|  |  |  |
| --- | --- | --- |
| **Problem Chosen** C | **2022 MCM/ICM Summary Sheet** | **Team Control Number** 2216714 |

**Your Paper's Title**

**Summary**

**Keywords:** keyword1; keyword2; keyword3; keyword4

Contents最后记得更新整个目录

[1 Introduction 3](#_Toc96244427)

[1.1 Problem Background 3](#_Toc96244428)

[1.2 Restatement of the Problem 3](#_Toc96244429)

[1.3 Our Work 4](#_Toc96244430)

[2 Assumptions and Justifications 4](#_Toc96244431)

[3 Notations 4](#_Toc96244432)

[4 Data processing and analysis 5](#_Toc96244433)

[5 Predicting price by ARIMA model 5](#_Toc96244434)

[5.1 Stationary Tests 6](#_Toc96244435)

[5.2 White Noise test 7](#_Toc96244436)

[5.3 Model Recognition and Parameter Estimation 7](#_Toc96244437)

[5.4 Model Calibration 9](#_Toc96244438)

[6 The name of model 2 9](#_Toc96244439)

[7 The name of model 3 10](#_Toc96244440)

[8 Sensitivity Analysis 10](#_Toc96244441)

[9 Model Evaluation and Further Discussion 10](#_Toc96244442)

[9.1 Strengths 10](#_Toc96244443)

[9.2 Weaknesses 10](#_Toc96244444)

[9.3 Further Discussion 10](#_Toc96244445)

[10 Conclusion 11](#_Toc96244446)

[11 Memo 12](#_Toc96244447)

[References 14](#_Toc96244448)

[Appendices 15](#_Toc96244449)

# Introduction

## Problem Background

Have you ever invested in stocks? As we all know, stocks in the stock market are extremely volatile, and investors need to take huge risks when buying and selling stocks. Investors aim to minimize investment risk and maximize their total return. Therefore, investors need to find a way to predict stock price trends as accurately as possible and formulate trading strategies as soon as possible to achieve the goal of maximizing returns with minimal risk. Two specific cases are gold and bitcoin.

From agricultural society to commercial society, gold, as a precious metal currency, acts as a general equivalent for a long time due to its low reserves, difficulty in mining, and stable intrinsic value. The gold market plays an important role in the world economy, and for many investors, gold can be used as a hedge against rising prices and other financial risks. However, the gold market is not only non-stationary and fluctuating but also affected by various factors, such as relevant market activity, political events, etc. Therefore, it is very challenging to predetermine the price of gold.

The number of bitcoins is limited, which avoids inflation. In addition, Bitcoin's liquidation properties give it the ability to move funds instantaneously. Therefore, Bitcoin can be traded around the clock. In contrast, the above-mentioned gold is only available for trading on working days. Since the Bitcoin trading market works longer than the gold trading market, the trading cost of Bitcoin is higher than that of gold. The reason why the bitcoin trading market attracts many investors to gradually participate in it is the characteristics of bitcoin.

## Restatement of the Problem

In this problem, we have $1000 and have a five-year trade period from November 9, 2016, to October 9, 2021. In the daily trade period, we will have an account containing cash, gold, bitcoin, and the initial state is [100, 0, 0]. The cost per trade  is of the trade amount. There is no cost to hold assets.

Given the background information and constraints identified in the problem statement, we will accomplish the following tasks according to the given data:

* Give the best strategy for daily trading based on the price data up to the forecast day.
* Calculate the total value owned on October 9, 2021 with an initial investment of $1,000.
* Prove that the trading strategy we provide is optimal, that is, to prove that in our strategy, traders can have higher returns with lower risk.
* Determining the sensitivity of our strategies to trading prices. Analyze the impact of trading price on our strategy and results.

## Our Work

In our work, **the first step** is data processing and analysis. We analyze the past five years of data on gold and bitcoin through Exploratory Data Analysis (**EDA**) and clean the data. **The second step** is to model the trading strategy. We start by building the Autoregressive Integrated Moving Average model (**ARIMA**) that allows us to predict the present price using only past daily prices. Then, we establish a linear programming model (**LP**) to solve the optimal trading strategy and obtain specific profit values. **The third step** is to analyze the results. Through the sensitivity analysis of the transaction price, we can understand the impact of the transaction price on the strategy and results. Finally, we carry out the evaluation and outlook of the model.

In summary, our work can be shown as follows:

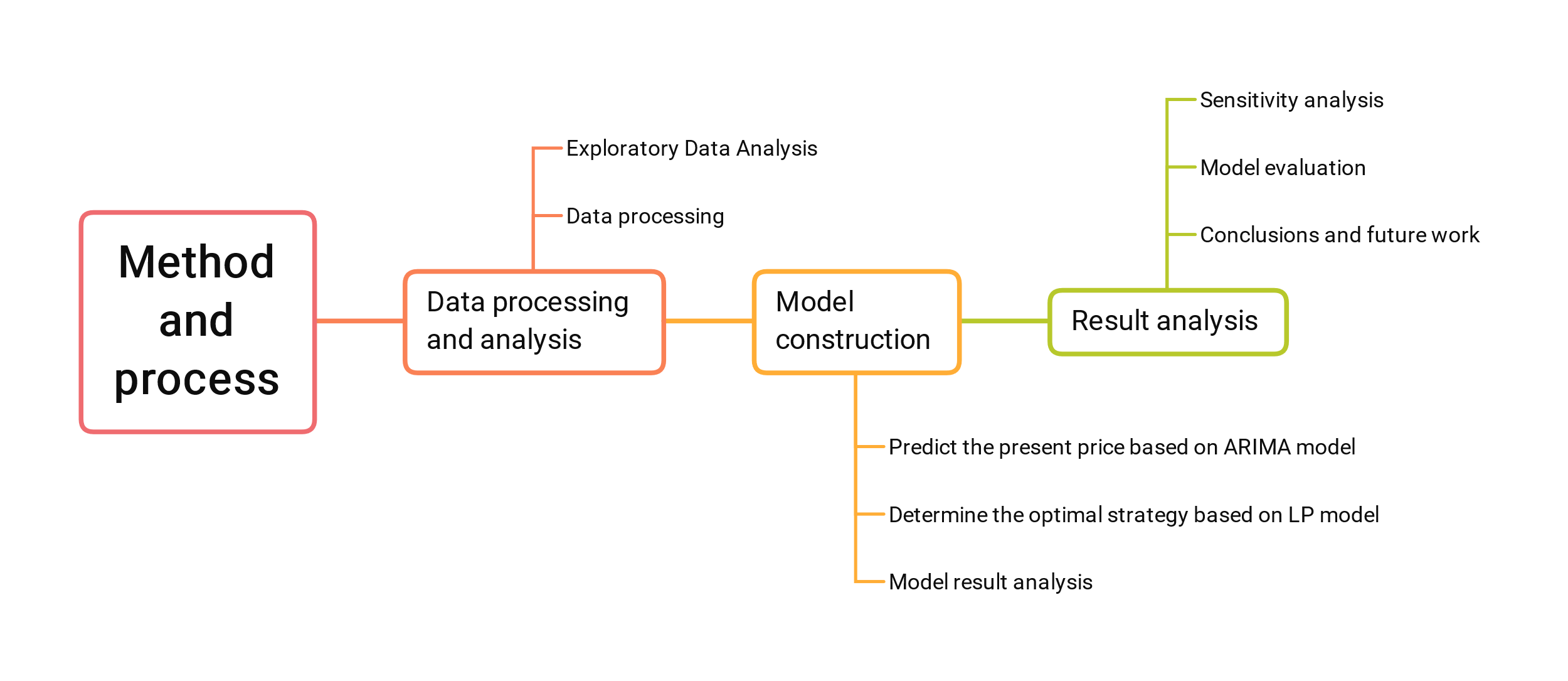


Figure Method and process

# Assumptions and Justifications

**1. We assume that when predicting the trend of stocks, observing the trading market in the first n days, that is, not buying and selling stocks, will not affect the final total assets.**

Although we may miss the opportunity to buy stocks at a low price, we can predict the future trend more accurately through the observation of these n days. So we can make better decisions and get higher profits. All in all, in the big picture, the gains that may be lost in the beginning are innocuous.

**2. We assume that funds cannot be borrowed to invest during the investment process. And the rate of return on cash in the account is 0.**

Generally speaking, the trading market is very volatile. When considering total investment return, we also need to be aware of risk. To avoid losing everything, we need to control risk. It is worth noting that the most important thing is to know how to invest and not to care about the investment capital. In addition, the cash held in the account will only change depending on the purchase and sale of stock. The cash in the account will not change. This is our assumption based on the characteristics of the trading market.

# Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

|  |  |
| --- | --- |
| **Symbol** | **Description** |
|  | Trade cost of gold as a ratio of trade value |
|  | Trade cost of bitcoin as a ratio of trade value |

# Data processing and analysis

Exploratory Data Analysis (**EDA**) [1] is used by data scientists to analyze and investigate datasets and summarize their main characteristics, usually using data visualization methods. It helps determine how to most efficiently process data sources to get the answers they need, making it easier for data scientists to spot patterns, spot anomalies, test guesses, or test hypotheses.

We can observe the data from the Nullity matrices for gold and bitcoin, as shown in Figure 2. The Nullity matrix is a data-dense display that lets us quickly visually pick out patterns in data completion. First, we observe from the axis on the left side of the picture that gold has 1265 data and Bitcoin has 1826 data. This is because Bitcoin can be traded every day, but gold can only be traded when the market is open. Second, we observe that there are blank horizontal lines only in the Nullity matrix representing the value of gold. This shows that some gold transaction price data is missing. Specifically, 10 pieces of data are missing, and the missing rate is 0.8%. Gold has no missing date value, and Bitcoin has no missing date value nor transaction price. Additionally, neither gold nor bitcoin has their respective dates repeated. In other words, there is no such case as two trading prices on the same date.

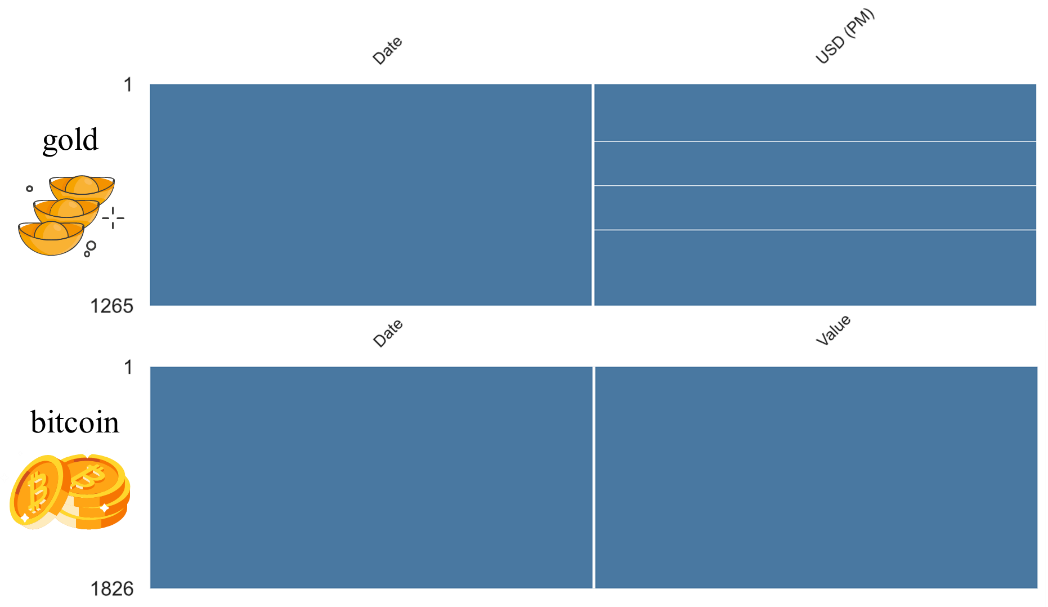


Figure Observing data integrity through the Nullity matrix

We process the data based on our knowledge of Bitcoin and Gold transaction data. The date is first formatted, after which the golden missing data mentioned above is removed.

# Predicting price by ARIMA model

To solve task one, we need to develop an optimal strategy for trading each day, based on price data up to that day. We first need to predict the price of gold and bitcoin on that day to develop a more accurate strategy. Time series forecasting is an important area of forecasting, where past observations of the same variable are collected and analyzed to develop models that describe underlying relationships. The model is then used to extrapolate the time series into the future. The Autoregressive Integrated Moving Average (ARIMA) model is one of the most important and widely used time series models. [] In this section, we use the ARIMA model to predict the volatility price of gold and bitcoin.

The structure of our ARIMA model is shown in Figure 3. The steps of the model are as follows:

* **Step 1** Do the white noise test on the time series. Generally speaking, only those sequence values have a close correlation, and historical data has a certain influence on future development, can it be used for modeling to mine effective information in historical data and predict future development. Purely random sequences have no analytical value.
* **Step 2** Check the stationarity of the sequence. Only stationary time series can be modeled with ARIMA. If the sequence is non-stationary, in general, the stationarity of the time series can be achieved by the first-order difference method, and sometimes the second-order difference is required.
* **Step 3** Building the ARIMA model. It includes three processes: model identification, parameter estimation, and model checking.
* **Step 4** Using the constructed fitting model, dynamic structural analysis can be carried out to examine the operation law of the research object, adjust the input variables to control the development of the research object, and make sequence predictions for future changes.

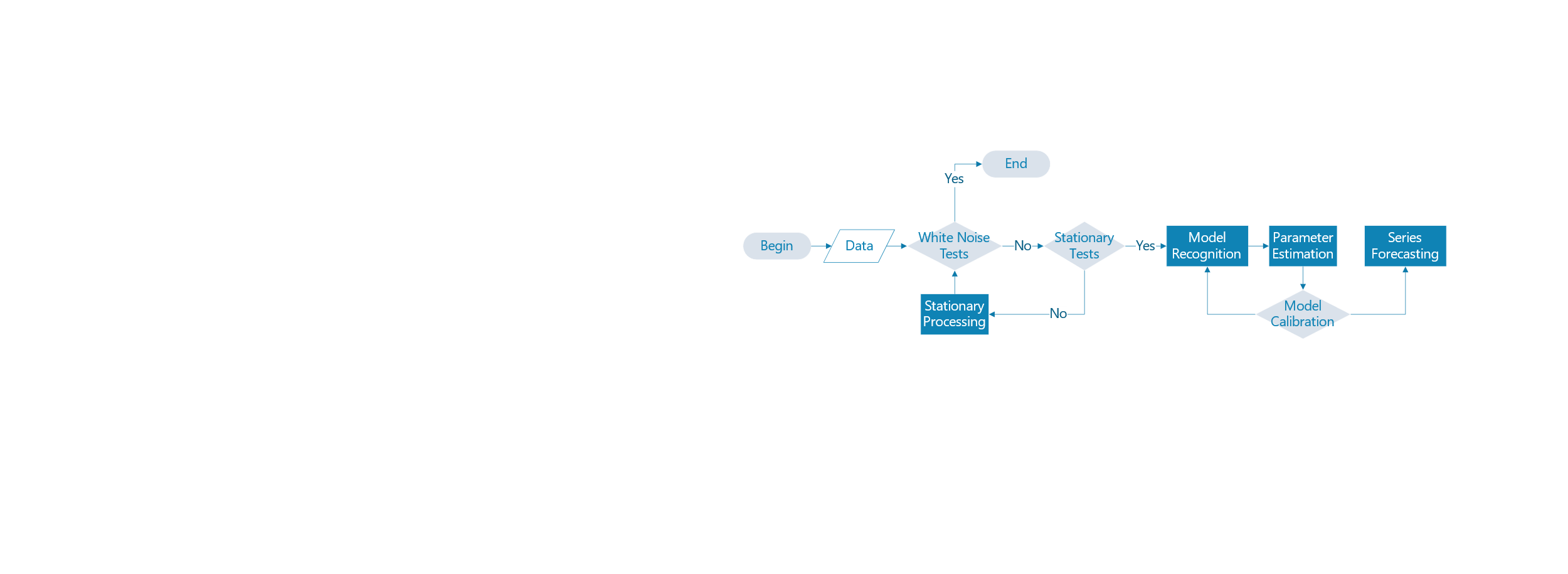


Figure The structure of the ARIMA model

## Stationary Tests

Series stationarity is a prerequisite for time series analysis. So how to determine whether the series is stationary? A rigorous statistical test method is the Augmented Dickey-Fuller test (ADF). The ADF test is to determine whether the sequence has a unit root: if the sequence is stationary, there is no unit root; otherwise, there is a unit root. Therefore, the ADF test hypothesizes that there is a unit root. The hypothesis of the ADF test is that there is a unit root. If the obtained significance test statistic is less than three confidence levels (10%, 5%, 1%), it corresponds to (90%, 95, 99%) confidence to reject the null hypothesis.

We can see from Table 1 that the t-statistic value of the original Bitcoin is -0.238 which is greater than the critical value (1%) -3.434, and the p-value is 0.934 which is greater than 1%. In addition, the p-value represents the probability value corresponding to the T statistic. So the hypothesis cannot be rejected, that is, the sequence is not stable. The same is true for the original golden sequence. The t-statistic value of -0.434 is greater than the critical value (1%) -3.434, and the p-value of 0.904 is greater than 1%. So the golden sequence is also unstable.

So we need a way to make a non-stationary time series stationary—calculate the difference between adjacent observations. This method is called differencing. Its formula is as follows:

|  |  |
| --- | --- |
|  | () |

Where  refers to the price on the day ,  refers to the price on day , and  refers to the difference between the two. Thus we construct a relatively stationary difference sequence.

Therefore, we need to perform differential processing on the data, and then perform an ADF test. The test method is the same as above. Both gold and bitcoin pass the test, which shows that the series after differencing is stationary.

Table The results of the ADF test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Inspection item | Gold test result | Bitcoin test results | Gold test result | Bitcoin test results |
| Number of difference | 0 | 0 | 1 | 1 |
| Test Statistic Value | -0.434 | -0.238 | -8.159 | -8.535 |
| p-value | 0.904 | 0.934 |  |  |
| Critical Value(1%) | -3.436 | -3.434 | -3.436 | -3.433 |
| Critical Value(5%) | -2.864 | -2.863 | -2.864 | -2.863 |
| Critical Value(10%) | -2.568 | -2.568 | -2.568 | -2.567 |

## White Noise test

We perform a white noise test on gold and bitcoin. The results are shown in Table 2. We found that the P-value of the two statistics is much smaller than the significance level of 0.05, then the null hypothesis can be rejected with a 95% confidence level, and the sequence is considered to be a non-white noise sequence. That is to say, it is not a randomly generated sequence and has a temporal correlation.

Table White Noise test

|  |  |  |
| --- | --- | --- |
| Inspection item | Gold test result | Bitcoin test results |
| Test Statistic Value | -3.321 | -3.114 |
| p-value |  |  |
| Critical Value(1%) | -3.668 | -3.669 |
| Critical Value(5%) | -2.964 | -2.964 |
| Critical Value(10%) | -2.620 | -2.621 |

## Model Recognition and Parameter Estimation

The identification problem of the model is the problem of order determination, which is mainly to determine the three parameters p, d, and q in the ARIMA model, and the order d of the difference can be obtained by observing the graph. We determined p and q by ACF and PACF.

The autocorrelation function ACF (autocorrelation function) describes the linear correlation between the time series observations and their past observations. Calculated as follows:

|  |  |
| --- | --- |
|  | () |

where k is the number of lag periods.

Partial autocorrelation function PACF (partial autocorrelation function) describes the linear correlation between time series observations expected to past observations given intermediate observations.

The result is shown in Table 3. From the graphs, we can see that with more than 95% probability,  and  fall in the range of .

Table The result of AFC and PAFC

|  |  |  |
| --- | --- | --- |
|  | **Goal** | **Bitcoin** |
| **AFC**  based on the original sequence | 图片包含 条形图  描述已自动生成 | 图表, 条形图  描述已自动生成 |
| **AFC**  Based on First Order Difference Sequence | 应用程序, Word  描述已自动生成 | 图表  描述已自动生成 |
| **PAFC**  Based on First Order Difference Sequence | 图片包含 图表  描述已自动生成 | 图表  描述已自动生成 |

Based on the above analysis, we get multiple sets of  combinations. We determine the best set of  values according to the information criteria of AIC and BIC. The smaller the information criterion value, the better the model. The calculation formulas of AIC and BIC are as follows:

|  |  |
| --- | --- |
|  | () |
|  | () |

Where  is the maximum likelihood under the model,  is the number of data, is the number of variables in the model.

Finally, we get:

|  |  |
| --- | --- |
|  | () |
|  | () |

## Model Calibration

We restore the prediction sequence and evaluate the model intuitively by comparing the predicted data with the real data. We compare the two data together. The red curve is the real data, and the green curve is our predicted data. Analyzing Figure 4, we can find that the predicted data is very close to the real data whether it is the overall trend or the local data. From this point of view, our model predicts well.

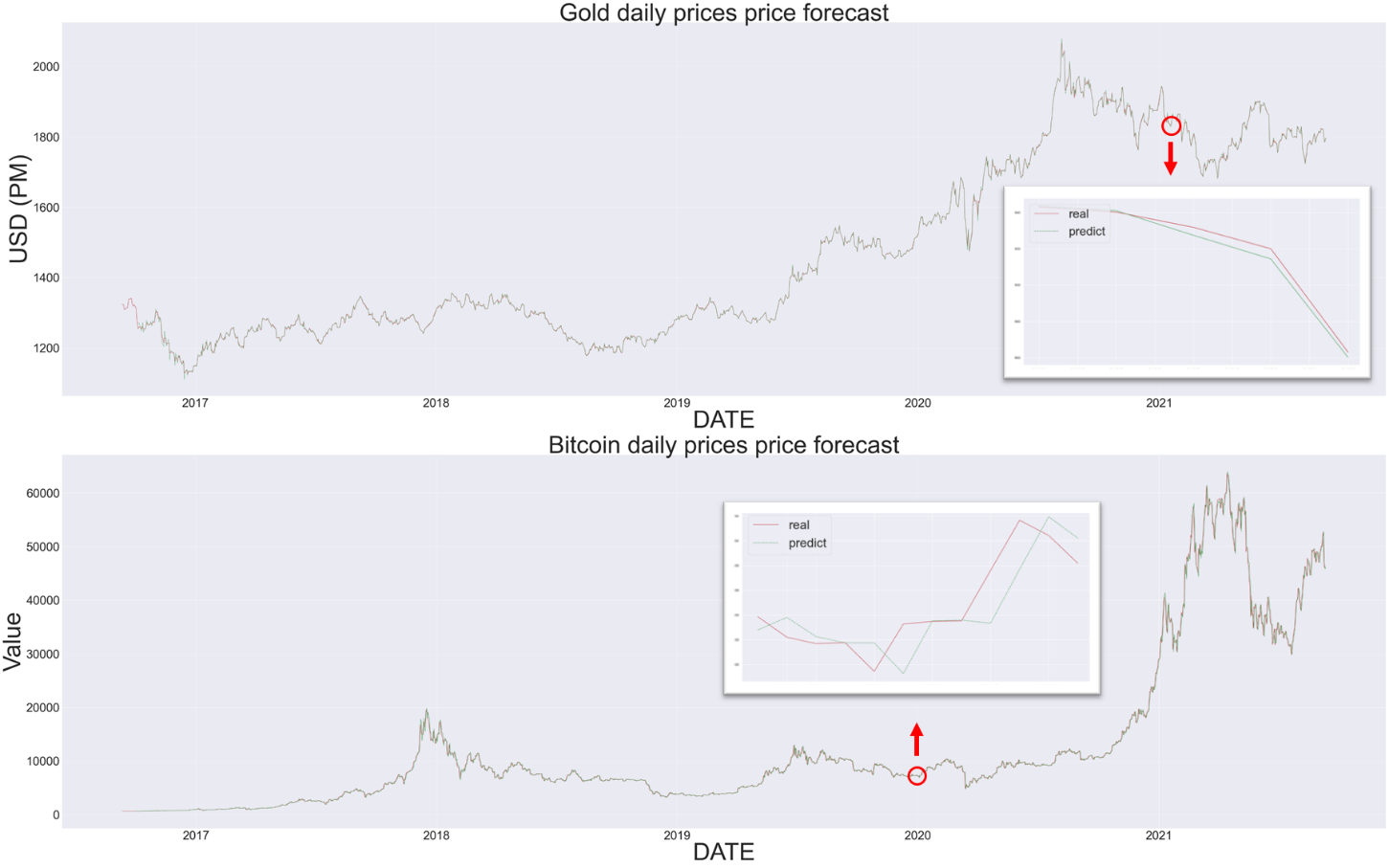


Figure Comparison of predicted data with actual data

# Develop strategies by LP model

Before planning, we analyze the products in the account. We can divide it into two categories. One is risk-free assets such as cash. The change in its price  can be expressed as:

|  |  |
| --- | --- |
|  | () |

The other category is risk assets such as gold and bitcoin. The change of its price  obeys geometric Brownian motion,[] namely:

|  |  |
| --- | --- |
|  | () |

where  is the expected return,  is the volatility coefficient, and  is the Brownian motion increment. and u is a constant.

和上一个部分类似的写法。

|  |  |
| --- | --- |
| 这里插入公式 | () |

|  |  |
| --- | --- |
| 这里插入公式 | () |

# The name of model 3

和上一个部分类似的写法。

注意：大多数美赛优秀论文都是对每个问题或者每个模型作为一个一级标题，就像我们上面的这种布局；也有一部分论文建立一个大的一级标题，取名为“Models and Solutions”，然后在这个大的标题下设计每个问题或者模型对应的二级标题，这一种排版布局在国赛中用的更多。

# Sensitivity Analysis

在我们的决策模型之中，模型受两个参数

From Figure 5, we can tell that total assets are sensitive to A and B.

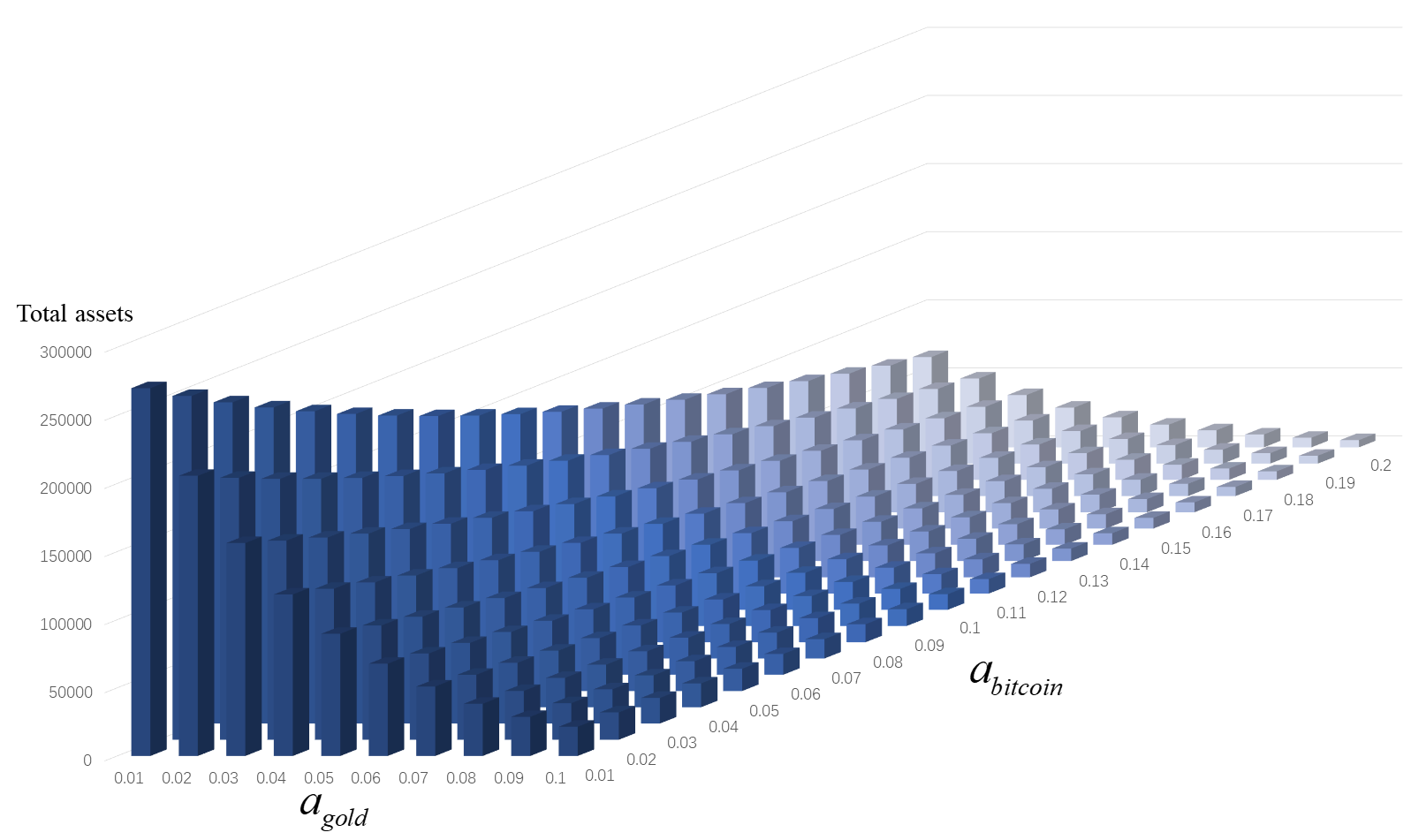


Figure :Influence of  and  on total asserts

# Model Evaluation and Further Discussion

注：本部分的标题需要根据你的内容进行调整，例如：如果你没有写进一步讨论的话，就直接把标题写成模型的评价。（优缺点一定要写）

## Strengths

这里写论文或者模型的优点

## Weaknesses

这里写缺点：缺点写的个数一般要比优点少

## Further Discussion

进行进一步的讨论，这里可以写模型的改进和拓展：

模型的改进：主要是针对模型中缺点有哪些可以改进的地方；

模型的拓展：将原题的要求进行扩展，进一步讨论模型的实用性和可行性。

# Conclusion

结论部分，这个部分在国赛论文很少见到，但在美赛中出现的频率很高。

这个部分可以是论文中心思想的重申、研究结果或主要观点的归纳，也可以是某些启示性的解释或考虑。

有些论文把“Model Evaluation and Further Discussion”的内容放到了结论部分，这也是可以的，大家可以灵活调整。

# Memo

**MEMO**

**TO: The market traders**

**FROM: Team # 2216714**

**DATE: 2022.2.19**

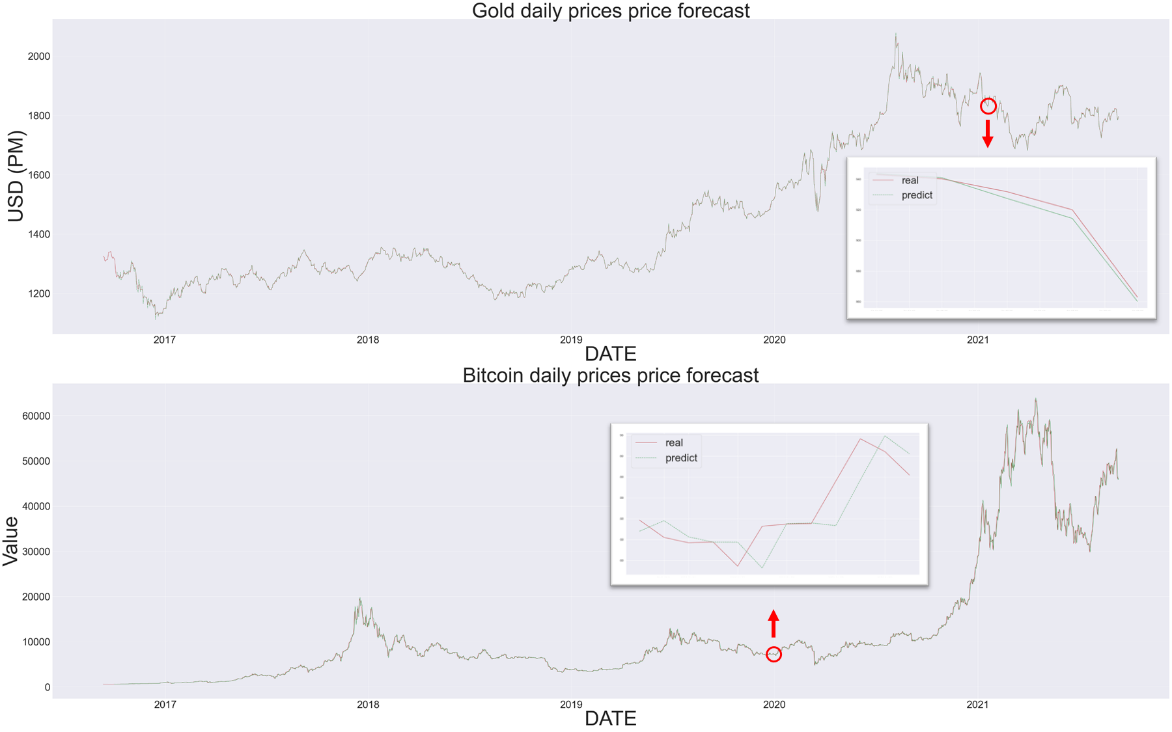
**SUBJECT: The best strategy for trade**

Dear Sir or Madame

We are honored to inform you of our achievement after performing data analysis and modeling.

The best strategy we came up with was based on the two models we built. One is a model that predicts stock volatility, and the other is a model that calculates how buying and selling can maximize total returns.

First, we introduce you to our predictive model. Our price prediction model is based on the price data as of that day to predict the stock price for that day. Our analysis predicts gold and bitcoin price volatility over five years. We can see from the figure that the prediction effect of our model is very good.



Second, based on our forecasts, we present to you an optimal strategy for how to buy, hold or sell the assets in your portfolio. Our decision models are rigorously mathematically validated. Our strategy can achieve high profits with low risk. Another highlight of it is that you can propose the most suitable strategy for you according to the principal you have. The specific implementation is as follows:

(其次，基于我们的预测，我们向你提出一个如何购买、持有或出售你们投资组合中的资产的最佳策略。我们的决策模型经过了严格的数学验证。我们的策略可以实现以较低的风险获得高额利润。它另外一个亮点在于可以通过您所拥有的本金不同，提出最适合于您的策略。具体实施方案如下所示：)

We sincerely wish you to achieve your goals.

Please contact us if you have any problems.

# References

[1] Tukey, J. W. (1977). Exploratory data analysis (Vol. 2, pp. 131-160).

[] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175. doi:10.1016/s0925-2312(01)00702-0

# Appendices

|  |
| --- |
| Appendix 1 |
| Introduce: 这里放上附录1的介绍 |
|  |

|  |
| --- |
| Appendix 2 |
| Introduce: 这里放上附录2的介绍 |
|  |

本部分是附录部分，美赛对于附录不是特别看重，今年还限制了论文的页数（从第二页开始编号，不能超过25页）。

一般新起一页列出附录。

在不超过页数限制的条件下，附录中可以包括下面内容：

* 你们写的代码；
* 某一问题的详细证明或求解过程；
* 自己在网上找到的数据；
* 比较大的流程图；
* 较繁杂的图表或计算结果。