-to-Flow Model**, were positive additions to the models and could be good additions to other models already created. Some things to be wary of when modelling Bitcoin also came up. Accounting for seasonality in Bitcoin's price is complex and might not be possible. Future model builders will be able to look at this project and see what does and does not work when constructing their own Cryptocurrency models.

## Next Steps

There are a few steps to be taken from here. More complex models such as Amazons DeepAR can be used for modelling as it is specifically built to analyze multiple features for a single prediction. Additionally, more complex seasonal decompositions can be applied to further reduce the variance in the timeseries residuals and apply this seasonality to multi-seasonal models. More features unique to Bitcoin and Cryptocurrency can be evaluated, including metrics that might be unique to coins that are not Proof of Work, which Bitcoin is.

## References

Facebook. (2017). *Seasonality, Holiday Effects, And Regressors.* Retrieved from https://facebook.github.io/prophet/docs/seasonality,_holiday_effects,_and_regressors.html

Forest-Baldini, J. (2020). *Grid Search for SARIMAX Parameters.* Retrieved from https://towardsdatascience.com/grid-search-for-sarimax-parameters-adbd5104cb8b

Jay. (2020). *Predicting Bitcoin Price Using Stock To Flow Model & Linear Regression.* Retrieved from https://pythoninoffice.com/predicting-bitcoin-price-stock-to-flow-python/

Khandelwal, R. (2019). *Time Series Prediction using SARIMAX.* Retrieved from https://medium.datadriveninvestor.com/time-series-prediction-using-sarimax-a6604f258c56

Manning. (2022). *Timeseries Forecasting in Python.* Retrieved from https://livebook.manning.com/book/time-series-forecasting-in-python-book/chapter-9/v-8/64

Schmuddle, D. (2019). *Facebook Prophet: Diagnostics.* Retrieved from https://nextjournal.com/fb-prophet/facebook-prophet-diagnostics

TOTH, D. J. (2021). *Multi seasonal time series analysis: decomposition and forecasting with Python.* Retrieved from https://medium.com/analytics-vidhya/multi-seasonal-time-series-analysis-decomposition-and-forecasting-with-python-609409570007

Yang, S. (2020). *Time series analysis using Prophet in Python — Part 2: Hyperparameter Tuning and Cross Validation.* Retrieved from https://medium.com/analytics-vidhya/time-series-analysis-using-prophet-in-python-part-2-hyperparameter-tuning-and-cross-validation-88e7d831a067

April 1st, 2022