Cryptocurrencies: A Time Series Analysis Through Logistic Regression, Prophet, Cointegration, Clustering, and Geometric Brownian Motion

By Santino Luppino, Yitao Huang, Zheng Li, Patrick Besser, Luke Ditton

Advisor: Professor Khashanah

Overview

- Research Question
- Introduction
- Literature Review Summary
- Data
- Methodology & Results
- Conclusion & Further Research

Research Question

Is it possible to accurately predict the price movements of cryptocurrencies?

The goal of this research is to identify relationships between cryptocurrencies and the side factors typically used to evaluate the equities market through different time series methods. Knowing more about the predictability of cryptocurrencies could help investors make more informed decisions and boost investor confidence in the market.

Introduction

- Cryptocurrencies
  - Investment opportunity
  - Alternative to traditional monetary system
  - Traditionally higher returns compared to the equity market
- Overall goals:
  - Analyze possible technical trends apparent from the results of the models
  - Advise on investment strategies pertaining to the results of the study
  - Better define the risk factors for cryptocurrency as compared to other equities
  - Determine if a number of external components have an impact on the price of a cryptocurrency
  - Determine which models are applicable to predicting the returns of the cryptocurrency market

Literature Review

Kwon, et al. “Time Series Classification of Cryptocurrency Price Trend Based on a Recurrent LSTM Neural Network.”	● Applied the long short-term memory (LSTM) model to classify the time series for cryptocurrency ● Showed that LSTM wasn’t the best model
Göttfert, J. (2019). Cointegration among cryptocurrencies : A cointegration analysis of Bitcoin, Bitcoin Cash, EOS, Ethereum, Litecoin and Ripple (Dissertation).	● Used Johansen cointegration test and the Engle-Granger two step analysis for cointegration ● Price of Bitcoin has a statistically significant long-run effect on the prices of Bitcoin Cash, Ethereum, Litecoin, and Ripple
Pichl, Lukáš, and Taisei Kaizoji (2017). “Volatility Analysis of Bitcoin Price Time Series.”	● Used Heterogeneous Autoregressive model to realize the volatility of Bitcoin ● Suggested that more sophisticated machine learning methods are necessary for higher prediction accuracy
Yenidogan, et al(2018). “Bitcoin Forecasting Using ARIMA and PROPHET”	● Used Prophet and ARIMA to predict bitcoin ● Prophet appears more accurate and useful

Data

- The adjusted close prices for Bitcoin (BTC-USD), Ethereum (ETH-USD), Binance Coin (BNB-USD), XRP USD (XRP-USD), Litecoin USD (LTC-USD), & DogeCoin (DOGE-USD) will be used in this study
  - Downloaded using the quantmod library in R from Yahoo Finance
  - The data is free to obtain
  - Frequency of the data is daily
- SPY, VIX, IEF, DWAS, VV, UUP, GLD, IAU, USO
- The data will be taken over the period from 2018-01-01 to 2021-12-31 due to the availability of the data

Data Snapshot

btc_ret	bnb_ret	eth_ret	xrp_ret	ltc_ret	doge_ret	vix_ret
0.092589260	0.04906510	0.13514466	0.03689716	0.11007621	0.02614528	-0.065562587
0.014505086	0.07602694	0.08480342	0.22451147	-0.04118322	0.01895535	0.007621158
0.025858428	-0.03433855	0.01873036	0.02896426	-0.01642810	0.03417333	0.000000000
0.110944530	0.48179193	0.01697970	-0.04737858	0.03220964	0.23239141	0.032019811
0.005578376	0.42248091	0.04311748	0.01473651	0.17335145	0.20014753	0.057158414
-0.061740690	-0.19878518	0.10167986	0.08777251	-0.02736691	0.13950156	-0.026132140

spy_ret	tnx_ret	dxy_ret
0.006305321	-0.007329023	0.007732806
0.004206057	0.002448981	0.024673659
0.006641921	0.009332590	0.013151380
0.001826693	0.026264940	0.014017912
0.002261105	0.001569859	0.007793437
-0.001531257	-0.007478878	-0.013875187

Methodology

This study will explore the following time series models:

1. Logistic Regression
2. Clustering
3. Cointegration
4. Geometric Brownian Motion
5. Prophet

Logistic Regression Overview

- Logistic Regression is a statistical regression model that uses a logistic function to model a binary variable
    - Logistic Regression is used when the dependent variable is categorical
    - The dependent variable will be binary for 1 (if the logarithmic return of the cryptocurrency is greater than 0) and 0 otherwise, this will be set as our decision boundary
- The inputs of this Logistic Regression will be the lagged returns for the chosen cryptocurrency
    - The amount of days for the lagged returns will be set to 5 days
    - An autoregressive model will not be used because of the possibility of autocorrelation problems and the fact that the logistic regression model does not take models as inputs
- The Logistic Regression model in this particular study has the following equation:

$$log(\frac{p_{1,t}}{1-p_{1,t}}) = \beta_0 + \beta_{1,1} r_{1,t-1} + \beta_{1,2} r_{1,t-2} + \beta_{1,3} r_{1,t-3} + \beta_{1,4} r_{1,t-4} + \beta_{1,5} r_{1,t-5}$$

- After our model is created, the accuracy of this model and the confusion matrix will be computed
    - A confusion matrix demonstrates the predictions the model is making relative to its actual value

Logistic Regression Overview (cont.)

- The data will be split into an even 50% Testing/Training set split
- A data frame of the direction for each crypto along with the 5 lagged returns will be used for the model
- The confusion matrix for each cryptocurrency will be shown

Logistic Regression Confusion Matrices

BTC:

```
y.logistic.pred   0   1
              0  54  54
              1 305 315
```

ETH:

```
y.logistic.pred   0   1
              0 128 149
              1 225 226
```

XRP:

```
y.logistic.pred   0   1
              0 245 201
              1 138 144
```

BNB:

```
y.logistic.pred   0   1
              0 251  89
              1 107 281
```

DOGE:

```
y.logistic.pred   0   1
              0 248 206
              1 140 134
```

LTC:

```
y.logistic.pred   0   1
              0 236 240
              1 118 134
```

Logistic Regression Results

- BTC, BNB, ETH, XRP, LTC, & DOGE are 50.69%, 73.08%, 48.63%, 53.43%, 50.82%, and 52.47%

- BNB is the only crypto that performed will in terms of accuracy and prediction
  - BNB has the highest Kurtosis among the group of cryptocurrencies
    - Suggests a higher frequency of extreme (positive or negative) returns
  - Moreover the Kurtosis computed for BTC, BNB, ETH, XRP, LTC, & DOGE is 2.74, 14.91, 1.95, 11.32, 8.19, 9.10

Clustering Overview and Methodology

- Clustering method meant to finding subgroups, or clustering clusters, in a data set.
  - When we cluster the observations of a data set, we seek to partition them into distinct groups so that the observations within each group are quite similar to each other, while observations in different groups are quite different from each other.
  - Clustering looks to find homogeneous subgroups among the observations.
- K-mean clustering
  - Randomly assign a number, from 1 to K, to each of the observations. These serve as initial cluster assignments for the observations.
  - For each of the K clusters, compute the cluster centroid. The kth cluster centroid is the vector of the p feature means for the observations in the kth cluster
  - Assign each observation to the cluster whose centroid is closest , where closest is defined using Euclidean distance.

Clustering Overview and Methodology

- Formula: Clustering method can be express as:

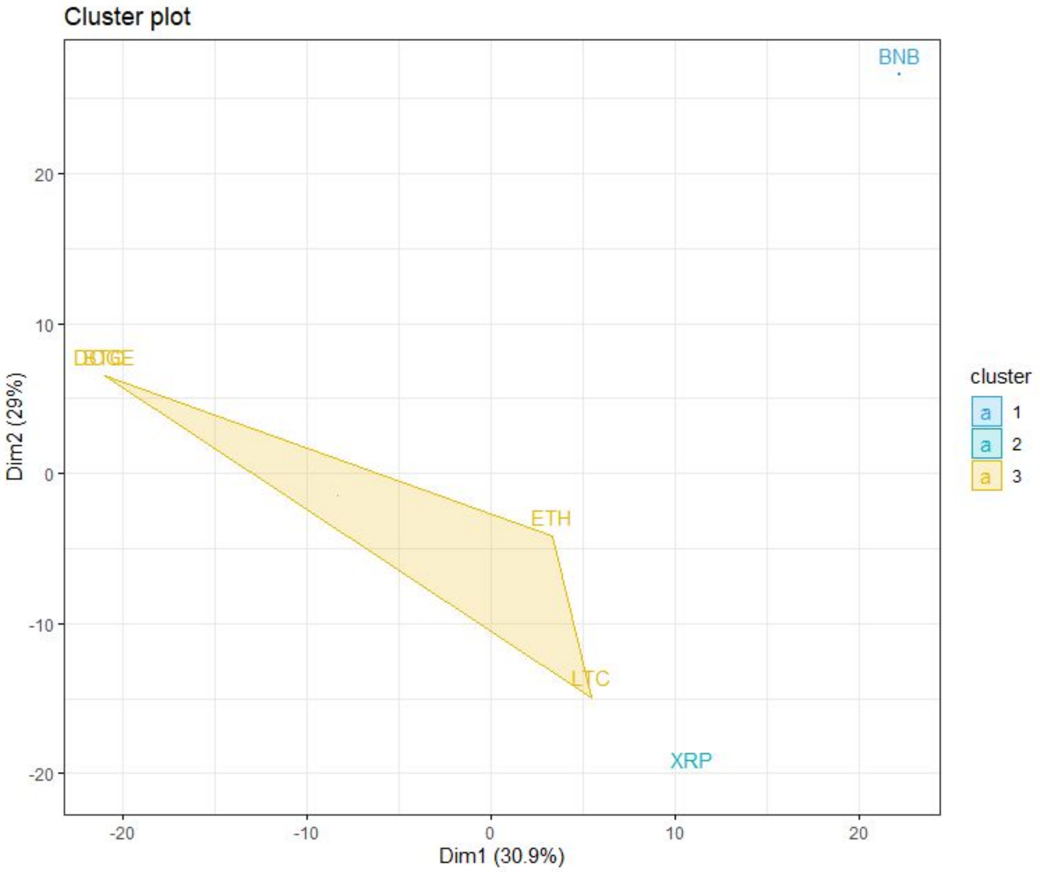
$$\sum_{i \in C_k} \sum_{j=1}^{P} (x_{ij} - \bar{x}_{kj})^2$$

- ❖ Ck is the Cluster
- ❖ X bar is the mean of each cluster.
- ❖ Xij-Xkj is Euclidean distance for each data point.

Clustering Results part 1

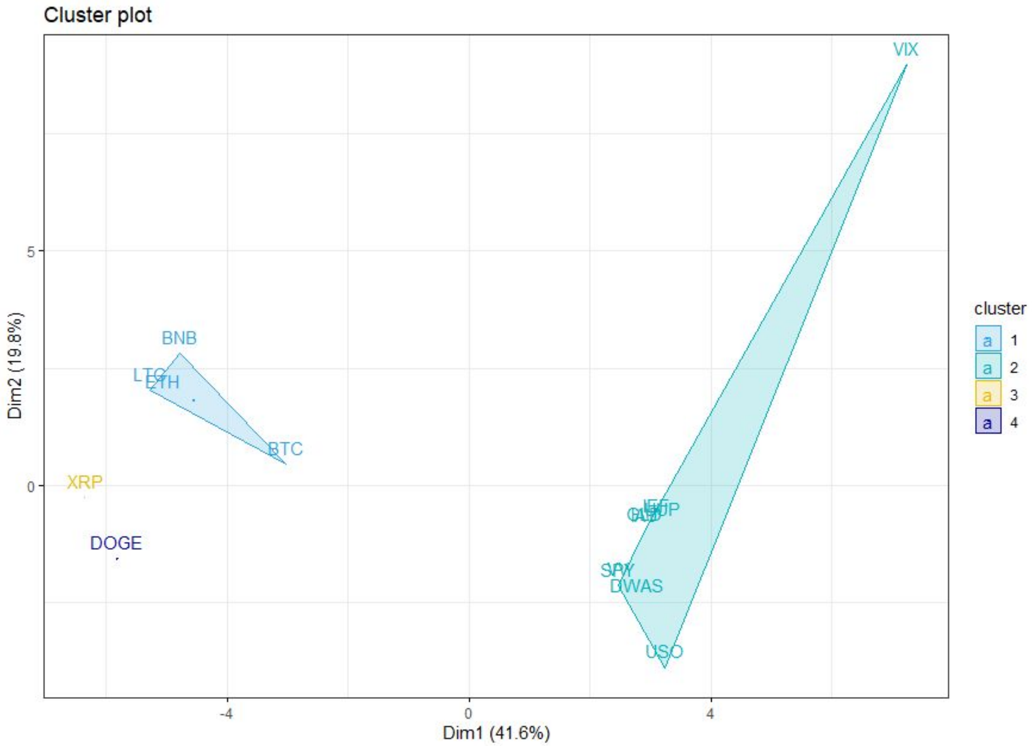
- Log daily returns
- Includes 6 cryptocurrencies
- Clustered into 3 groups
- BNB and XRP are in their own groups
- In group 3, BTC and DOGE are almost identical on the graph

BNB	XRP	BTC	ETH	LTC	DOGE
1	2	3	3	3	3



Clustering Results part 2

- Log monthly returns
- Included S&P 500 ,volatility index, US Dollar Index(UUP), Treasury Yield(IEF), Small-Cap Index (VB), Large-Cap Index (VV), GLD and IAU as gold index, and USO(Oil Index) as additional factors
- DOGE AND XRP in different category
- All side factors separate from crypto
- If there were 5 groups, VIX would probably be separate



BTC	BNB	ETH	LTC	SPY	VIX	IEF	DWAS	VV	UUP	GLD	IAU	USO	XRP	DOGE
1	1	1	1	2	2	2	2	2	2	2	2	2	3	4

Correlation Matrix

	meanOFcrypto	SPY	VIX	IEF	DWAS	VV	UUP	GLD	USO
meanOFcrypto	1.000000000	0.3107878	-0.20634462	-0.11817267	0.276650916	0.3093427	-0.17234882	-0.001411725	0.1402788
SPY	0.310787837	1.0000000	-0.76282201	-0.36550828	0.819802098	0.9990576	-0.34753971	0.130133006	0.4380860
VIX	-0.206344616	-0.7628220	1.00000000	0.27385097	-0.589369724	-0.7539920	0.21333565	-0.027383734	-0.3019969
IEF	-0.118172673	-0.3655083	0.27385097	1.00000000	-0.362985533	-0.3562320	0.02127489	0.334121152	-0.4209741
DWAS	0.276650916	0.8198021	-0.58936972	-0.36298553	1.000000000	0.8262991	-0.30135083	0.007620942	0.5614410
VV	0.309342701	0.9990576	-0.75399199	-0.35623202	0.826299062	1.0000000	-0.35623653	0.131370237	0.4363869
UUP	-0.172348824	-0.3475397	0.21333565	0.02127489	-0.301350831	-0.3562365	1.00000000	-0.565456512	-0.2298004
GLD	-0.001411725	0.1301330	-0.02738373	0.33412115	0.007620942	0.1313702	-0.56545651	1.000000000	-0.1359788
USO	0.140278756	0.4380860	-0.30199686	-0.42097412	0.561440971	0.4363869	-0.22980039	-0.135978816	1.0000000

- Correlation Matrix rank the side factor as SPY, VIX , UUP and USO correlation compare with other side factors

Cointegration - Overview

- Used to determine if there is a correlation between several time series in the long term
  - two or more non-stationary time series integrated together in a way that they cannot deviate from equilibrium in the long term
- Stationarity of residuals:

$$A(L)\,\Delta y_t = \gamma + B(L)\,\Delta x_t + \alpha(y_{t-1} - \beta_0 - \beta_1 x_{t-1}) + \nu_t .$$

- Johansen Trace test
  - Multiple time series data with large sample sizes
  - Allows for multiple cointegration relationships

Cointegration - Data

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
r <= 10 |	318.72	7.52	9.24	12.97
r <= 9 |	653.28	17.85	19.96	24.60
r <= 8 |	1028.71	32.00	34.91	41.07
r <= 7 |	1414.23	49.65	53.12	60.16
r <= 6 |	1820.05	71.86	76.07	84.45
r <= 5 |	2249.31	97.18	102.14	111.01
r <= 4 |	2691.39	126.58	131.70	143.09
r <= 3 |	3145.53	159.48	165.58	177.20
r <= 2 |	3657.57	196.37	202.92	215.74
r <= 1 |	4182.25	236.54	244.15	257.68
r = 0 |	4801.52	282.45	291.40	307.64

Eigenvectors, normalised to first column:
(These are the cointegration relations)

	btc_ret.l3	bnb_ret.l3	eth_ret.l3	xrp_ret.l3	ltc_ret.l3	doge_ret.l3	vix_ret.l3	spy_ret.l3	tnx_ret.l3	uup_ret.l3	uso_ret.l3	constant
btc_ret.l3	1.000000000	1.000000000	1.000000e+00	1.000000000	1.0000000000	1.00000000	1.000000000	1.0000000000	1.000000000	1.0000000000	1.0000000000	1.000000000
bnb_ret.l3	-0.035033535	-0.607895788	4.145395e-01	0.366817490	-1.2016746129	-5.43590048	-0.361034008	0.2176594945	-0.897419438	-0.2789140576	-0.1952965692	-0.195718550
eth_ret.l3	-0.654726702	1.498889922	-2.775622e-01	-0.698453341	0.7951800620	8.47672404	0.156180699	-0.1915548287	-1.994979819	-0.5738769241	-0.7946813081	1.814149075
xrp_ret.l3	-0.132871784	-0.309229599	4.296056e-01	0.135108880	0.0327407452	-6.73380990	0.764017657	-0.5481615831	0.094419811	0.0437985759	0.1147399286	1.064631743
ltc_ret.l3	-0.222625894	-1.975123562	-1.122528e+00	-0.818240988	0.3405388734	-2.33007767	-0.502328630	0.1423993398	1.253651385	-0.3737955027	0.6495300015	-2.598899852
doge_ret.l3	0.128158583	0.189547508	-2.522700e-01	0.193337742	0.1764352410	-1.61972244	-0.276358824	-0.0511081643	0.130623345	0.2412296422	-1.2588289425	0.098174234
vix_ret.l3	1.693084604	-0.332772459	2.712334e-02	0.040634335	0.1487116887	4.39349320	0.048904517	-0.0489410466	-0.083263884	-0.2759644391	-0.1093898396	0.003320656
spy_ret.l3	11.626015945	1.326791949	-1.989880e-02	0.608896130	4.5652822783	-23.16947800	-2.338767093	-0.6794750291	-0.968285936	-3.3881202952	2.3172885523	2.133027067
tnx_ret.l3	-0.807250745	-0.752655761	-4.855999e-01	0.836720616	0.8785273272	5.35396254	-0.156016331	-0.1559453139	-0.562975836	-0.0460399624	0.4719947731	0.155937906
uup_ret.l3	11.926540901	4.748340924	-1.387422e+01	0.818900350	-3.3718022800	-68.02530655	1.034333603	-1.0604385397	-6.766587390	4.7870618584	9.1354177362	-2.529587155
uso_ret.l3	-0.421659311	0.480854297	-1.084083e+00	0.639116834	-1.2976132931	6.50184920	1.128172043	0.2694503533	1.030766650	-1.1297775521	-0.8148348153	-0.476523099
constant	-0.009876074	-0.002845894	4.290144e-04	-0.001750973	-0.0008004578	0.02562688	0.002686991	-0.0008819081	0.005763132	0.0007090879	0.0004638725	-8.084157046

Cointegration - Results & Analysis

- Side Factors:
  - (Absolute values) UUP, SPY, USO, TNX, VIX

	BTC	BNB	ETH	XRP	LTC	DOGE
VIX	1.69	-0.33	0.03	0.04	0.15	4.39
SPY	11.63	1.33	-0.02	0.61	4.57	-23.16
TNX	-0.81	-0.75	-0.49	0.84	0.88	5.35
UUP	11.93	4.75	-13.87	0.82	-3.37	-68.01
USO	-0.42	0.48	-1.08	0.64	-1.30	6.50

- These relationships give investors more insight into the price movements and behaviors of these cryptocurrencies

Cointegration - Results & Analysis

- All 6 currencies are cointegrated
- Introducing DOGE has caused the model to skew due to extreme fluctuations in value

	BTC	BNB	ETH	XRP	LTC
BTC	1.00	1.00	1.00	1.00	1.00
BNB	-0.04	-0.61	0.41	0.37	-1.20
ETH	-0.65	1.50	-0.28	-0.70	0.80
XRP	-0.13	-0.31	0.43	0.14	0.03
LTC	-0.22	-1.98	-1.12	-0.82	0.34

<u>Influence Rank:</u>

Without DOGE	With DOGE
1. BTC 2. LTC 3. ETH 4. BNB 5. XRP	1. ETH 2. BNB 3. XRP 4. LTC 5. BTC 6. DOGE

	BTC	BNB	ETH	XRP	LTC	DOGE
BTC	1.00	1.00	1.00	1.00	1.00	1.00
BNB	-0.04	-0.61	0.41	0.37	-1.20	-5.43
ETH	-0.65	1.50	-0.28	-0.70	0.80	8.47
XRP	-0.13	-0.31	0.43	0.14	0.03	-6.73
LTC	-0.22	-1.98	-1.12	-0.82	0.34	-2.33
DOGE	0.13	0.19	-0.25	0.19	0.18	-1.62

Geometric Brownian Motion Overview

- Geometric Brownian Motion (GBM) will be performed on cryptocurrencies in this study
- The logarithmic returns will be used as inputs
- The formula for GBM:

$$F_t = F_0 * e^{\left(\mu_p - \frac{\sigma_p^2}{2}\right) * t + \sigma_p Z_t}$$

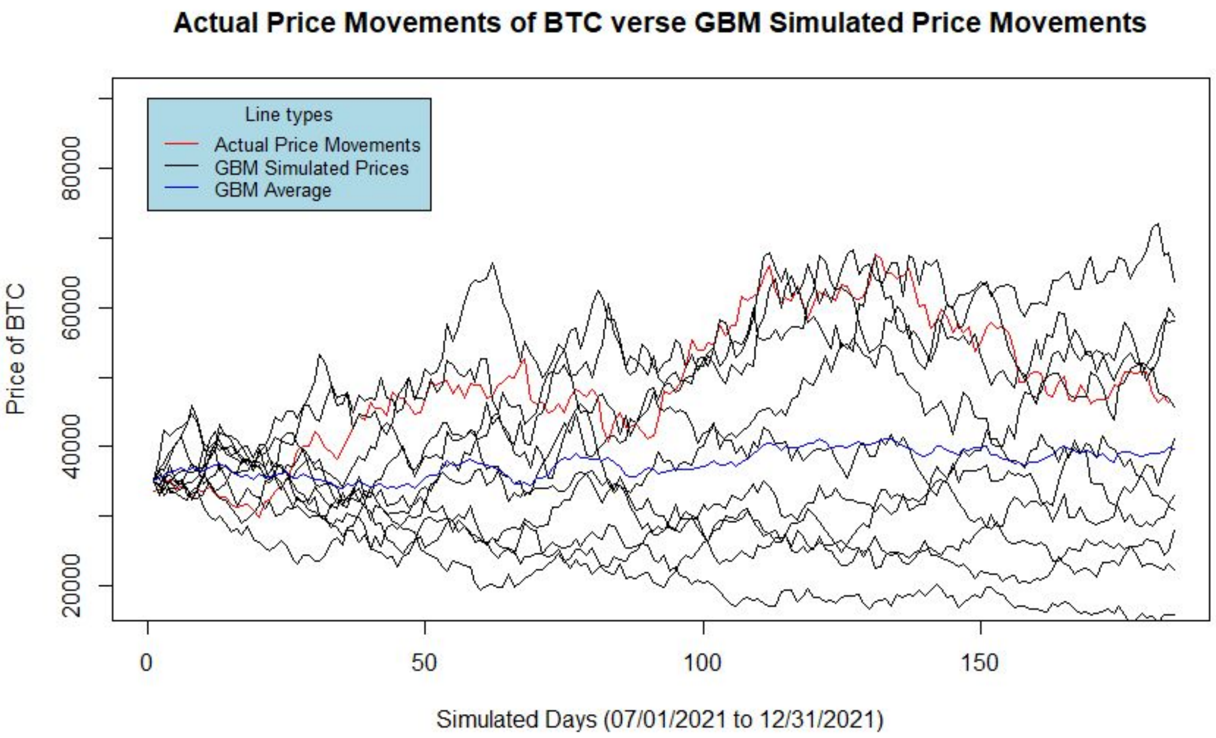
- Calibration technique will be used to minimize the discrepancy between theory and practice

$$\sigma = \sqrt{365} * s \text{ and } \mu = 365 * m + \frac{\sigma^2}{2}$$

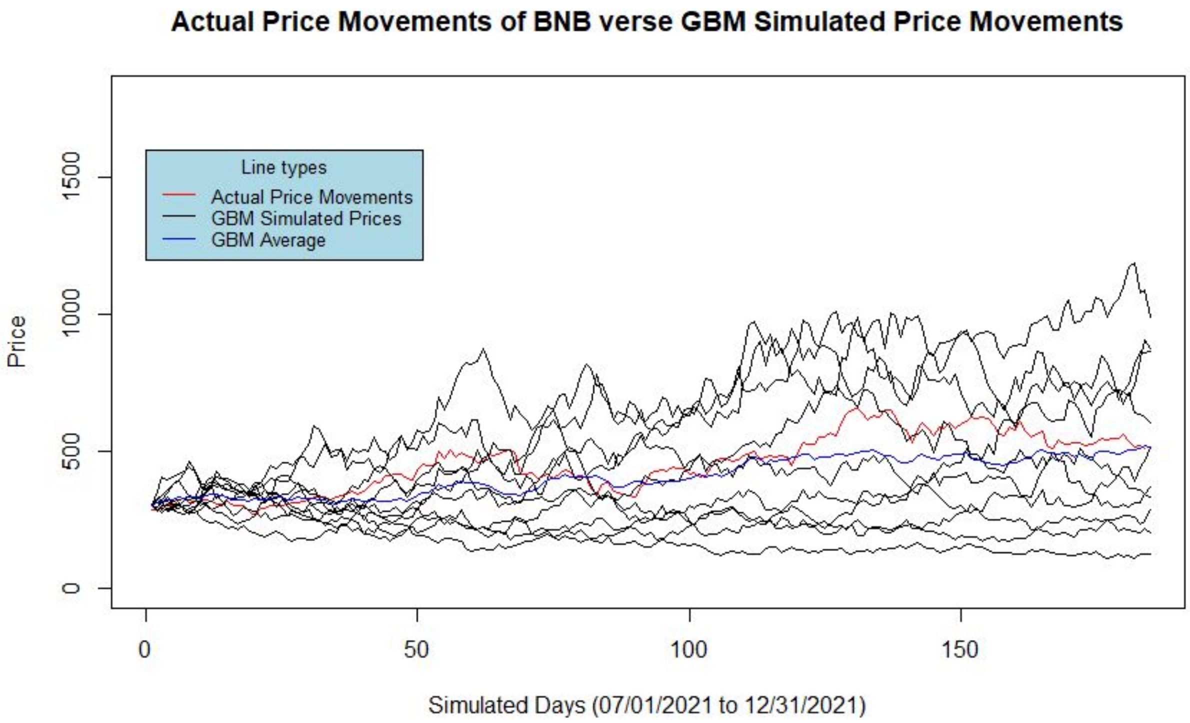
Geometric Brownian Motion Overview (cont.)

- 10 simulations will be created for each of the 6 selected cryptocurrencies
    - Each simulation will run from the last price in the training set
- The mean of each day in the simulation will be computed
    - This will be viewed as the “best fit” simulation
    - Potentially deliver a “theoretical average” of the simulations
- The mean square error between the returns of the simulated prices and actual returns will be calculated
- The Jarque-Bera Test for normality will be used to determine whether the assumptions for the GBM model are plausible

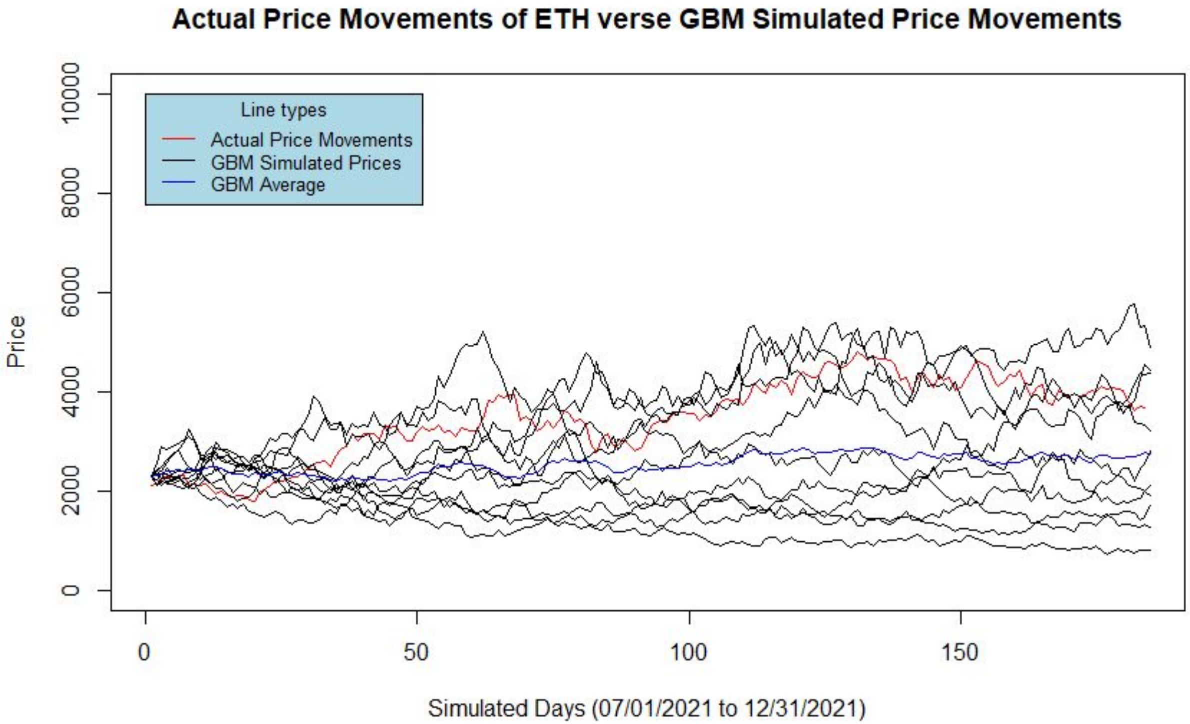
BTC Simulation



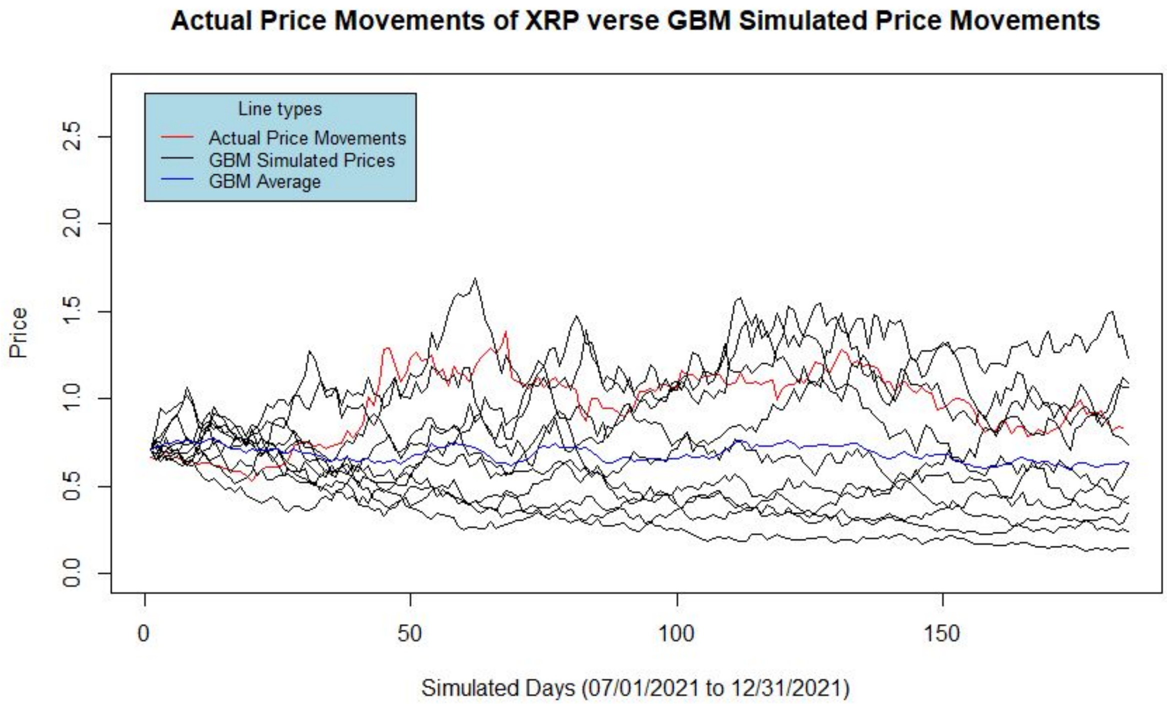
BNB Simulation



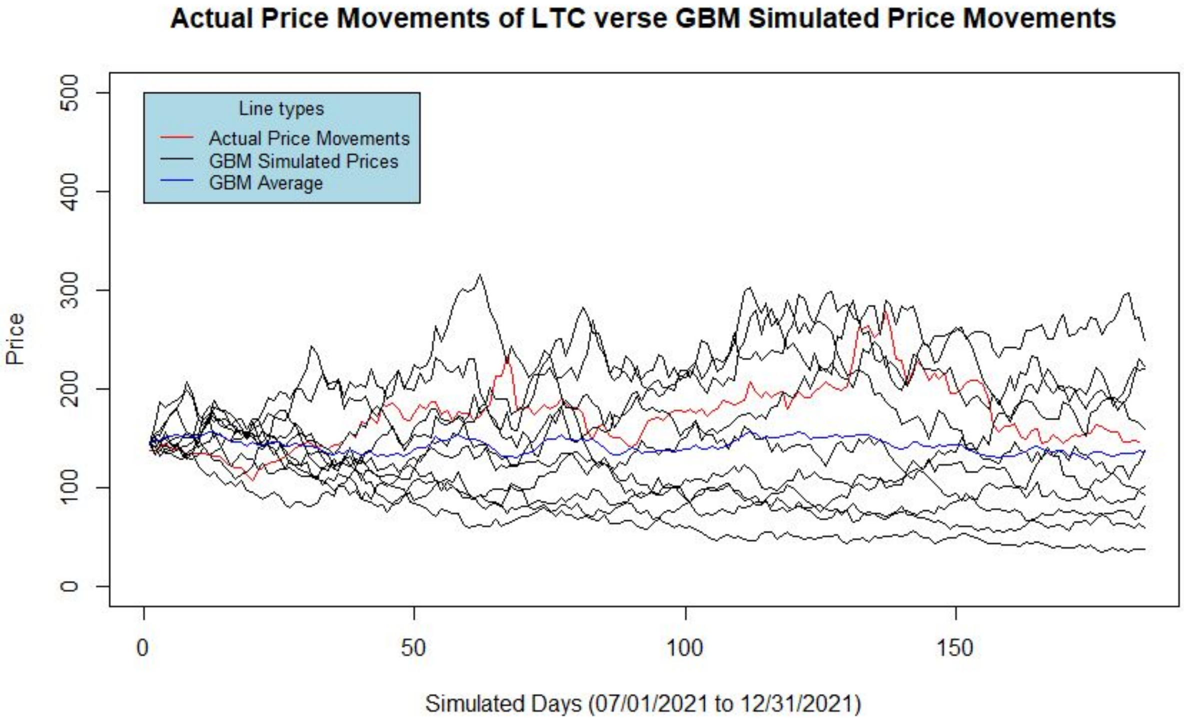
ETH Simulation



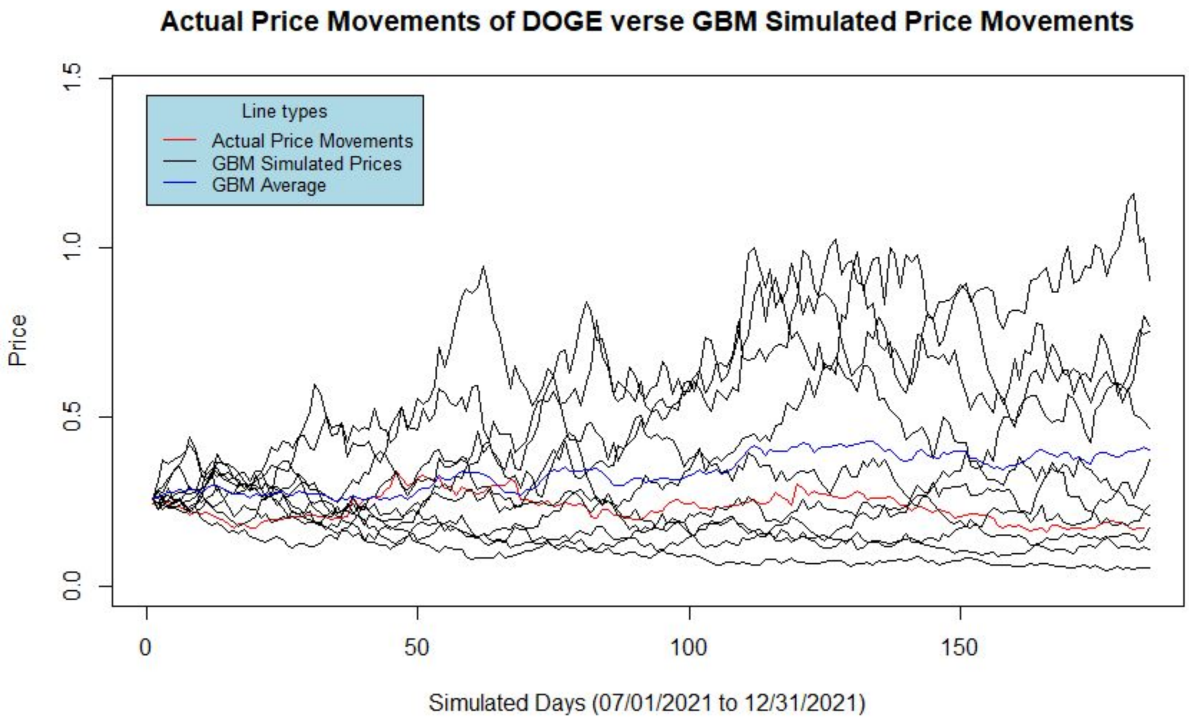
XRP Simulation



LTC Simulation



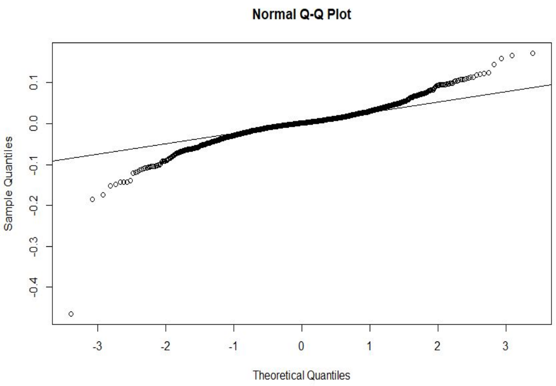
DOGE Simulation



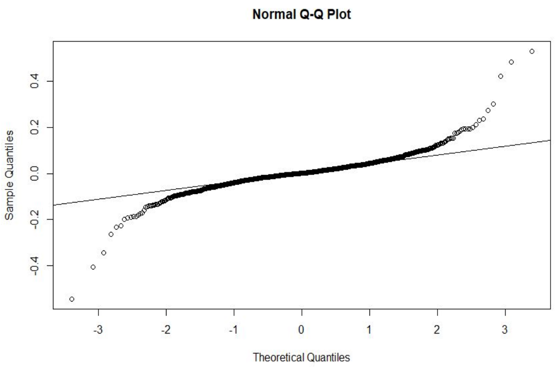
Results and Test for Normality

- The mean square error for BTC, BNB, ETH, XRP, LTC, & DOGE is 0.001297681, 0.002241178, 0.002164144, 0.003012867, 0.002740599, & 0.003690543 respectively
- All distributions were tested for normality using Jarque-Bera test
- All of the cryptocurrencies had statistically significant p-values
  - Due to kurtosis in the distributions

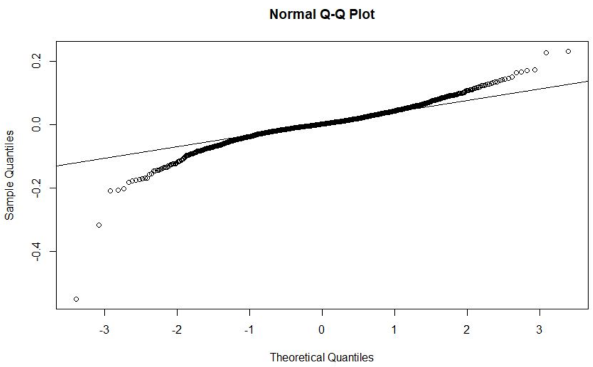
BTC:



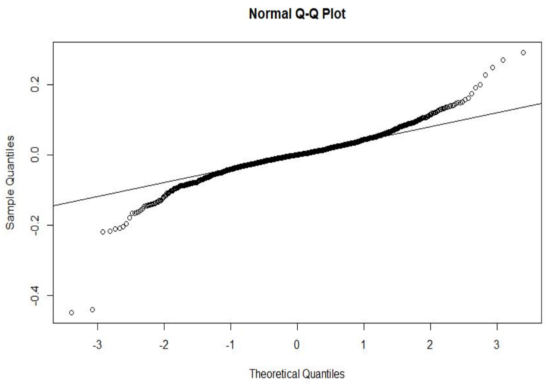
BNB:



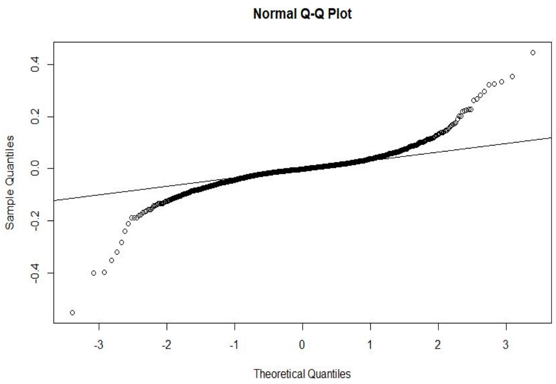
ETH:



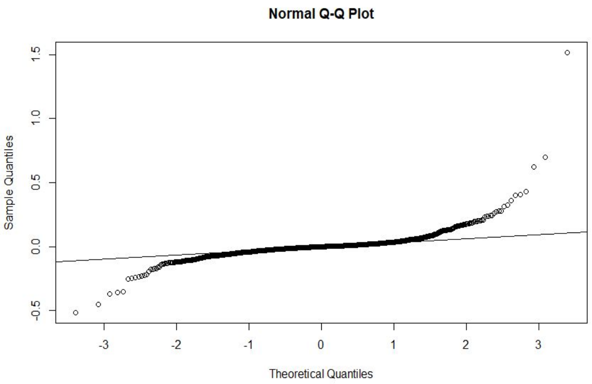
LTC:



XRP:



DOGE:



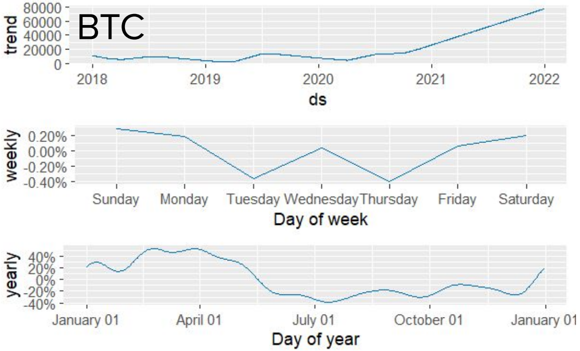
Prophet Overview

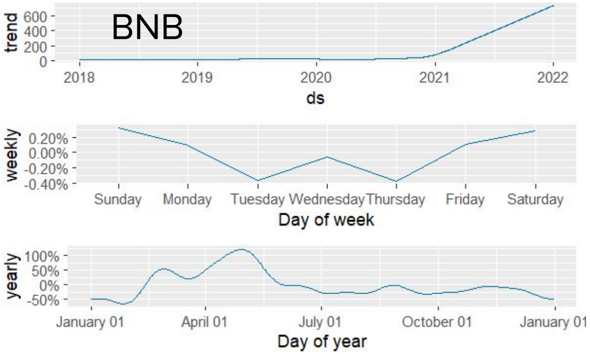
- Developed by Benjamin Letham and Sean Taylor at Meta
- Great for forecasting large scale time series with seasonalities, trends, and missing data
- The Prophet model can be expressed as a function:

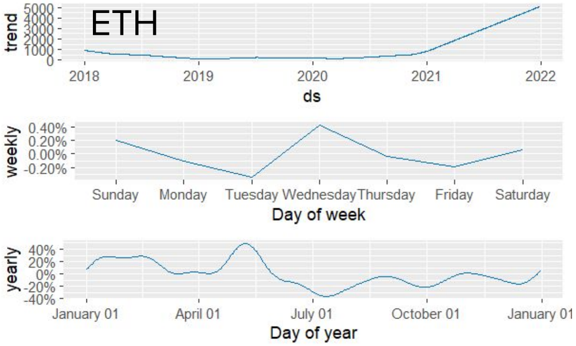
$$y(t) = g(t) + s(t) + h(t) + e(t)$$

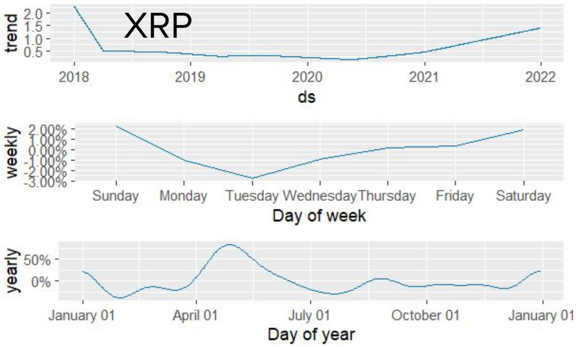
  - 4 components:
    - Growth(g): Linear, Logistic, Flat
    - Seasonality(s): Basically a Fourier Series, and it takes in consideration of time
    - Holiday(h): A built-in list of US holidays which the model will check for the effect on the input data
    - Error(e): Catch the randomness that is not accommodated by the model.

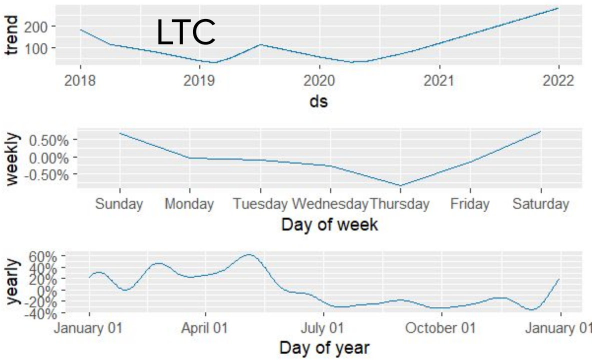
Trends and Seasonalities of the Six Cryptocurrencies

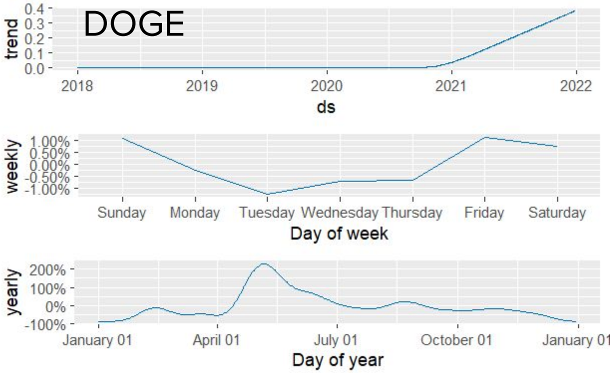












Prophet Overview (cont.)

Parameters:

- daily_seasonality = False
- Seasonality_mode = “multiplicative”

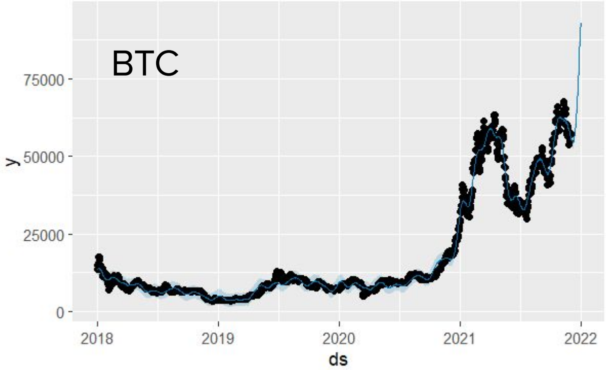
Side Factors:

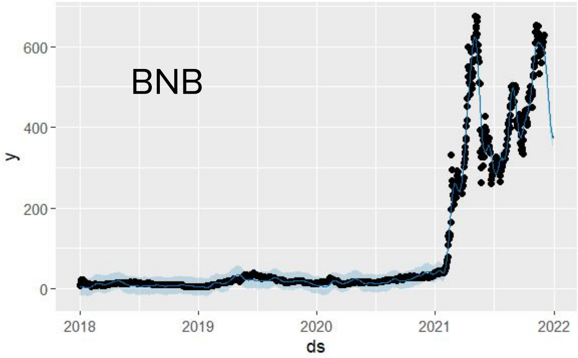
- UUP, SPY, USO

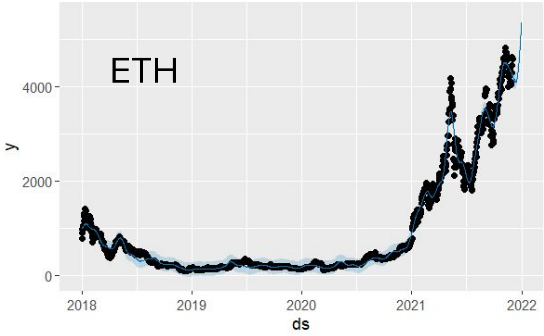
From here, the model was constructed in **three** different ways for each cryptocurrency:

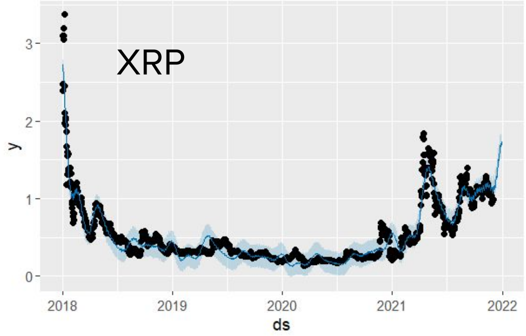
- General Prophet Fitness
- Comparison between Prophet with Side Factors and without
- Prophet Log Return prediction with GARCH’s volatility prediction

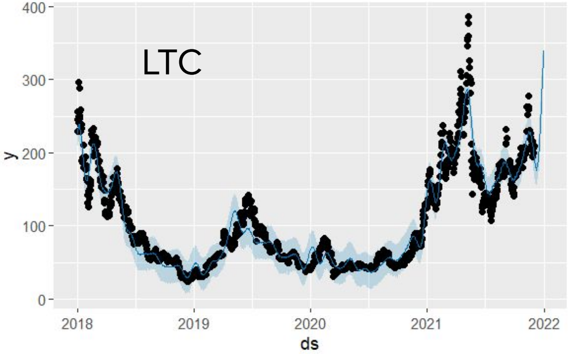
Prophet Results

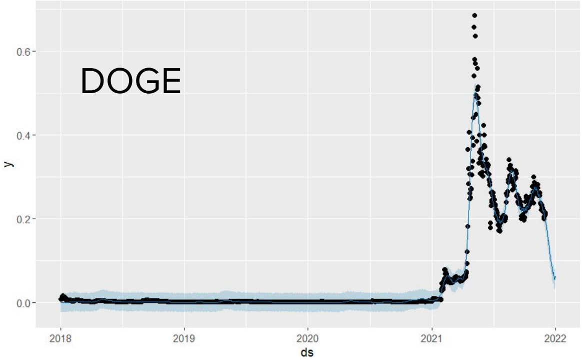






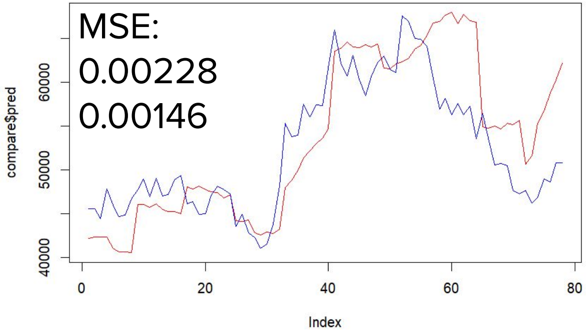




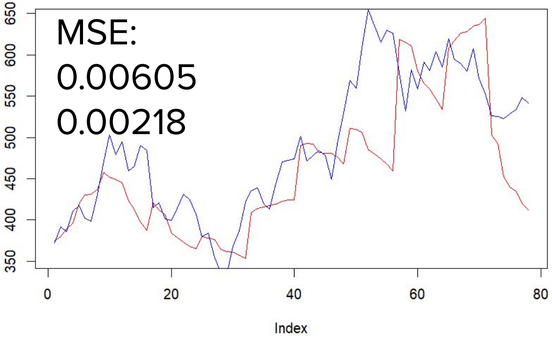


Prophet Results

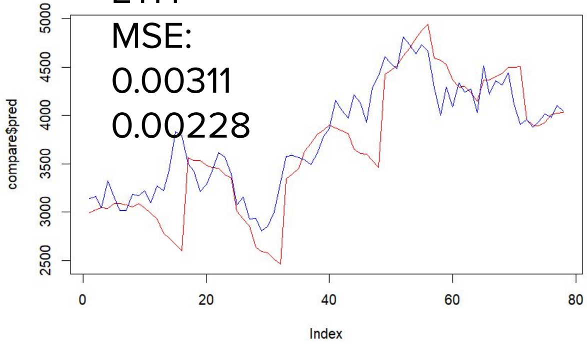
BTC
MSE:
0.00228
0.00146



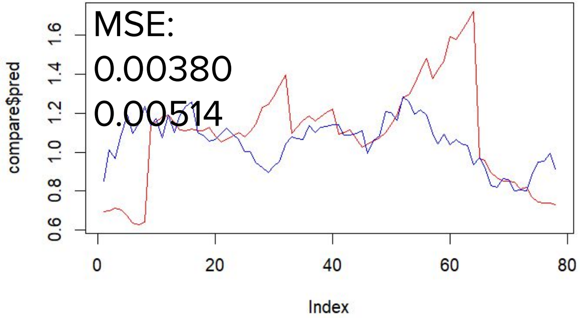
BNB
MSE:
0.00605
0.00218



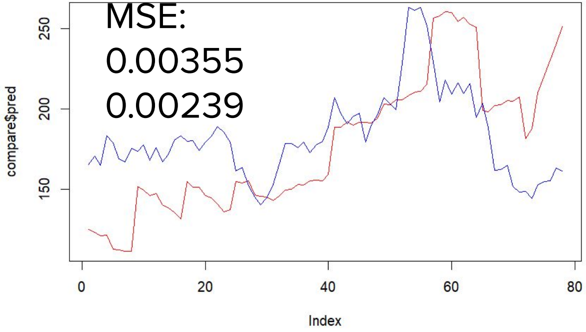
ETH
MSE:
0.00311
0.00228



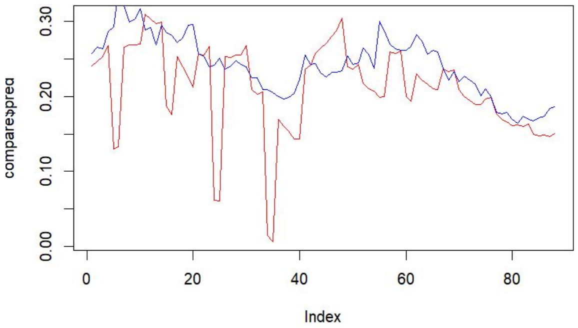
XRP
MSE:
0.00380
0.00514



LTC
MSE:
0.00355
0.00239

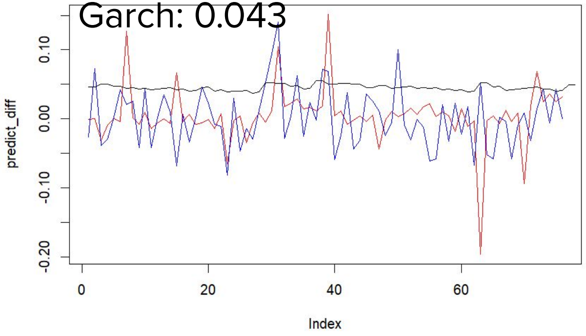


DOGE
MSE: 0.02037 0.00349

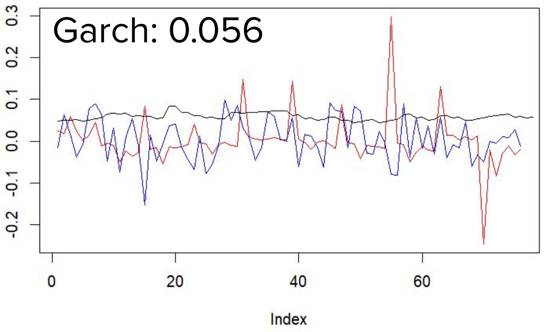


Prophet Results with Garch Model

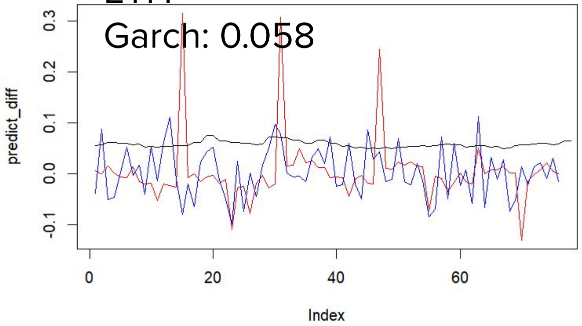
BTC
Garch: 0.043



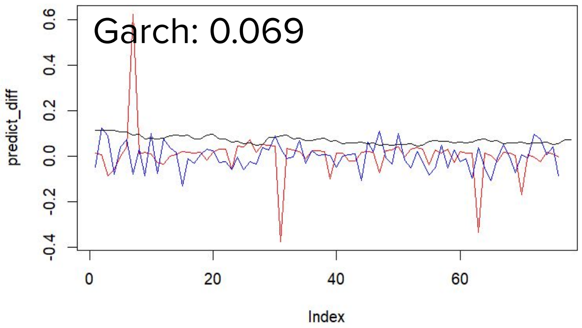
BNB
Garch: 0.056



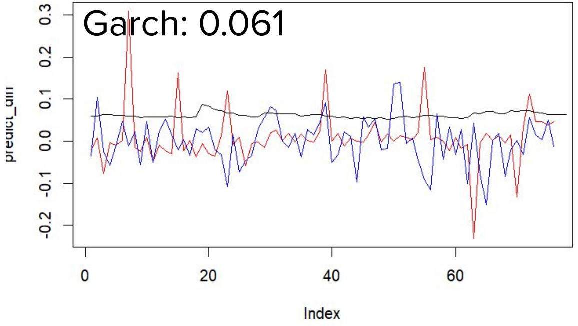
ETH
Garch: 0.058



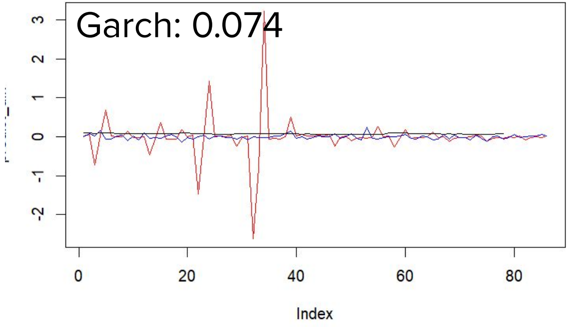
XRP
Garch: 0.069



LTC
Garch: 0.061



DOGE
Garch: 0.074



Conclusion

1. Extreme returns for the cryptocurrencies are more likely than other equities
    a. BNB had the highest kurtosis among the group, yet the highest accuracy in logistic regression
    b. GBM displayed that the distribution of the logarithmic returns of the cryptocurrencies do not follow a normal distribution
    c. Risk-averse investor should not invest cryptocurrency.
2. Side factors
    a. Data: Included S&P 500 ,volatility index, US Dollar Index(UUP), Treasury Yield(IEF), Small-Cap Index, Large-Cap Index (VV), GLD and IAU as gold index, and USO(Oil Index).
    b. Clustering: Method shows that they are not in the same cluster with Cryptocurrency, but by the correlation Matrix, we still find that they have relationship. SPY, VIX , UUP and USO
    c. Cointegration: Most influential side factors - UUP, SPY, USO, VIX
    d. Prophet: UUP, SPY, USO helped Prophet to have a better prediction with lower MSE
3. Comparison between GBM and Prophet
    a. Prophet’s MSE is 5x smaller than GBM’s MSE on average