Final Report: Searching the price movements of digital assets.

Naoya Ohara e0395606@u.nus.edu, Goh Zhen Hao (Kenneth) e0486543@u.nus.edu

Abstract*—We analyzed digital assets, widely known as cryptocurrencies, their trends and their relations with traditional assets. We collected and cleaned data, and implemented exploratory analysis with data visualization and preliminary analysis. Then, we presented further analysis by Regression, K nearest neighbours(K-NN), Time-series analysis, and Event studies to reach our goal to perform whole process of data analytics and reach to the prediction of future prices of cryptocurrencies. Also, we implemented and evaluated ARCH/GARCH models to predict the volatility of price.*

Keywords*—Data analytics, crypto currencies, regression, clustering, K-NN, time-series analysis, event studies.*

I. Introduction

1.1 Motivation

In the recent periods, bitcoin is seen to have an exponential surge of its price, such that the importance of cryptocurrencies as new asset class has been emerged. With this project, we explore and analyze the ways to predict the price movements of the digital assets.

1.2 Problem Statement

We sought to predict price movements of digital assets from various feature parameters such as price movements of traditional assets, price movements of correlated digital assets and how a current price can be predicted by historical prices.

II. Datasets

(Relevant python notebook file: DATACLEANING.ipynb)

2.1 Digital assets

We choose 4 digital assets from what is known to most digital traders as reliable source, CoinGecko. We picked digital asset classes with at least 4-5 years of data and were the top 10 digital assets in terms of market capitalization. As results, we collected historical data of Bitcoin (btc), Ethereum (eth), XRP (xrp), and Litecoin (ltc). For each crypto currencies, we obtained the closing price, closing market capitalization, closing volume and the date (1).

2.2 Traditional Asset Classes

To search inter-relationship between crypto currencies and traditional asset classes, we gathered following data of asset classes: Gold, S&P500, Nasdaq100, US10 year government bond yield, and dollar index. For each asset classes, we obtained date, closing price, and daily volume.

SPDR gold index is the most popular gold ETF (exchange-traded fund).(2)

The S&P 500 is one of the most commonly followed equity indices that measures the stock performance of 500 large companies listed on stock exchanges in the United States.(3)

The NASDAQ-100 is a stock market index of the major technology companies, which contain 100 of the largest non-financial companies listed on the Nasdaq stock market.(4)

The federal reserve provides various historical data for the US government bond yield. While the government issues several different terms of bonds, we chose US 10 year government bond yield for its popularity in representing interest rate.(5)

The U.S. Dollar Index is an index of the United States dollar relative to a basket of foreign currencies, often referred to as a basket of U.S. trade partners' currencies. The Index goes up when the U.S. dollar gains "strength" when compared to the other currencies (6). (Note: For Dollar index, there

```
[ ] price_df.describe()
```

	btc_price	eth_price	xrp_price	ltc_price	spx_price	ndx_price	gold_price	dxy_price
count	2405.00000000	1692.00000000	2321.00000000	2405.00000000	2405.00000000	2405.00000000	2405.00000000	2405.00000000
mean	4114.28265648	223.00454604	0.19810178	39.34510170	2414.57232848	5946.41632017	1336.12286902	93.04728280
std	4704.05143120	232.05712269	0.29800721	49.92772753	510.04224550	2269.27085741	195.56265532	6.41422624
min	68.08310000	0.43297860	0.00268621	1.14885101	1568.25000000	2830.50000000	1049.40000000	79.13999900
25%	416.85000000	12.66217706	0.00688700	3.75071485	2011.50000000	4267.00000000	1223.50000000	90.26999700
50%	1078.27471109	180.68903581	0.03866760	20.35770000	2341.75000000	5347.50000000	1283.80000000	95.16999800
75%	7538.55768728	304.61586605	0.29413900	57.07559991	2810.75000000	7376.50000000	1346.25000000	97.41999800
max	28837.28852896	1410.00021451	3.39845000	360.66176169	3748.75000000	12885.50000000	2067.15000000	103.29000100

is no daily volume as we cannot obtain it.)

As a bottom line, each data properties are shown as above table of price_df.describe().

2.3 Data cleaning

We cleaned up the missing data such as “N/A”, “HOLIDAY” and “NYSE Closed” values, which were replaced with numpy’s nan values in our traditional assets price datas. After correcting the dataset, we did a forward fill to replace the nan values with the previous day’s prices because when plotting a time-series analysis it would not be ideal to have a missing value of the specific period. Also, we encountered miscellaneous matters for data cleaning. For example, some crypto currencies such as Ethereum only have data after July 2015. We conducted qualitative research as well, to clean data (See 7,8). Right picture is snapshot of this process.

```
pd.set_option("display.max_rows", None, "display.max_columns", None)
#checking num of null data
cryptos_df.isnull().sum()

day             0
btc_price       0
btc_mktcap      0
btc_volume      0
eth_price     831
eth_mktcap    831
eth_volume    831
xrp_price      98
xrp_mktcap     98
xrp_volume     98
ltc_price       0
ltc_mktcap      0
ltc_volume      0
```

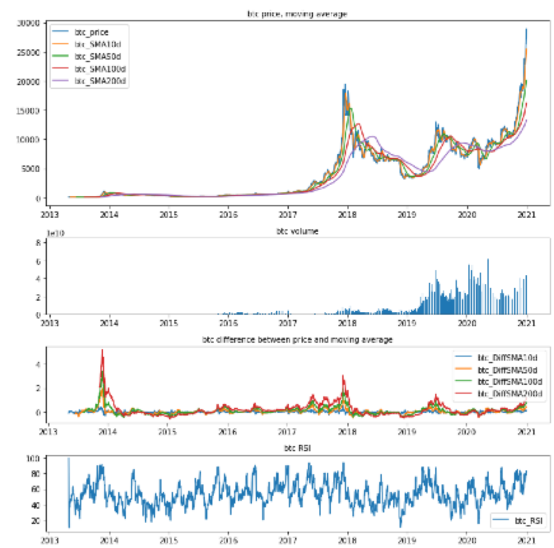
2.4 Data Transformation

We also processed and transformed data as follows.

- Daily % change of price (pct_change)
- SMA (Simple moving average), 10/50/100/200 days.
- Difference between price and SMAs
- RSI (Relative Strength Index)

First, we added pct_change, because we are basically interested in price movement.

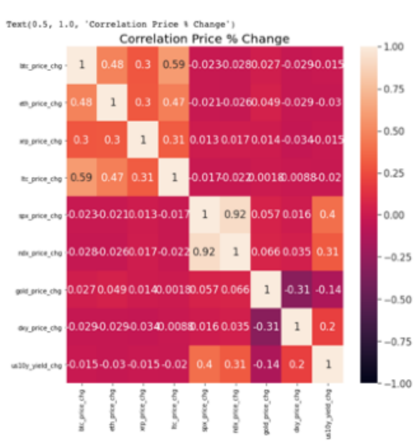
We also formulated SMA (9) from the `rolling` & `mean` methods in pandas, the indicator which is the average price over a given number of time periods. Then, we defined the golden-cross and death-cross. The golden-cross occurs when the price goes over SMA to the upside and is interpreted as buying opportunities. Similarly downside moving average crossover constitutes the death-cross and is understood to signal as downturn in a market(10). We also calculated the difference between price and each SMA, the indicator which is often used in trading strategy. If the price is far above the moving average, it indicates that the asset is "overbought" and vice versa.

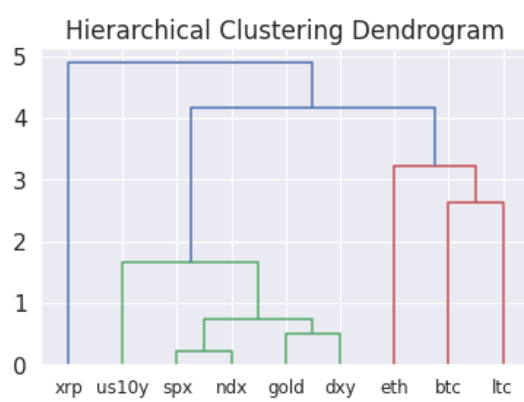


We made use of the `pandas_ta` library to calculate RSI of the prices for each assets class(11). The RSI is a momentum indicator used in technical analysis to evaluate overbought or oversold conditions in the price. If the values are 70-80 or above indicate that a security is becoming overbought. An RSI reading of 20-30 or below indicates an oversold.

Finally, we combined all datas into one singular csv file with same data format. We can see the data shown as above picture, which is the example for btc. We could see that the volume of crypto has increased since year 2019. Also, we could see many trading signals such as golden/death crosses from SMA, and RSI above 70-80 and below 20-30. We examined further regarding those findings, later in this report.

III. From exploratory analysis: Correlation heat map and Clustering





(Relevant python notebook file: ALLWORKS_IN_MILESTONE1.ipynb)

In this section, we introduce the important analysis and results in the exploratory analysis. With the exploratory analysis of correlation heat map and clustering, we could find that the relationship between traditional asset classes and cryptos are not strong, moving separately. Also, xrp moves totally separately from any other asset classes and cryptos. On the other hands, btc, ltc and eth have relatively high correlation and fall under same category. Especially, btc and ltc show closer relationship. Such phenomena can stem from the history and structure of each crypto currencies. For example, Litecoin has similar mechanics with Bitcoin (12).

From those findings, we decided to exclude traditional asset classes from our further research scope this time.Also, we decided to examine interrelationships among eth, btc, and ltc.

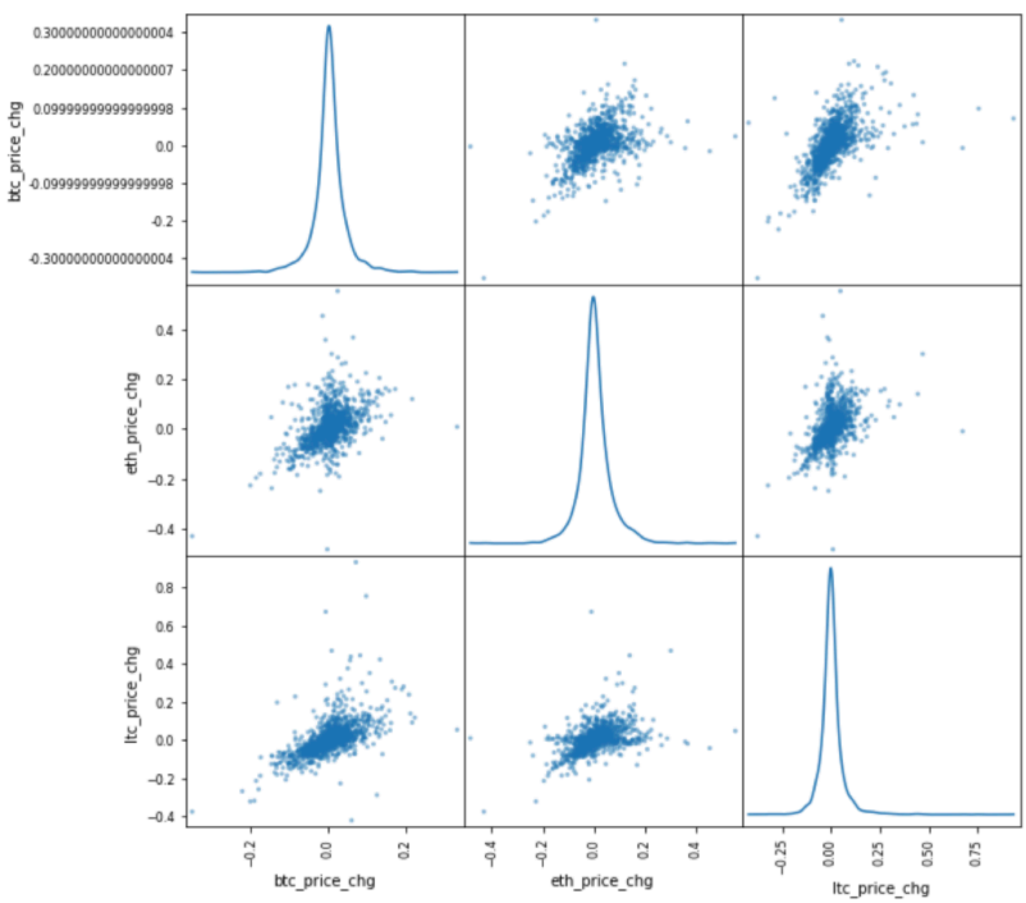
IV.Further researches and methodologies for prediction

In this section, we'll introduce several researches to predict the price of cryptocurrencies. Our dataset is time-series data, such that we worked on the time-series analysis. We also implemented other methodologies such as linear regression and K-NN technique. Also, we originally searched and implemented Event-study to examine whether the generally used technical indicators such as golden/death crosses of SMA and overbought/oversold of RSI can work as trading strategy. Finally, we implemented ARCH/GARCH to predict the volatility of bitcoin price.

4.1 Regression

(Relevant python notebook file: LINEAR_REGRESSION.ipynb)

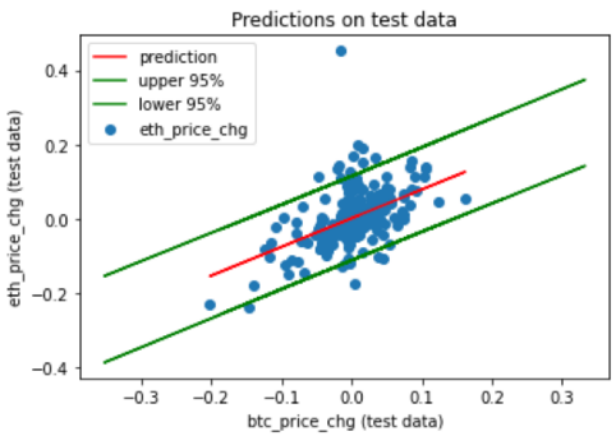
4.1.1 Scatter plots and kernel density estimation

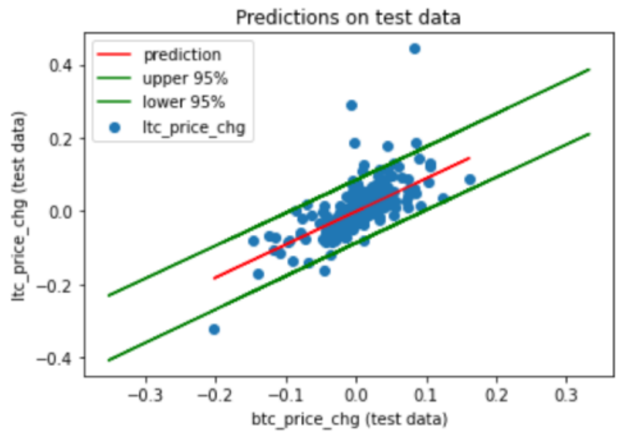


As the preliminary step of regression, we took scatter plot with kernel density estimate (KDE) from price % changes of btc, eth, and ltc. We could observe that there's a positive relationship between the digital assets. From the

KDEs, we found that those are centred around 0% change with density of about 0.3 for btc, >0.4 and >0.8 for eth and ltc respectively, the results which indicate that the bitcoin has flatter distribution with higher volatility than others.

4.1.2 Linear Regression Modeling





We implemented linear regression of the price percentage change of eth (left hand side of graph above) and ltc (right hand side) in relation to btc, and whether predicting price movements based on the relationship will be possible. We set btc_price_chg as explanatory variable x while eth and ltc price % changes as target variable y. The reason is that the bitcoin is the most actively traded with the highest market capitalization, such that we thought that predicting minor currencies such as eth and ltc from very competitively traded crypto of bitcoin is reasonable.

The results of x=btc_price_change, y=eth_price_change:

- R-square is not much high, 0.221.
- Regression coefficient of btc_price_chg is 0.772. This means that, on average, when btc_price_chg is +1%, eth_price_chg can move by +0.772%.
- Regarding the btc_price_chg coefficient, t-stat is very high with 19.582 and p-value is very low, nearly zero.
- Also, F-statistic is high 383.5, such that prob(F-stat) is very low.
- Regarding residual plot, while we observe some outliers such that the line is not straight, we can see that the line is totally horizontal around the points where many data is available (around -0.15 < btc_price_chg < +0.15).
- About the prediction performance in test dataset, while we see some outliers, many test dataset can fall between upper and lower 95% confidence interval.

```
                    OLS Regression Results
==============================================================
Dep. Variable:     eth_price_chg  R-squared:            0.221
Model:                      OLS  Adj. R-squared:        0.221
Method:           Least Squares  F-statistic:           383.5
Date:          Tue, 20 Apr 2021  Prob (F-statistic):  2.35e-75
Time:                  13:08:06  Log-Likelihood:       1941.9
No. Observations:          1352  AIC:                  -3880.
Df Residuals:              1350  BIC:                  -3869.
Df Model:                     1
Covariance Type:      nonrobust
==============================================================
                 coef   std err      t    P>|t|  [0.025  0.975]
--------------------------------------------------------------
Intercept      0.0026    0.002   1.635   0.102  -0.001   0.006
btc_price_chg  0.7718    0.039  19.582   0.000   0.695   0.849
==============================================================
Omnibus:              479.281  Durbin-Watson:          2.027
Prob(Omnibus):          0.000  Jarque-Bera (JB):   10629.448
Skew:                   1.108  Prob(JB):                0.00
Kurtosis:              16.557  Cond. No.                25.2
==============================================================
```

The results of x=btc_price_change, y=ltc_price_change:

- R-square is higher than that in case of eth, 0.403.
- Regression coefficient of btc_price_chg is 0.9024. This means that, on average, when btc_price_chg is +1%, eth_price_chg can move by +0.902%.
- The btc_price_chg coefficient, t-stat is very high with 30.17 and p-value is very low, nearly zero.
- Also, F-statistic is high 910.2, which is higher than the case in eth, such that prob(F-stat) is very low.
- Regarding residual plot, while we observe some outliers such that the line is not straight, we can see that the line is totally horizontal.
- About the prediction performance in test dataset, while we

```
                    OLS Regression Results
==============================================================
Dep. Variable:     ltc_price_chg  R-squared:            0.403
Model:                      OLS  Adj. R-squared:        0.402
Method:           Least Squares  F-statistic:           910.2
Date:          Tue, 20 Apr 2021  Prob (F-statistic): 2.85e-153
Time:                  13:08:07  Log-Likelihood:       2315.0
No. Observations:          1352  AIC:                  -4626.
Df Residuals:              1350  BIC:                  -4616.
Df Model:                     1
Covariance Type:      nonrobust
==============================================================
                 coef   std err      t    P>|t|  [0.025  0.975]
--------------------------------------------------------------
Intercept     -0.0012    0.001  -1.034   0.301  -0.004   0.001
btc_price_chg  0.9024    0.030  30.170   0.000   0.844   0.961
==============================================================
Omnibus:             1392.675  Durbin-Watson:          1.893
Prob(Omnibus):          0.000  Jarque-Bera (JB):  198698.039
Skew:                   4.568  Prob(JB):                0.00
Kurtosis:              61.683  Cond. No.                25.2
==============================================================
```

see some outliers, many test dataset can fall between upper and lower 95% confidence interval. 95% confidence interval is narrower than that of eth case. Also, it can fit to test data nicer than the case in eth.

As bottom line, although we have not so high r-square, we still see that the regression have a decent goodness of fit with very low p-values of coefficients and high F-statistics, and we are able to provide decent predictions based on the 95% confidence interval (Regarding the discussion around r-squares and model validity, please refer (13)). Thus, as expected, we can see the relationship between btc_price_chg (explanatory variable x) and eth_price_chg and ltc_price_chg (target variable y) in a same day. It is just the same day's relationship between different currencies, such that our findings does not mean "the prediction of tomorrow". However, we could obtain the sensitivity of btc_price_chg to eth and ltc_price_changes, the model which can work in test data as well. Therefore, we could utilize our findings for intra-day's short term trading, such as following:

- When the intra-day bitcoin price increased sharply while eth and ltc increased very little, i.e. less than β*btc_price_chg (β=0.772 for eth, β=0.9024 for ltc), we can buy eth and ltc and can sell eth and ltc when those price reaches around β times btc_price_chg.
- When the intra-day bitcoin price decreased moderately while eth and ltc decreased sharply, i.e. more than β*btc_price_chg, we can buy eth and ltc and can sell eth and ltc when those price reaches around β times btc_price_chg.

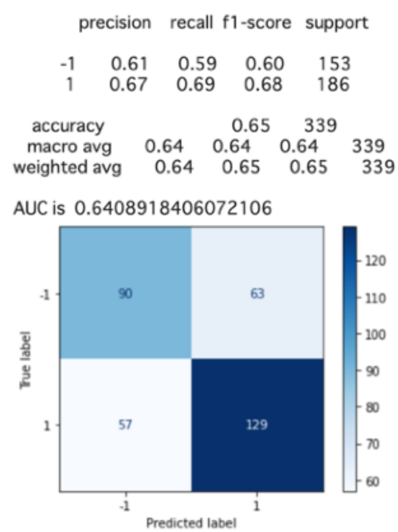
Because above trading strategy is "day-trading" (i.e. intra-day trading in finance term), we need 5 minutes tick data to inspect the actual efficacy of above strategies, the analysis which can be beyond our scope this time. However, as the future direction, it can be interesting topic.

4.2 K Nearest Neighbour (K-NN Classifier)

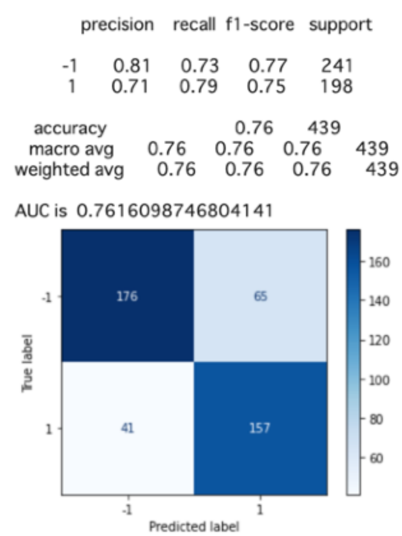
(Relevant python notebook file: K-NN.ipynb)

K nearest neighbour is one of the most intuitive methodologies for pattern recognition (14). Our aim of implementing K-NN is to test our assumption that we could utilize bitcoin's information other than price, such as volume change and market capitalization change, to predict the price changes of eth and ltc. We tried to predict whether eth/ltc will go up/down in price, based on the feature vectors of btc. The vectors used in our prediction are the btc's trade volume, market cap and price.you will see that we are trying to predict the price percentage direction (up/down, 1/-1) of eth and ltc based on those feature vectors of btc. The prediction results is shown as follows.

- Eth: We could predict precision=0.67, recall=0.69, and f1_score=0.68, which is better than 50/50 coin tossing. AUC is 0.64, better than 0.50. It implies that we can utilize bitcoin's volume and market cap information to predict eth price % changes.
- Ltc: We could predict precision=0.71, recall=0.79, f1_score=0.75, which is better than 50/50 coin tossing and eth's prediction. AUC is 0.76, better than 0.50 and the result of eth. It reflects the similarity of currency structure and high correlation between bitcoin and ltc.



ETH prediction by btc features



LTC prediction by btc features

- Additional features: We also tried to create model by increasing features regarding each of eth/ltc's volume and market capitalization information. In this case, while the training accuracy is relatively good, better than 50%, the results from test data set were far below the above results with AUC around 0.50. It implies that we encountered over-fitting by adding features.

As results, while our model just forecast today's eth/ltc price from today's btc's market information such as price, volume, and market cap % changes i.e. it's not the "prediction for tomorrow", at least we could show that we could capture the daily price direction of eth and ltc, by utilizing bitcoin's market information.

We may be able to utilize this insight for "day trading" i.e. intra-day trading in finance term. For example, at the 12 o'clock of daily market, we run this model, then, we may be able to obtain the insight about afternoon and night market direction. If we had more time, we could try such analysis as well.

4.3 Time series analysis

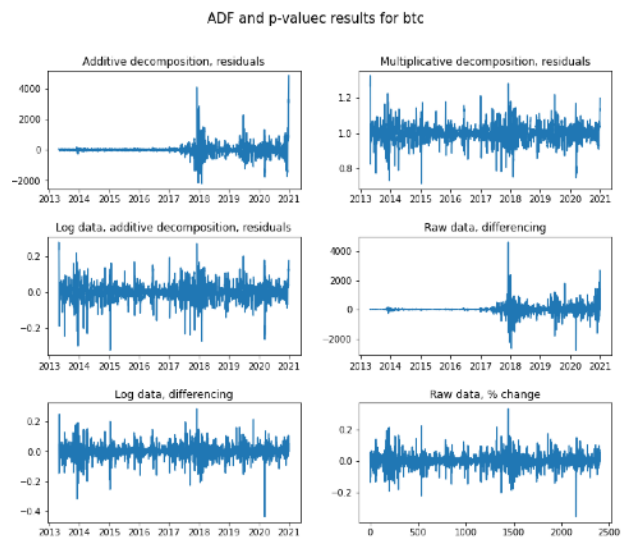
(Relevant python notebook file: TIMESERIES.ipynb)

First, we considered and implemented how we can obtain the stationarity of data. Then, we applied ARIMA model to cryptocurrencies to make the prediction of future price of cryptocurrencies. To optimize the (p, d, q) parameters of ARIMA model, we took two different approaches that are introduced in the tutorial, that is, 1. Graphical method and 2. Grid-search. Finally, we visualize our prediction in the graph. While our prediction took very wide in range, for example, it could capture the current price surge of bitcoin.

Stationarity check of data:

We checked how we can obtain the data stationarity for all cryptocurrencies of btc, eth, xrp, and ltc, in terms of following ways. For inspection methodologies, we applied ADF test and graphical method.

- Additive decomposition and taking residual.
- Multiplicative decomposition and taking residual.
- Taking log of data, then implement additive decomposition.
- Differencing by raw data.
- Differencing by log taken data.
- Taking daily price % change.



Key take-aways are as follows. Probably, those results stem from the historical data properties of cryptocurrencies which surged the price exponentially.

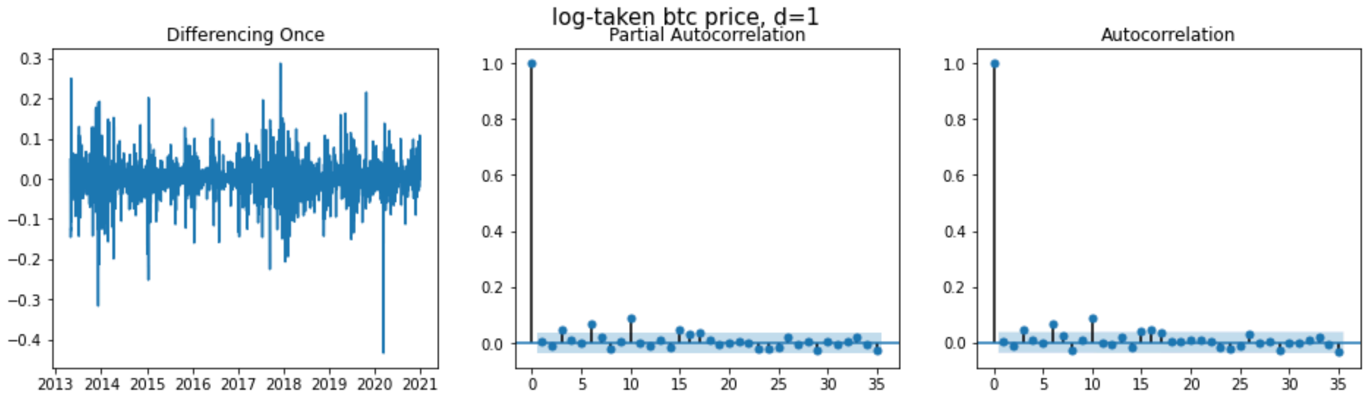
- In total, additive decomposition and differencing on raw data don't work well, by looking at graphs.
- Totally, below 3 are good to obtain stationarity data:
  1. Residuals on multiplicative decomposition
  2. Residuals on additive decomposition after taking Log
  3. differencing of log data.
- % change of raw data is very simple, but not bad by looking at graphs. Also, in btc, ADF test shows very good results.

Applying ARIMA model:

As we can see the above results, differencing of log data can serve stationarity relatively well for all cryptos, such that we convert raw price data into log then apply ARIMA model. To optimize the (p, d, q) parameters of ARIMA model, we took two different approaches that are introduced in the tutorial, that is, 1. Graphical method by ACF and PACF, and 2. Grid-search. We implemented with both methods, and reached to the best model of bitcoin prediction as ARIMA(10,1,10) by graphical methods and ARIMA(5,1,5) by grid-search method.

Graphically checking PACF and ACF:

First, we optimized parameters (p,d,q) of ARIMA model by graphically seeing ACF and PACF. First, we examined 1st and 2nd differencing of log-taken data (above is the results of bitcoin). As we tried it for all cryptocurrencies, 1st order of differencing was enough to obtain stationarity data. Therefore, we applied d=1.

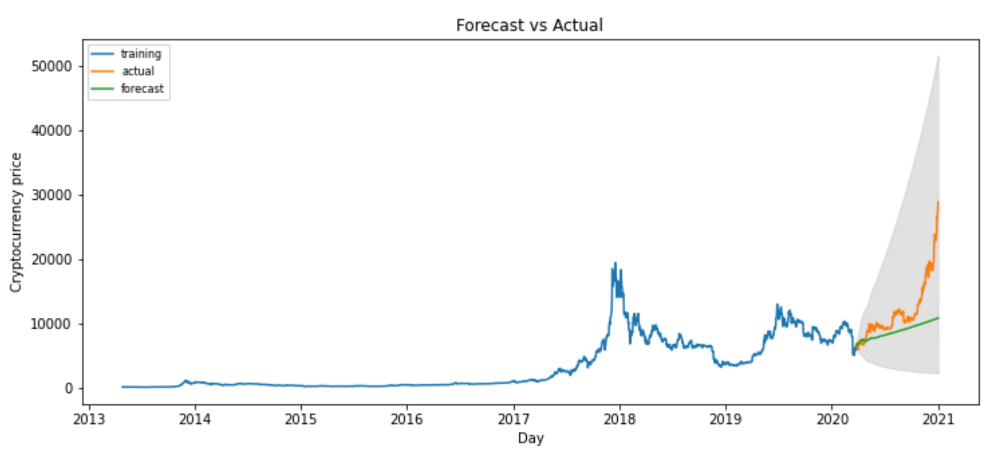


Then, we confirmed p by checking PACF, and q by checking ACF shown as above (above is the example of btc). From our observation of PACF and ACF graphically, below results can be obtained. For example, we can confirm ARIMA(10,1,10) as the best model for btc.

- btc: p=6,10, q=6,10 are looking good.
- eth: p=1,7, q=1,7 are looking good.
- xrp: p=2,5,8,9, q=2,5,8,9 are looking good.
- ltc: p=6, q=6 is looking good.

Grid-search:

We implemented grid-search by p=1-10, d=1, q=1-10. For the measurement of model validity, we utilized AIC (Akaike Information Criterion), which is popularly used for model selection. AIC could penalizes if we increase the number of parameters, such that we could find the balanced model between accuracy and simpleness/robustness of the model, by avoiding overfitting(15,16). From grid-search, for example, ARIMA(5,1,5) was the best model for btc.



As results, while the prediction takes very wide in range, for example, the ARIMA(10,1,10) of bitcoin could capture the current price surge of bitcoin price, shown as graph above.

4.4 EVENT STUDY

(Relevant python notebook file: EVENTSTUDY.ipynb)

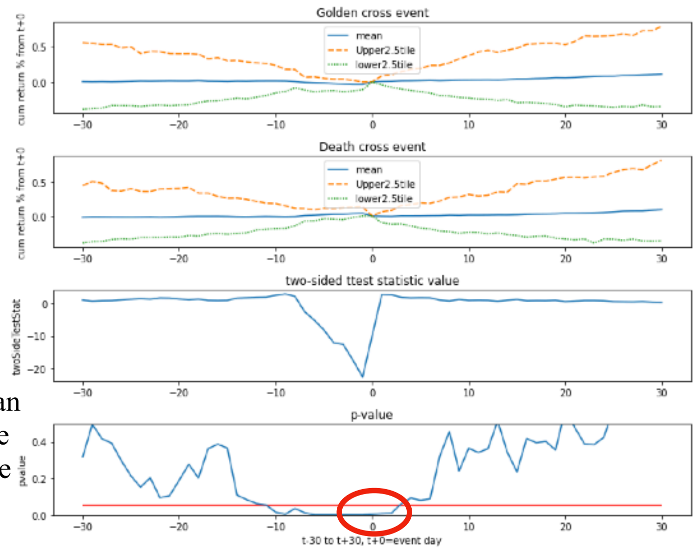
We implemented "Event Study" to investigate whether "golden-cross (=buy signal)", "death-cross (=sell signal)" of SMA and the selling and buying signals from RSI (i.e. RSI>75 = sell due to overbought, RSI<25 = buy due to oversold) really work or not as many technical analysts mention.

The event study is often used in econometrics/finance area, to examine how the price move before and after x days from the "event" day (t+0) i.e. the day when the golden/death crosses happen. We can calculate cumulative mean return and standard deviation after t+x later from the event, then we can apply two-sided t-test to examine whether the mean cumulative return of t+x day between buying and selling signals is different. Null hypothesis of two-sided t-test is that, “mean cumulative return between after-buy-signal and after-sell-signal at t+x day is not different”. Regarding the details of event study, please refer (17-19).

Event study of SMA for each cryptocurrencies:

After investigating all cryptocurrencies of btc, eth, xrp, ltc, regarding all golden/death crosses of 10/50/100/200 SMAs, we found that for the bitcoin in SMA 10 days, the mean returns of t+1 and t+2 have shown statistically different performance between golden-cross and death-cross. In t+2, while mean cumulative return after the golden-cross was +0.70%, that after the death-cross was -0.60%. It implies that we can buy bitcoin at golden-cross while we can sell bitcoin at death-cross, such that we can make the spread of profit by +1.30% in average, as prescriptive solution.

On the other hands, other than the result above, we could not reject null hypothesis, the hypothesis that the mean returns after t+x days of buying signals and selling signals are same. Right picture shows the example of our output for the bitcoin in SMA 10 days.



Event study of RSI for each cryptocurrencies:

After investigating all cryptocurrencies of btc, eth, xrp, llc, regarding all overbought signals (RSI>75) and all oversold signals (RSI<25), we found below insights and prescriptive solutions.

- **Bitcoin:** t+1 shows p-value of 0.029, which can reject null hypothesis and imply statistically significant difference in performance between RSI overbought events and RSI oversold events. As we can see, the mean return on t+1 at overbought event is +0.91% while the mean return on t+1 at oversold event is +4.30%. It implies that we can buy bitcoin at RSI<25 while we can sell bitcoin at RSI>75, such that we can make the spread of profit by +3.39% in average.
- **Eth:** We could not reject null hypothesis in all t+1 - t+30 time windows.
- **Xrp:** We could reject null hypothesis between t+4 and t+9. Especially, in t+7, we could obtain the lowest p-value with 0.0226. However, contrary with the consensus of how to use RSI, the mean cumulative return after t+7 of overbought event is +40.0%, which is significantly higher than that of oversold event i.e. +3.66%. This result indicates that we can BUY after the overbought event of RSI>75 while we can SELL after the oversold event of RSI<25, such that we can make the spread return of +36.3% in average.
- **Ltc:** We could see the p-value < 0.05 on t+25 and t+26. Especially, we could obtain the lowest p-value of 0.037 in t+26. However, contrary with the consensus of how to use RSI, the mean cumulative return after t+26 of overbought event is +44.0%, which is significantly higher than that of oversold event i.e. -5.30%. This result indicates that we can BUY after the overbought event of RSI>75 while we can SELL after the oversold event of RSI<25, such that we can make the spread return of +49.3% in average.

4.5 Volatility estimate by ARCH/GARCH models.

(Relevant python notebook file:ARCH_GARCH.ipynb)

In this section, we implemented further analysis of time-series analysis, ARCH/GARCH to predict the volatility of cryptocurrencies after the milestone 1. As we implemented exploratory analysis in milestone 1, we found interesting phenomena in which the volatility of cryptocurrencies may follow some pattern. The pattern is that, periodically we see the spike of the volatility, then the high volatility continues for a while, and finally it can calm down again.

To capture such volatility spike by time-series analysis, ARCH/GARCH model are popularly used. This time, we implemented ARCH(1) and GARCH(1,0,1) as the generally applied starting point. Then, we also applied ARCH(5)/

GARCH(5,1,5), and ARCH(10)/GARCH(10,1,10), by obtaining parameters p,o,q from the ARIMA model's evaluation by the graphical method and grid-search method. (Regarding ARCH/GARCH parameters determination from ARIMA model, please refer (20).)

	Data	GARCH(1,0,1)	ARCH(1)	GARCH(5,1,5)	ARCH(5)	GARCH(10,1,10)	ARCH(10)
Information Criterion	AIC	13,528.000	13,888.700	13,502.500	13,575.700	13,499.200	13,514.400
	BIC	13,551.200	13,906.200	13,578.000	13,616.400	13,632.900	13,584.100
p-values	mu	0.014	0.105	0.113	0.013	0.113	0.061
	omega	0.002	0.000	0.402	0.000	0.165	0.000
	alpha 1	0.000	0.000	0.090	0.000	0.122	0.001
	alpha 2	nan	nan	0.656	0.008	0.663	0.090
	alpha 3	nan	nan	0.727	0.189	0.653	0.309
	alpha 4	nan	nan	0.502	0.029	0.102	0.063
	alpha 5	nan	nan	0.818	0.003	0.659	0.028
	alpha 6	nan	nan	nan	nan	1.000	0.524
	alpha 7	nan	nan	nan	nan	0.563	0.054
	alpha 8	nan	nan	nan	nan	1.000	0.555
	alpha 9	nan	nan	nan	nan	0.762	0.237
	alpha 10	nan	nan	nan	nan	0.950	0.155
	gamma 1	nan	nan	0.209	nan	0.262	nan
	beta 1	0.000	nan	1.000	nan	1.000	nan
	beta 2	nan	nan	1.000	nan	1.000	nan
	beta 3	nan	nan	0.076	nan	1.000	nan
	beta 4	nan	nan	0.928	nan	1.000	nan
	beta 5	nan	nan	1.000	nan	1.000	nan
	beta 6	nan	nan	nan	nan	1.000	nan
	beta 7	nan	nan	nan	nan	1.000	nan
	beta 8	nan	nan	nan	nan	1.000	nan
	beta 9	nan	nan	nan	nan	1.000	nan
	beta 10	nan	nan	nan	nan	0.317	nan

*Note: Orange highlights show $p<0.05$, green highlights show $p<0.10$.

As results, we could obtain following insights. Generally saying, GARCH can work well with very simple GARCH(1,0,1) with low AIC/BIC and $p<0.05$ for all coefficients, while it doesn't show much improvement after our tuning of GARCH(5,1,5) or GARCH(10,1,10). Therefore, when we have limited information about time series data, we can utilize GARCH(1,0,1) as good starting point to predict the volatility. We understood this is why GARCH(1,0,1) is generally used to forecast volatility as the financial industry custom.

On the other hands, ARCH could not work well with simple ARCH(1) with high AIC/BIC and mu with $p>0.05$. However, when we have information about the parameter from ARIMA modeling, we could improve the model drastically by applying those results i.e. in our case, ARCH(5) and ARCH(10). ARCH tends to improve AIC/BIC and p-values as we raise the parameter. Therefore, when one already has much information regarding time-series data, he/she can apply ARCH to improve the predicting quality.

The volatility estimate is very important for option pricing and risk management in the finance. Therefore, our trial can contribute to those area, while the derivative market is not fully developed in the cryptocurrencies field. For details of ARCH/GARCH and those implementation, please refer (21-23) too.

V. Conclusion

We started from gathering and cleaning data of cryptocurrencies and traditional asset classes. From exploratory analysis of correlation heatmap and clustering, we found that the correlation between cryptocurrencies and traditional asset classes are totally low, such that it's difficult for us to predict the price movement of cryptocurrencies from traditional asset classes. Therefore we limited our research scope only to cryptocurrencies and decided to apply regression, K-NN, ARIMA model, event study, and ARCH/GARCH toward cryptocurrencies.

As results, although the prediction is not "the forecast for tomorrow", we could make prediction of eth and ltc from same day's price changes of bitcoin from regression, such that it could utilize our findings toward intra-day trading. Also, from K-NN, we could obtain implication that we could utilize other than price information such as trading volume and market capitalization. In addition, while the predictions took very wide in range, we could formulate the price prediction of cryptocurrencies by ARIMA model. Furthermore, from event study, we could find prescriptive solutions of trading strategies by utilizing SMA and RSI. Finally, we also tried to predict the volatility of bitcoin with ARCH/GARCH, such that we could find we could use GARCH(1,1) as nice starting points and could tune ARCH utilizing the results from ARIMA modeling.Through the whole process of this research, we could gain practical expertise of data analytics from data collection to result analysis and validation of the model.

For future direction, if we had more time, we could try additional analysis as followings. First, while the difficulty of prediction can become much higher, we could try “the prediction for tomorrow” by utilizing regression, K-NN, and another advanced methodologies such as machine learning, especially RNN and LSTM which can be applicable to time-series data.

Also, we could take larger amounts of cryptocurrency data by taking 5 minutes of price ticks rather than daily closing price, such that we could investigate trading strategies based on 5 minutes data, the strategy which could have more profiting opportunities, including intra-day trading that we’ve touched in this report.

VI.References

1. https://www.coingecko.com/en
2. https://www.spdrgoldshares.com/usa/historical-data/
3. https://www.nasdaq.com/market-activity/futures/es/historical
4. https://www.nasdaq.com/market-activity/futures/nq:us/historical
5. https://www.federalreserve.gov/datadownload/Download.aspx?rel=H15&series=bf17364827e38702b42a58cf8eaa3f78&lastObs=&from=&to=&filetype=csv&label=include&layout=seriescolumn&type=package
6. https://finance.yahoo.com/quote/DX-Y.NYB/history?period1=1264464000&period2=1611619200&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true
7. https://en.wikipedia.org/wiki/Ethereum
8. https://en.wikipedia.org/wiki/Ripple_Labs
9. https://www.investopedia.com/ask/answers/122314/how-do-i-use-moving-average-ma-create-forex-trading-strategy.asp#:~:text=Moving%20averages%20are%20a%20frequently,100%2C%20and%20200%20day%20periods
10. https://www.investopedia.com/ask/answers/121114/what-difference-between-golden-cross-and-death-cross-pattern.asp
11. https://www.investopedia.com/terms/r/rsi.asp
12. https://en.wikipedia.org/wiki/ Litecoin
13. http://www.stat.cmu.edu/~cshalizi/mreg/15/lectures/10/lecture-10.pdf
14. https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html
15. https://coolstatsblog.com/2013/08/14/using-aic-to-test-arima-models-2/
16. https://otexts.com/fpp2/arima-estimation.html
17. John Y. Campbell, Andrew W. Lo, A. Craig MacKinley, Econometrics of Financial Markets, Prinston University Press, 1997, chapter 4, p149-180
18. https://www.eventstudytools.com/introduction-event-study-methodology
19. https://www.investopedia.com/terms/e/eventstudy.asp#:~:text=An%20event%20study%20is%20an,react%20to%20a%20given%20event
20. https://medium.com/auquan/time-series-analysis-for-finance-arch-garch-models-822f87f1d755
21. https://readthedocs.org/projects/arch/downloads/pdf/latest/
22. https://arch.readthedocs.io/en/latest/univariate/introduction.html
23. https://arch.readthedocs.io/en/latest/univariate/univariate_volatility_modeling.html