Gold Price Prediction Using Linear Regression and ARIMA Model

Name: Jinrong Zhang

Matric number: S2111224

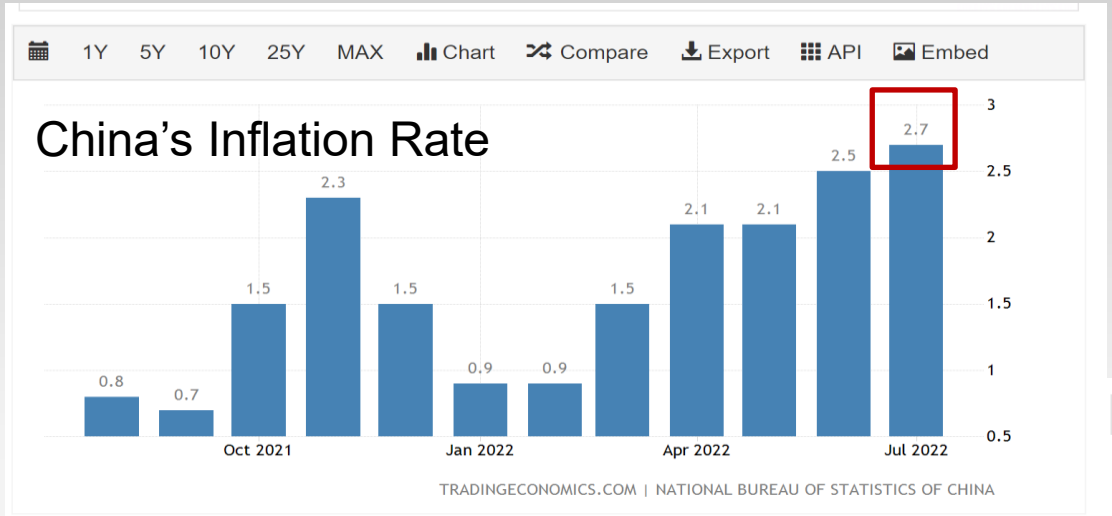
Supervisor: Assoc. Prof. Dr Ang Tan Fong

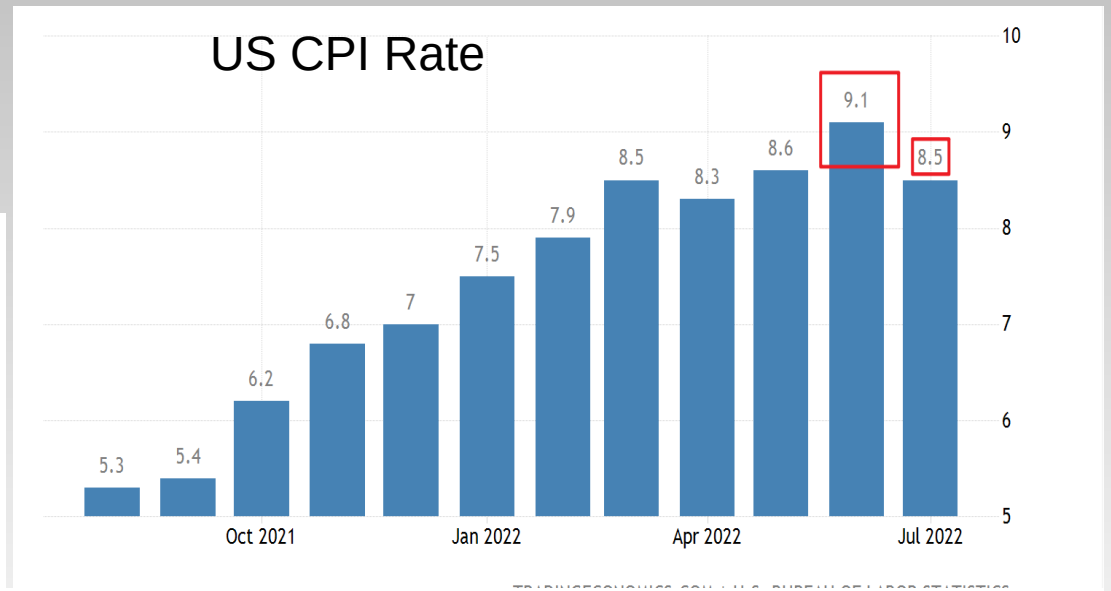
Contents

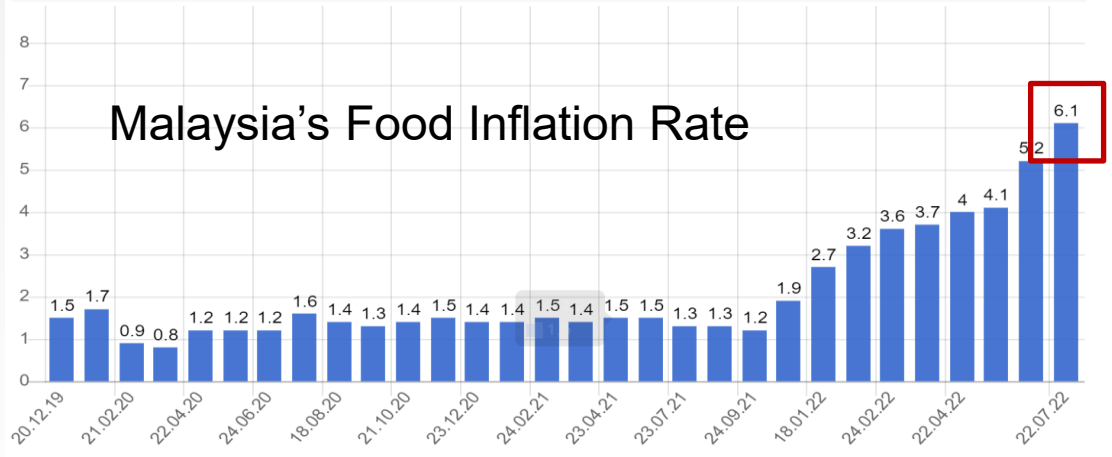
Introduction

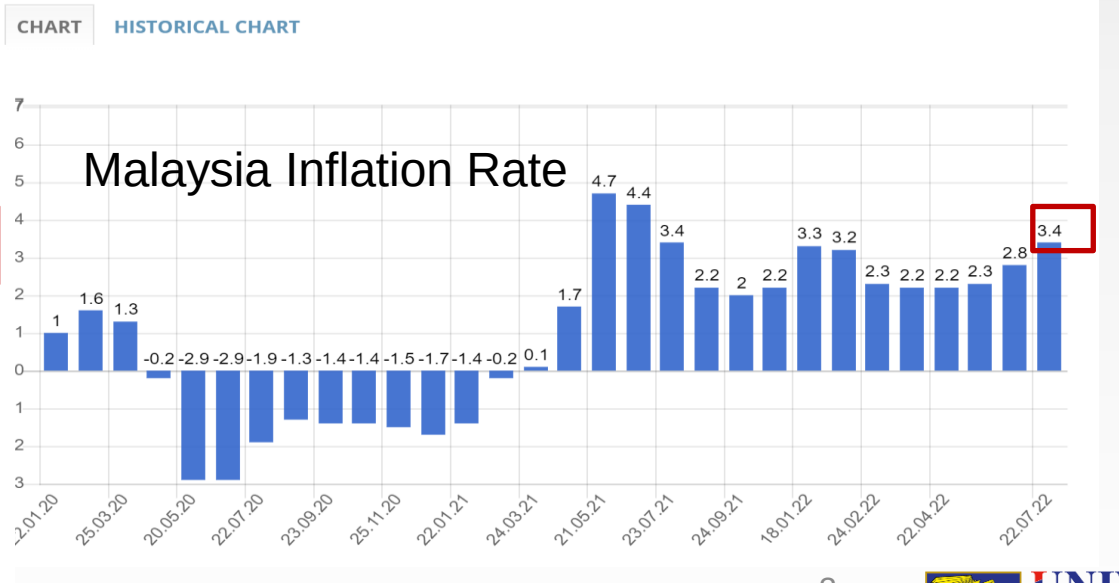
Background

Inflation Rate in US, Malaysia, and China









Reference:https://tradingeconomics.com/china/inflation-cpi

UNIVERSITI MALAYA

Introduction

Gold prices had a weak start to the year as rate hike expectations increased (in the US and other countries), **US real yields** rose (less negative), and the **US dollar** performed well.

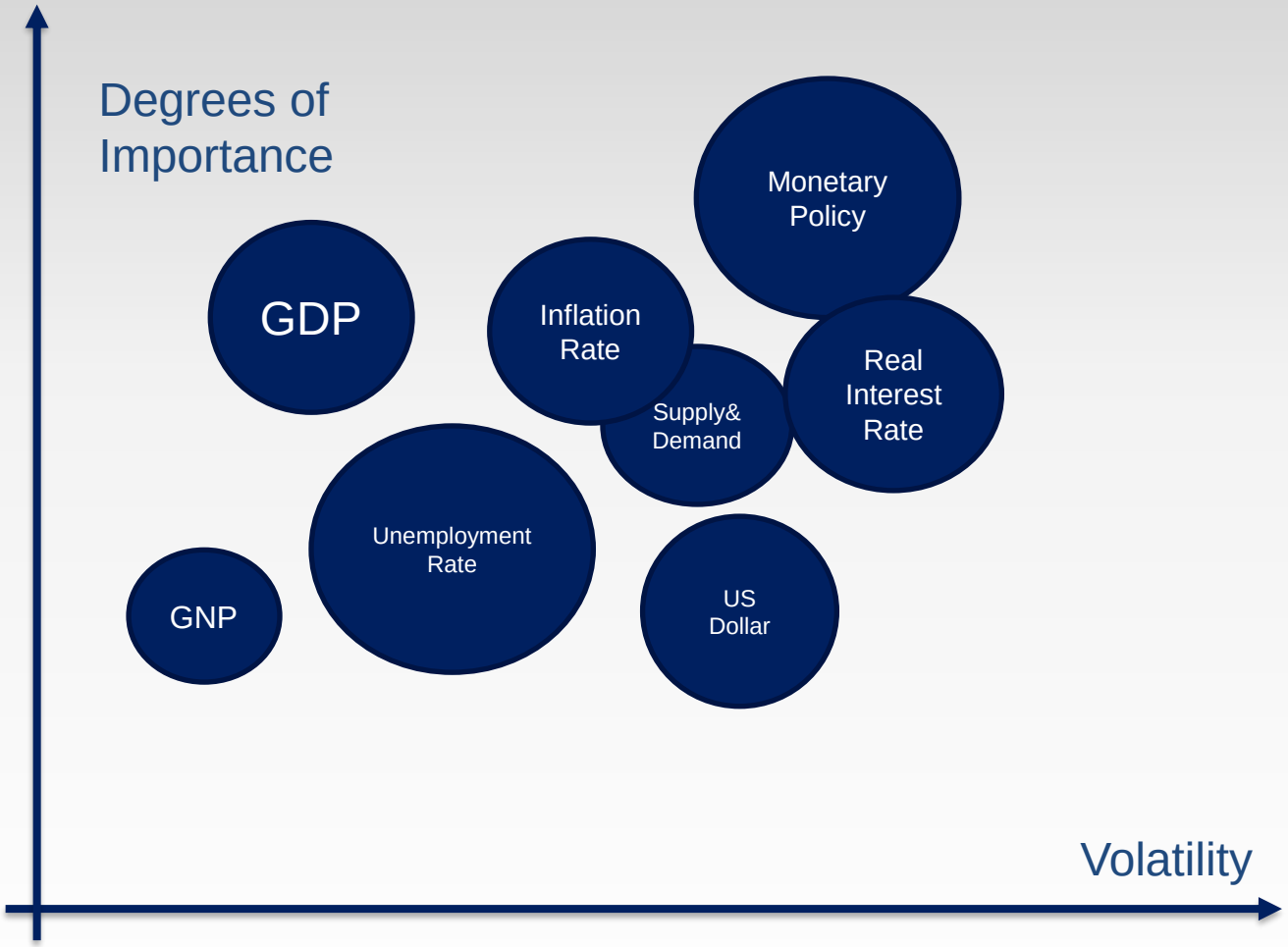
However, when the tensions between Russia and Ukraine became to the fore, gold prices started to bounce higher. Since the end of January, gold prices have rallied by **10%**.

Reference: https://jiaoyixia.com/

UNIVERSITI MALAYA

Introduction

Factors that impact gold price



Why Gold is important

20-40%

Central banks in countries value it so highly that most major central banks hold 20-40% of their foreign exchange reserves in it.

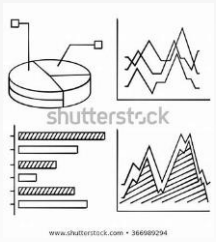
UNIVERSITI MALAYA

Problem statement

1. Gold price fluctuates over the course of a day, it is very difficult to forecast due to (factor)... Cite the paper

2. research q-(related to the model, two many
-There are too many models that used to forecast the gold price, however most of the models are (find disadvantages)-cite the paper

Linear Regression



ARIMA Model

Objectives

- To investigate the existing factors that affect the gold price.

- To develop a system that forecasts the price of gold using different models.

- To compare the accuracy of different models.

UNIVERSITI MALAYA

Research Questions

1. Does US dollar index have an impact on gold price?
2. What are the existing models that used to forecast gold price?
3. Which model can better forecast gold price?

UNIVERSITI MALAYA

Literature Review

Authors	Contents of Paper	Models	Pros	Cons
Syed Ali Raza, Nida Shah, Muhammad Ali, Muhammad Shahbaz.(2021)	This study examined the relationship between exchange rate return and gold price return association in G7 countries. The standard linear Granger causality test is applied, which shows that no causal association exists between exchange rate returns and gold prices.	Regression Model	the purpose of this study is to examine the gold price and exchange nexus in the context of G7 countries by using the nonparametric causality approach.	This study has some limitations pertaining to the sample period and size.
Jian Chai, Chenyu Zhao, Yi Hu, Zhe George Zhang--year	Based on the constructed SVAR model, the impulse response function and forecast error variance decompositionare used to analysis the dynamic relationship between gold, crude oil, the US dollar index and VIX.	SVAR	the study compared different prediction model for the gold price returns, the volatility of the US dollar index with less contribution is removed, then the gold price forecasting model is compared and analyzed through two perspectives.	Dimensionality problem.
Maria Immanuvel, and Daniel Lazar	The results reveal that average demand during the study period is highest for India, followed by China. The skewness is positive and near zero for India, and for other countries data have a slight variation.	VAR.	It is concluded from the empirical results that the international benchmark prices LBMA AM and PM fix prices are influenced by the gold demand of all the major goldconsuming countries.	from a certain aspects to predict gold price
Hanen Atri, Saoussen Kouki, Mohamed imen Gallali	The empirical results provide substantial evidence of the impact of the COVID-19 pandemic on commodity prices. We find that COVID-19 has opposite effects on oil and gold prices.	ARDL	This study is probably the first to examine the impact of the news, the panic and the media coverage of the major pandemic that rocked the world on commodity prices.	Given the novelty of the Covid-19 event and its significant implications for the global economy, we are limited by a short but sufficient study period.

Literature Review

Authors	Contents	Model Adopted	Pros	Cons
Maria Immanuvel, and Daniel Lazar(2022)	The results reveal that average demand during the study period is highest for India, followed by China. The skewness is positive and near zero for India, and for other countries data have a slight variation.	VAR.	It is concluded from the empirical results that the international benchmark prices LBMA AM and PM fix prices are influenced by the gold demand of all the major goldconsuming countries.	from a certain aspects to predict gold price
Hanen Atri, Saoussen Kouki, Mohamed imen Gallali(2021)	The empirical results provide substantial evidence of the impact of the COVID-19 pandemic on commodity prices. We find that COVID-19 has opposite effects on oil and gold prices.	ARDL	This study is probably the first to examine the impact of the news, the panic and the media coverage of the major pandemic that rocked the world on commodity prices.	Given the novelty of the Covid-19 event and its significant implications for the global economy, we are limited by a short but sufficient study period.
Syed Ali Raza, Nida Shah, Muhammad Ali, Muhammad Shahbaz.(2022)	This study examined the relationship between exchange rate return and gold price return association in G8 countries.the standard linear Granger causality test is applied, which shows that no causal association exists between exchange rate returns and gold prices.	Regression Model	the purpose of this study is to examine the gold price and exchange nexus in the context of G8 countries by using the nonparametric causality approach.	This study has some limitations pertaining to the sample period and size.
Yanhui Liang, Yu Lin, Qin Lu(2022)	This paper constructs a new hybrid model the ICEEMDAN-LSTM-CNN-CBAM to predict the gold futures and spot prices. ICEEMDAN, LSTM, CNN and CBAM are connected together to establish a cooperative relationship between the models. So the deep features can be extracted and the prediction accuracy can be improved.	ICEEMDAN-LSTM-CNN-CBAM(hybrid model)	This hybrid model can enhance the ability of model simulation.	The limitation of this paper is that there is no multivariable prediction and multi-step prediction.
Tirimisiyu F. Oloko, Ahamuefula E. Ogbonna, Abdulfatai A. Adedeji, Noman Lakhani(2021)	This study investigated gold price – inflation rate cointegrating persistence to analyze how the effect of shocks to gold price would persist on inflation persistence of developed and developing gold exporting countries.	CVAR model	The study drawn implications for inflation rate persistence, which distinguished it from earlier studies on gold price – inflation rate cointegrating relationship, which drawn evidence for hedging potential of gold in the face of high inflation.	The result shows that the gold price – inflation rate cointegrating persistence of countries adopting free-floating exchange rate regime is low (on the average of 0.176), while the gold price – inflation rate cointegrating persistence of countries adopting limited floating regime is high.

Literature Review

Authors	Contents	Model Adopted	Pros	Cons
Liya A, Qian Qin, Hafiz Waqas Kamran, Anusara Sawangchai, Worakamol Wisetsri, Mohsin Raza.(2021)	The research studies the macroeconomic factors that impact the gold price. These factors include GDP, GNP, inflation rate, the real interest rate, and unemployment. It generates a linear regression model by using these variables. The results show an optimistic and significant relationship between IR and gold prices.	Regression Model	The pros of the study is that it makes macroeconomic indicators an quantitative data to investigate, which are usually considered hard to explore.	The cons of this is that is didn't consider other factors such as US dollar and US stock market indexes.
Themba G. Chirwa, Nicholas M. Odhiambo(2020)	The study found that if structural breakpoints are considered in financial time series, parameter estimates are bound to be different for each breakpoint regime signifying the importance of accounting for multiple breakpoints in a time series. stocks have a significant positive association with gold price movements, and most stocks have gold mining companies listed.	ARDL	The investigation considered the identification of structural breakpoint regimes using methodology in the analysis to ensure parameter consistency in the time series data used before employing the ARDL-based error correction framework.	-
Syed Ali Raza, Nida Shah, Muhammad Ali, Muhammad Shahbaz.(2021)	This study examined the relationship between exchange rate return and gold price return association in G7 countries.the standard linear Granger causality test is applied, which shows that no causal association exists between exchange rate returns and gold prices.	Regression Model	the purpose of this study is to examine the gold price and exchange nexus in the context of G7 countries by using the nonparametric causality approach.	This study has some limitations pertaining to the sample period and size.
Jian Chai, Chenyu Zhao, Yi Hu, Zhe George Zhang(2021)	Based on the constructed SVAR model, the impulse response function and forecast error variance decompositionare used to analysis the dynamic relationship between gold, crude oil, the US dollar index and VIX.	SVAR	the study compared different prediction model for the gold price returns, the volatility of the US dollar index with less contribution is removed, then the gold price forecasting model is compared and analyzed through two perspectives.	Dimensionality problem.

UNIVERSITI MALAYA

Research Methodology

Data Description

1. The data used for research collected from **Meta-Trader 4.0**, a trading platform where forex brokers registered on the platform so that traders can order and view candle stick graphs through the software.

S&P 500	EURUSD
USIDX	GBPUSD
WTIUSD	USDCAD
XAGUSD	USDCHF
XAUUSD	USDJPY
DJ30	VIX

MetaTrader 4

Property	Type	Date	Time	Open	High	Low	Close	Tickvol
Metals	XAUUSD	2019.01.02	4:00:00	1282.21	1286.44	1282.01	1286.2	2377
Metals	XAUUSD	2019.01.02	8:00:00	1286.2	1288.7	1284.55	1286.98	8759
Metals	XAUUSD	2019.01.02	12:00:00	1286.98	1287.87	1283.27	1284.22	9373
Metals	XAUUSD	2019.01.02	16:00:00	1284.13	1287.63	1282	1284.02	12372
Metals	XAUUSD	2019.01.02	20:00:00	1283.92	1284.94	1278.64	1284.6	6390
Metals	XAUUSD	2019.01.03	0:00:00	1287.68	1290.13	1285.06	1288.27	5757
Metals	XAUUSD	2019.01.03	4:00:00	1288.27	1290.23	1287.53	1289.95	5676
Metals	XAUUSD	2019.01.03	8:00:00	1289.95	1292.26	1287.27	1288.03	11285
Metals	XAUUSD	2019.01.03	12:00:00	1288.03	1290.09	1284.5	1287.72	10842
Metals	XAUUSD	2019.01.03	16:00:00	1287.72	1291.3	1286.76	1290.36	17289
Metals	XAUUSD	2019.01.03	20:00:00	1290.36	1294.96	1290.03	1293.97	6469
Metals	XAUUSD	2019.01.04	0:00:00	1293.66	1298.52	1292.69	1296.92	7070
Metals	XAUUSD	2019.01.04	4:00:00	1296.93	1298.05	1293.91	1295.43	6442
Metals	XAUUSD	2019.01.04	8:00:00	1295.38	1295.73	1289.34	1290.06	7769
Metals	XAUUSD	2019.01.04	12:00:00	1290.17	1292.09	1285.77	1287.61	10055
Metals	XAUUSD	2019.01.04	16:00:00	1287.7	1287.7	1276.54	1284.3	23524

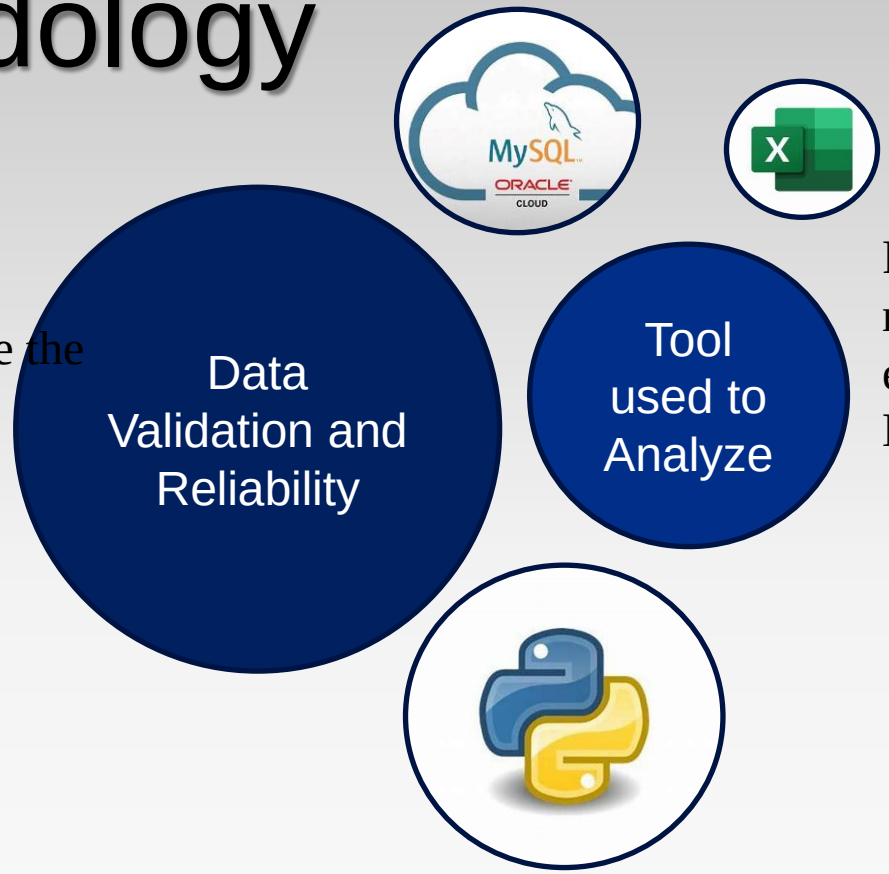
2. The researched time span ranges **from January 2019 to July 2022**. Besides, the data collected include a specific period during which the Russia-Ukraine war took place.

UNIVERSITI MALAYA

Research Methodology

Research Tools

I adopt two well-known measures to evaluate the accuracy of predictions which are the mean absolute error (MAE) and the mean squared error (RMSE). Smaller MAE and RMSE values give higher prediction accuracy, and small improvements in RMSE or MAE can significantly affect the quality of recommendations. Experimental results show that the datasets has high confidence.

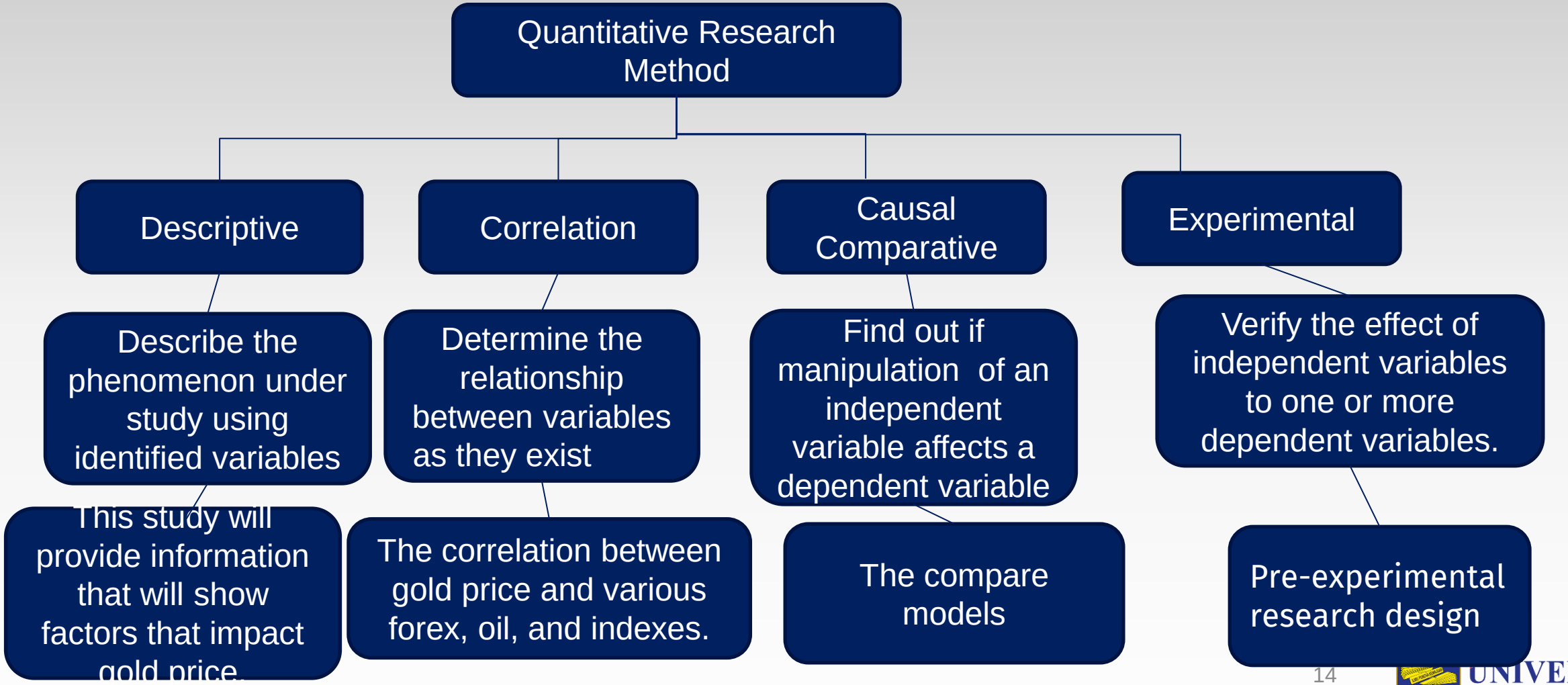


Python(numpy, pandas, matplotlib, etc.), Mysql, Microsoft Excel

UNIVERSITI MALAYA

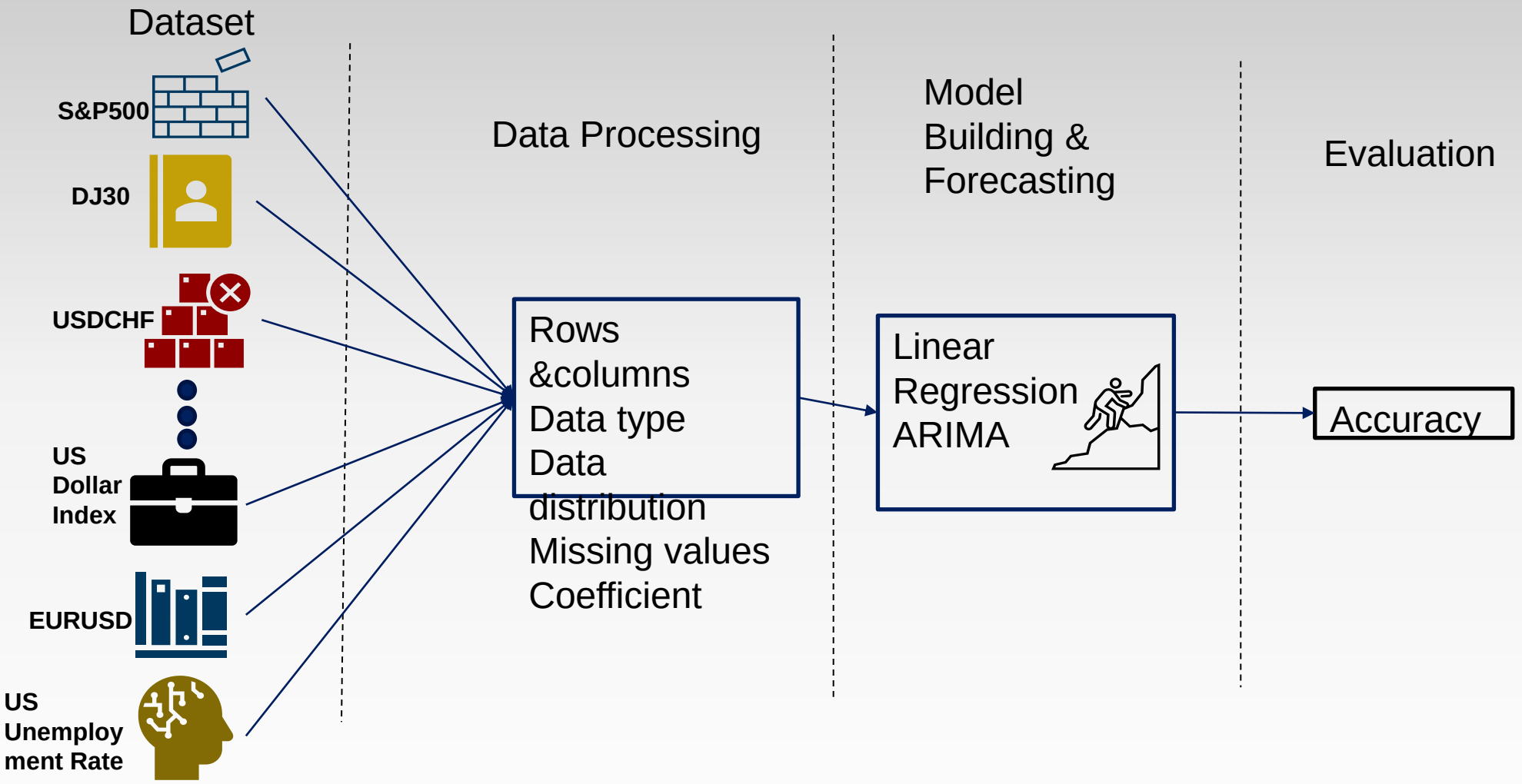
Research Methodology

Quantitative &Qualitative Research Methods



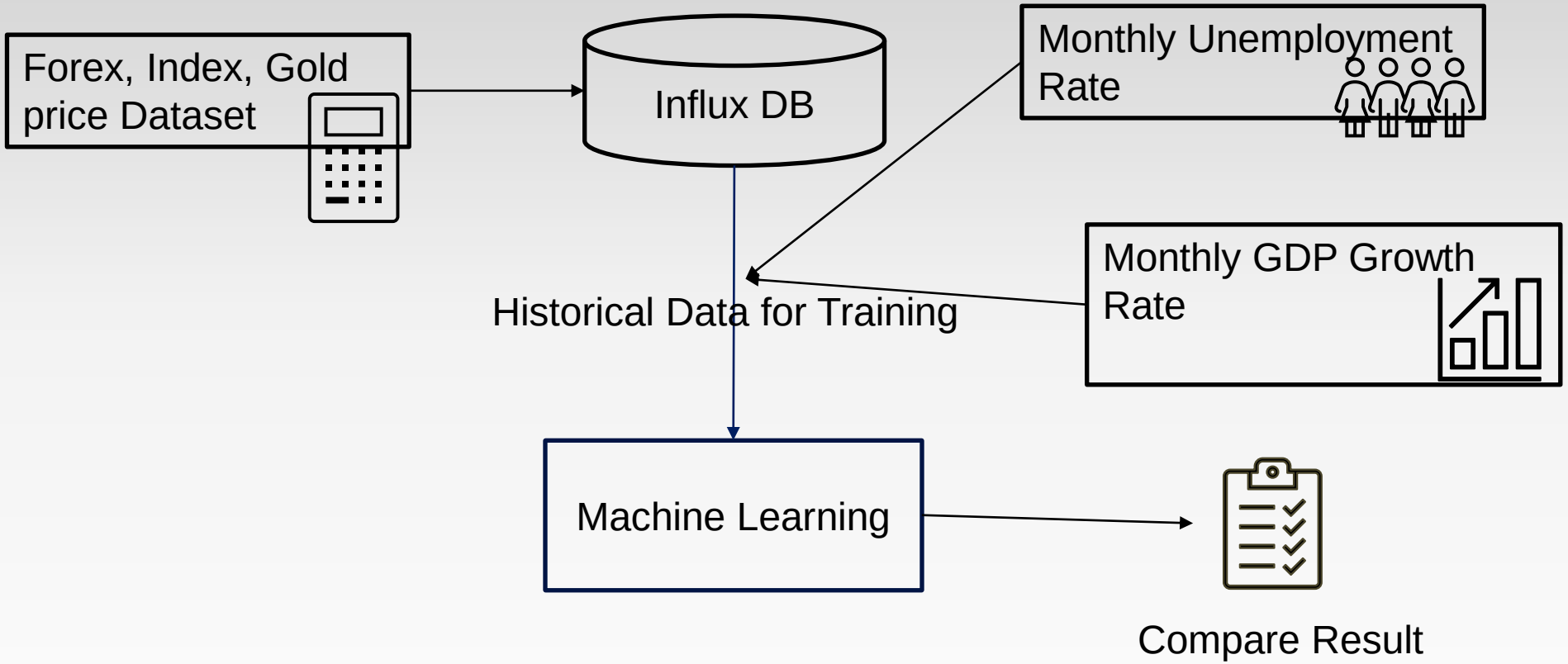
UNIVERSITI MALAYA

System Architecture



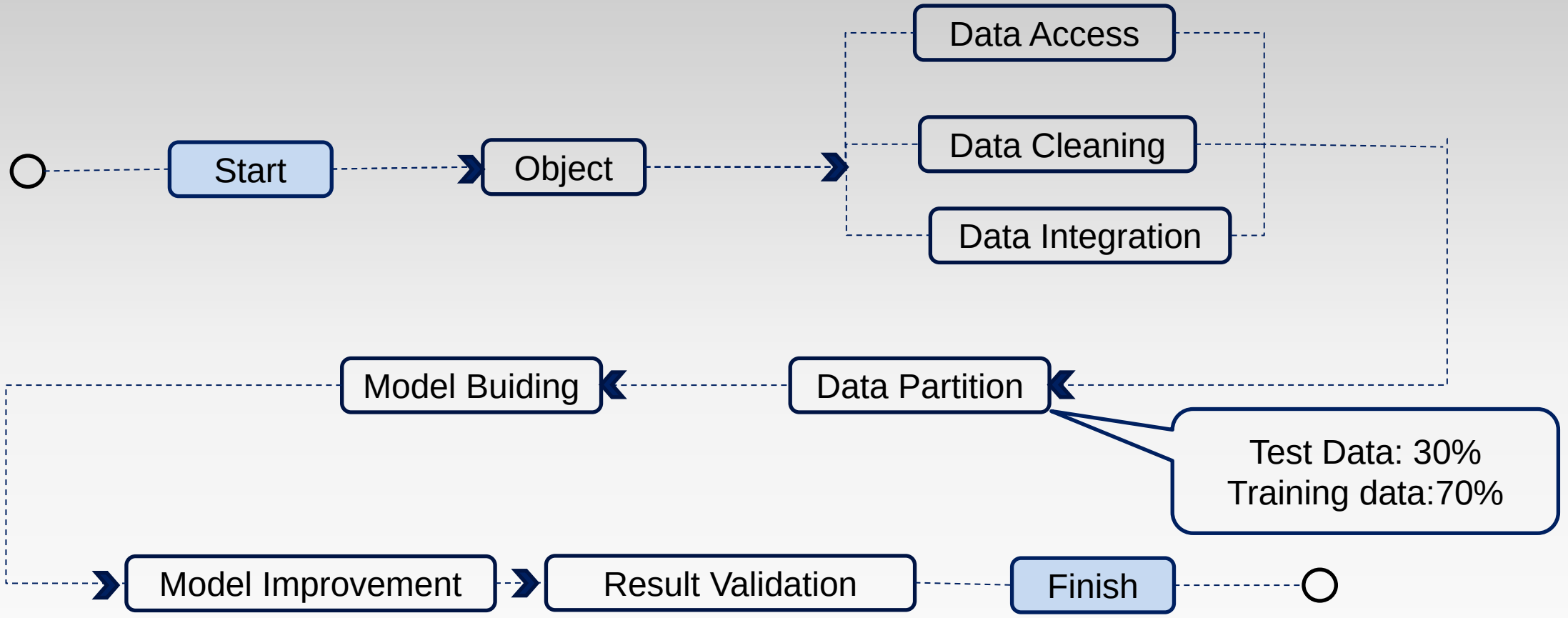
UNIVERSITI MALAYA

Model Design / Algorithm



UNIVERSITI MALAYA

Experiment Setup



Dataset

The dataset change into using daily candle data ranging from March 26th 2018 to October 14th 2022

Daily


```
Data columns (total 15 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   TIME SERIES    1177 non-null   datetime64[ns]
 1   XAUUSD_CLOSE   1177 non-null   float64
 2   DJ30_CLOSE     1177 non-null   int64
 3   USIDX_CLOSE    1177 non-null   float64
 4   EURUSD_CLOSE   1177 non-null   float64
 5   GBPUSD_CLOSE   1177 non-null   float64
 6   SP500_CLOSE    1177 non-null   float64
 7   USDCAD_CLOSE   1177 non-null   float64
 8   USDCHF_CLOSE   1177 non-null   float64
 9   USDJPY_CLOSE   1177 non-null   float64
 10  WTIUSD_CLOSE   1177 non-null   float64
 11  XAGUSD_CLOSE   1177 non-null   float64
 12  month          1177 non-null   int64
 13  year           1177 non-null   int64
 14  month_year     1177 non-null   period[M]
dtypes: datetime64[ns](1), float64(10), int64(3), period[M](1)
memory usage: 138.1 KB
```

	Unnamed: 0	Time Serie	XAUUSD_Close	DJ30_Close	USIDX_Close	EURUSD_Close	\
0	0	2018/3/27	1344.68	23867	88.91	1.22785	
1	0	2018/3/28	1324.76	23889	89.70	1.22394	
2	0	2018/3/29	1325.15	24117	89.68	1.22810	
3	0	2018/4/2	1341.29	23639	89.61	1.23209	
4	0	2018/4/3	1332.77	24020	89.76	1.23550	

	GBPUSD_Close	S&P500_Close	USDCAD_Close	USDCHF_Close	USDJPY_Close	\
0	1.40800	2607.2	1.27644	0.96080	106.773	
1	1.40044	2636.3	1.27517	0.96347	107.385	
2	1.40910	2582.1	1.27742	0.95917	106.942	
3	1.41319	2613.6	1.26965	0.95617	106.747	
4	1.41750	2645.7	1.25988	0.95693	107.200	

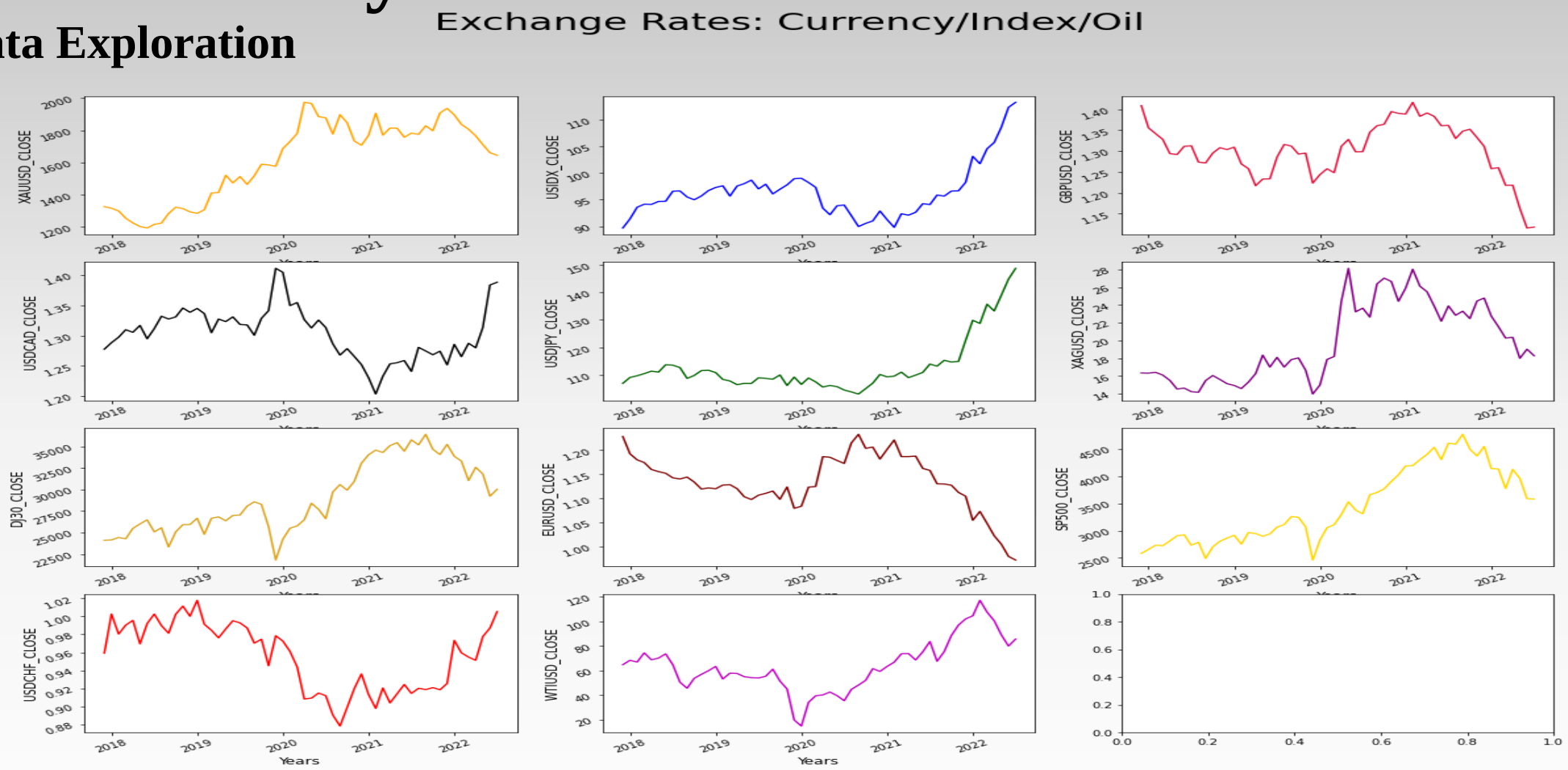
	WTIUSD_Close	XAGUSD_Close
0	64.62	16.509
1	64.56	16.274
2	64.82	16.340
3	62.98	16.575
4	63.52	16.396

- Date are divided into three parts, which are year, month, and month-year.
- All the prices use the daily closing prices to reflect the changes.

UNIVERSITI MALAYA

Result Analysis

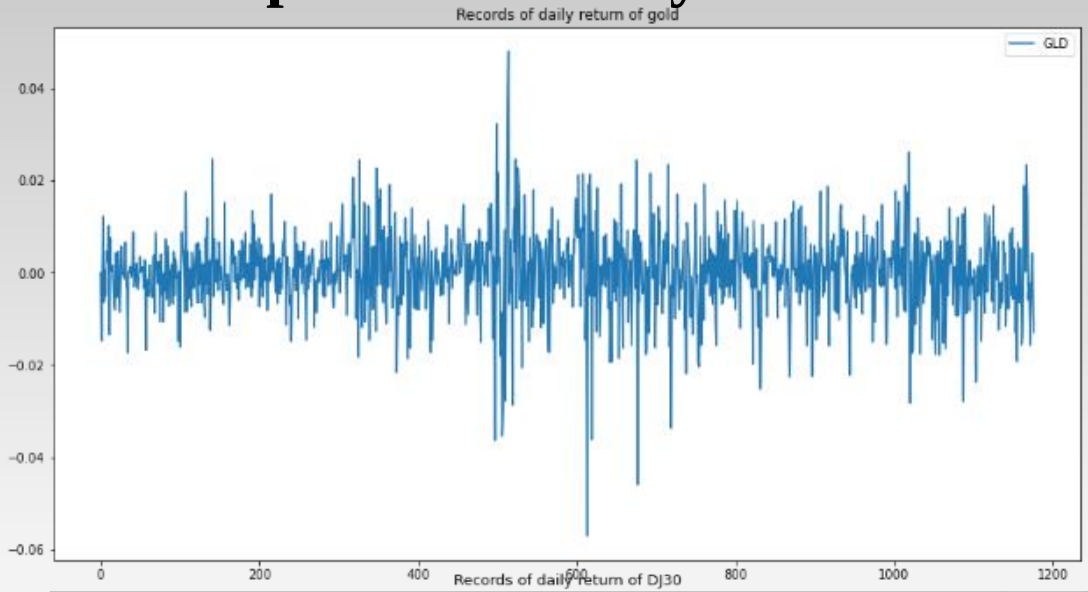
Data Exploration

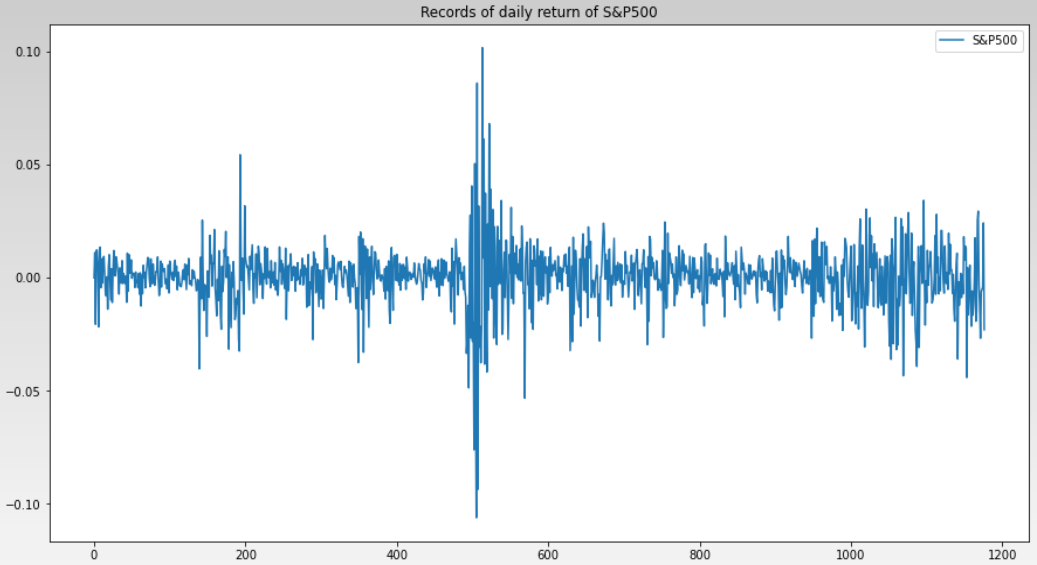


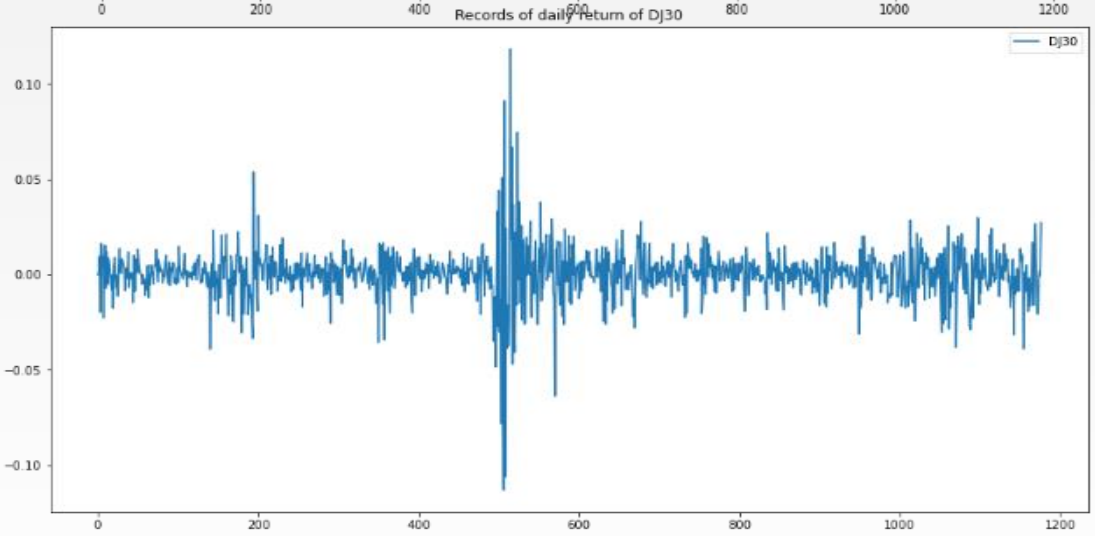
UNIVERSITI MALAYA

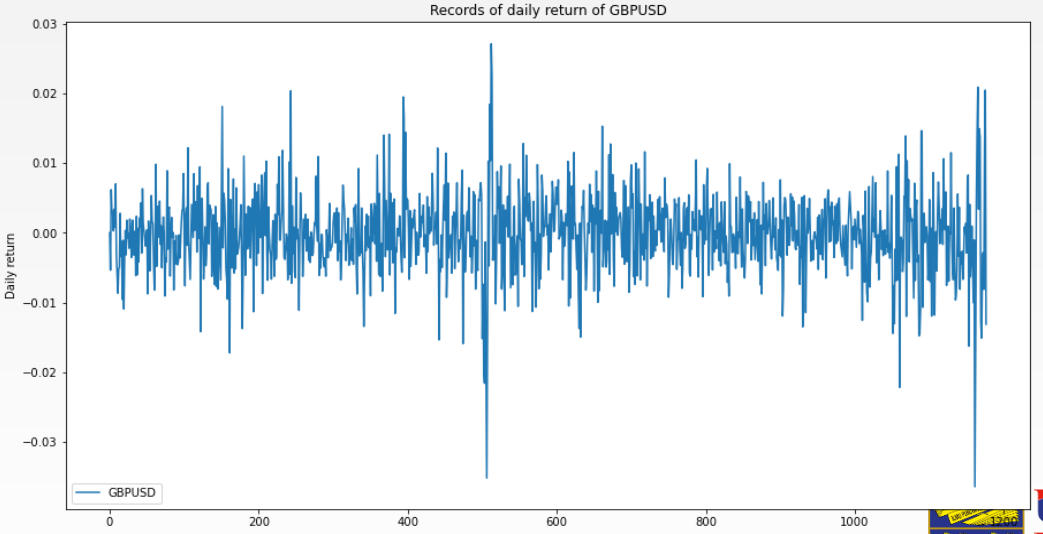
Result Analysis

Data Exploration- Daily Return of Variables





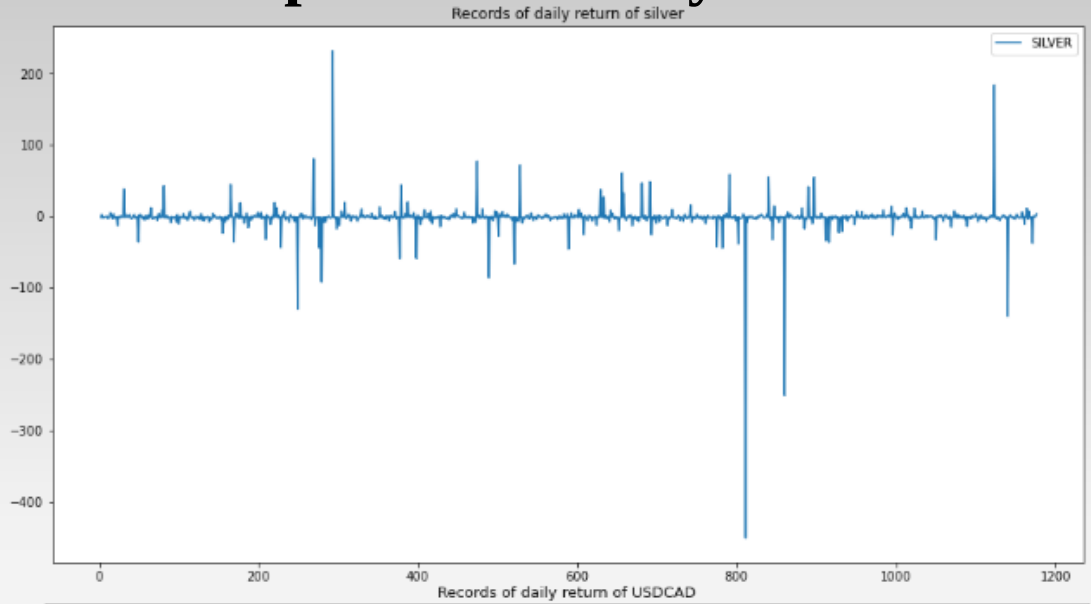


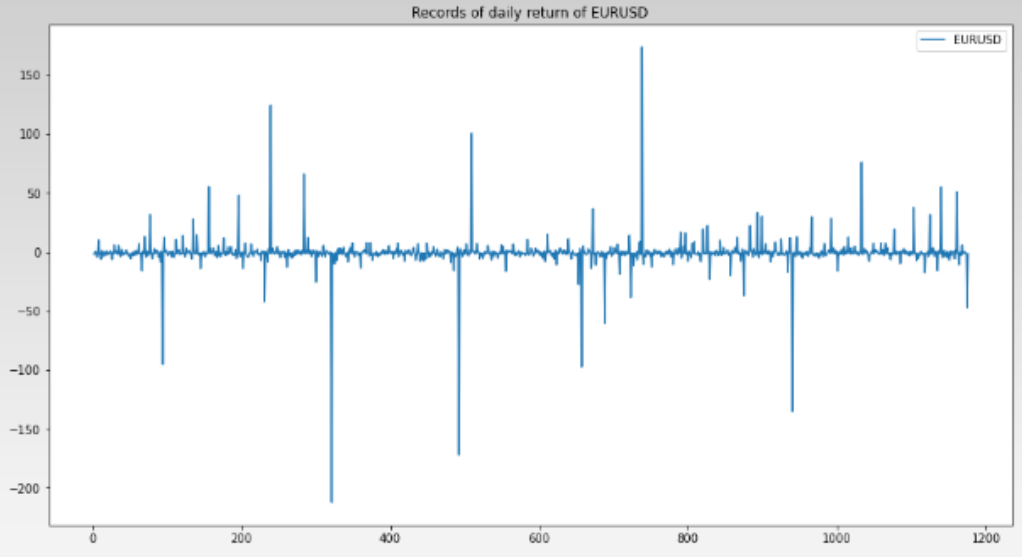


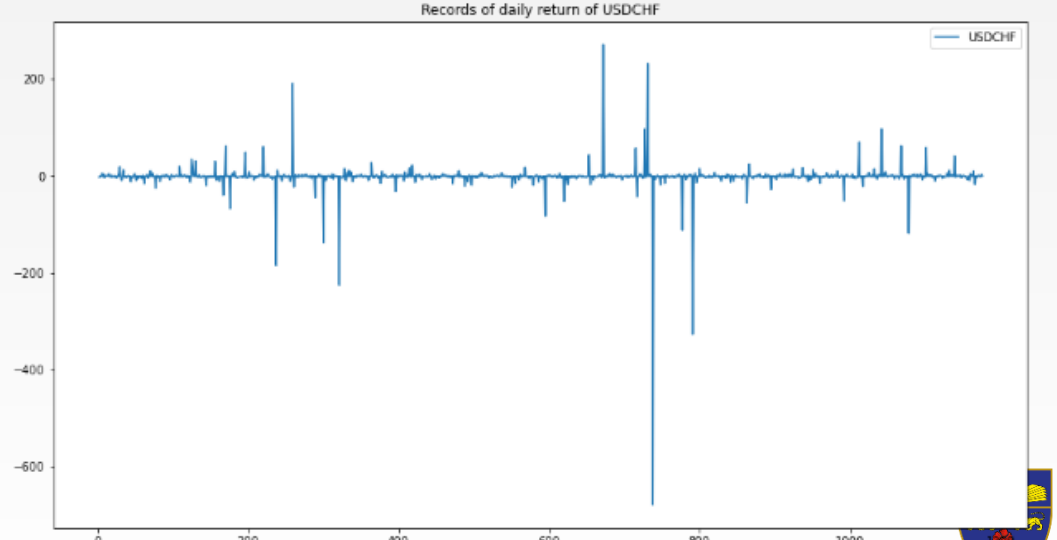
UNIVERSITI MALAYA

Result Analysis

Data Exploration- Daily Return of Variables



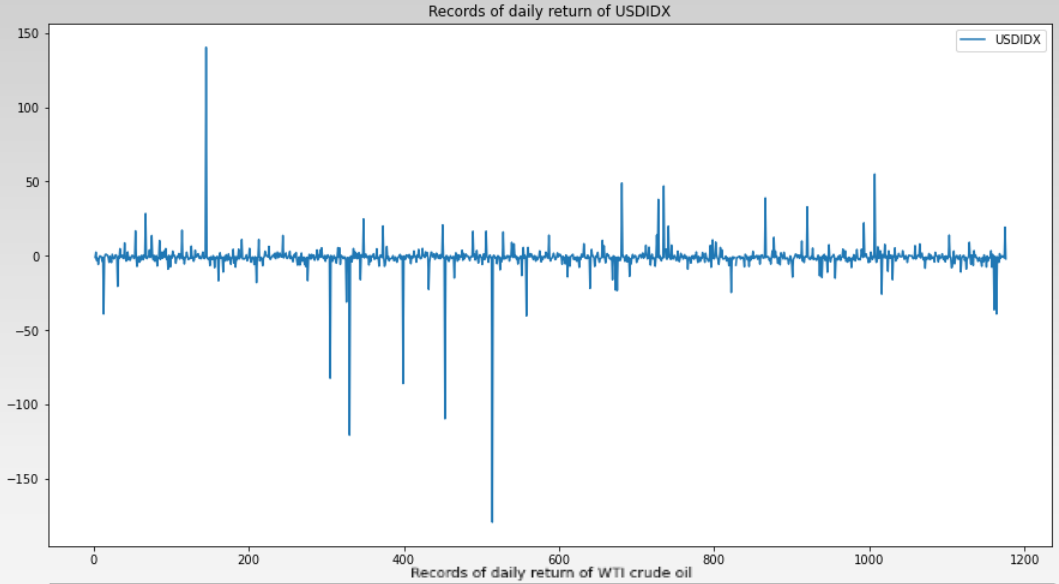


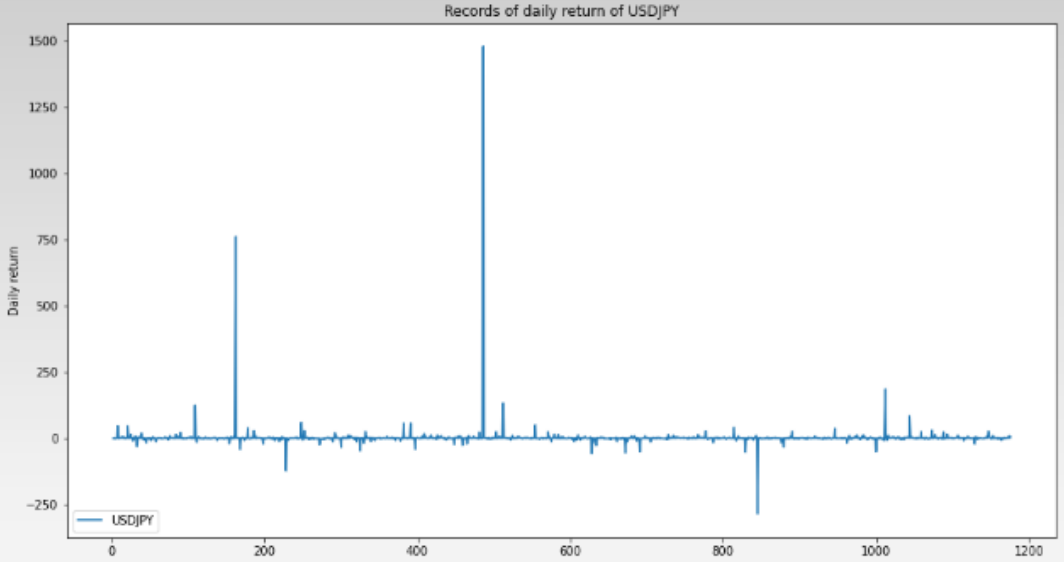


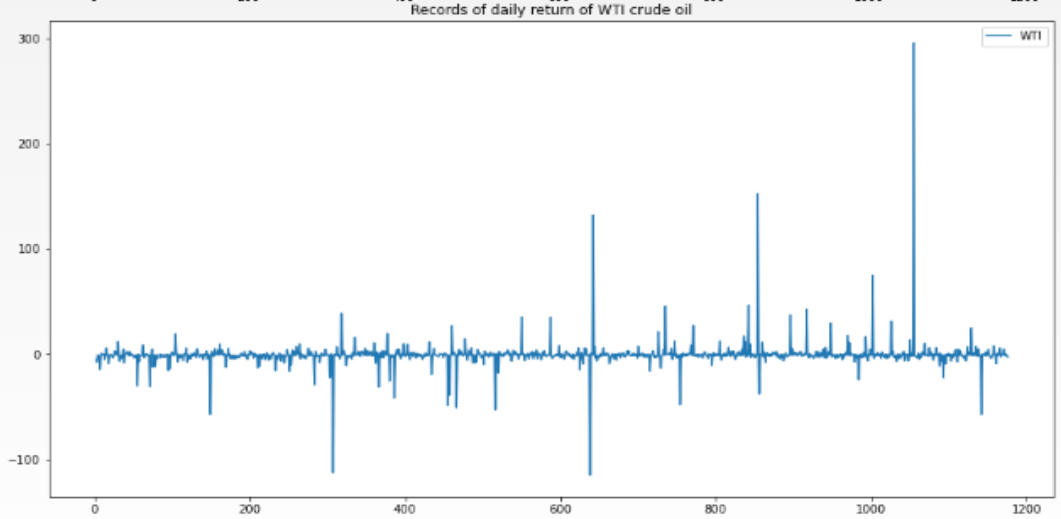
UNIVERSITI MALAYA

Result Analysis

Data Exploration- Daily Return of Variables



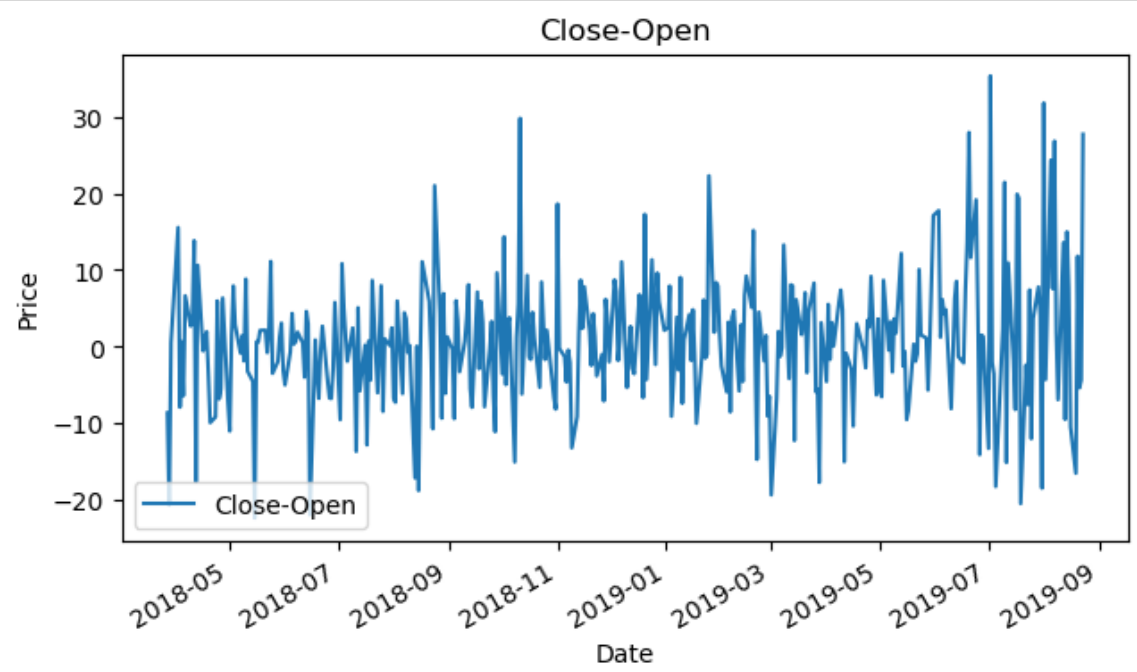


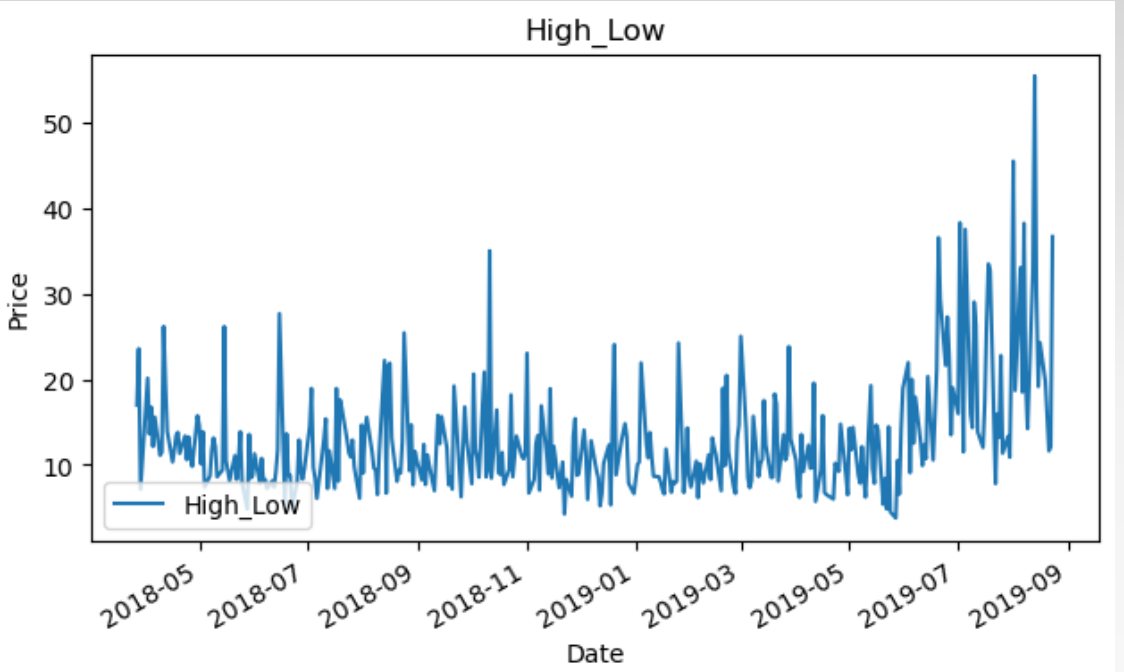


UNIVERSITI MALAYA

Result Analysis

Data Exploration-Fluctuation of Gold

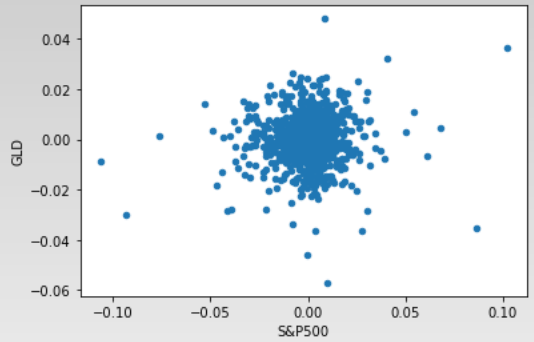


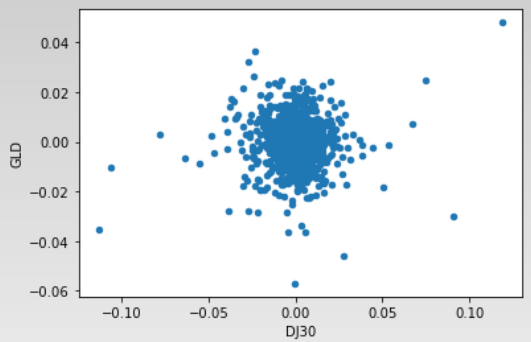


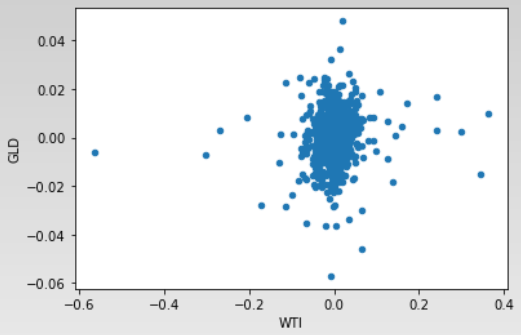
UNIVERSITI MALAYA

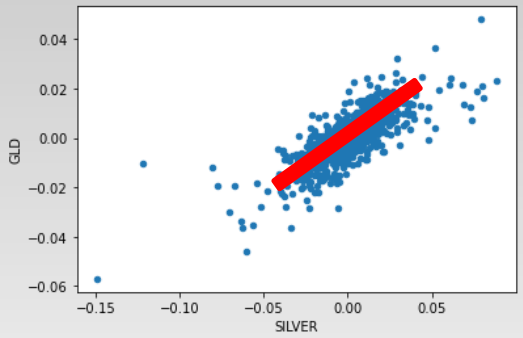
Result Analysis

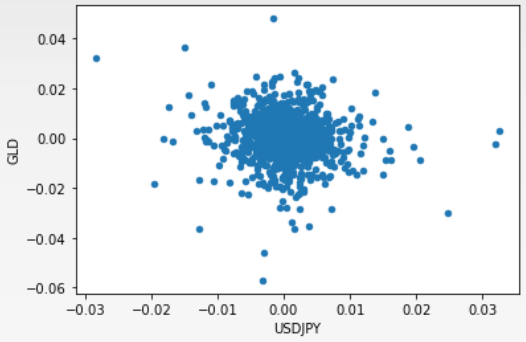
Data Exploration-Scatter Plot

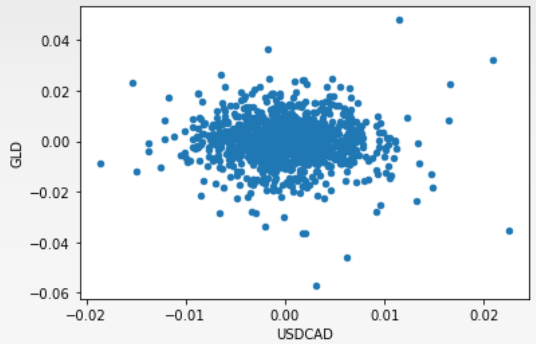


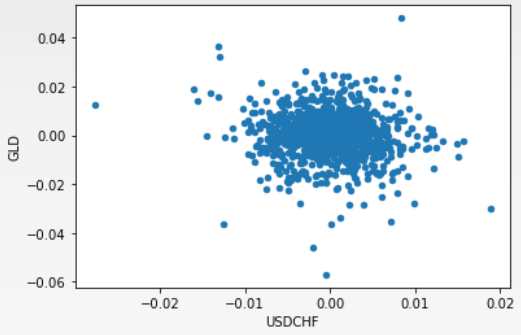


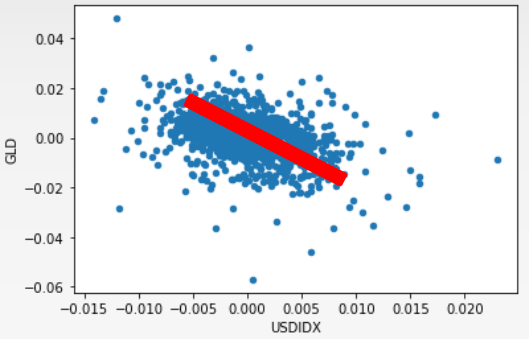


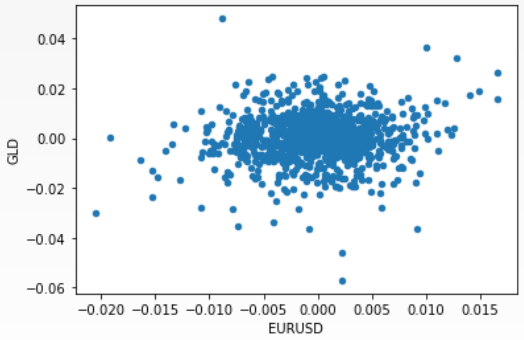


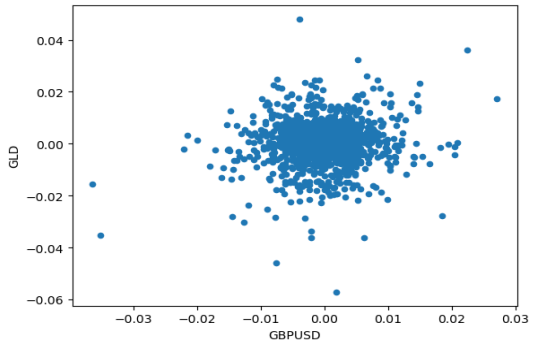








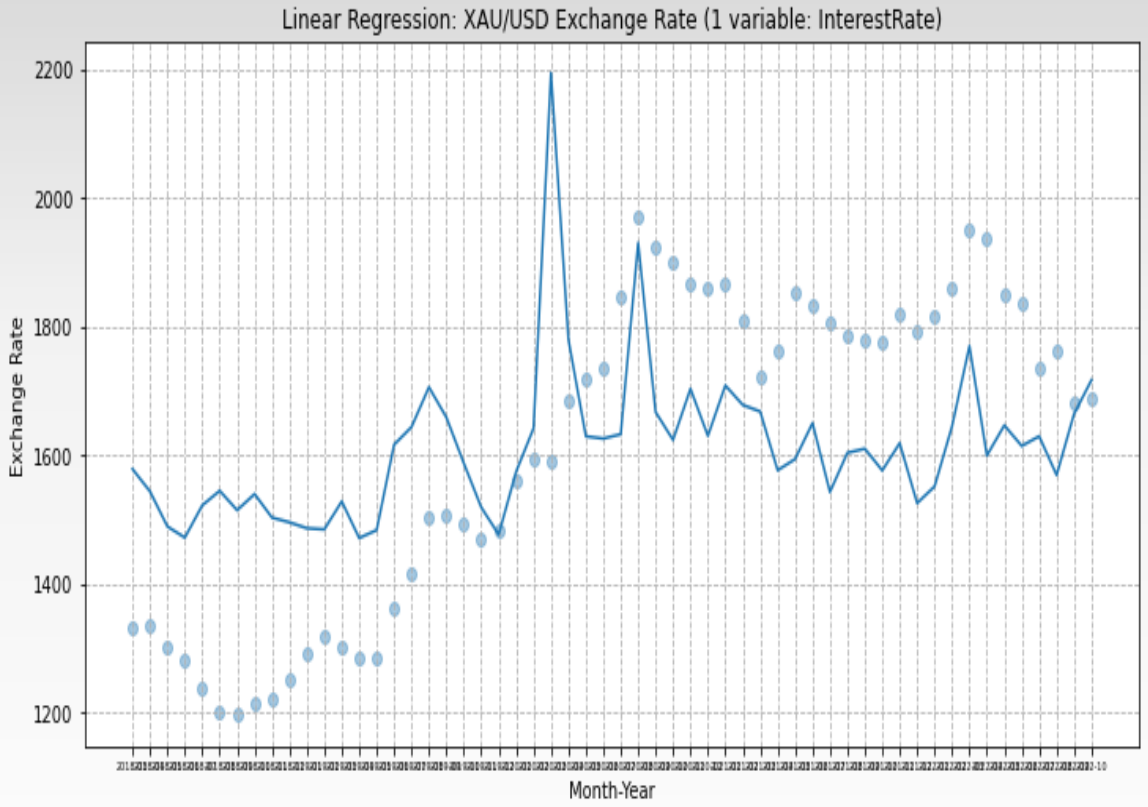


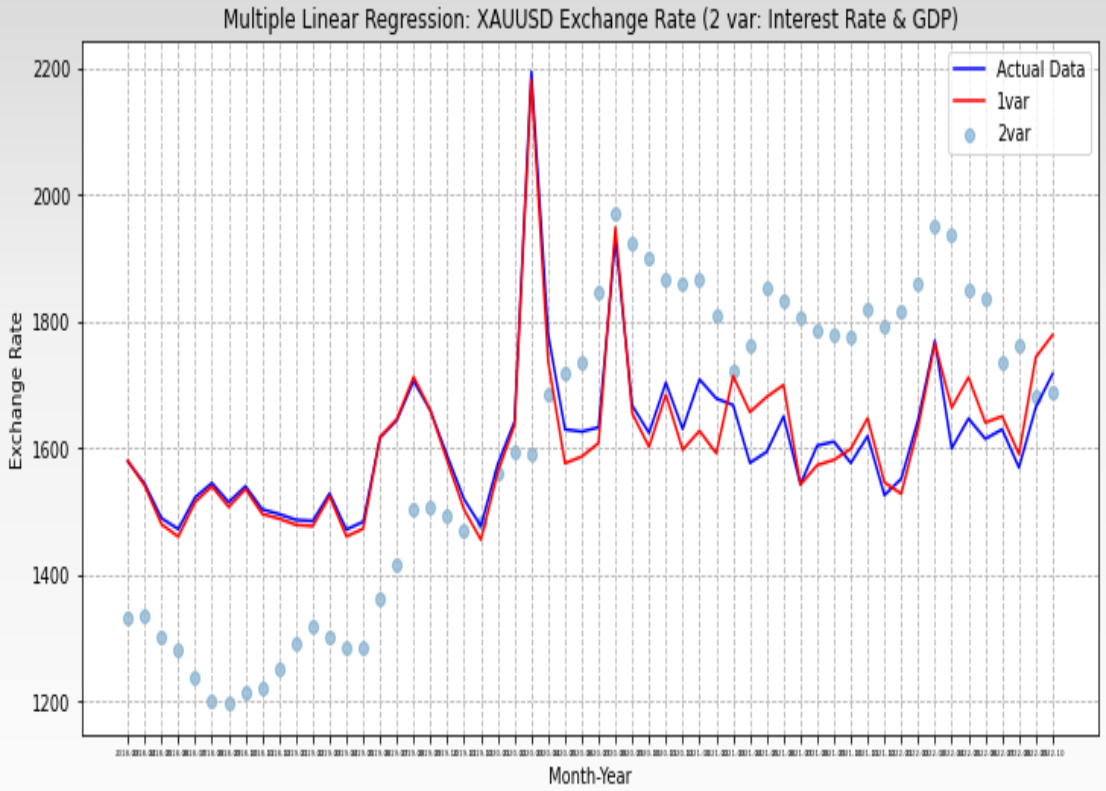


UNIVERSITI MALAYA

Result Analysis

Linear Regression Implementation

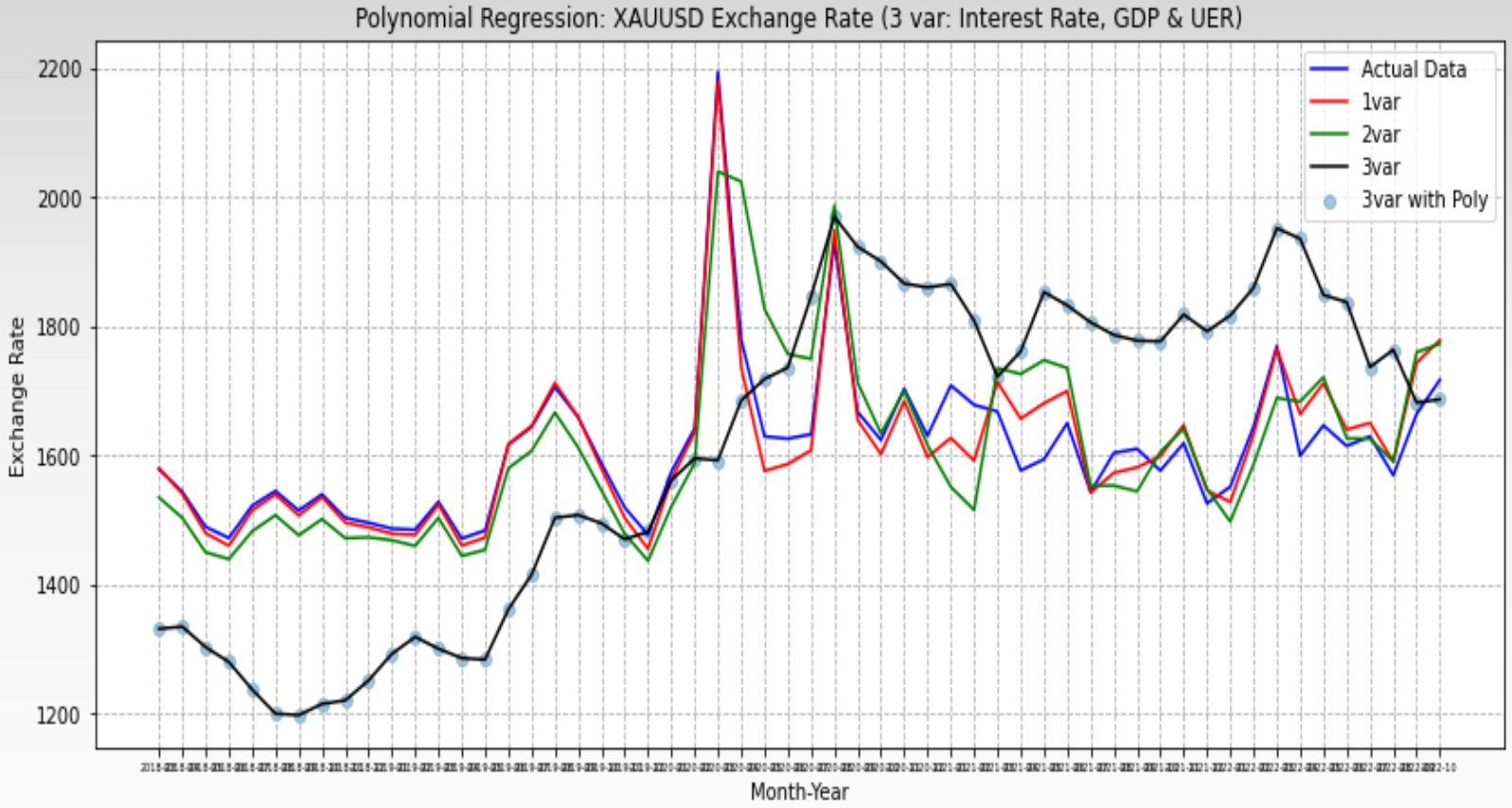




UNIVERSITI MALAYA

Result Analysis

Linear Regression Implementation



UNIVERSITI MALAYA

Result Analysis

ARIMA Model Implementation

Stationary Test

Adfuller

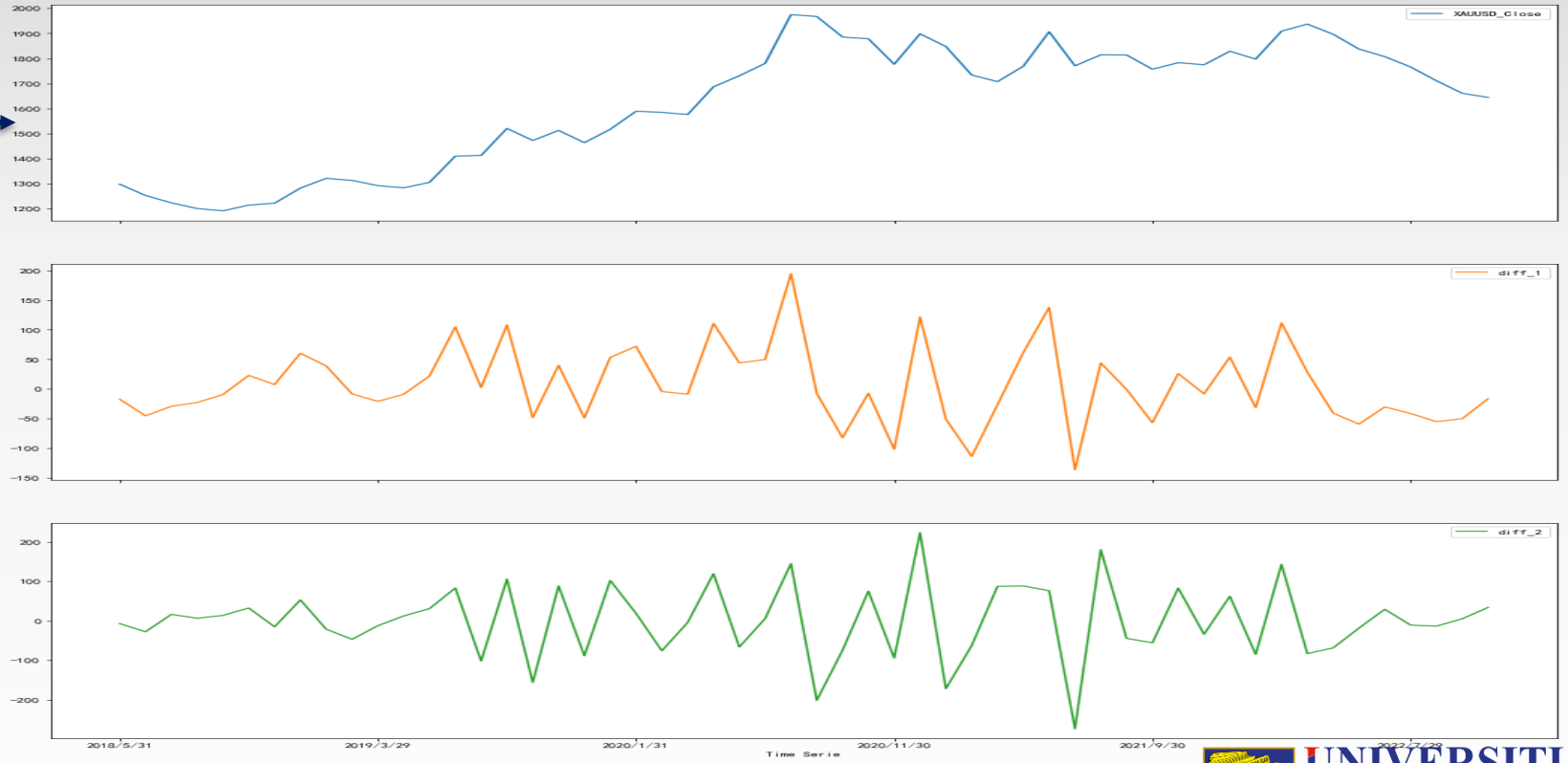
(-1.3052683213610279, 0.6267586166796115, 0, 55, {'1%': -3.5552728880540942, '5%': -2.9157312396694217, '10%': -2.5956695041322315}, 498.574
58490913194)

Diff_1

(-7.438280993635395, 6.100830050008523e-11, 0, 54, {'1%': -3.55770911573439, '5%': -2.9167703434435808, '10%': -2.59622219478738} 492.16896
827271995)

Diff_2

(-6.447790863736818, 1.5502880861470936e-08, 3, 50, {'1%': -3.568485864, '5%': -2.92135992, '10%': -2.5986616}, 492.3090072866694)



UNIVERSITI MALAYA

Result Analysis

ARIMA Model Implementation

White Noise Test

White noise test is also known as pure randomness test

P-value: the smaller the better

	lb_stat	lb_pvalue
1	0.057478	0.810527
2	0.086567	0.957640
3	0.092347	0.992740
4	0.119740	0.998278
5	0.802504	0.976873
6	2.168063	0.903611
7	2.304207	0.941105
8	3.414719	0.905707
9	3.473593	0.942534
10	4.723424	0.908871
11	4.723673	0.943807
12	5.431182	0.942007
13	6.448389	0.928302
14	6.570113	0.950017
15	7.105478	0.954656
16	7.324545	0.966487

Diff_1

	lb_stat	lb_pvalue
1	15.948201	0.000065
2	16.003637	0.000335
3	16.009784	0.001129
4	16.351603	0.002582
5	18.386178	0.002499
6	21.237395	0.001663
7	23.337652	0.001488
8	24.937045	0.001593
9	24.960327	0.003015
10	25.936684	0.003826
11	26.916728	0.004730
12	28.652815	0.004435
13	30.197961	0.004410
14	30.502713	0.006503
15	31.096731	0.008527
16	32.483613	0.008645

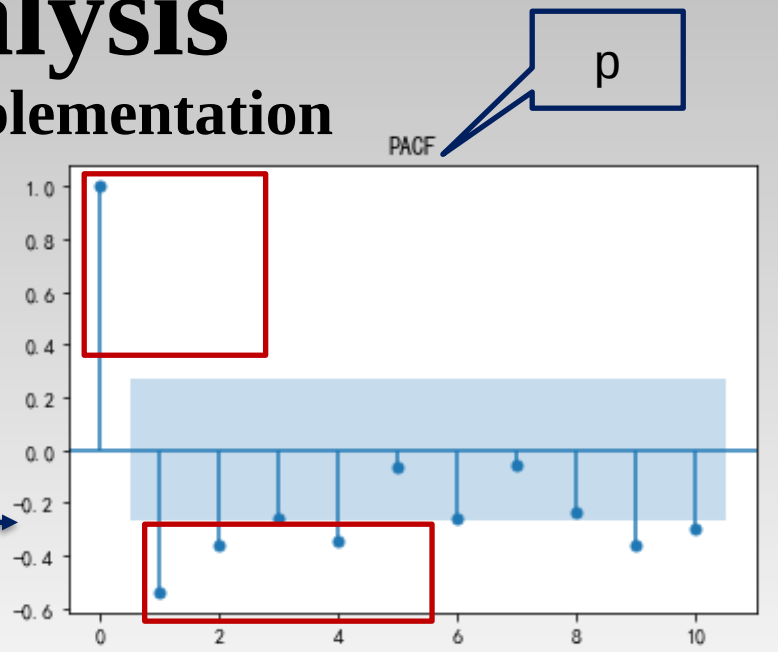
Diff_2

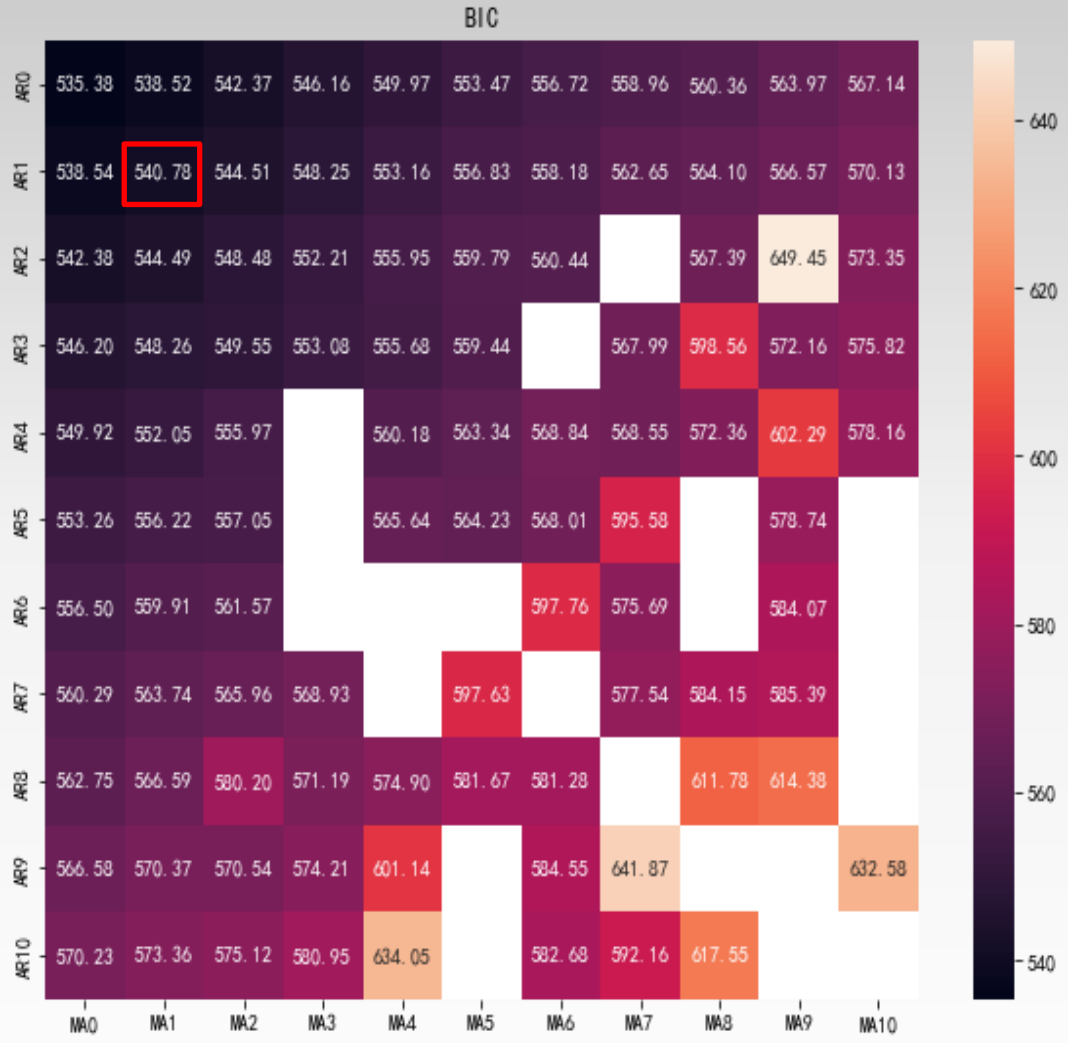
UNIVERSITI MALAYA

Result Analysis

ARIMA Model Implementation

Parameter Selection





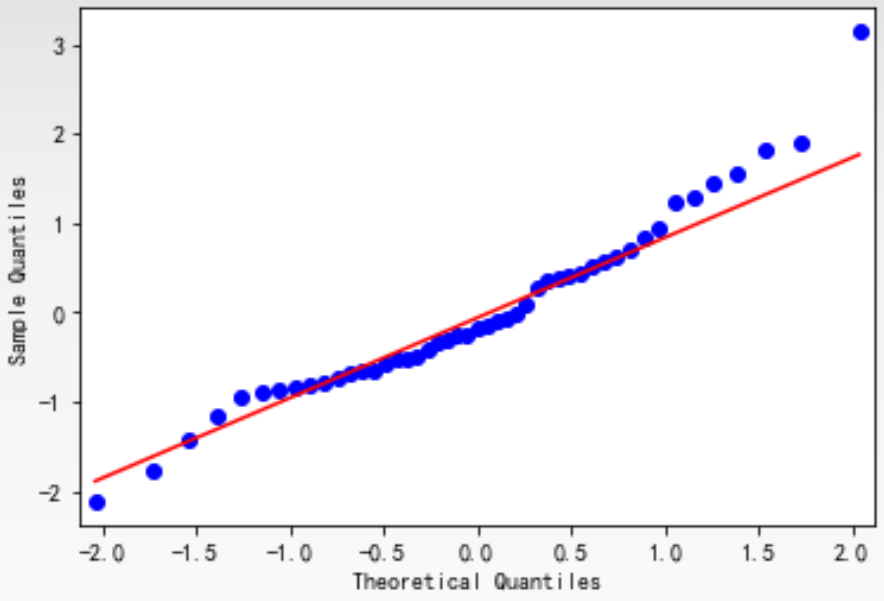
The best parameter for p is 1, and q is 1.

UNIVERSITI MALAYA

Result Analysis

ARIMA Model Implementation

Residual Test



Normality test (shapiro)

ShapiroResult(statistic=0.9576875567436218, pvalue=0.08140654116868973)

Autocorrelation Test (Durbin-Watson)

1.6867592419226185 ≈ 2

residual white noise test

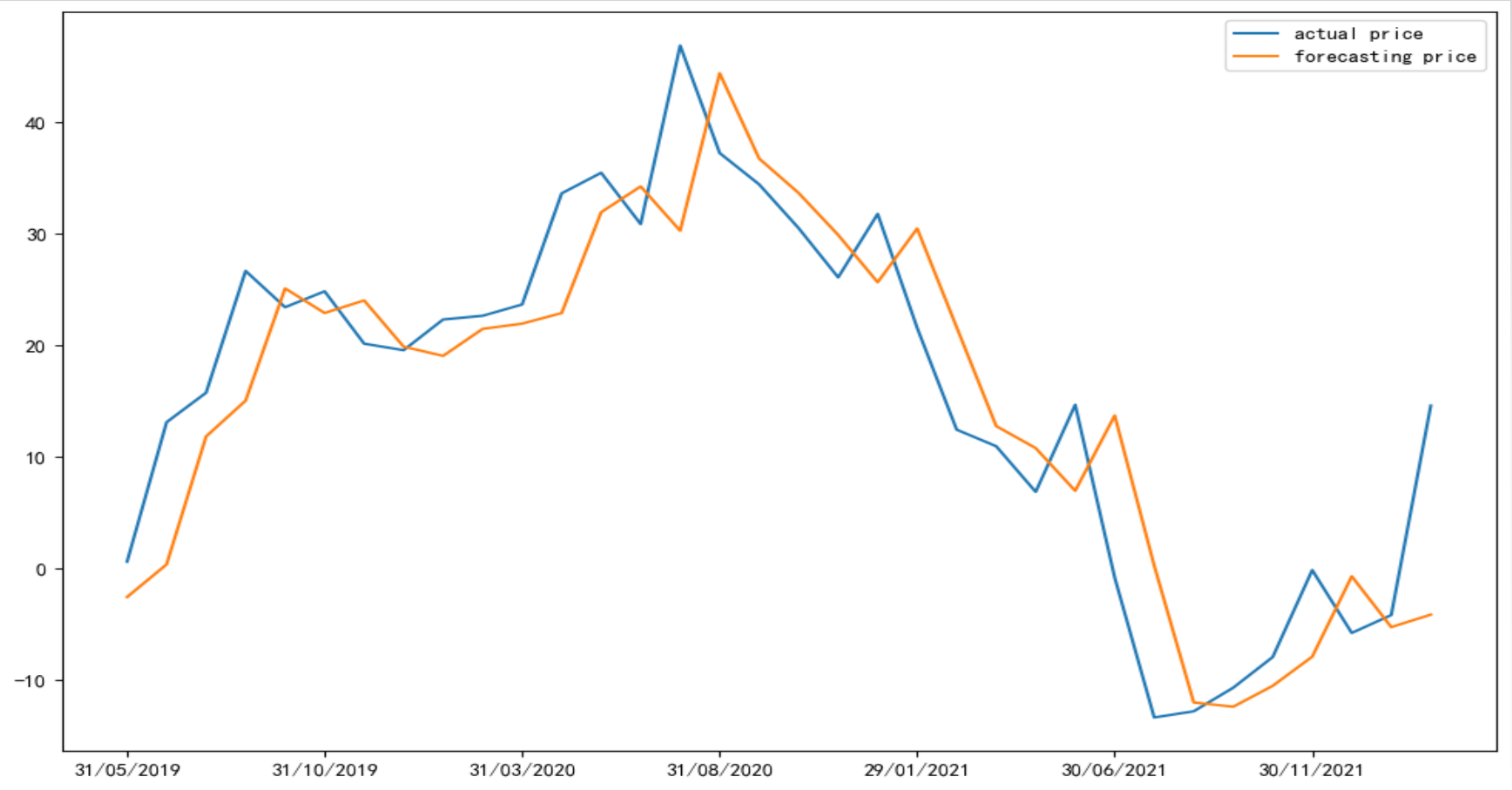
lag	AC	Q	Prob(>Q)
1.0	0.000147	0.000001	0.999162
2.0	0.041355	0.089232	0.956365
3.0	0.053046	0.239308	0.971007
4.0	0.030106	0.288746	0.990529
5.0	0.127408	1.194756	0.945379
6.0	−0.049747	1.336173	0.969626
7.0	0.035637	1.410514	0.985247
8.0	−0.099271	2.001798	0.980957
9.0	0.001475	2.001932	0.991437
10.0	0.048345	2.149549	0.995060
11.0	0.003014	2.150139	0.997895
12.0	0.123193	3.161914	0.994289
13.0	−0.052211	3.348837	0.996354
14.0	−0.066134	3.657571	0.997198
15.0	−0.195961	6.450355	0.971150
16.0	0.056230	6.687495	0.978825

Residual passes the white noise test.

UNIVERSITI MALAYA

Result Analysis

ARIMA Model Implementation



UNIVERSITI MALAYA

Result Discussion

01 Silver, unemployment rate, GDP rate, and interest rate have impacts on gold price.

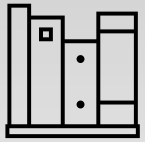
02 From the scatter plot, some forexes seem to have no correlation with gold price.

03 The accuracy may be impacted by the complexity of models

	Linear Regression	ARIMA
MAE	23.1688	5.8718
MAPE	0.0142	2.4224
RMSE	32.3447	7.6297

UNIVERSITI MALAYA

Conclusion



01

The factors, including US dollar index, unemployment rate and the monthly GDP growth rate, that may impact gold price have been explored.

02

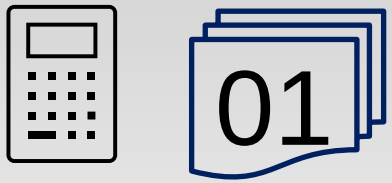
A system that forecast the price of gold using different models has been developed.

03

Two models have been evaluated based on accuracy.

UNIVERSITI MALAYA

Future Works

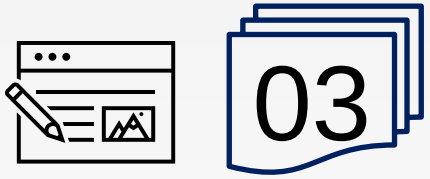


Quantitative analysis of fundamentals by machine learning methods

-such as using machine learning algorithms to extract investor sentiment indicators from text, etc.

Gold Price Forecast Based on Improved BP Neural Network

-to find the optimal BP network structure. Using the improved model, the gold futures price has achieved high-precision simulation.



Establish an early warning mechanism for gold price fluctuations

-The deterioration of the macroeconomic situation will increase the volatility of the gold market, preventing investors from facing greater market risks

UNIVERSITI MALAYA

References

[1] Liang, Y., Lin, Y. and Lu, Q., 2022. Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. *Expert Systems with Applications*, 206, p.117847.
Liang, Y., Lin, Y. and Lu, Q., 2022. Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. *Expert Systems with Applications*, 206, p.117847.

[2] Atri, H., Kouki, S. and Gallali, M., 2021. The impact of COVID-19 news, panic and media coverage on the oil and gold prices: An ARDL approach. *Resources Policy*, 72, p.102061.
[3] Bildirici, M. and Sonustun, B., 2021. Chaotic behavior in gold, silver, copper and bitcoin prices.*Resources Policy*, 74, p.102386.

[4] Chai, J., Zhao, C., Hu, Y. and Zhang, Z., 2021. Structural analysis and forecast of gold price returns. *Journal of Management Science and Engineering*, 6(2), pp.135-145.

[5] Chirwa, T. and Odhiambo, N., 2020. Determinants of gold price movements: An empirical investigation in the presence of multiple structural breaks. *Resources Policy*, 69, p.101818.

UNIVERSITI MALAYA

References

[6] Kanjilal, K. and Ghosh, S., 2017. Dynamics of crude oil and gold price post 2008 global financial crisis – New evidence from threshold vector error-correction model. *Resources Policy*, 52, pp.358-365.

[7] Mishra, A., Ghate, K., Renganathan, J., Kennet, J. and Rajderkar, N., 2022. Rolling, recursive evolving and asymmetric causality between crude oil and gold prices: Evidence from an emerging market. *Resources Policy*, 75, p.102474.

[8] Ning, Y., Kazemi, H. and Tahmasebi, P., 2022. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Computers & Geosciences*, 164, p.105126.

[9] Raza, S., Shah, N., Ali, M. and Shahbaz, M., 2021. Do Exchange Rates Fluctuations Influence Gold Price in G7 Countries? New Insights from a Nonparametric Causality-in-Quantiles Test. *Zagreb International Review of Economics and Business*, 24(2), pp.37-57.

[10] S, M. and Lazar, D., 2022. Does Volume of Gold Consumption Influence the World Gold Price?. *Journal of Risk and Financial Management*, 15(7), p.273.