Gold Price Prediction Using Linear Regression and ARIMA Model

by Jinrong Zhang

Submission date: 30-Jan-2023 06:23PM (UTC+0800)
Submission ID: 2001462220
File name: Gold_Price_Prediction_Chap1-6-final_ok.docx (3.06M)
Word count: 16749
Character count: 88503

Gold Price Prediction Using Linear Regression and ARIMA Model

Jinrong Zhang

18
FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA KUALA LUMPUR

2023

Gold Price Prediction Using Linear Regression and ARIMA Model

Jinrong Zhang

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF DATA SCIENCE

FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA KUALA LUMPUR

2023

ABSTRACT

The international gold price has been wildly erratic as the gold market has been evolving. Following a tumultuous drop from 1981 to 1999, the price of gold has been steadily increasing since 2000. From US$250 per ounce in 1999 to roughly US$1,600 per ounce in 2012, the price of gold fluctuated. Even in August 2011, the highest spot price of gold reached US$1,900 per ounce, an increase of nearly 7 times. The role of gold in maintaining and increasing value has become more apparent in light of the 2008 U.S. subprime mortgage crisis, which contributed to the global financial crisis and had an impact on the European debt crisis as well. The cost of gold has risen to $2,070 per ounce in the event that Russia invades Ukraine in March 2022.

In this situation, it is important to think about what causes the price of gold to fluctuate. Are the elements influencing the ups and downs in the price of gold consistent, and what direction will the price of gold trend in the future? This paper starts with a thorough analysis of the several variables that affect gold's price, using the ARIMA and linear regression models to do so. It then applies the method of empirical measurement to create a pricing model for the global gold market. It examines the precise elements that have an impact on gold's price and quantifies those elements' contributions.

Key Word: ARIMA; Linear Regression; Gold price Forecasting;

Ramalan Harga Emas Menggunakan Regresi Linear dan Model ARIMA

ABSTRAK

Harga emas antarabangsa sangat tidak menentu kerana pasaran emas telah berkembang. Berikutan kejatuhan yang teruk dari tahun 1981 hingga 1999, harga emas telah meningkat secara berterusan sejak tahun 2000. Dari AS$250 seauns pada tahun 1999 kepada kira-kira AS$1,600 seauns pada tahun 2012, harga emas berubah-ubah. Malah pada Ogos 2011, harga spot emas tertinggi mencecah AS$1,900 per auns, peningkatan hampir 7 kali ganda. Peranan emas dalam mengekalkan dan meningkatkan nilai telah menjadi lebih jelas berikutan krisis gadai janji subprima A.S. 2008, yang menyumbang kepada krisis kewangan global dan turut memberi kesan kepada krisis hutang Eropah. Kos emas telah meningkat kepada $2,070 setiap auns sekiranya Rusia menyerang Ukraine pada Mac 2022.

Dalam keadaan ini, adalah penting untuk memikirkan apa yang menyebabkan harga emas turun naik. Adakah unsur-unsur yang mempengaruhi naik turun harga emas konsisten, dan ke arah mana arah aliran harga emas pada masa hadapan? Kertas kerja ini bermula dengan analisis menyeluruh beberapa pembolehubah yang mempengaruhi harga emas, menggunakan model ARIMA dan regresi linear untuk berbuat demikian. Ia kemudiannya menggunakan kaedah pengukuran empirikal untuk mencipta model harga bagi pasaran emas global. Ia mengkaji unsur-unsur tepat yang mempunyai kesan ke atas harga emas dan mengukur sumbangan unsur-unsur tersebut.

Kata Kunci: ARIMA; Linear Regression; Ramalan harga emas;

Acknowledgements

I want to sincerely appreciate and thank my supervisor, Associate Professor Dr. Ang, for his support while I worked to finish my thesis. And thanks to Dr. Liew for his support.

My completion of this project could not have been accomplished without the support of their lectures. My sincere gratitude and appreciation to all lecturers involved in Master of Data Science coursework, Dr Saw, Associate Professor Dr Salimah, Professor Dr The Ying Wah, Associate Professor Dr Shivakumara who had lectured and guided me in all papers in completing my master courses.

Finally, I would like to thank my family: my parents who gave birth to me and support my education and my elder brother who always encourages me to be a dreamer and think bigger. Thank you to my family for their understanding when I am not able to spend quality time with them throughout this study period.

Table of Contents

List of Figures

List of Tables

Chapter1: Introduction

Gold entered the market pricing phase as a result of the collapse of the Bretton Woods regime. Gold's price begins to be influenced by an increasing number of variables, and its vibrational intensity also sharply increases. The US dollar no longer has exclusive power over gold's price. The sharp rise and fall in gold's price in a few of decades caught many market participants off guard.

Because of this, it is essential for gold research to accurately determine how often prices fluctuate and dynamically study the characteristics of gold pricing. By developing an econometric model based on theoretical analysis of the factors influencing gold price, which will serve as the basis for the research of changes in gold price, it is possible to measure the amount of effect of various elements on the price of gold.

The price of gold has always been a primary concern for everyone involved in the gold market, including producers, consumers, and investors, and even a small shift in the price can have a big effect on how they act. The recent decline in gold's price has led many individuals to believe that gold trading will enter a "bear market" period. By examining the factors that affect gold price and speculating on its future course, this article can help participants in the gold market effectively measure the gold price and

swiftly adjust their strategies in the stormy market. Small gold customers, gold producers, and the administration of gold reserves across entire nations will all gain significantly.

1.1 The Dual Properties and Main Functions of Gold

1.1.1 The Dual Properties of Gold

Gold is a unique item with two characteristics: the commodity property and the money property. These two characteristics are constant and will always exist. However, the characteristics that gold emphasises vary depending on the historical stage of development.

1.1.1.1Commodity Properties of Gold

Gold has the value and usage value of common commodities before it has any purchasing power. Even though it later acquired the role of currency, its commodity characteristics persisted. Marxist political economy theory holds that the use value of gold is reflected in the various specific uses of gold, such as jewellery that is common in daily life, electronic industry, medical gold, commemorative gold coins, etc., while the value of gold is reflected in the abstract and undifferentiated human labour used in the production of commodity gold.

1.1.1.2 Monetary Properties of Gold

Gold and silver are naturally money, but gold and silver are not money by nature. Gold becomes money when it is designated as the general equivalent of the medium of exchange. Gold still retains value and is still used as money. Its own value changes due to the abstract and undifferentiated general human labour used to produce currency gold, and its use value also changes. This is primarily reflected in the value as a general equivalent, used to express and measure the value of other commodities, fulfilling the requirements of the international financial system.

The definition of money, how much it is worth, how it is used to make payments, how it is stored, and what it does. During the time of the gold standard, gold's monetary qualities were most completely apparent, and while serving as the standard currency, it served a variety of monetary purposes. The era of credit money has arrived since the Jamaica system was established, but gold's status as a reserve asset and currency has not changed. In fact, gold has become even more significant during the current moment of global political and economic unrest. function, particularly in asset storage and financial investments.

1.2 The Pattern and Elements of International Gold Market

In addition to being a crucial investment and financing tool, gold acts as a significant component of the global reserves. The gold market is one of many financial

marketplaces that is significant. In addition to being a crucial investment and financing tool, gold acts as a significant component of the global reserves. The gold market is one of many financial marketplaces that is significant.

Since the growth of the gold market, a complex multi-level global structure has emerged. The spot and forward markets, which are primarily divided into two categories and centred on the London exchange market under the control of the London Bullion Market Association. The London Bullion Market Association (LBMA), the biggest over-the-counter gold market in the world, has an effect on gold market prices all over the world. The other element is a market on an exchange where futures options are king.

The Mumbai Commodity Exchange (MCX), Shanghai Futures Exchange, Tokyo Commodity Exchange (TOCOM), the Chicago Board of Trade's CBOT division, and Indian commodities make up the majority of this market (SHFE). Because of the close linkages that exist between the two parts of financial institutions and the existence of cross-market arbitrage, the prices of gold on the spot and futures markets won't dramatically vary.

To meet the diverse trading needs of gold merchants, numerous gold trading venues have been established globally. Three types of venues can be distinguished among them: European-style, American-style, and Asian-style. Regardless of the manner of exchange or whether gold dealers or banks purchase or sell on their own or as agents,

there are only minor variations in specific forms and procedures. The fundamental principles are the same, and the goal is to provide easy gold dealing to meet the needs of various gold dealers throughout the world.

1.3 Gold Pricing Mechanism in Major International Gold Trading Markets

1.3.1 London Gold Market and Its Pricing Mechanism

More than 300 years ago, the London Gold Market began trading. The hub of the global gold trade shifted to London in 1804, taking the place of Amsterdam, the Netherlands. The London Gold Market was first formed in 1919 and traded mostly spot gold until the London Gold Futures Market opened in 1982. There was activity in gold futures. Currently, London is the world's largest gold trading market, and South Africa is its primary gold source.

1.3.2 Zurich Gold Market and Its Pricing Mechanism

Following World War II, the Zurich Gold Market was created. Only the London International Gold Market is superior to the Zurich Gold Market. A free and secure environment is provided for gold transactions through Switzerland's unique banking system and supporting transaction services. Switzerland is not just a recent addition to the world; it also has advantageous arrangements with South Africa and collects gold from the former Soviet Union. the biggest private gold storage facility and gold transfer station in the world.

The three biggest Swiss banks, UBS, Credit Suisse, and UBS, handle clearing and settlement for the Zurich gold market, which lacks a formal organisational structure. The three biggest banks not only carry out their own gold transactions but also act as agents for their customers.

Unlike the London gold trading market, the Zurich gold market does not have a fixed price, but determines the gold price for the day's trading according to the supply and demand situation. This price is the official price of Zurich gold.

1.3.3 New York Gold Market and Its Pricing Mechanism

The New York Gold Market was founded in 1975 and specialises in trading gold futures. When the United States declared in 1975 that its citizens might own and sell gold, a sizable global gold market was created. After the summit in Jamaica, the U.S. Treasury Department and the International Monetary Fund held a sizable gold auction because New York is one of the biggest global financial hubs. The domestic demand for gold in the United States at that time was growing. As a result, the New York gold futures market quickly replaced other gold futures markets throughout the globe. significant component of the market.

The majority of participants make more money by short selling than they do from actually delivering goods when they expire. In order to deal with the accumulation of

several contracts by a small number of criminals, the market simply needs to keep a fixed amount of merchandise on hand. The massive volume of futures transactions causes the gold price in the New York market to occasionally be more valuable than the pricing in the gold markets in London and Zurich.

1.3.4 Hong Kong Gold and Silver Exchange and Its Pricing Mechanism

The founding of the gold and silver trading market in Hong Kong gave rise to a market with a history spanning more than a century. Since its founding in 1910, the Hong Kong Gold and Silver Exchange has gone by the name Gold and Silver Exchange. It wasn't given the name "Gold and Silver Exchange" formally until 1918. The Hong Kong Gold and Silver Exchange has developed into a major global hub for gold trading as a result of the company's ongoing growth. As of 2011, Hong Kong imported more than 300 tonnes of gold, and more than 30 million taels were exchanged in the market.

The Hong Kong Gold and Silver Exchange has retained a distinctive transaction technique for more than 90 years, namely an open outcry and a bookmaker system. No contracts are ever signed; all agreements are made verbally. You can purchase 2,000 SIMA TALES for a single fee if you accept the dealer's offer (the gold transaction specification of the Hong Kong Gold and Silver Exchange Field is five sima taels for a standard gold bar).

1.4 Analysis of the Relevant Influencing Factors of the International Gold Price

1.4.1 From A Supply and Demand Perspective

When there is a bigger worldwide supply of gold than there is a demand for it, the price of gold will decrease; otherwise, the price will trend upward. This is similar to how supply and demand affects the price of other commodities. But because of its uniqueness, gold is not exactly equivalent to commonplace goods, and its price is less obviously affected by supply and demand than are other goods. Additionally, the relationship between gold's supply and demand has remained largely steady, meaning that both the supply and demand are largely unaffected by one another.

1.4.2 From a Macroeconomic Perspective

The macro economy will undoubtedly have an impact on the price of gold because of its tight relationship to the macroeconomic development level. The US dollar index, GDP figures, inflation rate, real interest rates, and policy considerations are examples of macroeconomic factors.

1.4.3 Other Influencing Factors

1.4.3.1 Stock Market Volatility

The stock market is frequently used as a "barometer" of the health of the economy. The price of gold tends to be in a downward range when the stock market is performing well, which is known as a bull market in the stock market, and vice versa. This is primarily due to investment expectations for future economic development: if most

people have positive expectations for the future of the economy, then they will invest more money in the stock market. In contrast, they will invest less money in the gold market, which will eventually result in a decline in the price of gold.

1.4.3.2 Price Volatility in the Crude Oil Market

Crude oil is a significant source of essential consumption energy. The price of crude oil is increasing, which will increase the cost of making a variety of chemicals that use crude oil as a raw material. As a result, many things will become more expensive across society, adding to the pressure on inflation. In addition, gold is frequently recognised as a powerful asset against inflation, which will undoubtedly enhance investor demand for it. The rise in demand will drive up the price of gold.

To maintain and increase asset value while distributing the risks associated with investments, oil-producing nations will sell their oil and make more money if the price of oil 23 exhibits an upward trend for a while. These nations will encourage the rising of gold prices by raising the proportion of gold in their international reserves in order to spread the risk of keeping a big amount of dollars.

1.4.3.3 Global Economic and Financial Environment

The global economic and financial environment and gold price swings are intricately intertwined. Gold has come to the attention of central banks and regular investors as

the last "life-saving straw" when the global economy is experiencing severe turmoil and a recession.

1.5 Machine Learning

According to Kevin Jolly's 2018 book Machine Learning with Scikit: Learn Quick Start Guide, machine learning is a group of techniques that can find patterns in data and utilise those patterns to predict the future. From banking to healthcare, a wide range of businesses have discovered tremendous benefit in machine learning.

Broadly speaking, machine learning can be categorised into three main types.: supervised learning, unsupervised learning, reinforcement learning.

In the case of supervised learning, a set of labels or a numerical goal variable are given to our data. This method, which also includes support vector machines, logistic regression, k-nearest neighbour, nave bayes, and tree-based algorithms, can address both classification and regression problems.

The majority of the time, unsupervised learning methods are employed to cluster data points based on distance. Customers are divided into several categories based on a range of factors using the K-means algorithm.

1.5.1 ARIMA Model

The ARIMA model is a term for analysing time series data (Auto-Regressive Integrated Moving Average). For more than 50 years, it has been used to analyse time series data. It is regarded as a robust and user-friendly tool for data processing.

Auto-regression, or AR, is produced using prior data points; I integrated the general trend in the data; and MA denotes terms of error or noise based on prior data. The ARIMA model is made up of these three components.

It is common to see the ARIMA model in stock price forecasts. However, there are several requirements to meet before using this potent algorithm to choose the best data processing technique. It is appropriate for erratic and outlier-filled seasonal data. Additionally, the facts about the mean are inconsistent and vary.

When we utilise the ARIMA model, we must supply the three parameters (p, d, and q). The Auto-Correlation Function determines the p value (ACF). The term q, which illustrates how effectively the current time series value is connected to the earlier ones, is found using ACF. The PARAF, or partial auto-correlation function, is used to define the q parameter.

1.5.1.1 Stationary Test

A stationary test describes the statistical properties of a process that generates time series that do not change over time. The ARIMA model can be used to generate predictions if the time series is not stationary. If the data being recognised are not steady, the ARIMA model cannot implement the method. For establishing whether the data is stationary, two methods are shown.

The rolling test, which visually displays the moving average or moving standard deviation of a certain time period, looks at whether these statistic values move over time. It is referred as as Augmented Dickey Fuller the following (ADF). ADF testing confirms the non-stationarity of time series as the null hypothesis by statistical means. We reject the null-hypothesis if the test statistic result is less than crucial values at various confidence levels. It is therefore regarded as immobile.

There are three way to convert data into time series stationary, which are detrending, differencing, and decomposition.

For instance, WTI crude oil company frequently laments that the price of oil is frequently impacted by exceptionally cold winter weather and mechanical breakdown due to a lack of regular maintenance. It produces non-stationary data as a result, which is useful for predicting. But this can be handled with certain strategies. The first method is transformation, which reduces bigger values into smaller ones by applying a log,

square, or cubic root. The second strategy for eradicating a time series' dependence on time and stabilising its mean is differencing. The next step is to take averages over a specific time frame, like weekly or monthly data.

1.5.1.2 White noise data

Unpredictable random independent number sequences make up white noise data. If there are white noise data, a set of time series data could be modelled. A statistical method called Ljung-Box is used to examine white noise data. The data series is assumed to be white noise, which means it is independently distributed, thereby setting the null hypothesis. The alternative theory is that there is serial correlation and that the data series is not randomly dispersed.

1.5.2 Linear Regression Model

The link between two variables can be ascertained using linear regression by applying a linear equation to the observed data. The terms "dependent variable" and "independent variable" relate to two different types of variables. Linear regression is frequently used in predictive analysis. How strongly two variables are related to one another can be seen by looking at the correlation coefficient. The range of the coefficient runs from -1 to +1. The correlation coefficient shows how closely the two variables and the observed data are related.

1.6 Forecasting

This thesis seeks to forecast gold prices using ARIMA and linear regression models. The ARIMA model will be used to forecast gold prices over the long term, while the linear regression model will be used to forecast gold prices over the short term. Historical gold prices and other economic variables known to affect gold prices, such as inflation and interest rates, will be utilised as the data for the models. The models' outputs will then be compared and assessed to determine how well they predicted gold prices.

1.7 Problem statement

According to an empirical study by Yang Liuyong and Shi Zhentao from 2004, the federal funds rate, the Dow Jones price index, the U.S. inflation rate, and the nominal exchange rate of the U.S. dollar all have a long-term impact on the price of gold. Li Ting (2009) examined the gold price and the US dollar index from 1986 to 2009 using a cointegration test and a VAR model.

Whenever an economic crisis occurs, there is unquestionably a positive correlation that occurs. In their study from 2010, Hao Yuzhu and Wang Qiuying looked at the factors from three different angles, including the time when there was a high level of global risk aversion, the fact that both the US dollar and gold were utilised as hedging tools, and the fact that gold was priced in US dollars. Ma Chengping discussed the mechanism by which the US dollar affects the price of gold in 2012.

There are sporadic exceptions to the US dollar index and gold prices' overall negative link, according to Wang Yihong's (2013) research. Li Xi (2015) established the long-term relationship between gold and the US dollar by utilising the VECM model to analyse the relationship between the price of the US dollar and the price of gold futures. Yang Lijun and Zhang Kun (2016) investigated the connection between the US dollar index and the price of gold using the theory of adaptive expectations. They came to the conclusion that adaptive expectations had a considerable impact on the US dollar index and that the two variables were considerably adversely connected. According to Kunkler and Macdonald's (2016) split of the gold dollar price into two portions, there is no correlation between the fluctuation of the world gold price and the variation of the global dollar price.

1.8 Research Objectives

Given that the price of gold is closely correlated with the degree of macroeconomic development, the macro economy will surely have an effect on it. Since gold is priced in US dollars as a financial asset, its price has a strong inverse relationship with the US dollar exchange rate. Gold has always been thought of as the best investment against inflation. The cost of gold will increase during an inflationary time, much like other basic items. Under normal circumstances, there is consistency between the direction of change in the gold price and the rate of inflation.

Interest rates, or the cost of money, are a reflection of the tension between the supply and demand for money. It is not the same as holding money to hold gold. In addition to not paying any interest income to investors, holding it costs money. This opportunity cost is the discrepancy between the nominal interest rate and the inflation rate, or real interest rate. The price of gold and interest rates typically have an inverse relationship. When interest rates drop, profits from capital investments fall but income from gold investments rises, supporting an increase in the price of gold;

The price of international crude oil will significantly affect the price of gold, according to past history. On the one hand, both inflation and the price of crude oil rise in an effort to counteract gold's role as an inflationary force. Naturally, gold is thought of as a strong asset to combat inflation, and its value will inevitably rise. Gold is in high demand among investors, thus as a result of this growth in demand, its price will climb.

Contrarily, in order to reduce investment risk and maintain or even increase asset value. Oil-producing nations will sell their oil and earn more money when oil prices rise. These nations will raise the percentage of gold in their foreign reserves, which will encourage the increase in gold prices, in order to spread out the risk of keeping a huge amount of dollars.

The one-hour data of gold prices from January 1, 2018 to October 14, 2022 were utilised as the data source for this article's analysis. The ARIMA model and the Linear Regression model were then used to predict the gold price. By contrasting each model's forecast accuracy, the most effective prediction model is chosen.

1.9 Research Questions

1.9.1 US dollars

With the ongoing changes and expansion of the global economy, gold is the only remaining shelter for wealth. Its dual status as a currency and a commodity makes it a crucial financial tool for the preservation and growth of value, and it is also a reserve that is acknowledged internationally. Relationship of substitution. Dollars are used as the unit of exchange for gold on the international gold market. The US dollar index will fall if the supply of US currencies rises. The price of gold rises as more US dollars are exchanged for less gold, provided that the gold increment is constrained. Therefore, there is a strong correlation between gold and the US dollars.

1.9.2 Existing Models That Used to Forecast Gold Price

1.9.2.1 ARIMA model

Box and Jenkins introduced the Autoregressive Integrated Moving Average Model, or ARIMA model, as a time series forecasting method in the early 1970s. A model known as an ARIMA model is produced by regressing the dependent variable only on its lag value as well as the present value and lag value of the random error component in order

to convert a non-stationary time series into a stationary time series. Time series models are suitable for short-term forecasting as the previous change patterns of statistical series have not significantly changed.

The core idea behind the ARIMA approach is to interpret the data sequence that the prediction object generates over time as random and to use a particular mathematical model to approximately explain this sequence. Once discovered, the model may forecast future values based on the past and present values of the time series. Businesses can successfully anticipate future sales using modern statistical methods and econometric models.

1.9.2.2 LSTM Model

In recent years, financial sequence prediction problems have slowly been resolved using the LSTM model (long short-term memory network model). Researchers now frequently simulate and forecast the stock market using the LSTM model. Chen et al. used it in 2015 to The prediction effect of LSTM has been significantly enhanced when compared to other prediction models in the stock price forecast of my nation's stock market. There have been more successful financial sequence forecasts using the LSTM model, however there are not many study findings that have been applied to the forecast of the gold futures market.

1.9.2.3 Grey Sequence

Grey system theory's differential equations are the foundation of the forecasting technique known as grey forecasting. Its main function is to do a correlation study on a time-related grey process that fluctuates within a particular range. The original data is first added up or subtracted to create a grey sequence, or sequence with some regularity, and the grey sequence is then applied to prediction.

1.9.2.4 Neural Network

One of the most popular and effective neural networks is the BP neural network. A complicated nonlinear dynamic system that can represent nonlinear processes is the BP neural network. The fundamental purpose is to achieve the desired output objective by adjusting the neural network's weights and thresholds utilising backward feedback from training errors.

A BP neural network typically has three layers or more in the input layer, hidden layer, and output layer. There are several levels in the structure, and inside each layer there are many neurons connected by weights. Information must be delivered to the neurons in the layer below since information cannot go between neurons in the same layer. The information that the input layer has gathered from the outside world is transmitted to the hidden layer, which uses the weight and activation functions of the neurons to transmit it to the output layer. The result is then output by the output layer.

1.9.2.5 Combined Models

In general, a single prediction model can only accurately predict certain aspects of the system, and it also has some limitations that prevent it from reflecting all of the system's data. In contrast, a combined model can combine multiple single models and combine their advantages to more accurately predict the system. reliable data.

The weighted output value of each individual prediction model serves as the output result of the combined prediction model. The combined weight of each individual prediction model is derived using the least mean square error.

1.9.3 Better model to forecast gold price.

Although the ARIMA model has a high level of prediction accuracy, it works best for short-term forecasting. The forecast effect diminishes as the number of predicted days rises. The model's forecast precision is mediocre. Even if a limited sample size is needed for modelling, exponential growth models can use it. The single-factor BP neural network model has better prediction accuracy and can better fit a variety of nonlinear functions, but it requires a large amount of data, and the error is large under small samples and does not include other Influencing factors. Alternatively, the prediction effect of fluctuating data is poor. The four main elements influencing the daily price of gold are included in the multi-factor BP neural network model, which has

a decent predictive ability. However, there are still issues with the model, such as a huge sample size and a sluggish convergence rate.

1.10 Thesis Organization

This thesis is being organized in the following chapters:

Chapter 1 gives an overview of the thesis, including the research background, and general description of the topic. In parallel, this chapter also provide introduction to Machine learning and exploration into of various types of forecasting model before focusing into the proposed algorithm.

Chapter 2 covers the literature review to study the industry requirement and specification of 5G network, as well as its supporting technologies which enable 5G requirements. This chapter also studies into previous research in the domain of computer network, as well as the research in machine learning focusing on network usage forecasting. This chapter ends with the gap analysis to identify the focus area for this research.

Chapter 3 outline the research methodology to address the research objective and research question. The chapter also explains method used in the experiment, explains how the dataset is being used and the method used to compare the result between 2 different algorithms.

Finally, the Appendix sections contain raw data result captured from the experiment

CHAPTER 2: LITERATURE REVIEW

2.1How macroeconomic indicators influence gold price management.

The violent fluctuations in the price of gold have attracted widespread attention from monetary authorities in various countries and global investors. In the past ten years, the gold market has ushered in a super bull market, which rose from a fluctuation of US$252/oz in August 1999 to US$1,920/oz in September 2011, with a maximum increase of 662%. Since then, the price of gold has been on the decline due to the sluggish recovery of the US economy and the strengthening of the US dollar. Even "Chinese aunts" went into a buying frenzy once gold prices fell. Gold is not just a significant component of a nation's official reserves; it is also linked to the financial security of a nation, a goal for investments to thwart inflation, and the critical interests of investors. Large and consistent swings in the price of gold have increased market risks and presented difficult obstacles for investors.

Therefore, it is important for investors in the gold market, monetary authorities and risk managers in the spot market of gold futures to effectively identify the factors that affect the fluctuations in the gold market. theoretical and practical significance.

2.1.1 The Linkage between the International Gold Price and the Price Fluctuations of China and the US Stock Market

Zhang Zhen (2016) investigated the correlation between the stock market and the international gold price. The correlation coefficient between the price of gold and the Chinese and American stock markets was, in general, empirically determined from the standpoint of static analysis, i.e. According to the gramme indices, tiny variations will follow small fluctuations and huge fluctuations will follow massive swings. The clustering properties allow for a clear observation of the three markets' variations. The Shanghai Composite Index is followed by the Nasdaq Index in terms of long-term persistence of price swings, and the international gold price has the lowest duration, indicating that the global gold market is crucial in maintaining the value of financial assets between nations.

2.1.2 Analysis of the Linkage between Oil and Gold Price

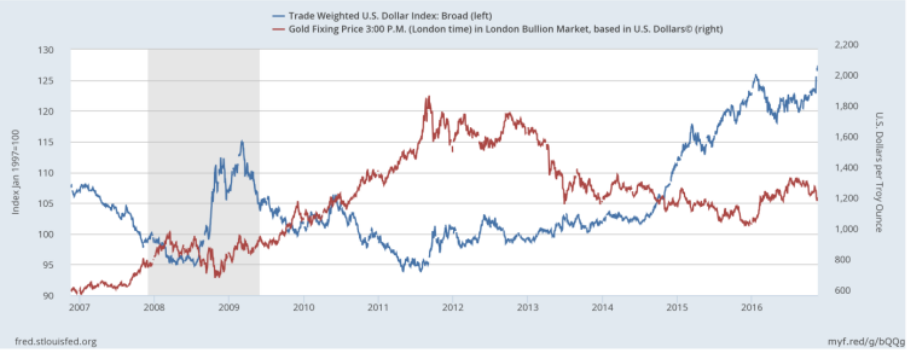
Liu Jie (2017) looked at the factors affecting fluctuations in the price of gold as well as the current dynamics of oil supply and demand. The analysis's conclusions show that there is normally a positive relationship between gold and oil prices. The exchange rate of the US dollar, inflation, how well oil-exporting countries are doing, and the condition of global politics all have an impact on it.

2.1.3 The Impact of Geopolitical Risk Events on the Price of Gold

Wang Bing (2022) believes that although the impact of major geo-risk events such as war on the price of gold exists, it is mainly concentrated in the short term and does not necessarily cause the price of gold to rise. On February 24, 2022, Russia announced a military operation in eastern Ukraine. According to the upward trend in gold prices since February, which has coincided with the development of tensions between Russia and Ukraine in Western nations. On February 24, the price of gold on the COMEX fluctuated considerably, reaching a high of $1,976 per ounce before ultimately falling by 0.26%. Since then, the price of gold has continued to rise, with a correction in the second week of March showing that the effect of the Russian-Ukrainian conflict on the price of gold is progressively waning. This pattern roughly follows the historical trajectory of how regional wars have affected gold prices.

2.1.4 The Linkage Between the US Dollar Index and the Gold Price

Based on the DCC-GARCH model, Lu Guoqing and Li Mingxue (2017) examined the relationship between the U.S. dollar index and gold prices. The U.S. dollar index and gold prices are less able to absorb fresh shocks, but they are still affected by volatility for a significant amount of time after the effect, according to the empirical findings. There is a two-way causal relationship between the US dollar index and the global gold price, which has primarily shown a negative association for a long period. The trend will only exhibit a positive correlation when financial market pressure is present and both the US dollar and gold are viewed as safe-haven assets (such as the 2008 economic crisis). The US dollar index and gold prices also displayed a very evident positive link in 2015 and 2016.



2.2 Machine Learning in Economic Data Forecasting

Economic data mining, economic indicator analysis, and economic policy evaluation are the main topics of discussion in economic research. Under various assumptions, traditional econometric models typically explain the linear relationship between

economic variables, but frequently less so than the extracted information As a result, using machine learning techniques in economic research can considerably increase the accuracy of problem analysis and therefore serve as a more useful resource for decision-makers.

2.2.1 Common Machine Learning Algorithms and Their Pros and Cons

Comparison of Machine Learning Methods

Methos	Advantages	Disadvantages	Application
SVM	High prediction accuracy, non-random method, strong robustness	Sensitive to parameter selection, it is difficult to solve the problem of multiple classification	Solve classification problems, forecasts
KNN	Characterize nonlinear effects without need explicitly assume that the data is distributed. High tolerance for outliers and noise	Inefficient and prone to dimensionality disaster	Solve the classification problem; Short-term liquidity forecasts
Decision Tree	Simple to use, able to handle classified and continuous data, running fast	The accuracy of the model has average performance, and it is prone to overfitting.	Solve classification problems, forecasts

Random Forest	Integrated machine learning methods provide less prediction error and robustness to outliers	Average performance	Solve classification problems, forecasts
Naïve Bayes	Simple for calculation and use, few parameters, reduce the risk of data snooping	Sensitive to the representation of the input data, a priori probabilities are required	Solve classification problems
XGBoost	Less study time and more flexibility	High operation complexity, more memory occupied, high operational complexity, and more memory	Solve classification problems and learn multiple classifiers
Neural Network	Characterize nonlinear effects and can be used for multivariate input problems and can be used to process large amounts of data	The accuracy of the model has average performance, and it is prone to overfitting.	Series prediction
K-means	Unsupervised learning, simple and easy to implement, works well when categories are obvious	Sensitive to outliers and noise, the initial selection of clustering center points	Category division

Table 2.1 Comparison of Machine Learning Methods

2.2.2 Application of Machine Learning Methods in GDP Forecasting

Lin Qi (2009) proposed a GDP forecasting method based on the least squares support vector machine in Fujian Province. The radial basis kernel function was used for simulation, and a high-precision forecasting model was established through parameter selection. According to the prediction findings, the training sample's average relative error of fitting is just 0.84%, the test sample's relative error is 1.7%, and the anticipated fitting accuracy is 8.3%. Elmira (2016) developed a simulation model for the GDP forecast for Turkey using the genetic algorithm and SRA algorithm. The findings demonstrate the GA-SVR model's strong estimation accuracy and its applicability as a

tool for simulating GDP, the primary indicator of economic growth.Chen Jinhao et al. (2019) analyzed the U.S. power industry by fusing power generation data, established a support vector machine model, and proposed a new research method for predicting GDP through related net power generation data. The correlation coefficient of the experimental results was 99.77%.

2.2.3 The Application of Machine Learning Methods in the Labor Market

An essential factor in determining a nation's level of development is unemployment. Social stability and economic growth will suffer from excessive unemployment. Spain has the second highest unemployment rate in the European Union, so Luna-Romera et al. (2019) used two cluster analysis methods, K-means and hierarchical clustering with average links, to match the labor market in Spain, from geographic and occupational perspectives. Learn about employment options for the workforce in times of economic crisis and economic recovery. Xu Zhengli et al. (2021) used web crawler technology and K-means and other methods to analyse the demand for labor positions, quantified the text information on the recruitment website, and used the demand matrix to indicate the skill requirements of different positions, which intuitively showed the labor market’s demand for talents.

2.2.4 Application of Machine Learning Methods in Policy Evaluation Research

Comparing the development situation before and after the policy's adoption is required to assess the policy's impact. Typically, researchers create an econometric model to

assess the impact. The evaluation of policies can also include machine learning techniques.

Gao Huachuan and Bai Zhonglin (2019) examined the causal relationship between the British "Brexit referendum" and the exchange rate of the British pound. They did this using the time-varying LASSO approach. By comparing the dollar exchange rate of the pound and the synthetic pound, the time-varying weight coefficient and the policy impact of the Brexit referendum on the exchange rate of the pound are determined. The estimation accuracy of this method is 20%–30% higher than typical counterfactual estimation.

2.3 Machine learning in forecasting gold price

2.3.1 ARIMA, LSTM, and Prophet model

Three time-series forecasting ML techniques were utilised by Yanrui Ning, Hossein Kazemi, and Pejman Tahmasebi (2021) to estimate the typical oil decline curve of a well in an unconventional shale resource. It comes to the conclusion that ARIMA is better suited for forecasting short-term data span. Particularly, LSTM and Prophet techniques can spot odd changes that can't be found using conventional methods. The Prophet model can identify seasonal variations that are helpful for investors to avoid potential market shocks that have already occurred.

2.3.2 LSTM model

The LSTM neural network, according to Liu Lu et al. (2021), not only carried over the advantages of RNN in handling sequence problems but also significantly enhanced the high-precision prediction outcomes of time series data. By comparing the experimental results and exploring a suitable two-way network model to realise the COMEX Analysis and prediction of gold futures price series, by observing the performance of RNN model, SVR model, double-layer LSTM model, and bidirectional LSTM neural network, the parameters of the LSTM neural network were adjusted after normalising the gold futures price data and other preprocessing operations.

It has been discovered that the bidirectional LSTM model has a promising future for use in predicting the price of gold futures. The next step is to investigate ways to incorporate additional models into the LSTM network and create a hybrid model to boost prediction accuracy even more.

2.4Risk alerting system for gold price changes

Zhang Xiaomei (2018) used risk early warning theory based on factor analysis to identify alternative factors, futures factors, and macro-environmental factors that affect gold prices, and established a linear relationship between gold prices and common factors. Zhang Xiaomei (2018) set the gold price fluctuation alarm limit according to the historical volatility of gold price. The regression model identifies the linear relationship between the price of gold and the common factor, offers a system for early

detection of the risk of changes in the price of gold, and then makes suggestions for workable risk management techniques.

2.5 How Covid-19 impact gold price

Guiye He (2021) During times of emergency, the price of gold will trend upward, and the quantity of newly discovered COVID-19 cases will benefit the spot price of gold. Demand-wise, the crisis itself will induce panic among the populace following the start of the new crown plague and its widespread spread, raising the demand for investments in gold, a "hard money." The new crown pneumonia outbreak has decreased the productivity of the gold supply in terms of production and transportation, which has resulted in a decrease in supply from the standpoint of supply. In the end, the pandemic's shift in supply and demand led to months of shortages and an increase in gold prices on the spot market.

2.6 How monetary policies impact gold price.

According to Zhang Cong et al. (2022), unexpected crises accompanied by the introduction of relaxing measures lack the pre-deliberation and pricing processes that can encourage the development of a persistent rising trend in gold prices. The Federal Reserve frequently performs expectation management prior to the start of the tightening policy in order to minimise the impact on the economy and has a lengthy decision-making process. This time frame lines up with the complete price process. As a result, the price of gold tends to fluctuate against the basic logic when the real policy is applied because of the high degree of pricing.

2.7 Literature Review Summary

This part summarizes the literature reviews that has been mentioned and researched above.

Author	Year	Analysis	Summary
Jiang Feng, Zhang Wenya	2022	Understand how machine learning changes the way of economic research	This article reviews the application of machine learning methods in economic research in recent years, from inflation, exchange rates and currencies,GDP, labor market, social stability, policy evaluation and other perspectives are summarized
Yanrui Ning, Hossein Kazemi, Pejman Tahmasebi	2021	Understand how time-series forecasting ML methods to predict a typical well's oil decline curve.	it's more appropriate in time-series forecasting because of the historical fluctuations in production well and reservoir operations.
Liya A, Qian Qin et al.	2021	understand How macroeconomic indicators influence gold price management	The results show an optimistic and significant relationship between IR and gold prices.
Liu Jie	2017	Understand the relationship between oil price and gold price.	In most cases, the price of oil and the price of gold are positively linked.
He Guihua	2021	Understand how Covid-19 impact gold price	The number of new confirmed cases of COVID-19 has a positive impact on the spot price of gold
Lu Guoqing, Li Mingxue	2017	Understand the relationship between US dollars and gold price.	The dollar index has a certain negative correlation with the price of gold in the long run
Cheng Ming	2017	Understand methods used to forecast gold price.	The BP neural network model has a wider applicability, and a combinatorial prediction model based on genetic algorithms is proposed
Wang Bin	2022	Understand how geopolitical risks impacts gold price	The impact of major geo-risk events such as war on gold prices is mainly concentrated in the short term and does

			not necessarily promote the rise in gold prices
Zhang Xiaomei	2018	Understand generate risk alerting system for gold price changes	Establish an early warning system for the risk of gold price fluctuations, and then put forward effective risk prevention measures.
Zhang Cong et al.	2022	Understand gold price volatility under the Fed's tightening policies.	The change in the Fed's monetary policy expectations corresponds to a turning point in the gold price trend
Lei Yufei	2021	Understand how monetary policies impact gold price.	The impact of monetary policy on the price of bulk gold is analyzed from the perspective of loose monetary policy and tight monetary policy
Liu Lu et al.	2021	Understand how to forecast gold price based on LSTM	The bidirectional LSTM model is superior to all contrast models and achieves good predictions

Table 2.2 Summary of Literature Reviews

2.8 Chapter Summary

The literature reviews from earlier studies on gold price forecasting are discussed in this chapter. We discuss the research done by academics on the variables that affect the price of gold. Economic indices like the GDP, unemployment rate, and productivity are predicted using machine learning techniques, and these techniques have been shown to be successful. Numerous academics have used SVM, ARIMA, and LSTM, among other machine learning techniques, to study the forecast of the price of gold.

We are able to develop new findings from a new perspective thanks to the earlier studies. We discussed how geopolitical considerations affect the price of gold in light of the past six months since Russia's invasion of Ukraine. Additionally, we looked at the literature on risk mitigation and developed a mechanism to prevent past occurrences.

In the next chapter, the methodology that we used will be discussed in detail.

Chapter 3: METHODOLOGY

This chapter introduce the methodology used to explore data mentioned in chapter 1. This part will give a detailed explanation of how these models work and will explain the computation formulas and the role of each parameter played in the formulation. In the end, each model will be tested for accuracy in a hope for finding the best model that fits the forecasting. Namely, the evaluation methods include MAE, RMSE, MAPE and R-Squared.

74

3.1 Conceptual Framework

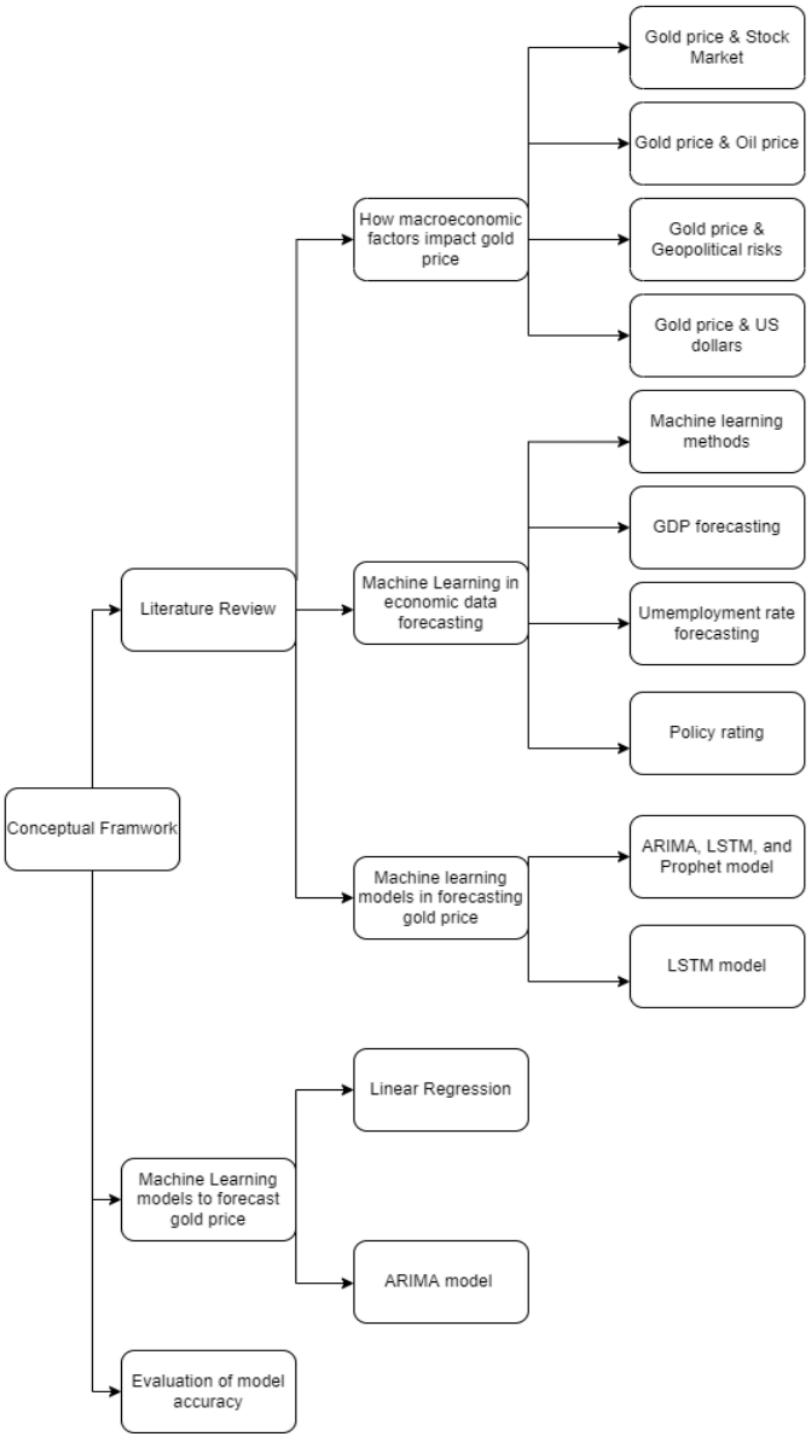

Conceptual Framwork
Literature Review
How macroeconomic factors impact gold price
Gold price & Stock Market
Gold price & Oil price
Gold price & Geopolitical risks
Gold price & US dollars
Machine Learning in economic data forecasting
Machine learning methods
GDP forecasting
Umemployment rate forecasting
Policy rating
Machine learning models in forecasting gold price
ARIMA, LSTM, and Prophet model
LSTM model
Machine Learning models to forecast gold price
Linear Regression
ARIMA model
Evaluation of model accuracy

Figure 3.1 Conceptual Framework of the Thesis

The structure depicts how gold price are forecasted. Through literature review on factors that have impacts on gold price changes and the forecasting models, the thesis proposed forecasting methods based on data given. The forecasting model are Linear Regression, and ARIMA model. In the end, several evaluation methods are selected to rate the accuracy of the model.

3.2 A quantitative method is proposed in forecasting gold price.

The forecasting model makes use of the ARIMA model and linear regression since prior research has shown that these two models are suitable for time series forecasting. A quantitative approach is used to assess how accurate the model is.

3.3 Experiment Design

Experiment Objective:

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

Collect and prepare the data: Historical gold price data, economic indicators, and other factors will be collected and prepared for use in the model. Construct the model: The linear regression model and the ARIMA model will be constructed and combined into a single model. Test the model: The model will be tested using a variety of methods to determine its accuracy and effectiveness. Deploy the model: The model will be deployed and used to make predictions about future gold prices.The model's output will be examined in order to ascertain the precision of the forecasts.

3.3.1 ARIMA model

3.3.1.1 Time series

We call a series of sequences that do not consist of random variables called random sequences. If the random sequences are sorted by time, that is, the time intervals of each variable are equal, where *t* is an integer time variable representing equal intervals, which can be positive or negative. The *t* before the current moment is negative, and the *t* after the current moment is positive, then this random sequence can be called a time series.

3.3.1.2 Strictly Stationary Time Series

For any integer that *s* is less than *t* and positive integer that *n* is greater than 1, the random vectors $\{X_s, X_{s+1}, \dots, X_t\}$ and $\{X_{s+n}, X_{s+1+n}, \cdots, X_{t+n}\}$has the same joint probability distribution. $\{X_t, t \in T\}$is a strictly stationary sequence.

3.3.1.3 Wide Stationary Time Series

For any integer$\{X_t, t \in T\}$, if the following three conditions are met at the same time:

①$\forall \in T$ $EX_t^2 < \infty$

②$\forall \in T$, $EX_t^2 = \mu$

③$\forall t, s, k \in T, Cov(X_t, X_s) = Cov(X_k, X_{k+s-t})$

Then $\{X_t, t \in T\}$ is a wide stationary sequence。

The probability distribution of time series is difficult to determine, so it is not easy to meet the conditions of strictly stationary time series. The wide stationary time series does not need to satisfy that all numerical features in the time series do not change with the translation of time, but only requires that the first and second order moments in the time series do not change with the passage of time.

3.3.2 White Noise

If the following conditions are met by time series $\{\boldsymbol{\varepsilon_t}, \boldsymbol{t} \in \boldsymbol{T}\}$,

①$\boldsymbol{E(\varepsilon_t) = 0}$

②$\boldsymbol{Var(\varepsilon_t) = \sigma^2 < \infty}$

③$\mathrm{Cov}(\boldsymbol{\varepsilon_t, \varepsilon_s}) = \boldsymbol{0, t \neq s, t, s = 0, \pm 1, \pm 2, \ldots}$,

it is called white noise series.

3.3.2.1 Auto-Correlation Function(ACF)

$$\boldsymbol{\rho_{t,s} = \frac{Cov(X_t, X_s)}{\sqrt{Var(X_t)Var(X_s)}}, t, s = 0, \pm 1, \pm 2, \cdots}$$

3.3.2.2 Partial Autocorrelation Function (PACF)

$$\boldsymbol{\emptyset_k = \frac{E[(X_t - EX_t)(X_{t-k} - EX_{t-k})]}{E[(X_{t-k} - EX_{t-k})]}}$$

The partial autocorrelation coefficient refers to after eliminating the intermediate k-1 random variables.

3.3.2.3 AIC Criterion

$$AIC = -2ln(L) + 2k$$

The AIC criterion is the minimum information criterion, which is used to analyze the lag order determined by the autoregressive model. The formula is as follows:

where k is the number of unidentified parameters and L is the value of the maximum likelihood function. The accuracy and complexity of the model are inversely correlated, with accuracy increasing as L in the first portion of the formula increases and complexity decreasing. The model is the best model when L reaches the minimum since it has the most appropriate amount of parameters.

3.3.3 Stationary Time Series Model

3.3.3.1 Auto-Regressive

The p-order autoregressive model of the time series can be recorded as AR(p), and the formula is expressed as:

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varphi_3 X_{t-3} + \cdots + \varphi_p X_{t-p} + \varepsilon_t, t = 0, \pm 1, \pm 2, \cdots$$

Where $\varphi_0, \varphi_1, \varphi_3, \ldots . \varphi_p$ is the autoregressive function that needs to be estimated , $\varphi_p \neq 0, \varepsilon_t \sim WN(0, \sigma^2)$. When φ_0 equals to zero, we call AR(p) a decentralized model.

The back-shift operator B is added to the decentralized model, which is expressed as:

$$\boldsymbol{\Phi(B)X_t = \varepsilon_t}$$

where Φ(B) is the autoregressive polynomial , $\boldsymbol{\Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_P B^p}$

3.3.4 Moving Average(MA)

The q-order autoregressive model of the time series can be recorded as MA(q), and the formula is expressed as:

$$\boldsymbol{X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} - \cdots - \theta_p \varepsilon_{t-p}, t = 0, \pm 1, \pm 2, \cdots}$$

Where $\boldsymbol{\theta_0, \theta_1, \theta_3, \ldots . \theta_q}$ is the autoregressive function that needs to be estimated. When $\boldsymbol{\mu}$ equals to zero, we call MA(q) a decentralized model.

The back-shift operator B is added to the decentralized model, which is expressed as:

$$\boldsymbol{X_t = \Theta(B)\varepsilon_t}$$

where $\boldsymbol{\Theta(B)}$ is the moving average polynomial， $\boldsymbol{\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_P B^q}$

3.3.5 ARIMA Calculation

The model of autoregressive moving average The model, known as ARMA(p,q), combines the AR(p) and MA(q) models, expressed as:

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \ldots - \theta_p \varepsilon_{t-p}, t = 0, \pm 1, \pm 2, \cdots$$

After adding the post-shift operator B to the model, it is expressed as:

$$\Phi(B) X_t = \Theta(B) \varepsilon_t$$

Since most time series observed and the time series being investigated in real life are non-stationary, it is necessary to convert the non-stationary sequence into a stationary sequence before fitting the ARIMA (p, q) model.

3.3.6 Modeling Steps of the ARIMA Model

The ARIMA model's modelling steps are as follows:

①To determine if a time series is stationary, there are two primary criteria. A time series scatter plot is the first. It is a stationary time series if the original data values are close to a particular constant value; otherwise, it is a non-stationary time series. The ADF unit root comes in second. It is a stationary time series if the test result accepts the alternative hypothesis.

②If the time series is not stationary, it can be transformed into a stationary time series using a difference transformation or another transformation (such as taking the logarithm first, moving average first, and then differencing).

③The ARIMA model's order p, q should be known. There are two steps in particular. The first step is to observe the properties of the ACF and PACF diagrams for the

transformed new stationary time series, so as to preliminarily determine the parameters q and p of the model. The second part revolves around the preliminarily determined p, q values and then takes multiple pairs of nearby values, selects the order corresponding to the minimum value according to the AIC criterion and the PACF criterion, and accurately estimates the model order p, q.

3.3.7 Model Significance Test

①On the established model's coefficients, a t-test is run with a confidence level of 0.95. The parameter is significant if the p value is less than 0.05;

② The F-test is run on the entire developed model at the confidence level of 0.95. The model is significant overall if the p value is less than 0.05;

3.3.8 White Noise Test for Residual Sequences

The estimated ARIMA(p,d,q) model's interference term is examined, and the S statistic is created, with a confidence level of 0.95. The model information has been entirely extracted if the p value is greater than 0.05 and the interference term is completely random.

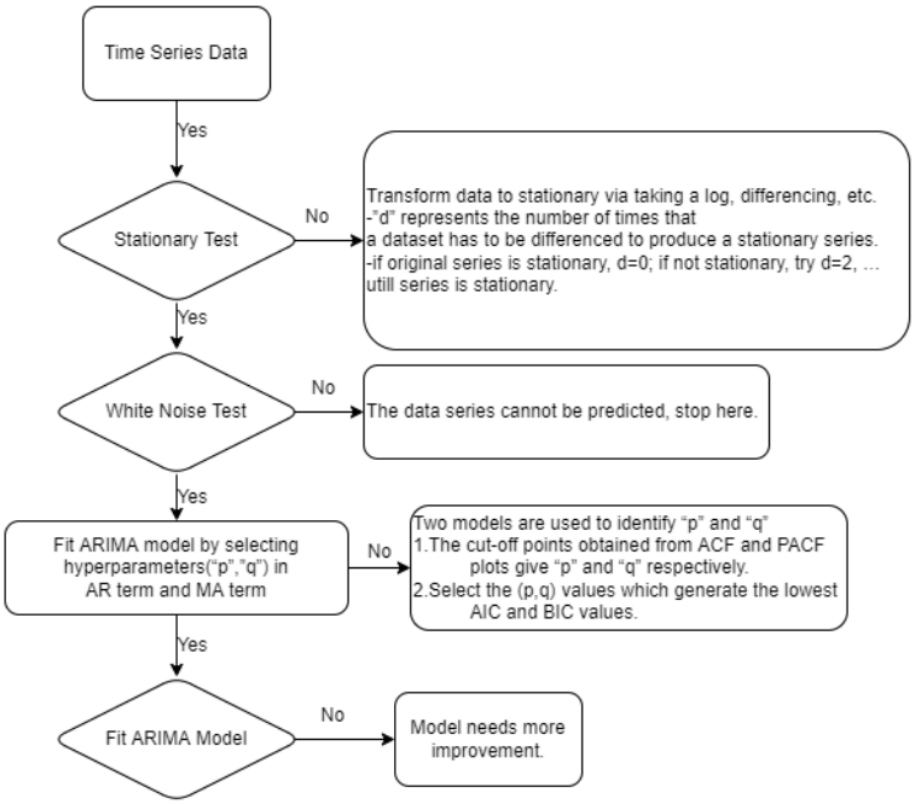


Figure 3.2 The Flow of ARIMA Model

The flow chart describes how ARIMA model work. When the data is given, it is required to judge whether the data is stationary. If not, we need to take log, differencing to make it stationary. The next step is to test whether it is the white noise data. If not, it cannot be predicted. The data can be used for prediction unless it is stationary and passes the white noise test. How to compute the p and q value has been illustrated in the above statement.

3.3.9 Linear Regression Model

A machine learning algorithm that belongs to supervised learning is the linear regression model. Simple linear regression, multiple linear regression, and polynomial regression are three different types of this model.

(1) Simple Linear Regression(SLR)

SLR assumes that the relationship is linear, which is written in mathematical notation as following.

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

where Y is the value of dependent variable. β_0 is the intercept, and ε is the random error.

(2) Multiple Linear Regression

To calculate the association between two or more independent variables and one dependent variable, utilise multiple linear regression. The following is a list of the formula:

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i,$$

where Y is the predicted variable, β_0 is the intercept, $\beta_3 X_{3i}$ is population slopes, and ε_i is the random error.

(3) Polynomial Regression

The relationship among variables are not always linear, which may result in poorly fit model. One way to solve this problem is to use polynomial regression, which takes the form:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^1 + \cdots + \beta_2 X^h + \varepsilon \ ,$$

where Y is the predicted variable, β_0 the intercept, X^h the nth degree polynomial, and ε is the random error.

3.4 Evaluating forecasting accuracy.

The discrepancy between the predicted value and the actual value represents a forecast's accuracy. The size of the forecast bucket affects forecast error. In general, the monthly forecast bucket has a lower forecast error than the weekly data.

A total of 5 indicators, including the root mean square error (RMSE), the mean absolute error (MAE), the mean error percentage (MAPE), and fit (R-Square), are chosen to evaluate the prediction performance of the experimental model for data such as gold futures prices that are not very large in scale and scope.

3.4.1 Error and Bias

3.4.1.1 Error

For each data, the error is equal to the forecast minus the actual value (assuming the sales forecast here, so the actual demand)

Error = forecast (f) - demand (d)

$$e_t = f_t - d_t$$

3.4.1.2 Bias

The overall deviation or error of the model is represented by the average value of Error in the test data, which is called Bias. The error's positive and negative values are a drawback. It may be erroneous when summed up since it is offset. The calculation is displayed as negative.

$$Bias = \frac{1}{n}\sum_{n} e_t$$

3.4.2 Mean Absolute Percentage Error (MAPE)

The expected value of the relative error loss serves as this indicator. The percentage between the real value and the absolute error is the so-called relative error. If there is an actual value (demand) that is very low, the percentage will be very large at this time, which will affect the final average value. The percentage has a greater impact. If the model is optimized based on the MAPE value, the overall predicted value will be smaller than the actual value.

$$MAPE = \frac{1}{n}\sum \frac{|e_t|}{d_t}$$

3.4.3 Mean Absolute Error (MAE)

MAE is a better average indicator since it measures the absolute difference between the expected value and the actual value. But there is a disadvantage, that is, because it is not a normalized value, it is not easy to compare. For example, MAE is 10, we don't know whether it is good or bad. If the average of the actual values is 1000, the error is low, but if the actual average is 1, the error is very large. Therefore, when comparing, it is often divided by the average of the actual value to make a normalization, so that it becomes a percentage for easy comparison.

$$MAE = \frac{1}{N}\sum |e_t|$$

$$MAE\% = \frac{\frac{1}{n}\sum |e_t|}{\frac{1}{n}\sum d_t} = \frac{\sum |e_t|}{\sum d_t}$$

3.4.4 Root Mean Square Error (RMSE)

The RMSE is obtained after rooting the MSE, and the root operation makes the error value consistent with the unit of the target variable. For example, when fitting the age, the value of the MSE indicator is the square of the age, and the unit of the RMSE is the age, which maintains the consistency of the dimension.

$$RMSE = \sqrt{\frac{1}{n}\sum e_t^2}$$

3.4.5 R-Squared

The R-square should be as near to 1 as possible. We need to include two additional parameters, SSR and SST, before calculating the coefficient since these two affect the coefficient of determination.

Sum of squares of the regression, or SSR, is the formula for calculating the difference between the mean of the original data and the projected data.

$$SSR = \sum_{i=1}^{n} w_i(\hat{y}_i - \bar{y}_i)^2$$

SST: Total sum of squares, that is, the sum of squares of the difference between the original data and the mean, the formula is as follows:

SST=SSE+SSR, *determination coefficient* is defined as the ratio of SSR and SST, so:

$$R - square = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

In actuality, the coefficient of determination is used to describe how well a match holds up to changes in the data. The usual value range of the "coefficient of determination" is [0 1], according to the expression above. The variables in the equation for y have greater explanatory power the closer it is to 1, and the more closely the model fits the data.

3.5 Chapter Summary

At the beginning of the chapter, it gives the structure how the thesis is organized and states that a quantitative method is employed based on the features of data given. This chapter mainly introduces the method proposed to forecast the gold price. There are

two machine learning methods, including ARIMA and Linear Regression which are viewed as good models for time series forecasting. Finally, MAE, MAPE, RMSE, and R-squared are used to evaluate the accuracy of the model.

Chapter 4: SYSTEM DESIGN IMPLEMENTATION AND TESTING

This chapter will give answer to question 1,2, and 3. The first part comes out the overall proposed solution, and how to achieve this goal after comparing results. The second part explain the prototype which is a concept to the whole design of the project. The last part is EDA processing and accuracy evaluation. The EDA processing depicts how the data performed and will give information for the model used as different types of data applies corresponding models. At the end of the day, the selected model will be used for forecasting and its accuracy will be tested.

4.1 Proposed solution

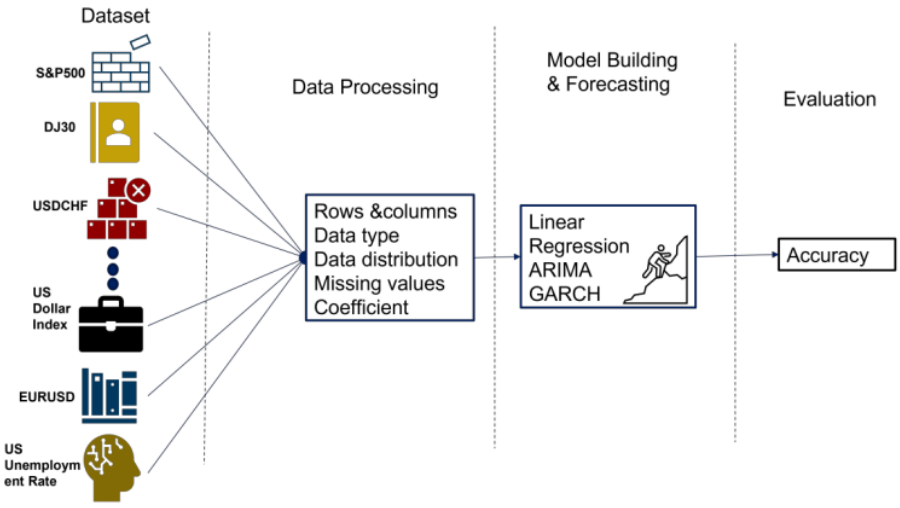


Figure 4.1 Proposed solution

As it is shown in Figure 4.1, the proposed solution for predicting gold prices using linear regression and ARIMA models is to construct a model that combines both methods. The ARIMA model will be used to forecast future gold prices based on the trend and any additional external factors that may have an impact on gold prices. The linear regression model will be used to estimate the trend of gold prices over time.

The model will be built using a variety of inputs, including historical gold price data, economic indicators, and other factors. Once the model is constructed, it will be tested and evaluated using a variety of methods to determine its accuracy and effectiveness. Finally, the model will be deployed and used to make predictions about future gold prices.

4.2 Prototype for Proposed Solution

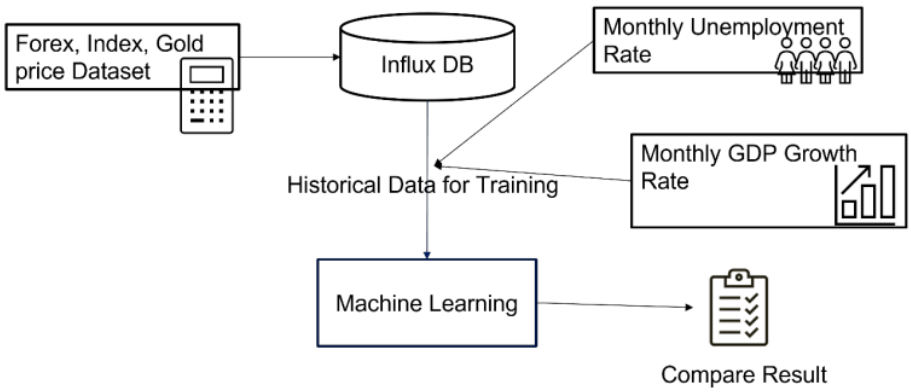


Figure 4.2 Prototype Flow System

The diagram shows the process of simulation system. At first, we use forex data, and index and gold price data during the EDA processing. When it comes to data training, monthly GDP growth rate and unemployment data from 2018 to 2022 are added to the model building. All these data are loaded to be modeled and produce the algorithm accuracy. To validate the accuracy of Linear Regression and ARIMA model, experiments are designed to compare the performance of these models. Furthermore, experiments are set to test parameters of ARIMA model, which has three parameters, namely d,p,q, so as to identify the optimum parameters to produce the best forecasting accuracy.

4.2.1 Data Set

As all the 1-hour candle data is in its availability except the US dollar index on the FOREX broker's website, so in the end, the dataset was changed into using daily candle data ranging from March 26th 2018 to October 14th 2022. Due to the data availability of different trading platforms, all the data are extracted from the following three platforms.

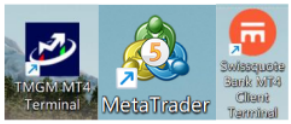


Figure 4.3 FOREX Platforms

The figure shows the trading platforms, namely TMGM MT4 Terminal, MetaTrader, and Swissquote MT4 Terminal.

	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

Figure 4.4 The Contents of Dataset

The figure shows columns of the data collected. All the prices use the daily closing prices to reflect the changes. The original contains 1177 rows and 13 columns in total. For further use, date are divided into three parts, which are year, month, and month-year.

```
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
```

Figure 4.5 Data Types of the Dataset

The figure displays data types of each column. There are 10 columns of float data, one integrated, one period and one datetime. In particular, only the column Time Series are datetime type can it be used for time series forecasting in following contents.

4.2.1.1 Data Processing

The dataset was merged as one file from 16 csv files. And it is updated as of October 14th 2022 considering that one of the US dollar index are not available. These files contain currencies that are directly related to US dollars and some metal categories like silver and WTI crude oil that are affected by changes of US dollars.

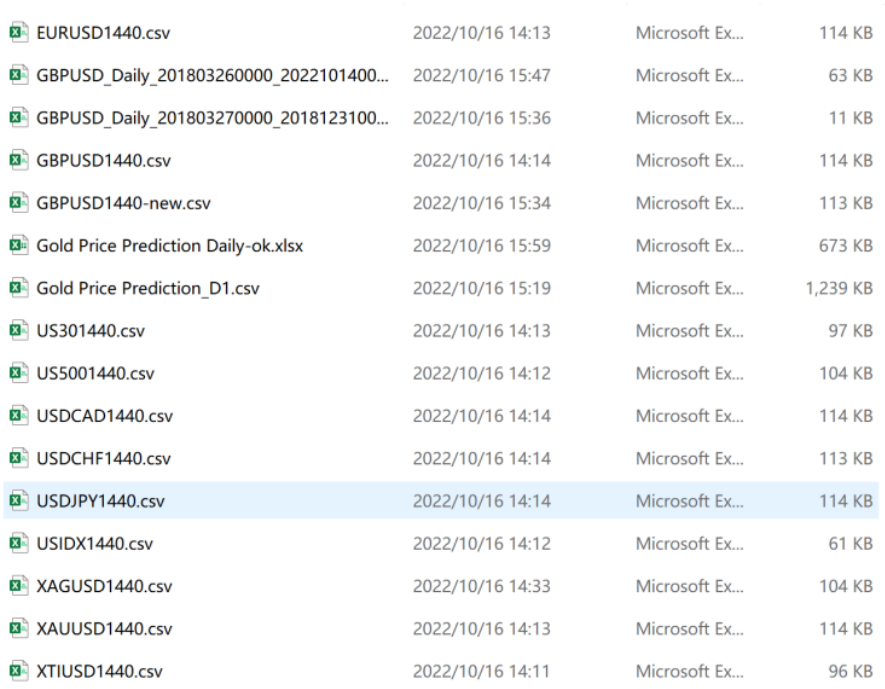


Figure 4.6 Dataset Files

```
: #Checking are there any NaN values are present or not
  df.isna().sum()

: TIME SERIES     0
  XAUUSD_CLOSE    0
  DJ30_CLOSE      0
  USIDX_CLOSE     0
  EURUSD_CLOSE    0
  GBPUSD_CLOSE    0
  SP500_CLOSE     0
  USDCAD_CLOSE    0
  USDCHF_CLOSE    0
  USDJPY_CLOSE    0
  WTIUSD_CLOSE    0
  XAGUSD_CLOSE    0
  dtype: int64
```

Figure 4.7 Null Colums

Figure 4.7 checks whether there are null cells. From the above diagram, no null data is detected.

4.2.1.2 Mapping of Dataset to Proposed Solution

For processing the data better, each variance such as XAUUSD_CLOSE are set to have an independent data frame by using the closing price of each month.

	XAUUSD_Close
Time Serie	
2018/3/29	1325.15
2018/4/30	1315.26
2018/5/31	1298.11
2018/6/29	1252.90
2018/7/31	1223.70
2018/8/31	1200.93
2018/9/28	1191.63
2018/10/31	1214.61
2018/11/30	1222.12
2018/12/31	1282.43

Figure 4.8 Example of Reset Data Frame for Gold

Figure 4.8 demonstrates the new data frame set using the algorithm.

4.2.2 Schema Design

Below are table of data columns, 11 of which are original data, the last three columns are split from column Time Series for the convenience of analysis. There are one datatime type, one period data type, three integers, and ten columns of float data.

Time Serie: researched time span

XAUUSD_Close: the closing price of gold price.

DJ30_Close: the closing price of Dow Jones 30.

USIDX_Close: the closing price of US dollar index

EURUSD_Close: the closing price of Euro versus US dollars

GBPUSD_Close: the closing price of Great Britain Pounds versus US dollars

S&P500_Close: the closing price of Standard & Poor 500

USDCAD_Close: the closing price of US dollars versus Canadian dollar

USDCHF_Close: the closing price of US dollars versus Swiss Franc

USDJPY_Close: the closing price of US dollars versus Japanese Yuen

WTIUSD_Close: the closing price of West Texas Intermedia versus US dollars

XAGUSD_Close: the closing price of silver versus US dollars

month: monthly data split from the column Time Series

year: yearly data split from the column Time Series

month_year: monthly and yearly data split from the column Time Series

	Date	Long Carry	Short Carry
0	2018-3	1.1878%	-0.7170%
1	2018-4	1.0515%	-0.0219%
2	2018-5	0.8373%	-0.0262%
3	2018-6	0.7696%	-0.1457%
4	2018-7	0.9665%	-0.0642%

Figure 4.10 Interest Rate Differential

	Date	eur_unemployment rate	usa_unemployment rate
0	2018-03-01	7.7	4.0
1	2018-04-01	7.6	4.0
2	2018-05-01	7.4	3.8
3	2018-06-01	7.4	4.0
4	2018-07-01	7.3	3.8

Figure 4.11 Unemployment Rate

	month_year	EUR_GDP	USA_GDP	GDP_diff
51	2022-06	100.7	98.4	2.3
52	2022-07	100.8	99.1	1.7
53	2022-08	100.7	98.4	2.3
54	2022-09	100.8	94.0	6.8
55	2022-10	100.7	96.0	4.7

Figure 4.12 GDP Growth Rate

Interest rate and unemployment rate are added for modelling to improve outcome predicting. Figures 4.10 and 4.11 illustrate the interest rate and related figures, Figure 4.12 details the GDP growth rates, and Figure 4.12 compares the unemployment rates in the US and the EU.

4.2.3 Prototype Solution Architecture

The structure contains four main components, which are data loading in to excel files, database storage, forecasting tools including Linear Regression, ARIMA, and data visualization tools by using Python.

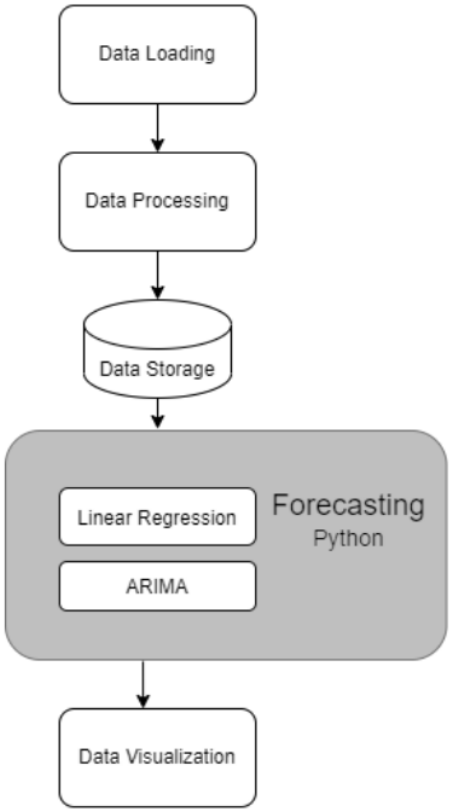


Figure 4.13 Simulation Architecture

4.2.3.1 Data Loader

We use python to analyse data. At first, we load the gold price via pandas library, and then we load other economic indicators into the data. Finally these data are merged together for forecasting.

4.2.3.2 Data Storage

An affordable, scalable alternative to storing files on in-house hard drives or storage networks is cloud storage. You can save data and files with cloud service providers at

an off-site location that you can access via the open internet or a dedicated private network connection. The supplier ensures that you have access to the data whenever we need it by hosting, securing, managing, and maintaining the servers and related infrastructure.

4.2.3.3 Forecasting Component

The forecasting components contains Linear Regression, ARIMA, and GARCH. These forecasting methods are suitable for time series data like stock prices and foreign exchange prices.

For the Linear Regression model, there are multiple linear regression and polynomial regression:

(1) Multiple Linear Regression

To calculate the association between two or more independent variables and one dependent variable, utilise multiple linear regression.

The formula for multiple linear regression takes the following forms:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon$$

Y refers to the predicted value of the dependent variable

B_0 refers to the value of y when all other parameters are set to 0

ε refers to model error

$\beta_n X_n$ refers to regression coefficient of the last independent variable.

(2) Polynomial Regression

In order to quantify the relationship between two variables, simple linear regression (SLR) is performed. However, SLR makes the following mathematical assumption in order to presume that the relationship is linear:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

The relationship between variables is not always linear, though, and this can lead to poorly fitted models. Thus, polynomial regression, which has the following form, can be used to solve this problem.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^1 + \cdots + \beta_2 X^h + \varepsilon$$

In this equation, h refers to the degree of the polynomial. The value h is able to fit nonlinear relationship. In practice, h will be no greater than 3, or 4. Beyond this figure, the model will risk of overfitting and becomes too flexible.

The first stage in the ARIMA algorithm is to determine whether the data is steady. The ADF (Augmented Dickey-Fuller) test and the KPSS (Kwiatkowski-Phillips-Schmidt-Shin) test are two approaches that are prepared for the test.

(1) ADF Test

H0: dataset is not stationary H1: dataset is stationary.

The null hypothesis is accepted if P is more than 0.05 and the statistic value exceeds any of the crucial values. The dataset is therefore non-stationary. The dataset will be stationary if the null hypothesis is not accepted.

(2) KPSS Test

The null hypothesis is rejected if P value is less than or equal to 0.05 and the statistic value exceeds any of the crucial thresholds. The dataset is therefore non-stationary. The dataset is steady if the null hypothesis is believed to be true.

4.2.3.4 Data Visualization

Data visualization makes information presented more understandable. Visual displays include common graphs such as charts, plots, infographics, and even animations.

Charts displayed in the project are drawn through Python libraries such as seaborn and matplotlib.

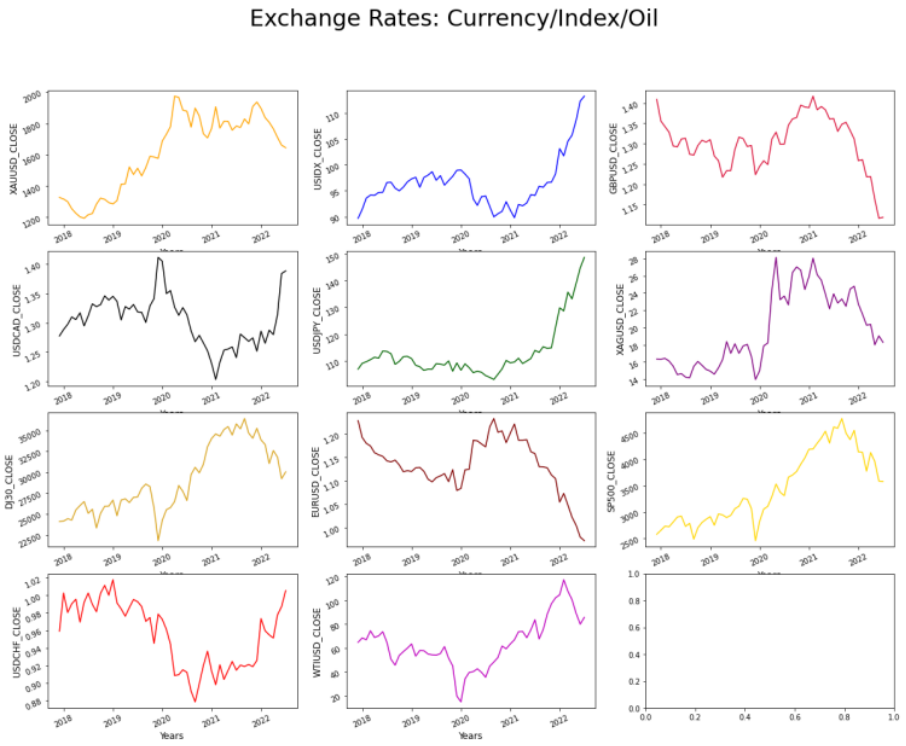


Figure 4.14 Trend of All Variables

Figure 4.14 shows trends of all viables. Obviously, great fluctuations took place during these three years.

4.3 Experiment Implementation

This section shows how the data are performed for time series forecasting using Linear Regression, and ARIMA model.The experiment is conducted to compare 3 months of forecast data. Errors will be used for evaluation model accuracy.

4.3.1 ARIMA Implementation

(1) ADF stationary test

The property of time series known as stationarity in the stationarity test asserts that the value of the variable does not vary over time, i.e., variation in time does not cause changes in the value of a variable.

If it is non-stationary, the results will be skewed due to the presence of price variation brought on by the passage of time.

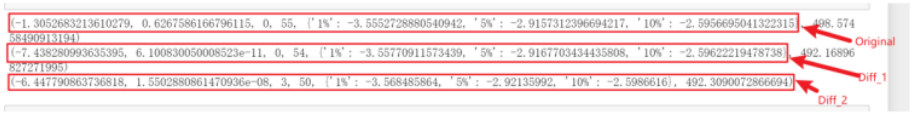


Figure 4.14 ADF Test Result

1%, 5%, and 10% compare the ADF Test result and the statistical value of the null hypothesis to various degrees. The hypothesis is strongly rejected if the ADF Test result is less than 1%, 5%, and 10% all at once. In this instance, the ADF test yields a

p value that reads 0. This value is greater than the values in 1%, 5%, and 10% but less than 0.05. The data are stationary, so we reject the null hypothesis.

Figure 4.14 shows that the t statistic for the ADF is the first value. The p value, which denotes the probability value corresponding to the t statistic, is the second value. The critical ADF test result under the confidence intervals of 99%, 95%, and 90% is the value from the final second. Simply examine the p value and the ADF test value for each of the three confidence intervals.

P value must be less than 0.05; in other words, the lower the number, the better. Any one of the three ADF test values must be less than the t statistic value. Figure 4.14 demonstrates that the first difference and second difference values of the closing gold price can both satisfy the criteria. For further investigation, the goal value has been transformed into stationary data.

(3) Stationary Test Using Seasonal Decompose

A time series is divided into sections using the time series decomposition approach, with each part standing for a pattern category, trend, seasonality, and noise.Long-term patterns, seasonal variations, cyclic fluctuations, and irregular fluctuations typically make up a time. A longer amount of time is required for a phenomena to continue to develop and alter, which is referred to as a long-term trend. Seasonal fluctuations are recurring adjustments in the stage of development of phenomena brought on by seasonal changes. Cyclical fluctuations are recurring,

continuous variations that occur throughout time without following any set criteria. The influence of time series caused by numerous unintentional variables is referred to as irregular fluctuation.

An additive model and a multiplicative model are further divisions of the decomposition model. The four components of the time series have the same dimension and are independent of one another thanks to addition. The seasonal, cyclic, and irregular variation items are all independent random variables that follow a normal distribution, and the output section of the multiplication model has the same dimension as the trend item.

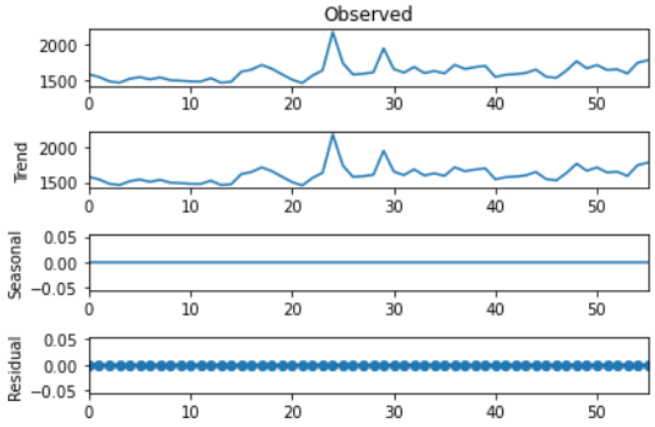


Figure 4.16 Seasonal Decompose

Figure 4.16 depicts the time series decompose to draw the seasonal, overall trend, and residual changes in the time span. The trend reaches a peak between 20 and 30 months, while the seasonal data as well as residual data stay around zero.

(4) Autocorrelation Function(ACF)

The ACF provides the autocorrelation value for any series with lag values. It is a complete autocorrelation function. It simply describes how closely the sequence's present value and its previous values are correlated. The elements of trend, seasonality, periodicity, and residuals are examples of time series components.

All of these factors are taken into account by ACF while searching for correlations. ACF provides information on both direct and indirect correlations when describing the autocorrelation between two observations.

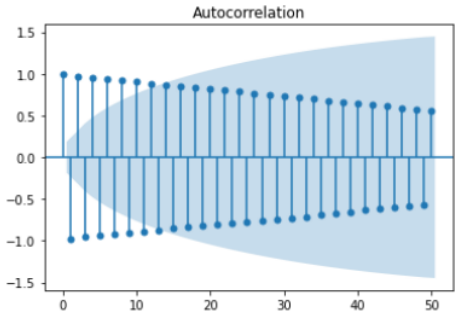


Figure 4.17 ACF

The error band is indicated by the blue area, and the data there are thought to be inconsequential. The data autocorrelation before 12 months, as shown in Figure 4.17, is not substantial; the data are negatively correlated every 2 months and positively correlated every 2 months. A pattern in which the amplitude of the ACF is gradually reduced generally means that there is an autocorrelation, and over which periods of the autocorrelation, PACF can help us find it.

(5) Partial Autocorrelation Function(PACF)

With the exception of one variable, the P in PACF stands for partial, which has the same meaning as the P in the partial derivative. While PACF just takes into account the correlation of one particular period, ACF takes into account the correlation of all periods, which offers us a decent place to start when figuring out the autocorrelation period.

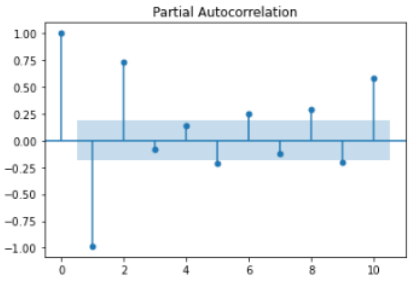


Figure 4.18 PACF

In the above Figure 4.18, for example, 1, 2, 10 can be used as candidates for the autocorrelation time span. When determining the AR part (p) of the ARMA model, consider the value that is not 0 in the PACF figure (the unexpected value of Error Band in the blue area, below the same number of cycles). Similarly, when determining the MA part (q) of the ARMA model, consider the number of cycles in the ACF graph that are not zero.

(6) Choosing Parameters for ARIMA

From the above PACF graphs, we may know the range of parameter p. 1, or 2 is beyond the blue area, which could be chosen as candidates for parameters. The ACF diagram can be used to decide the parameter q. However, the blue color covered area

is considered as insignificant. Thus, candidates for q are 1, 2, and 3. However, only when the parameter p is set as 4, q that is equal to 3 is stationary.

Parameters for ARIMA		
p	**d**	**q**
1	0	1
1	0	2
2	0	1
3	0	1
4	0	1
4	0	2
4	0	3

Table 4.1 Table 4.1 Parameters for ARIMA

Table 4.1 lists six types of combines of parameters in the ARIMA model. Parameter d is always set as 0.

ARMA Model Results

Dep. Variable:	FX	No. Observations:	56
Model:	ARMA(2, 2)	Log Likelihood	-341.426
Method:	css-mle	S.D. of innovations	106.486
Date:	Wed, 02 Nov 2022	AIC	694.852
Time:	10:37:30	BIC	707.004
Sample:	0	HQIC	699.564

	coef	std err	z	P>|z|	[0.025	0.975]
const	1613.0664	20.212	79.806	0.000	1573.451	1652.682
ar.L1.FX	-0.0456	0.165	-0.277	0.782	-0.369	0.277
ar.L2.FX	-0.5418	0.201	-2.695	0.007	-0.936	-0.148
ma.L1.FX	0.4369	0.119	3.678	0.000	0.204	0.670
ma.L2.FX	0.8250	0.143	5.769	0.000	0.545	1.105

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	-0.0421	-1.3579j	1.3586	-0.2549
AR.2	-0.0421	+1.3579j	1.3586	0.2549
MA.1	-0.2648	-1.0687j	1.1010	-0.2887
MA.2	-0.2648	+1.0687j	1.1010	0.2887

Figure 4.19 ARIMA Model Testing

P-values are all 0, and all are significant.

4.3.2 Linear Regression Implementation

The model will be built using a variety of inputs, including historical gold price data, economic indicators, and other factors. Once the model is constructed, it will be tested and evaluated using a variety of methods to determine its accuracy and effectiveness. Finally, the model will be deployed and used to make predictions about future gold prices.

4.4 System Testing

System testing is performed to ensure that the data are captured from preprocessing phase correctly into the system.

4.4.1 Data Exploration

Fluctuations of Variables

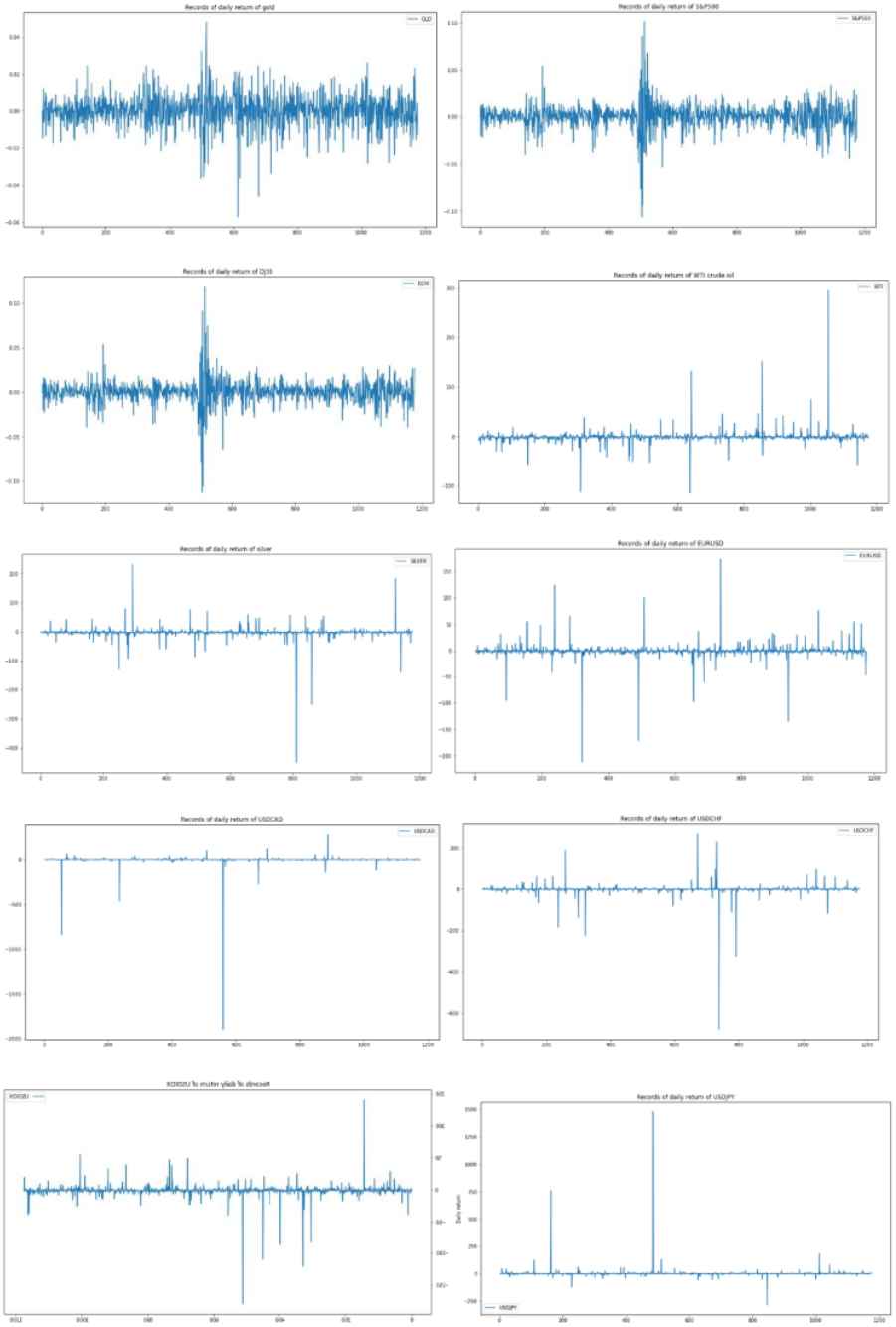
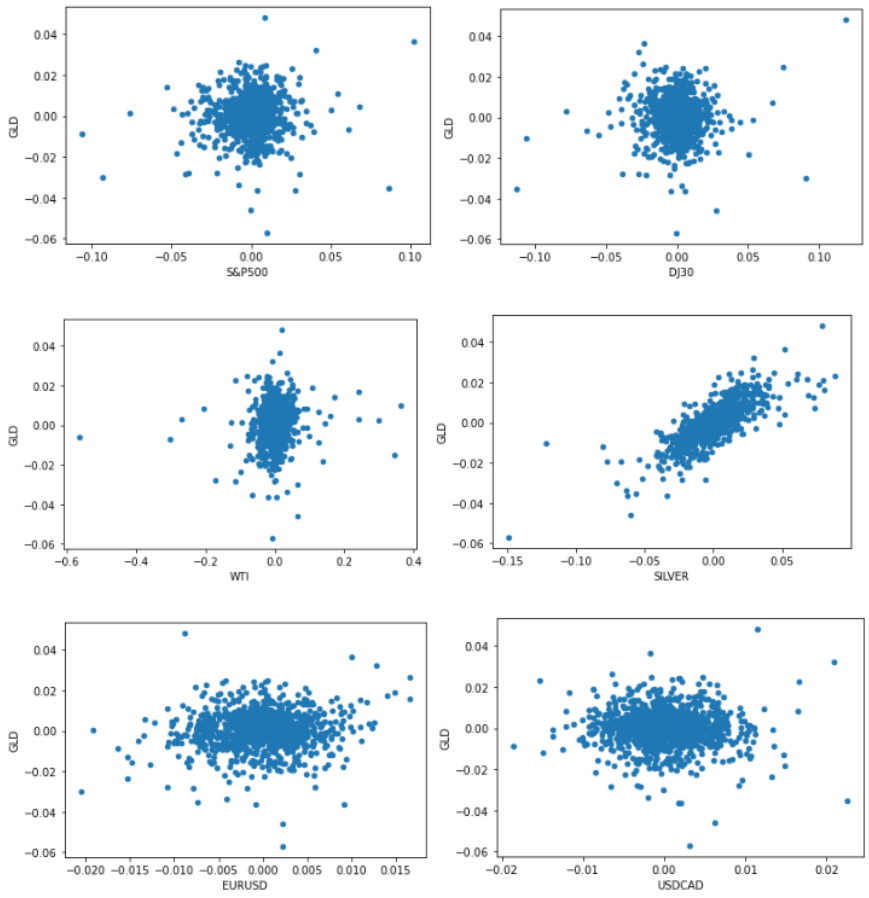


Figure 4.20 Fluctuations of Variables

Figure 4.20 shows the fluctuations of variables. From the graph, we have two findings:1)gold and DJ 30 and S&P 500 are most frequent traded currency or indicators. 2)other currencies like USDJPY, USDCHF are less traded by investors.

Scatter plot

From the scatter plot, we have one significant finding: the silver have positive correlation with gold price, while US index has negative correlation with gold price. However, other variables show no significant trend with gold price,



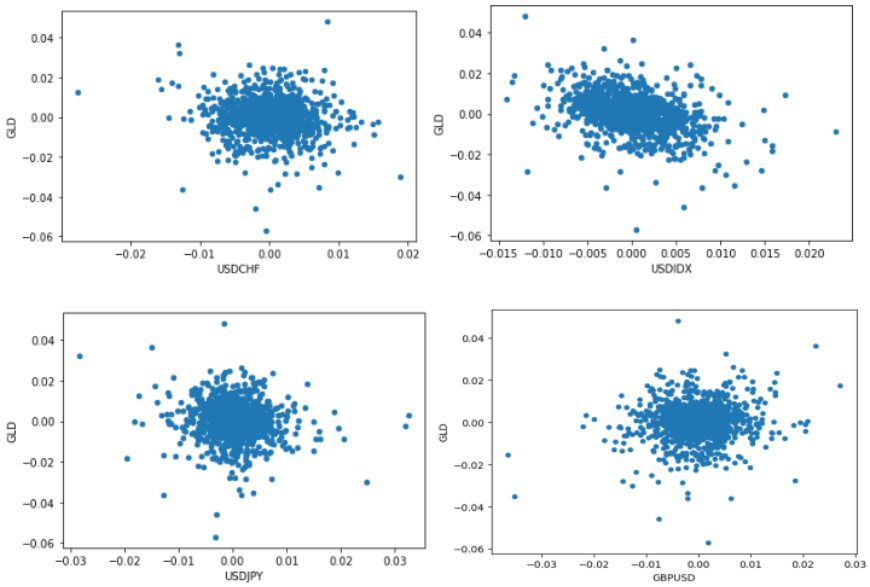


Figure 4.21 Scatter Plot of Variables

4.4.2 Data Exploration Conclusion

The scatter plot of the data shows some locations where there are positive correlations between gold and the US dollar index and between gold and silver, while other variables show no discernible correlations. According to the gold price fluctuation chart, the open-close price and high-low price both change by about 50 pip.

With regard to the daily return of all factors, the gold run increases by around 0.4 percent while decreasing by about 0.6 percent. In terms of daily return, S&P 500 and Dow Jones 30 are second only to gold. Other currency pairs and the index have not changed significantly.

4.5 Conclusion

This chapter explains how experiment is designed and how the model is set to execute the program. Data exploration helps to acknowledge how data is organized and to know some attributes of each variable. For example, the open-close price improves our understanding that to what extent the gold price will surge and dive.

In the model building, this chapter introduce in detail how the pre-steps are set in order to run the model. The most typical case is the ARIMA model. The model will experience stationary test and choose the most appropriate parameter for the formula before build the model in case of overfitting or underfitting. There are two parameter options, one of which is 1, 0, and 1 for parameter d, p, and q respectively. Another is 2, 0, 1. However, the best is 1,0,1 for the model.

Chapter 5 Result and Discuss

This chapter will illustrate the experiment setup, the dataset used for the experiment. Then we will go through the Linear Regression model, and ARIMA model. The result of both models will be visualised and their performance will be estimate and using RMSE, and MAE etc. Besides, as the performance of the model is assessed by accuracy, thus improvement for the model, including some parameters will be adjusted to deliver the desirable result.

5.1Experiment Setup

Table 5.1 shows the dataset used for the experiment. The period used for determining the dataset and the experimental period are listed as below.

Data Set	Start Datetime	2022-07-01 00:00:00
	End Datetime	2022-10-14 00:00:00
Training Data Set Range	Start Datetime	2022-10-01 00:00:00
	End Datetime	2022-12-10 00:00:00
Testing Data Set	Start Datetime	2022-12-10 00:00:00
	End Datetime	2023-01-10 00:00:00

Table 5.1 Dataset Used for the Experiment

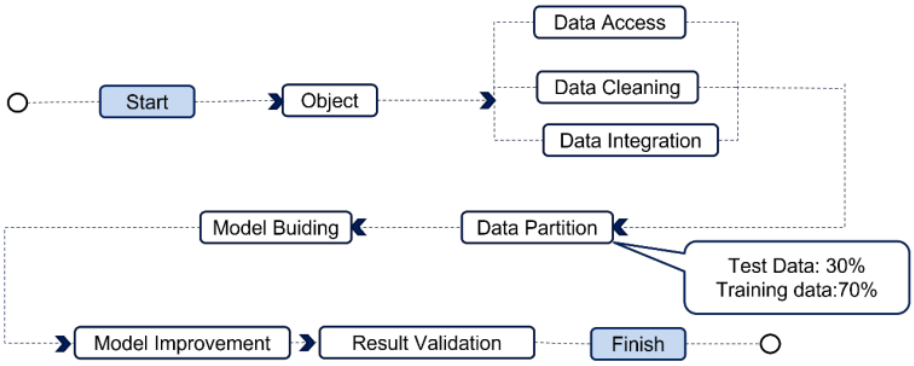


Figure 5.1 The experiment Setup

The experiment setup shows how the experiment is designed. The goals of the subject should be set at first. After that, we access data, clean data, and integrate data, use data partition strategy to separate dataset. Models will be built based on the processed data, the performance of models will be assessed and improved to deliver the desirable result.

5.1.1 Data Partition Strategy

For data partition strategy, as the data set applies to time series analysis, it will add bias to the result if we use random sample method to segregate the training data set and test data set. For this consideration, the first 70% of the dataset will be set as training data, while the rest data as test data.

The machine learning method will load historical data for training and creating model. Based on the historical data and the model, it will forecast the value. For the testing dataset, actual value will be used for the comparison with forecast value. Error is to be calculated between the forecasted value and actual value.

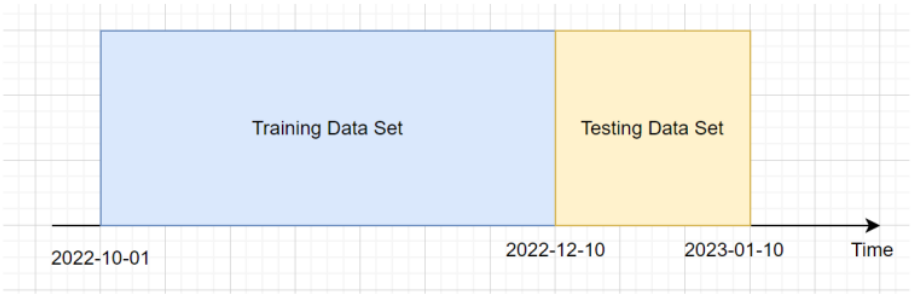


Figure 5.2 Data Partition Strategy

5.2 Experiment Result

This section aims to show the result of forecasting models, namely Linear Regression and ARIMA model. Observations and conclusions come from this section.

5.2.1 Linear Regression Result

Three different forms of linear regression models—Simple Linear Regression, Multiple Linear Regression, and Polynomial Linear Regression—are used to predict data using the linear regression model.

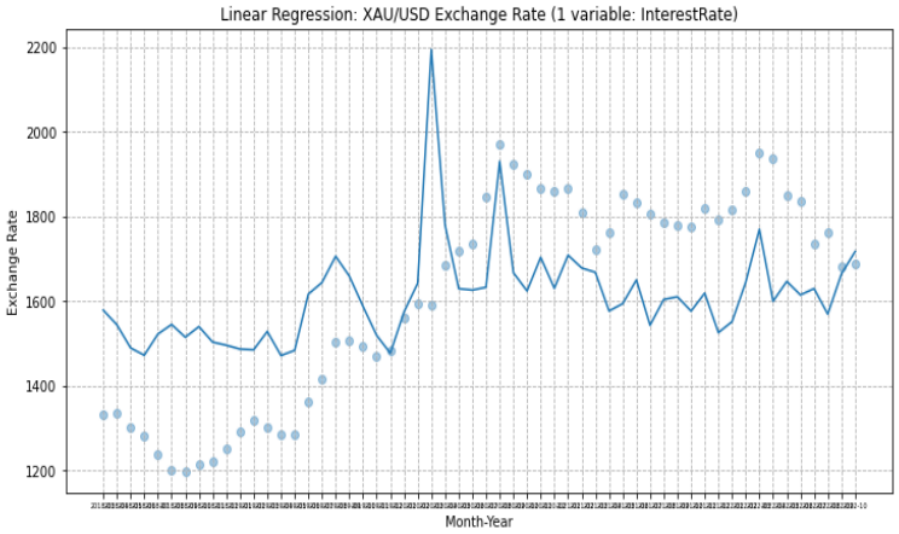


Figure 5.3 Linear Regression with one Variable

The simple linear regression model is displayed in Figure 5.3. The dotted line displays the predicted value, whereas the blue line displays the actual data. The US interest rate is the predicting variable.

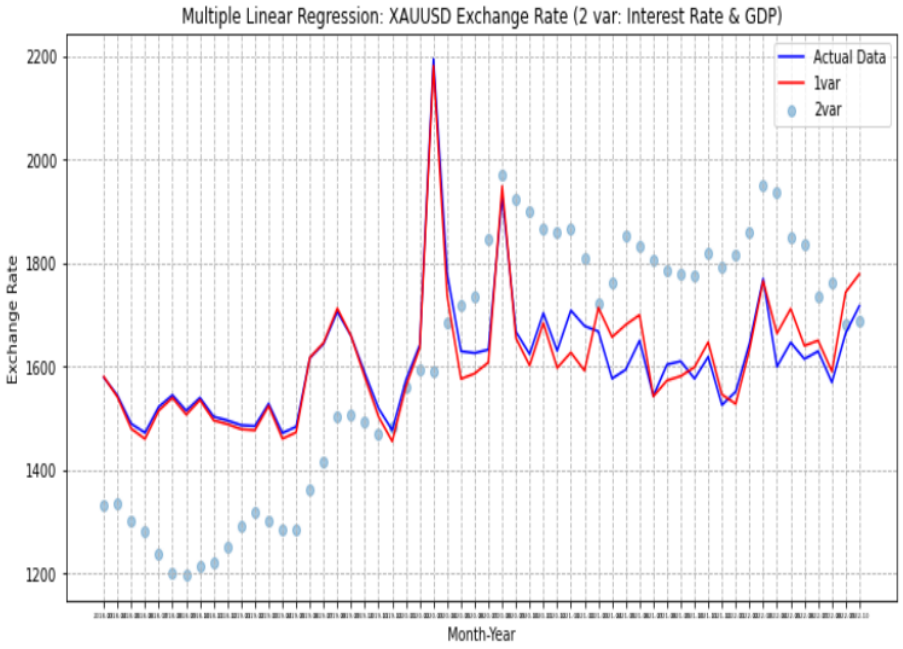


Figure 5.4 Linear Regression with Two Variable

The multiple linear regression model with two variables—the US interest rate and the US monthly growth rate—is depicted in Figure 5.4. The dotted line shows the US monthly interest rate, the red line the US monthly predicted GDP value, and the blue line the actual number. The projected value is near to the actual value, however in a later section of the section, the RMSE, MAE, and MAPE must be used to evaluate accuracy.

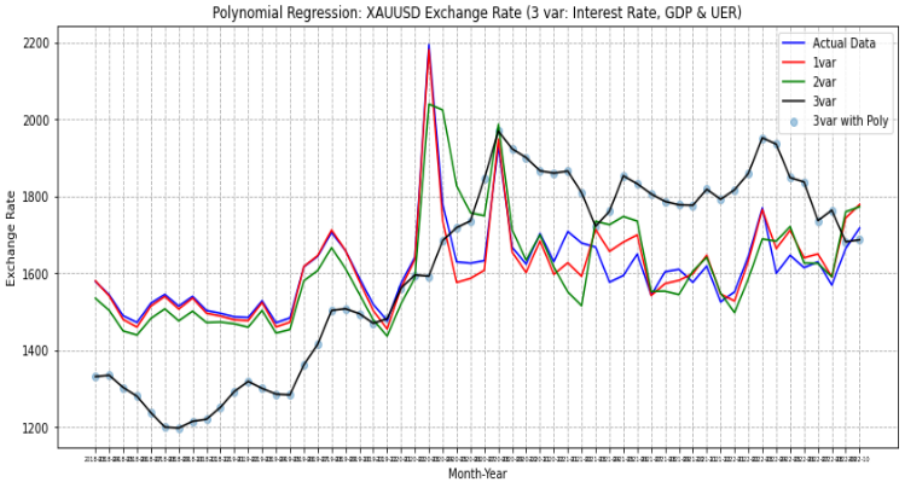


Figure 5.5 Linear Regression with Three Variable

Polynomial Linear Regression prediction involved three variables. The actual number is shown by the blue line, the US monthly GDP growth rate by the red line, the US unemployment rate by the green line, and the US monthly interest rate by the dotted line. In terms of interest rate, there is obviously a large discrepancy between the real value and the forecasted value.

5.2.1.1 Experiment Summary Linear Regression

The results give an overview on the accuracy of Linear Regression model. When it is added more factors into the historical data, the model also performs well. The interest rate element appears to have little impact on the gold price according to this model, as seen by the significant price discrepancy between anticipated and actual values.

5.2.2 ARIMA Result

The output of the ARIMA model is detailed in this section. Given the intricacy of the model, some model implementation and model change will also be demonstrated in this section.

5.2.2.1 Experiment Summary ARIMA

The first step for this model is to test whether the data is stationary. If it is not, we need to transfer the data into stationary one to avoid bias via differencing or logarithmic method.

Figure 5.6 shows the result after differencing. The orange line is the result after first differencing, while the green line is the second differencing. Obviously, the data after second differencing seems steadier than the first differencing. But whether we adopt the first differencing method or second differencing depends on the following step: white noise test.

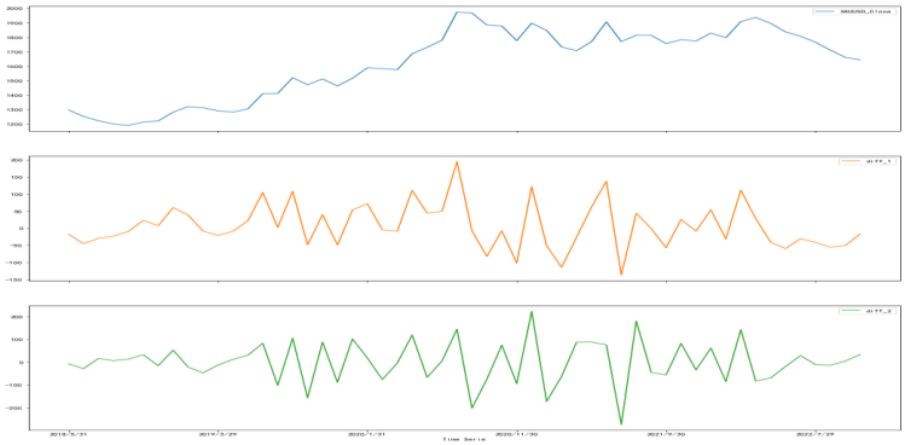
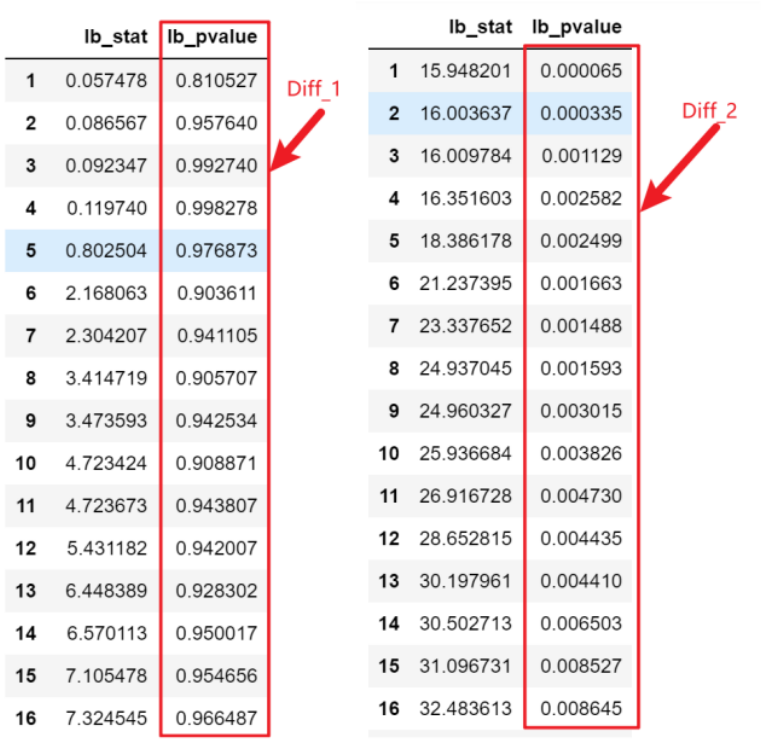


Figure 5.6 Data via Differencing Method



	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

	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

Figure 5.7 White Noise Test Result

The white noise test is conducted by to data set. This process produce the p value. The smaller the p value the better. The left side of Figure 5.7 is the result of white noise test after first differencing, while the right is the result after second differencing. It can be noticed that the p value after second differencing is much smaller than the one after first differencing. So the data set after second differencing will be adopted in later model building process.

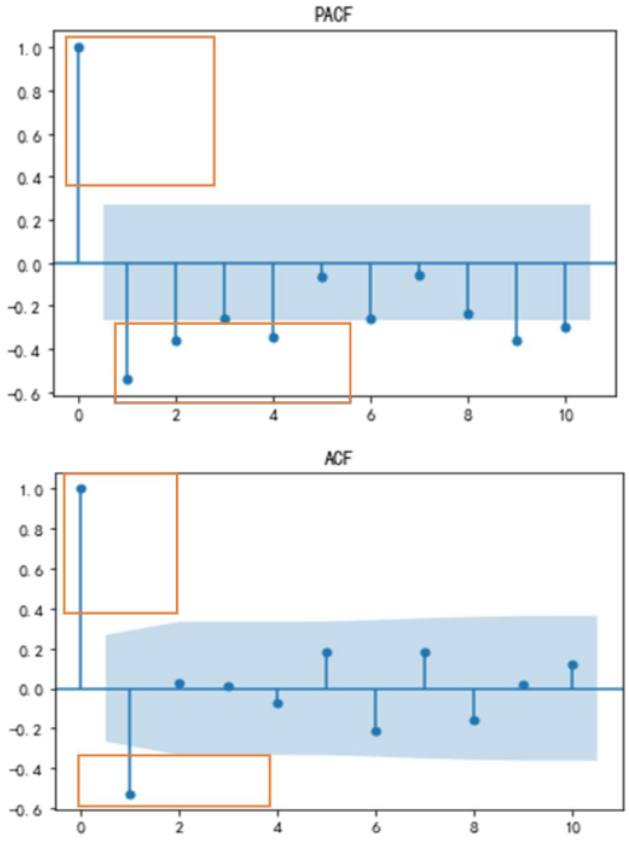


Figure 5.8 Parameter Choosing

There are three parameter for ARIMA model, namely p, n, and q. What we need to choose for the model is p and q. PACF is used to choose the parameter q, while ACF determines q. The blue color covered area is considered as insignificant. Beyond the area, it could be chosen as candidates for parameters. From Figure 5.8, it can be seen that red rectangular is the desirable area for parameter choosing. But it cannot be accurate if we just use naked eyes. Thus, an algorithm is designed to figure out the best p and q for the model. Figure 5.9 shows the heatmap for p and q. we choose a relative high value with the corresponding p and q value is 1 and 1 respectively.

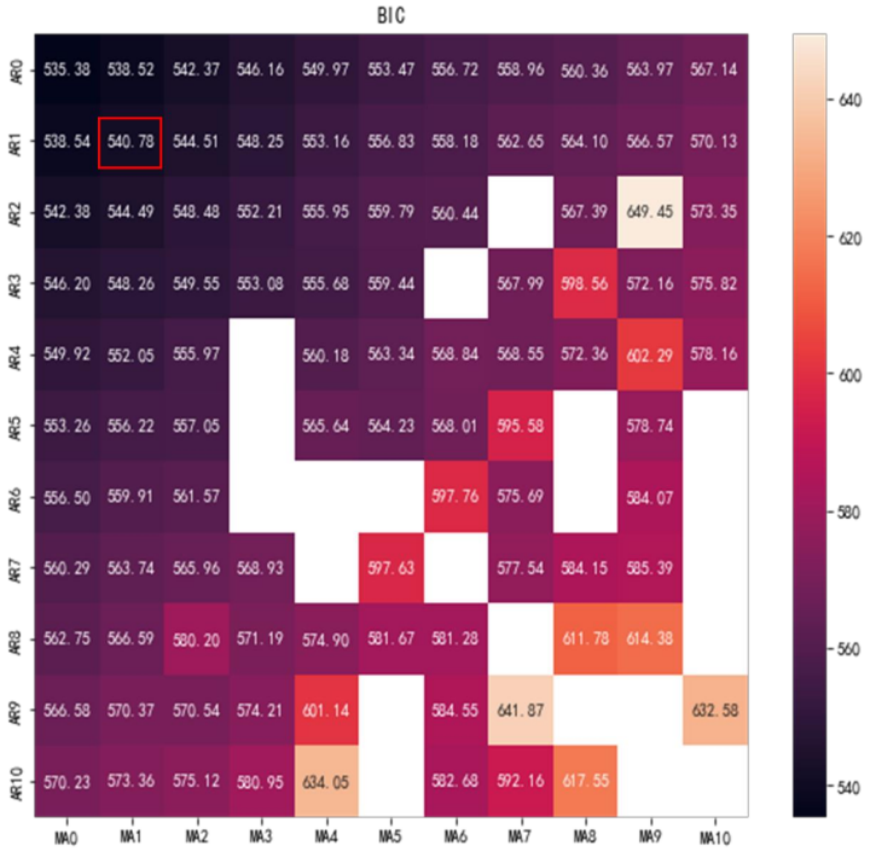


Figure 5.9 Heatmap for Parameter Choosing

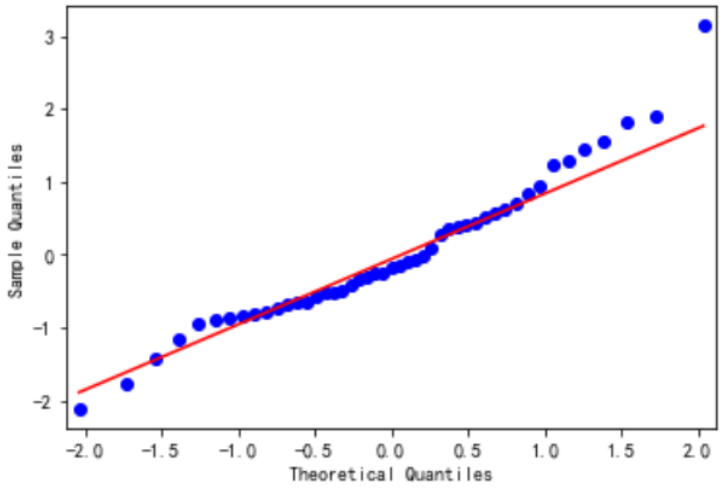


Figure 5.10 Residual Test

The residual also needs to pass the white noise test. It can pass the residual test unless two requirements are met. The first is normality test, and the second is the autocorrelation test. The normality result (p value) by using shapiro method reads 0.08(p value is acceptable when it exceeds 0.05). The Durbin-Watson is employed to conduct autocorrelation test. The value for autocorrelation test is 1.686 close to 2, which is acceptable in this instance.

```
ShapiroResult(statistic=0.9576875567436218, pvalue=0.08140654116868973)
```

Figure 5.11 The Normality Test Result

5.2.3 Experiment Summary ARIMA

From the above steps, we are able to make data become stationary, to conduct white noise test, to use algorithm to choose the best parameter for the model. Finally, the

residual also passes the white noise test. After all these steps, the model is ready to be built based on historical data. And a visualized graph is given depicting the predicted value and actual value.

Adopting a propriate parameter is no such a easy thing as it correlated with the performance of the model. Thus after several trials for parameters, we are able to choose the best pair parameter. Actually, the parameter choosing using algorithm takes a quite long time for each trial. The goal for the model is thus achieved through trials.

5.2.4 Linear Regression vs ARIMA

Based on the modelling process of two models, the result of both models are placed side by side for comparison.

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

Table 5.2 Comparison between Linear Regression and ARIMA

From the comparison table, we can see ARIMA outperforms Linear Regression from three indicators.

5.3 Forecasting Result Visualization

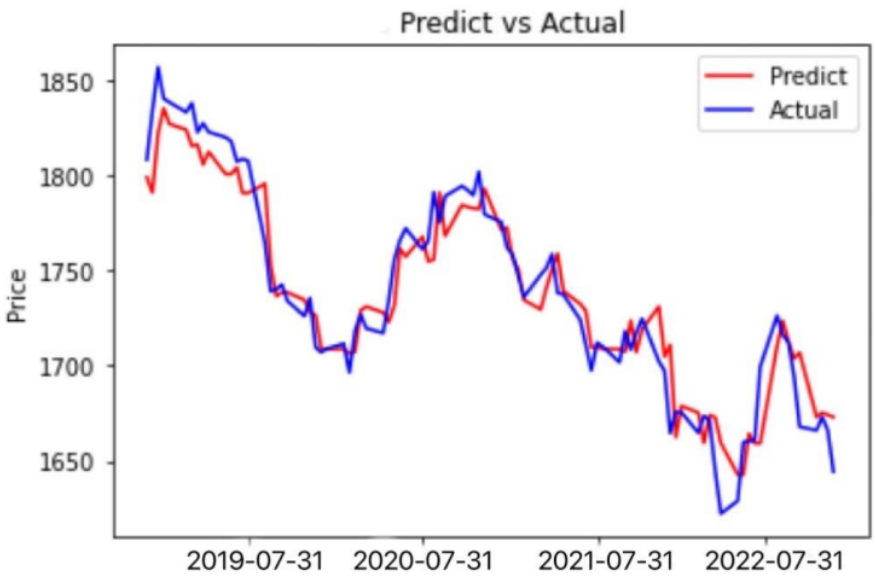


Figure 5.12 The Result of ARIMA Model

The proposed solution for predicting gold prices using linear regression and ARIMA models will be evaluated by creating a visualization of the results. This visualization will include a plot of the predicted gold prices and the actual gold prices over time. The plot will also include a line that shows the trend of the gold prices and any anomalies or fluctuations in the prices. This will allow for a quick comparison of the model's accuracy and effectiveness in predicting gold prices. Additionally, the plot will display any external factors that may have affected the gold prices. The visualization will be useful for assessing the performance of the model and can be used to identify any areas for improvement.

5.4 Experiment Summary and Discussion

5.4.1 Historical Data Size

Experiment shows that with small data, ARIMA yield better result than Linear Regression. If more data are given, Linear Regression outperforms the forecast result of ARIMA.

5.4.2 Training Time and Reusability of Model

The advantage that Linear Regression outperforms the ARIMA is its reusability for any time series data given. In the ARIMA model, it's not applicable for any types of data for forecasting. ARIMA involves quite a lot of steps. Each step needs to pass some corresponding test until it goes to the next step. In this instance, ARIMA requires much more time than Linear Regression in the process of parameter choosing and white noise test. In times of network traffic, ARIMA can be a trouble to react with the new forecasting.

5.5 Chapter Summary

In this chapter, we discussed results of two models. To make it more understandable of the model, we provide more detail explanation of how ARIMA model is implemented.

Each model chose for prediction experienced a lot of consideration. As our data set applies to time series forecasting, some time series models need to be considered in this model. Thus we chose Linear Regression and ARIMA model. Linear Regression

is a more universal method for numerical data, and also it is suitable for stock foreign exchange prices prediction. We want it to compare with a more specific forecasting model-the ARIMA to see which one performs better. As Linear Regression is a universal method for forecasting, it can be easy to implement the process. We suppose several factors, namely unemployment rate, monthly GDP interest rate, impose impact on gold price. So we employ simple, multiple, and polynomial linear regression to see which one weights most. As a result, the interest rate drop the queue, its forecasting performance worse than other factors.

Despite the complex process of ARIMA, we show how each significant process is figured out. Time also needs to spent in this model compared with Linear Regression model. Due to the stationary process, the data set that is originally not stationary needs to be change into stationary using differencing method. The second differencing method is the better choice for next step- the white noise test. The stationary data need pass the white noise test until it goes to choose two important parameters. The parameter choosing is also conducted via algorithm. It the parameter choosing that takes a lot of running time. The last step before the model is ready to be built is that the residual white noise test.

Finally, the two models were put side by side for comparison. Evaluation methods such as MAE, MAPE, MSE for numeric data are employed in the process.

Chapter 6: THESIS SUMMARY

6.1 Summary

By examining the gold prices during the previous ten years, this thesis seeks to investigate the precision of linear regression and ARIMA models for gold price prediction. The study will evaluate the accuracy of both models' predictions of gold prices and pinpoint their advantages and disadvantages. The research will also include the creation of a data mining programme to precisely forecast future gold prices. The findings of this thesis will shed light on which model is more effective at forecasting gold prices as well as the efficacy of the created data mining method. In the end, this study aims to offer useful knowledge on how to accurately forecast gold prices in the future.

This thesis examines the use of linear regression and ARIMA models for predicting gold prices. It begins by introducing the concept of gold as an asset class and its historical price movements. The thesis then looks at the different types of predictive models, such as linear regression and ARIMA, that can be used to forecast gold prices. It explains how these models work and compares their effectiveness in predicting gold prices. The thesis then explores the data and presents a model for predicting gold prices using linear regression and ARIMA. Finally, the thesis presents

the results of the model and discusses the implications of the findings. The thesis concludes by providing recommendations for further research and development.

6.2 Thesis Contribution

The research contribution of this thesis is to provide insight into the accuracy of gold price prediction using linear regression and ARIMA models by analysing the past 3 years of gold prices. Furthermore, the research contributes to the development of a data mining algorithm which accurately predicts gold prices in the future. The results of this research will provide valuable information on the strengths and weaknesses of both models and which one is best suited for predicting gold prices in the future. This thesis will also provide valuable insight into the effectiveness of the developed data mining algorithm and how it can be used to accurately predict future gold prices.

The technical contribution of this thesis is to provide insight into the accuracy of gold price prediction using linear regression and ARIMA models. The research will explore the strengths and weaknesses of both models in predicting gold prices and how they can be used to accurately predict future gold prices. The research will also involve the development of a data mining algorithm to accurately predict gold prices in the future. The results of this research will provide insight into which model is best suited for predicting gold prices and the effectiveness of the developed data mining algorithm. Ultimately, this research seeks to identify the most suitable model for predicting gold prices in the future.

6.3 Future Work Suggestion

For future work, there are 3 areas for further exploring. They are:

- quantitative analysis of fundamentals by machine learning methods
- gold price forecast based on improved BP Neural Network
- establish an early warning mechanism for gold price fluctuations.

In the first area, fundamental quantitative investment, which combines quantitative investment (computer-driven) and value investment (human-driven), is an intelligent quantitative investment method that has attracted much attention in recent years.

As one of the innovative forms of financial technology, intelligent quantitative investment can greatly improve the efficiency of asset management by carrying out asset management business through artificial intelligence technology, and is becoming an important part of the high-quality development of the financial industry.

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

For the last area, as the deterioration of the macroeconomic situation will increase the volatility of the gold market, it will help prevent investors from facing greater market risks. Gold price fluctuations have brought price fluctuation risks to gold market participants, including the central bank as a country and ordinary investors as individuals. In order to effectively avoid the risks brought about by fluctuations in

gold prices, this section intends to establish a risk warning mechanism for risk prediction and prevention.

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.

[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.

[11]Kumar, V. (2018). Predicting gold prices in India using linear regression. *International Journal of Applied Business and Economics Research*, 16(1), 653-669.

[12]Vadlamani, S. (2017). Forecasting gold prices using linear regression and ARIMA models. *International Journal of Applied Business and Economics Research*, 15(1), 651-671.

[13]Li, X., Lu, Z., & Li, Y. (2016). Gold price prediction using linear regression and ARIMA

Gold Price Prediction Using Linear Regression and ARIMA Model

ORIGINALITY REPORT

13%	6%	7%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Yanrui Ning, Hossein Kazemi, Pejman Tahmasebi. "A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet", Computers & Geosciences, 2022 Publication	1%
2	Submitted to Higher Education Commission Pakistan Student Paper	<1%
3	Submitted to Bahrain Institute of Banking and Finance Student Paper	<1%
4	www.mdpi.com Internet Source	<1%
5	"Machine Learning for Cyber Security", Springer Science and Business Media LLC, 2023 Publication	<1%
6	Submitted to King's College Student Paper	<1%

7	Submitted to Queen's University of Belfast Student Paper	<1%
8	Submitted to University of Glasgow Student Paper	<1%
9	Submitted to Liverpool John Moores University Student Paper	<1%
10	Submitted to University of Southampton Student Paper	<1%
11	khatabook.com Internet Source	<1%
12	Submitted to Aston University Student Paper	<1%
13	Submitted to University of Sussex Student Paper	<1%
14	Submitted to RMIT University Student Paper	<1%
15	Fenghua Wen, Minzhi Zhang, Mi Deng, Yupei Zhao, Jian Ouyang. "Exploring the dynamic effects of financial factors on oil prices based on a TVP-VAR model", Physica A: Statistical Mechanics and its Applications, 2019 Publication	<1%
16	Submitted to essex Student Paper	<1%

17	Submitted to Piri Reis University Student Paper	<1 %
18	Submitted to University of Malaya Student Paper	<1 %
19	Submitted to La Trobe University Student Paper	<1 %
20	Guoyan Huang, Xinyi Li, Bing Zhang, Jiadong Ren. "PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition", Science of The Total Environment, 2021 Publication	<1 %
21	Submitted to University of Nottingham Student Paper	<1 %
22	www.diplomarbeiten24.de Internet Source	<1 %
23	www.diva-portal.se Internet Source	<1 %
24	Submitted to Heriot-Watt University Student Paper	<1 %
25	Jian Chai, Chenyu Zhao, Yi Hu, Zhe George Zhang. "Structural analysis and forecast of gold price returns", Journal of Management Science and Engineering, 2021 Publication	<1 %

26	Submitted to University of South Africa Student Paper	<1 %
27	stax.strath.ac.uk Internet Source	<1 %
28	www.researchgate.net Internet Source	<1 %
29	Liya A, Qian Qin, Hafiz Waqas Kamran, Anusara Sawangchai, Worakamol Wisetsri, Mohsin Raza. "How macroeconomic indicators influence gold price management", Business Process Management Journal, 2021 Publication	<1 %
30	Submitted to University of Newcastle upon Tyne Student Paper	<1 %
31	Qian Ding, Jianbai Huang, Wang Gao, Hongwei Zhang. "Does political risk matter for gold market fluctuations? A structural VAR analysis", Research in International Business and Finance, 2022 Publication	<1 %
32	link.springer.com Internet Source	<1 %
33	mantechpublications.com Internet Source	<1 %
	repositorio.iscte-iul.pt	

34	Internet Source	<1%
35	Lingxiao Zhao, Zhiyang Li, Leilei Qu. "Forecasting of Beijing PM2.5 with a hybrid ARIMA model based on integrated AIC and improved GS fixed-order methods and seasonal decomposition", Heliyon, 2022 Publication	<1%
36	acspublisher.com Internet Source	<1%
37	Submitted to University of Greenwich Student Paper	<1%
38	archive.cm.mahidol.ac.th Internet Source	<1%
39	dias.library.tuc.gr Internet Source	<1%
40	espace.curtin.edu.au Internet Source	<1%
41	f1000research.com Internet Source	<1%
42	fedetd.mis.nsysu.edu.tw Internet Source	<1%
43	repository.tudelft.nl Internet Source	<1%

44	"Trading with Intermarket Analysis", Wiley, 2012 Publication	<1 %
45	Laor Boongasame, Piboonlit Viriyaphol, Kriangkrai Tassanavipas, Punnarumol Temdee. "Gold-Price Forecasting Method Using Long Short-Term Memory and the Association Rule", Journal of Mobile Multimedia, 2022 Publication	<1 %
46	Submitted to RDI Distance Learning Student Paper	<1 %
47	Submitted to The University of Manchester Student Paper	<1 %
48	Submitted to University of Reading Student Paper	<1 %
49	research.vu.nl Internet Source	<1 %
50	Submitted to Hong Kong Baptist University Student Paper	<1 %
51	Submitted to International Medical University Student Paper	<1 %
52	Submitted to University of Cape Town Student Paper	<1 %
53	Submitted to University of Western Australia Student Paper	

		<1 %
54	jafmonline.net Internet Source	<1 %
55	repository.londonmet.ac.uk Internet Source	<1 %
56	Submitted to The Robert Gordon University Student Paper	<1 %
57	Submitted to University of Lancaster Student Paper	<1 %
58	Submitted to University of Liverpool Student Paper	<1 %
59	academic.csuohio.edu Internet Source	<1 %
60	Submitted to Bournemouth University Student Paper	<1 %
61	Submitted to Kaplan International Colleges Student Paper	<1 %
62	Submitted to Leeds Beckett University Student Paper	<1 %
63	Submitted to London School of Commerce Student Paper	<1 %
64	Submitted to University of Leeds Student Paper	<1 %

65	pure.coventry.ac.uk Internet Source	<1 %
66	rucore.libraries.rutgers.edu Internet Source	<1 %
67	www.strategicgold.com Internet Source	<1 %
68	Submitted to Napier University Student Paper	<1 %
69	Submitted to University of Hull Student Paper	<1 %
70	Submitted to University of Wales Swansea Student Paper	<1 %
71	Submitted to University of Wollongong Student Paper	<1 %
72	analyticsindiamag.com Internet Source	<1 %
73	core.ac.uk Internet Source	<1 %
74	erepository.uonbi.ac.ke Internet Source	<1 %
75	Submitted to HELP UNIVERSITY Student Paper	<1 %
76	Yang Li, Xiaojun Shen, Chongcheng Zhou. "Dynamic multi-turbines spatiotemporal	<1 %

correlation model enabled digital twin technology for real-time wind speed prediction", Renewable Energy, 2022
Publication

77	helion.pl Internet Source	<1 %
78	www.nature.com Internet Source	<1 %
79	"Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications", Springer Science and Business Media LLC, 2021 Publication	<1 %
80	"Genetic Programming Theory and Practice XII", Springer Science and Business Media LLC, 2015 Publication	<1 %
81	Lin Xu, Zhenwei Guo. "Effect of Regulation on the Increasing Price of Metals and Minerals to Meet the Challenges in Clean Energy Transitions: A Case Study of China", Sustainability, 2022 Publication	<1 %
82	László Vancsura, Tibor Tatay, Tibor Bareith. "Evaluating the Effectiveness of Modern Forecasting Models in Predicting Commodity	<1 %

Futures Prices in Volatile Economic Times", Risks, 2023
Publication

83	Shengwei Wang, Ping Li, Hao Ji, Yulin Zhan, Honghong Li. "Prediction of air particulate matter in Beijing, China, based on the improved particle swarm optimization algorithm and long short-term memory neural network", Journal of Intelligent & Fuzzy Systems, 2021 Publication	<1 %
84	Submitted to University of Birmingham Student Paper	<1 %
85	Submitted to University of Hertfordshire Student Paper	<1 %
86	Submitted to University of Hong Kong Student Paper	<1 %
87	eprints.hec.gov.pk Internet Source	<1 %
88	journals.sagepub.com Internet Source	<1 %
89	www.elixirpublishers.com Internet Source	<1 %
90	Graham Smith. "Tests of the random walk hypothesis for London gold prices", Applied Economics Letters, 2010	<1 %

Publication

91	Submitted to Hankuk University of Foreign Studies Student Paper	<1 %
92	Lawrence Madziwa, Mallikarjun Pillalamarry, Snehamoy Chatterjee. "Gold price forecasting using multivariate stochastic model", Resources Policy, 2022 Publication	<1 %
93	Mehmet Balcilar, Zeynel Abidin Ozdemir, Muhammad Shahbaz. "On the time-varying links between oil and gold: New insights from the rolling and recursive rolling approaches", International Journal of Finance & Economics, 2019 Publication	<1 %
94	Sahin, M.. "On the effects of viscoelasticity on two-dimensional vortex dynamics in the cylinder wake", Journal of Non-Newtonian Fluid Mechanics, 20041110 Publication	<1 %
95	Zhang, Qi, Jun Hai Ma, and Yan Wang. "Study on Forecasting of Gold Price Based on Varying-Coefficient Regression Model", Key Engineering Materials, 2011. Publication	<1 %
96	bth.diva-portal.org Internet Source	<1 %

97	deepai.org Internet Source	<1 %
98	dokumen.pub Internet Source	<1 %
99	dspace.cvut.cz Internet Source	<1 %
100	epdf.pub Internet Source	<1 %
101	mlforall.blogspot.com Internet Source	<1 %
102	repository.javeriana.edu.co Internet Source	<1 %
103	web.math.ku.dk Internet Source	<1 %
104	www.econstor.eu Internet Source	<1 %
105	www.environmentasia-2019.science.cmu.ac.th Internet Source	<1 %
106	Hakan Yildirim, Merve Boyaci Yildirim, Alihan Limoncuoğlu. " Escape from - 19 pandemic to safe haven ", Journal of Public Affairs, 2021 Publication	<1 %
107	Xinyi Li, Shuchen Bai, Xiaoyan Jiang. "Does gold future interact with West Texas	<1 %

Intermediate (Crude Oil) by using impulse response?", 2021 International Conference on Computer, Blockchain and Financial Development (CBFD), 2021
Publication

108	Zhongda Tian, Hao Chen. "Multi-step short-term wind speed prediction based on integrated multi-model fusion", Applied Energy, 2021 Publication	<1 %
109	www.studymode.com Internet Source	<1 %
110	"CS2-PC-22_HR", ActEd Publication	<1 %
111	Guangyong Zhang, Le Jiang, Lixin Tian, Min Fu. "Analysis of the gold fixing price fluctuation in different times based on the directed weighted networks", The North American Journal of Economics and Finance, 2021 Publication	<1 %
112	dissertations.umi.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		