MDS6212 Fintech Theory and Practice

Assignment 7

Yihang Li
220041006

1 Question1

In this question, you are expected to conduct some simple analysis on the price dynamics of Bitcoin and Ethereum. Coinmarketcap (https://coinmarketcap.com/) is a famous crypto-currency platform, which keeps track on price data on almost all listed crypto-currencies.

1.1 Part a)

Please collect the historical price data for Bitcoin and Ethereum from Coinmarketcap and plot the price time-series for both coins. [You may manually download the data from the website or directly call their official APIs, the documentation for which can be found at https://coinmarketcap.com/api/documentation/v1/]

1.1.1 Some Comments

In this Part, I tried to either manually download the data or use the official API but both failed. In the process of searching data, I noticed that there is a guy who said, "I recently found myself wanting to download historical cryto data in order to do some data analysis. Although they do have a public API, this does not seem to have a facility for downloading any historical data. There is also a notice that this API will be taken offline in Dec 2018 and replaced by a Professional API. The Professional API will allow for historical data to be downloaded but at the very steep price of $699/month for access to 12 months of historical data and you need to inquire for pricing for access to up to 5 years of historical data."

So, in the end, I have to directly scrap the data from the website by using Python's package, request, and using regular expression. To be more clear, I put all my scraping codes in a function, *Historical_Data_from_CMC(coin_name:str, start_date:str, end_date:str)*, by using it , I successfully collected the historical price data for Bitcoin and Ethereum from CMC and saved them both into CSV file separately.

1.1.2 plot the price time-series for both coins.

Figure 1 and Figure 2 are the price time-series for both coins.

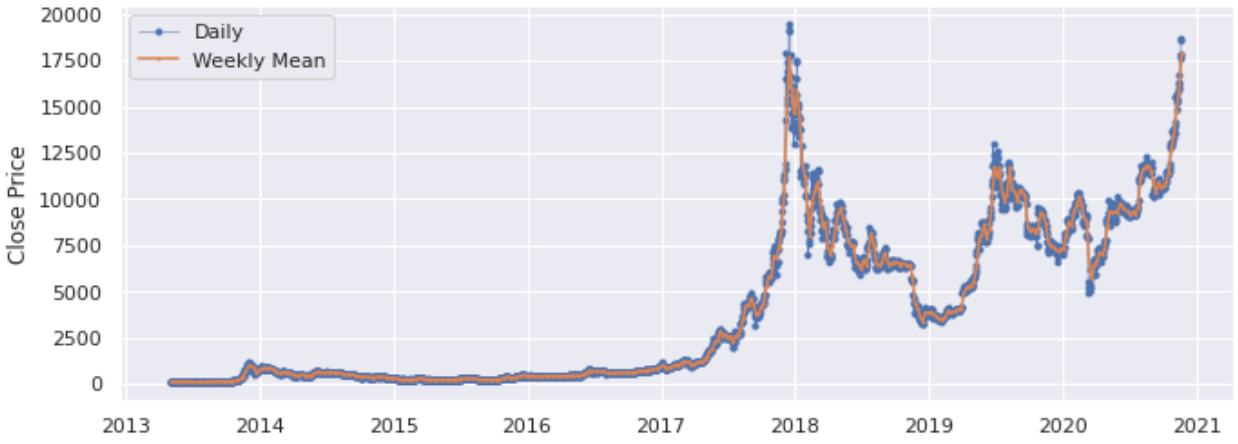


Figure 1: Price Time Series for Bitcoin

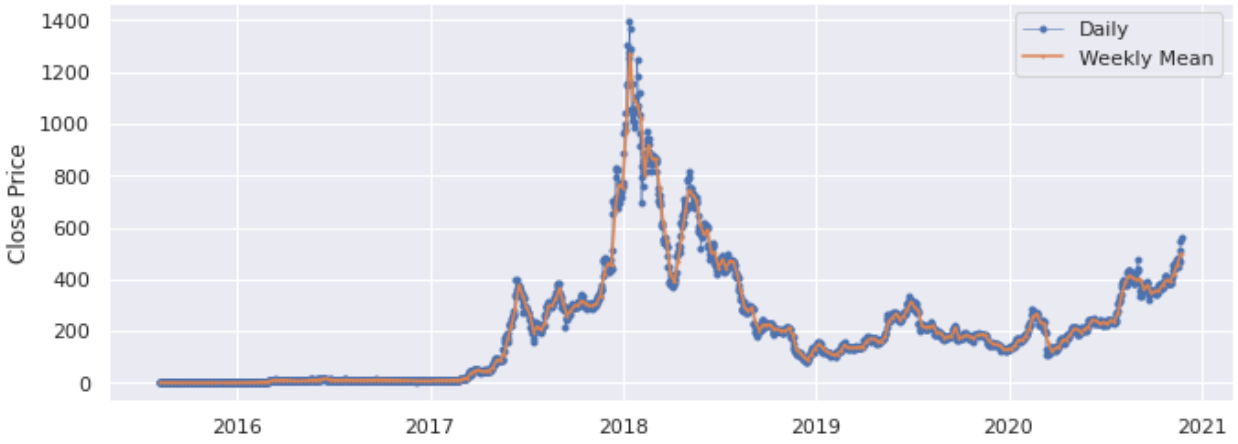


Figure 2: Price Time Series for Ethereum

1.2 Part b)

Calculate the correlation between Bitcoin price series and Ethereum price series. Summarize your findings.

1.2.1 Calculate Correlation

Recall that Cross-correlation is the comparison of two different time series while Auto-correlation is the comparison of a time series with itself at a different time.

So here we calculated the Cross-correlation between Bitcoin and Ethereum price time series by using the following Python code:

```
Bitcoin['Close'].corr(Ethereum['Close'], method="pearson")
```

And the result is: **0.7355659568999063**.

1.2.2 Summarize

As we see, the correlation is close to 1, which means the price series between Bitcoin and Ethereum are highly related. We can also see this from the trends in both Figure 1 and 2.

1.3 Part c)

Fit two price series with classical ARMA models. Report your regression results.

Take the time series of Bitcoin as an example. We illustrate the process as follows:

1.3.1 Stationary Test

There are some basic assumptions about classical time series models like ARMA. One of those is that the time series should be stationary.

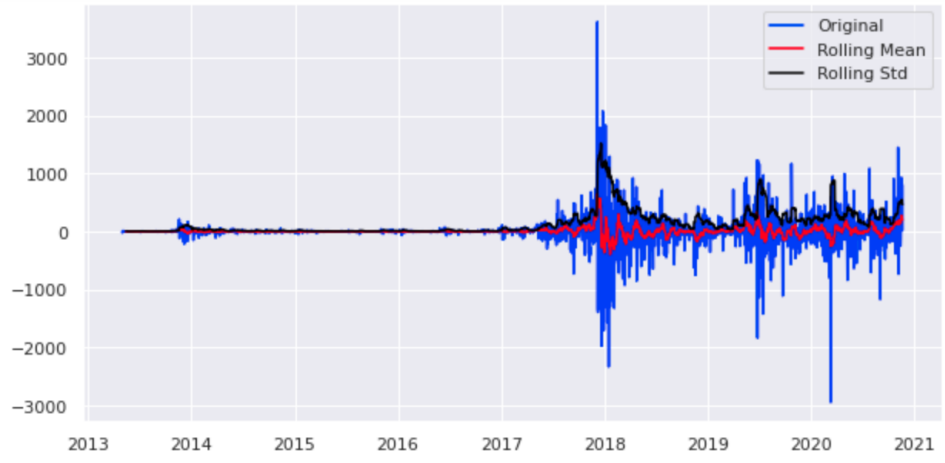
Figure 3 is the result of the stationary test. Obviously, this time series is not stationary since the p-value is much higher than 0.05, so we need to perform first order difference to try to make it stationary. After that, we test the df_ts, the time series of difference, again, get the result as Figure 4, now our time series is stationary.

1.3.2 Determine Parameters p and q

Before we start building ARMA model on our stationary series (which is equivalent to build ARIMA with $d = 1$ on the original time series), we need to make sure what our parameters p and q are. To do this, let's plot the first 70 ACF and PACF of our stationary series as Figure 5. From the plot, we can see $p = 1$ and $q = 1$ is roughly enough.


```
Results of Augmented Dickey-Fuller Test:
Test Statistic                    0.144093
p-value                           0.968946
#Lags Used                        0.000000
Number of Observations Used    2763.000000
Critical Value (1%)              -3.432719
Critical Value (5%)              -2.862587
Critical Value (10%)             -2.567327
```

Figure 3: Stationary Test for Bitcoin Time Series




```
Results of Augmented Dickey-Fuller Test:
Test Statistic                  -51.616279
p-value                           0.000000
#Lags Used                        0.000000
Number of Observations Used    2762.000000
Critical Value (1%)              -3.432720
Critical Value (5%)              -2.862587
Critical Value (10%)             -2.567327
dtype: float64
```

Figure 4: Stationary Test for Bitcoin Time Series after Difference

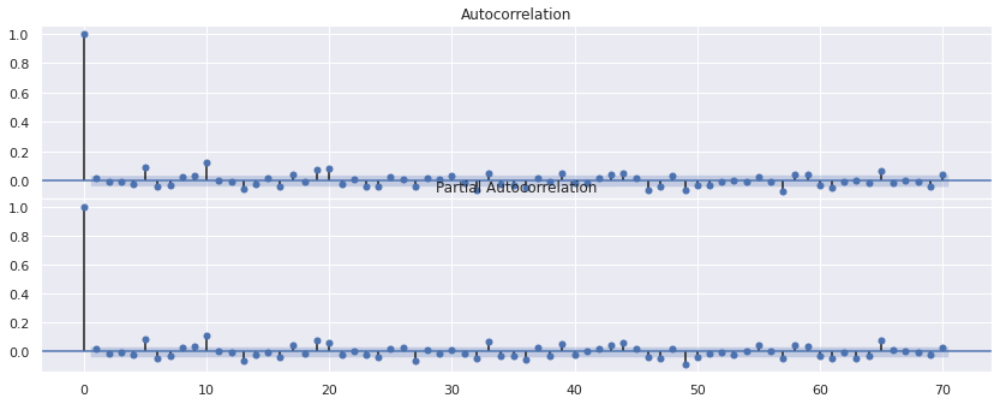


Figure 5: First 70 ACF and PACF

1.3.3 Modeling

By using the following Python codes, we build our ARIMA(1,1,1) Model. And Figure 6 is the regression results for Bitcoin. We can see all the p-values are significant(ignore the constant term).

Also, by following the same process, we build the ARIMA(6,1,6) Model for Ethereum. And Figure 7 is the regression results for it. Again, from the results, all the p-values are significant.

```
ts.index = pd.DatetimeIndex(ts.index.values,
                            freq=ts.index.inferred_freq)
model = ARIMA(ts, order=(1,1,1))
model = model.fit(disp=-1, method = 'mle')
```

2 Question2

Now lets evaluate the quality of ICO white papers, which plays a crucial role in the financing process. Two pieces of ICO white papers are provided here, with one from a success (Filecoin) and the other from a failure (Arqute). You are expected to analyse both documents.

2.1 Part a)

Please carefully read through two white papers and summarize the main differences in between.

2.1.1 Summarize the Main Differences

Filecoin wants to use the power of the blockchain to build a decentralized file storage service using free space on computer hard drives around the world. The Filecoin whitepaper demonstrates the technical proficiency of the founders. Theres a thorough explanation of the different incentive mechanisms that will keep the interaction between people storing data and storage providers running smoothly. The white paper

ARIMA Model Results

Dep. Variable:	D.Close	**No. Observations:**	2763
Model:	ARIMA(1, 1, 1)	**Log Likelihood**	-19287.783
Method:	mle	**S.D. of innovations**	260.291
Date:	Mon, 23 Nov 2020	**AIC**	38583.566
Time:	19:03:24	**BIC**	38607.263
Sample:	04-30-2013	**HQIC**	38592.126
	- 11-21-2020		

	coef	std err	z	P>|z|	[0.025	0.975]
const	6.6943	5.022	1.333	0.182	-3.148	16.536
ar.L1.D.Close	-0.7640	0.124	-6.140	0.000	-1.008	-0.520
ma.L1.D.Close	0.7888	0.118	6.675	0.000	0.557	1.020

Figure 6: Regression Results for Bitcoin

shows us the team is actively thinking about ways the storage ecosystem could be gamed and is working on provable solutions to these issues.

ARQUTE Global Animation Studio, the companionship which produces children animation content in the form of intellectual property, guaranteed by mass publication based on the blockchain. The decentralized Autonomous Organization is a new kind of organizational and legal form that can exist online and is based on blockchain technology. This technology regulates the lawful relationship of the producer and the authors of the animation product. The scheme is based on the science forbidden in physics about the luminiferous ether the all-penetrating average, vibrations that manifest themselves in the form of electromagnetic waves. Under different circumstances, heroes under the influence of ether receive supernormal powers, through which they strive with internal and outer enemies.

2.2 Part b)

Please apply textual analysis techniques (recall what we have learnt in Week 4) to derive two quantitative variables that can differentiate these two papers in view of readability and informativeness. [Note that two papers are given in the format of pdf. You may want to convert them into plaint texts with OCR in the first place.]

ARIMA Model Results

Dep. Variable:	D.Close	**No. Observations:**	1934
Model:	ARIMA(6, 1, 6)	**Log Likelihood**	-8357.180
Method:	mle	**S.D. of innovations**	18.206
Date:	Mon, 23 Nov 2020	**AIC**	16742.360
Time:	19:10:06	**BIC**	16820.303
Sample:	08-08-2015	**HQIC**	16771.028
	- 11-22-2020		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2851	0.466	0.612	0.541	-0.629	1.199
ar.L1.D.Close	1.1264	0.033	34.239	0.000	1.062	1.191
ar.L2.D.Close	-1.0620	0.062	-17.264	0.000	-1.183	-0.941
ar.L3.D.Close	0.8768	0.087	10.078	0.000	0.706	1.047
ar.L4.D.Close	-0.9990	0.082	-12.200	0.000	-1.159	-0.838
ar.L5.D.Close	1.0595	0.051	20.702	0.000	0.959	1.160
ar.L6.D.Close	-0.8738	0.030	-28.883	0.000	-0.933	-0.814
ma.L1.D.Close	-1.1240	0.037	-30.289	0.000	-1.197	-1.051
ma.L2.D.Close	1.0905	0.086	12.745	0.000	0.923	1.258
ma.L3.D.Close	-0.8713	0.125	-6.984	0.000	-1.116	-0.627
ma.L4.D.Close	0.9654	0.121	7.990	0.000	0.729	1.202
ma.L5.D.Close	-0.9414	0.079	-11.888	0.000	-1.097	-0.786
ma.L6.D.Close	0.8628	0.035	24.804	0.000	0.795	0.931

Figure 7: Regression Results for Ethereum

Figure 8: WordCloud for Filecoin

2.2.1 Convert pdf into plaint texts by using pdftotext

By using the following Python Codes, we convert the two pdf files into txt files.

```
import pdftotext
with open("Filecoin_whitepaper.pdf", "rb") as f:
    pdf = pdftotext.PDF(f)
with open('Filecoin.txt', 'w') as f:
    f.write("\n\n".join(pdf))
with open("Arqute_whitepaper.pdf", "rb") as f:
    pdf = pdftotext.PDF(f)
with open('Arqute.txt', 'w') as f:
    f.write("\n\n".join(pdf))
```

2.2.2 WordCloud

In this Part, we present their corresponding wordclouds as Figure 8 and Figure 9. Then we can see the Filecoin is mainly about Data Storage sort of things while Arqute is mainly focus on Studio Project sort of things.

2.2.3 Derive two quantitative variables that can differentiate these two papers in view of readability and informativeness.

Recall that fog index is the quality of being easy or enjoyable to read. So, we tried to calculate the fog index for them. By using textstat in Python, the fog index for Filecoin is: 16.87 while for Arqute is: 24.26.

Figure 9: WordCloud for Arqute

Since when fog index is more than or equal to 18, means the text is unreadable while when fog index is between 14 and 18, means the text is difficult to read. We can tell the readability of these two papers.