

Research on trading strategies based on time series - taking gold and bitcoin as examples

Feiyang Zheng*, Hao Chi

Guilin University of Electronic Technology Business School, Guilin, Guangxi, 541000, China

*Corresponding author: zfy030108@163.com

Abstract. In response to the problem of how to optimally invest in gold and bitcoin, this paper establishes a daily trading strategy model, and uses a large amount of price data to conduct in-depth research on how to conduct optimal transactions to obtain maximum value. Based on the daily prices of gold and bitcoin within 5 years, we first build an ARIMA time series forecast model, and use MATLAB programming to solve the forecast results to get the next day's gold and bitcoin prices. Using this prediction result, we established a daily trading strategy model, and used recursion to find the daily trading strategy and the gains obtained in turn. The application of the model will help financial practitioners to analyze and guide the investment of related products.

Keywords: ARIMA; The rebuttal method; Perturbation analysis.

1. Background

Quantitative Trading (quantitative investment) refers to the investment method of trading by using computer technology with the help of modern statistics, mathematics, machine learning, and other methods [1]. Quantitative trading selects a variety of "high probability" events that can bring excess returns from huge historical data to formulate strategies, verify and solidify these laws and strategies with quantitative models, and then strictly implement the solidified strategies to guide investment, to obtain sustainable, stable and higher than average returns. Quantitative trading usually needs to formulate reasonable and stable investment strategies and make decisions through the analysis of past data and the prediction of future data. Therefore, the formulation of a trading strategy is very important [2].

The gold trading market is based on the traditional futures trading market. The gold trading strategy is different from other financial products and has its unique characteristics, but at the same time, the gold trading strategy also has something in common with other transactions. According to the characteristics of gold trading and learning from the trading strategies of other products, it is inevitable to find the strategies to guide gold trading [3].

Bitcoin is a new trading product emerging in recent years. Asymmetric encryption technology, digital summarization technology, blockchain technology, and so on are used in the process of trading in the bitcoin network. Bitcoin transactions are essentially a process of inputting and outputting a pile of UTXOS [4]. With the consumption of old UTXOS, new UTXOS are generated, and bitcoin transactions are completed again and again. The transaction process is double protected by asymmetric encryption and hash algorithm. Bitcoin holders can safely complete the transaction without worrying about identity disclosure. Part of bitcoin is also consumed in the transaction process to reward miners for packaged transactions. Based on this new trading theory, bitcoin has increased dramatically in recent years, and countless people who master trading strategies have made a lot of profits. Task I: Model Building and Problem Solving

2. ARIMA Model

In the time series prediction model, the ARIMA model can use the historical time data of variables to predict the future and can convert non-stationary time series into stationary time series. It is a model established for non-stationary time series [5]. ARIMA model is recorded as $ARIMA(p,d,q)$. p is the number of autoregressive terms, q is the number of moving average terms, and d is the number of differences made to make it a stationary sequence [6]. In the prediction, the time sequence law cannot

be found due to the unstable data. Therefore, "difference" is the key step. The time-series data predicted by the ARIMA model must be stable [7]. To accurately predict the value of bitcoin and gold, we choose the ARIMA time series model to solve this problem [8].

ARIMA model consists of three parts, namely autoregressive model (AR), difference model (I), and moving average model (MA). It is a combination of the AR model and MA model [9]

Autoregressive model: describes the relationship between the current value and historical value, and must meet there quirements ofstationarity. ToestablishanARmodel, wefirstneedtodetermine an order p and use the historical values of several periods to predict the current value [10]. The formula of the p -order autoregressive model is:

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \epsilon_t \tag{1}$$

Moving average model: pay attention to the accumulation of error terms in the autoregressive model. The formula of q -order autoregression is:

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \tag{2}$$

y_t is the current value, that is, the value of investment products we will predict.
 μ is a constant term, p is the order, γ_i is the autocorrelation coefficient, ϵ_t is the error.

Difference: it refers to single integration. The time series must be stable before the econometric model can be established. The unit root test is carried out for the time series. If it is a non-stationary series, it needs to be transformed into a stationary series through difference before prediction.

The establishment of the ARIMA model includes three stages: model identification and order determination, parameter estimation, and model test. For the identification and order determination of the model, it is mainly to determine the three parameters $p, d,$ and q . For the difference number d , we can determine it by observing the sequence diagram; For order p and q , we determine it based on the following rules according to the ACF and PACF residual diagrams and the criteria for judging tailing and truncation:

The autocorrelation coefficient ACF divides a column of data into two columns according to the lag and describes the linear correlation between the time series observations and their past observations. The ACF formula is as follows, where k represents the number of lag periods.

Table 1. Determination of ARIMA (p, d, q) Order

| Model | ACF | PACF |
|-----------|---|---|
| AR(p) | The attenuation tends to zero | Censored after p-order |
| MA(q) | Censored after q-order | The attenuation tends to zero |
| ARMA(p,q) | The decay tends to zero after the q-order | The decay tends to zero after the p-order |

$$ACF(k) = \rho_k = \frac{COV(y_t, y_{t-k})}{Var(y_t)} \tag{3}$$

Partial autocorrelation coefficient PACF describes the linear correlation between time-series observations and expected past observations under the condition of given intermediate observations

There is a confidence interval in the ACF and PACF diagrams so that the order corresponding to the last value exceeding the confidence interval in the ACF diagram is n , and in the PACF diagram, the attenuation of the value after order n tends to zero, then n is the value of q , and similarly, the value of p can be obtained.

3. Build an ARIMA Model to Predict the Price

3.1 Gold

Buying a ticket preprocesses data on the five-year price of GOLD (LBMA-GOLD.csv). An outlier value of 0 was found in the data, and there was a missing value in the data set because gold was closed on weekends and holidays. We believe that when gold does not open (weekends), its value does not change, i.e. the missing value can be replaced by the value of gold when it closes on Fridays.

According to the formula (2), the preprocessed price data is automatically tested for stability. The data of formula calculation is too volatile to pass the stationarity test. After observing the sequence diagram, it was decided to conduct first-order difference processing on the data, and the stationarity test was conducted again on the result after difference, which could pass. Therefore, there was no need to attempt higher-order difference, so d was set to 1. • Participates in residuals test for white noise. ACF was calculated to find q value in formula (2), and PACF was calculated to find p value in formula (2). Residual graphs of ACF and PACF were obtained through SPSS running time series model, which shows in Fig.1. You can see that both graphs are stationary, so the model is reasonable. The sequence has the characteristic of becoming very small from a certain point in time and should therefore be regarded as a case of (p) ACF truncation. In the ACF diagram, the 0th posterior intercept is made, and in PACF the 5th posterior intercept is made. According to the above Table.1, $p=0$ and $q=5$ can be obtained, which is ARIMA (0,1,5) model.

The processed gold price data set is predicted by using the tested model. We choose to start forecasting from the 30th day, and get the gold price prediction curve and the prediction curve in the given data set, as shown in Figure 1.

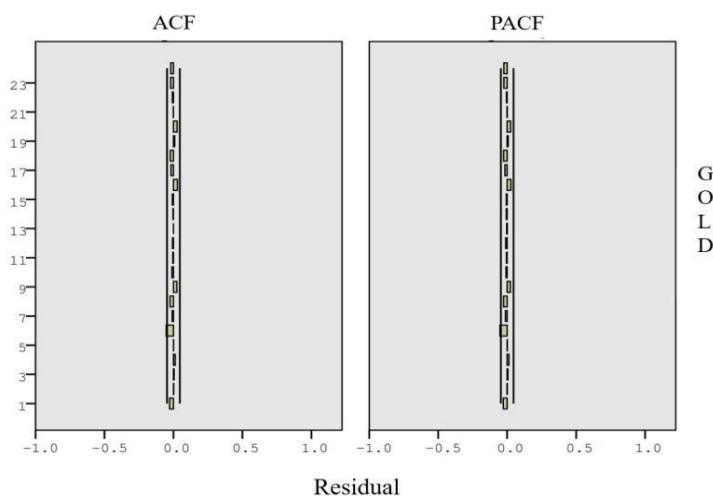


Figure 1. ACF & PACF for GOLD

3.2 Bitcoin

According to the formula (2), the 5-year price of Bitcoin price (bchain-mkpru.csv) is preprocessed, and the stability of the preprocessed data is tested. (All the formulas used here should be given above).After observing the time sequence diagram, it is decided to conduct first-order difference processing on the data, and the stationarity test is conducted again on the result after difference, which can pass. Therefore, d is 1.

Participates in residuals test for white noise. ACF was calculated to find q value in formula (2), PACF was calculated to find p value in formula (2). Residual graphs of ACF and PACF were obtained through SPSS running time series model. It could be seen that both ACF and PACF graphs were stable, so the model was reasonable. It can be seen from Fig.2 that this sequence has the characteristic of becoming very small from a certain point in time, so it should be regarded as the case of truncation of (partial) autocorrelation coefficient. The autocorrelation coefficient and partial autocorrelation coefficient of all lag order are not significantly different from 0. Therefore, this model is

ARIMA(0,1,0) model, which is one of the special cases of ARIMA model. In the process of predicting value changes, the value of each day is only related to the previous day.

ARIMA (0,1,0) is used to predict the bitcoin price from day 30 to day 1827, and the results are shown in Figure 2.

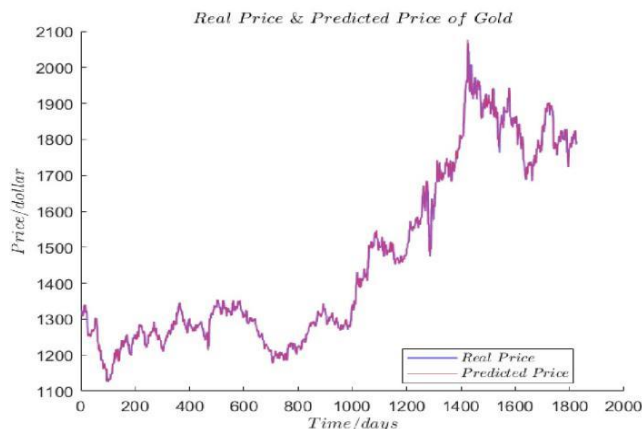


Figure 2. The Predicted Price & real Price for Gold

4. Model Checking

We can test the ARIMA model from two aspects: Significance of parameter estimation test (t test) q test is performed on residual of ARIMA (0,1,0) and ARIMA (0,1,5) models, and the p values obtained are 0.416 and 0.750 respectively, that is, we cannot reject the null hypothesis and consider residual as white noise sequence. Therefore, the two ARIMA models can well solve the prediction problem in this problem.

Test the randomness of residual sequence, that is, the residuals are independent. We choose DW (Dubin Watson) test method and use SPSS software to solve. Firstly, the autocorrelation coefficient is denoted as ρ , then the DW value can be approximated as $DW=2(1-\rho)$. Since the autocorrelation coefficient generally ranges from -1 to 1, DW values range from 0 to 4. According to the sample size N and the number of explanatory variables, the distribution table of DW can be checked, and the upper bound of critical value is d_U and the lower bound of critical value is d_L . When $DW=2$, the model has no autocorrelation. According to the final SPSS running results, it is obvious that the test has passed.

Let the current capital status be $[C, G, b]$, C, G and B represent the value of dollar, gold and bitcoin respectively, and the value is in dollar. On the first day (September 11, 2016), our initial account status is $[1000, 0, 0]$, that is, the fund distribution is \$1000, 0 gold and 0 bitcoin. At this time, the account value is calculated as $V_i[1000, 0, 0]$ (indicates subscript), that is, the account value is \$1000.

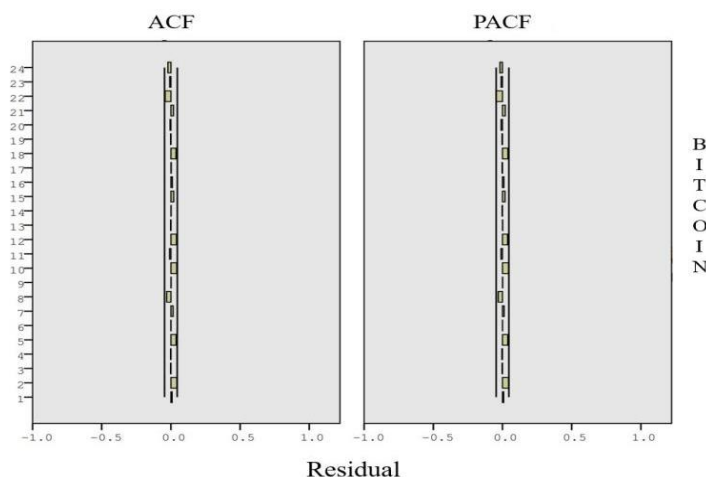


Figure 3. ACF & PACF for Bitcoin

Suppose we trade on day t and formulate a trading strategy on day t based on the predicted price of bitcoin and gold on day t+1. Given tomorrow's price, you can leave only one product in your hands today that maximizes your profit (as the next section shows). We assume that our prediction is sufficiently accurate (as will be proved in the next section) so that the account value for the day is highest only when [0,G,B] or [C,0,B] or [C,G,0]. In other words, two values of C,G, and B are zero and only one value greater than zero is the best choice. Thus, the problem can be reduced to the fact that the value of our account flows from one value of C, G, and B every other day.

The flow relationship of the three values C, G and B is shown in Fig.5. We make decisions based on the predicted value, so V_b represents the predicted value, $C^0G^0B^0$ respectively represent the value of dollar, gold and bitcoin under certain conditions according to the actual price of tomorrow under this decision. Under certain conditions, funds will flow from bitcoin to dollar through formula O1, The predicted value obtained at this time is $V_b = B - B \cdot \beta$. Under certain conditions, funds will flow from dollar to bitcoin through formula O2, The predicted value obtained at this time is $V_b = B - B \cdot \beta$, The measured value obtained at this time is $V = C \cdot Y - C \cdot \beta$, The real value of bitcoin tomorrow is $B^0 = C \cdot Y - C \cdot \beta$. Under certain conditions, the capital will flow from the dollar to gold through formula O3, and the predicted value obtained at this time is $V = C \cdot X - C \cdot \alpha$, The real value of gold tomorrow is $G = C \cdot X - C \cdot \alpha$, The real value of the dollar tomorrow is also $C = B - B \cdot \alpha$. Under certain conditions, bitcoin will flow to gold through formula O1O4. Under certain conditions, gold will flow to bitcoin through formula O2O3. The specific conversion conditions

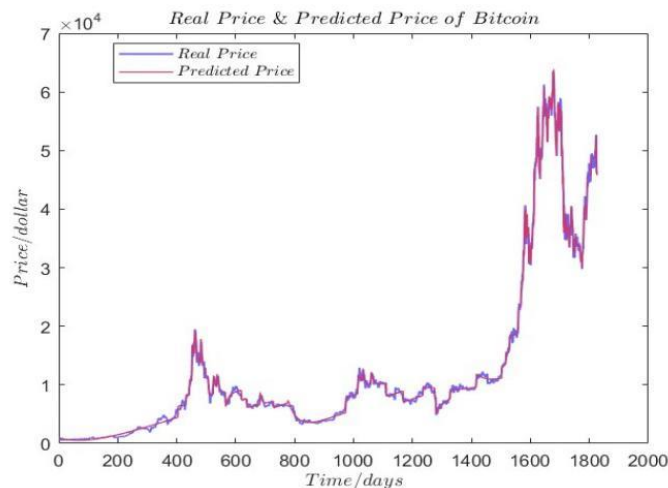


Figure 4. The Predicted Price & Real Price for Bitcoin

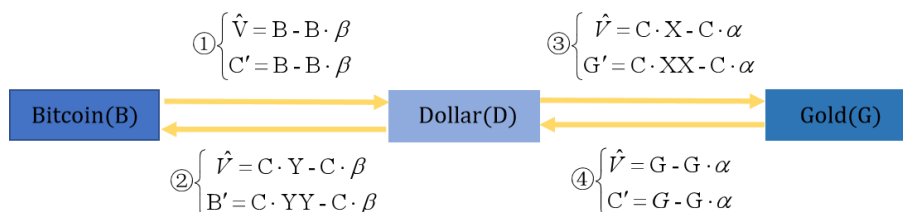


Figure 5. Value Conversion Chart for USD, Gold and Bitcoin

and the calculation of the obtained value will be given in the following formula.

When both gold and bitcoin predict the price decline on day t + 1, that is, $y \leq 1$ $x \leq 1$: if the resource in the account is dollar, our strategy is not to trade, and the value of the account is still C, as shown in formula (4); If the resource in the account is gold, compare the value of gold traded into dollar after deducting the Commission with the gold value of the account on day t + 1 calculated by the predicted price. If the former is large, all the gold in the account will be converted into dollar. On the contrary, no transaction will be carried out, as shown in formula (5); Bitcoin is the same, see formula (6).

$$\text{if } C > 0, V_{t+1} = C \quad (4)$$

$$\text{if } G > 0, V_{t+1} = \max\{X \cdot G - G \cdot \alpha, G\} \quad (5)$$

$$B > 0, V_{t+1} = \max\{Y \cdot B - B \cdot \beta, B^-\} \quad (6)$$

When the price of gold goes up on day t+1 and bitcoin goes down, i.e. $x > 1, y \leq 1$: If the resource in the account is dollar, compare the value of trading into gold with USD after deducting the commission and the dollar in the account on day t+1 calculated with the predicted price, if the former is large, then convert all the dollar in the account into gold, and vice versa, no trading, as in formula (7); if the resource in the account is gold, then our strategy is not to trade, as in formula (7); if the resource in the account resource is gold, then our strategy is not to trade, as in equation (8); if the resource in the account is bitcoin, we have to compare the size of three values, the value of converting bitcoin to dollar and deducting the commission, the value of converting bitcoin to dollar and then to gold and deducting the commission, and the value of bitcoin after the devaluation on day t+1, if the first value is the largest, then we convert all the bitcoin to dollar, if the second value is the largest, then we change it to gold, if the third value is the largest, then we choose not to trade, as in equation(9).

$$\text{if } C > 0, V_{t+1} = \max\{X \cdot C - C \cdot \alpha, C\} \quad (7)$$

$$\text{if } G > 0, V_{t+1} = G \quad (8)$$

$$\text{if } B > 0, V_{t+1} = \max\{C, X \cdot C - C \cdot \alpha, B\} \quad (9)$$

When gold falls in price on the day t+1 and bitcoin rises in price, our strategy can be derived from equation (10) (11) (12).

$$\text{if } C > 0, V_{t+1} = \max\{Y \cdot C - C \cdot \beta, C\} \quad (10)$$

$$\text{if } G > 0, V_{t+1} = \max\{C, Y \cdot C - C \cdot \beta, G\} \quad (11)$$

$$\text{if } B > 0, V_{t+1} = B \quad (12)$$

When both gold and bitcoin increase in price on day t+1, our strategy can be derived from equation (13) (14) (15).

$$\text{if } C > 0, V_{t+1} = \max\{Y \cdot C - C \cdot \beta, X \cdot C - C \cdot \alpha, C\} \quad (13)$$

$$\text{if } G > 0, V_{t+1} = \max\{C, Y \cdot C - C \cdot \beta, G\} \quad (14)$$

$$\text{if } G > 0, V_{t+1} = \max\{C, X \cdot C - C \cdot \alpha, B\} \quad (15)$$

In the works implemented in the above formula, we take into account that gold is not available for trading on some days, so before performing the calculation, we will first determine whether the market is open for gold and whether the day is stored in gold so far.

5. Analysis of the Results

Based on the above trading strategies and the gold and bitcoin prices that we predicted using the ARIMA model to provide the basis for our strategy. Write a C++ program in part 3.3 to solve for the optimal trading value and trading strategy for each day. The initial capital was set to \$1000, and the

daily predicted price of gold, the actual price of gold, the daily predicted price of bitcoin, and the actual price of bitcoin were entered into the program for calculation, and the results are shown in Figure 6.

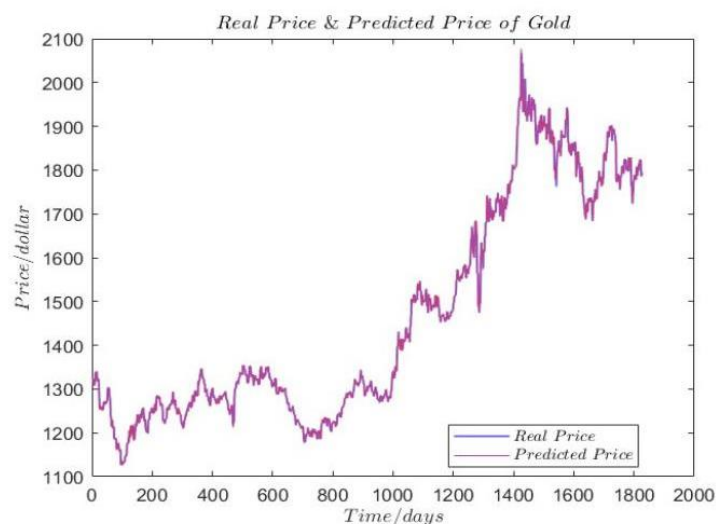


Figure 6. Projected Asset Value

Analysis of Fig.6 shows that we can get the best trading strategy for each day and invest according to our trading strategy. As can be seen from Fig.6, the initial account of \$1000 becomes \$1254568.91 after five years (2021.9.10). That's an increase of 1254 times. The daily trading strategy simply looks at which of the daily triples [C,G,B] is not 0 to know that all such products were bought the previous day. The daily gain is simply a matter of looking at what changed in value between today's triad and yesterday's triad. If you own \$x worth of G today and \$y worth of B yesterday, and $x < y$, then you bought bitcoin yesterday after buying it with gold and lost $y-x$. Analysis of the Results

In this problem, we will split the proof of the optimal trading strategy into two parts, the proof of the optimal prediction model and the proof of the optimal model of maximizing daily income.

In the part of proving the optimization of the prediction model, the essence of the BP algorithm is to find the minimum value of the mean square error function and analyze the training results. In the training process curve, the result of MSE is 21191.3, and the epoch at the inflection point is almost close to 0, and the subsequent training curve tends to be horizontal, indicating that the effect of reducing MSE after hundreds of training is very poor, BP neural network model is not the best model to solve this prediction problem; At the same time, by disturbing ARIMA model, it is proved again that the original prediction model performs best in the stability of prediction sequence, overall fitting degree, error value and so on.

In the part of proving the optimization of the daily income maximization model, we assume a conclusion contrary to the proposition conclusion and believe that this conclusion is obtained under the condition of following the proposition conditions. Then, starting from the trading results of the last day, the state of assets during daily trading is deduced against the time axis. Finally, it is found that it is inconsistent with the proposition conditions, and the conclusion of the original trading strategy is confirmed through theoretical derivation.

6. Conclusion

In response to the problem of how to optimally invest in gold and bitcoin, this paper establishes a daily trading strategy model, and uses a large amount of price data to conduct in-depth research on how to conduct optimal transactions to obtain maximum value. Based on the daily prices of gold and bitcoin within 5 years, we first build an ARIMA time series forecast model, and use MATLAB programming to solve the forecast results to get the next day's gold and bitcoin prices. Using this prediction result, we established a daily trading strategy model, and used recursion to find the daily

trading strategy and the gains obtained in turn. The application of the model will help financial practitioners to analyze and guide the investment of related products.

References

- [1] Conrad, J., & Kaul, G. (1998). An anatomy of trading strategies. *The Review of Financial Studies*, 11(3), 489-519.
- [2] Farmer, J. D., & Joshi, S. (2002). The price dynamics of common trading strategies. *Journal of Economic Behavior & Organization*, 49(2), 149-171.
- [3] Gilbert, K. (2005). An ARIMA supply chain model. *Management Science*, 51(2), 305-310.
- [4] Ho, S. L., & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. *Computers & industrial engineering*, 35(1-2), 213-216.
- [5] Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. *IEEE transactions on power systems*, 18(3), 1014-1020.
- [6] Zhang Rui, Song Hejia, Chen Qiulan, Wang Yu, Wang Songwang, Li Yonghong. Comparison of ARIMA and LSTM for prediction of hemorrhagic fever at different time scales in China. [J]. *PloS one*, 2022, 17(1).
- [7] Mancini S, Francavilla A, Graziuso G, Guarnaccia C. An Application of ARIMA modelling to air pollution concentrations during covid pandemic in Italy [J]. *Journal of Physics: Conference Series*, 2022, 2162 (1).
- [8] Ahmad T E, Rais A, Azhari D R, Rosalia A A, Hutapea R Y F. Fisheries forecasting, physical approach comparison between regression and autoregressive integrated moving average (ARIMA) [J]. *IOP Conference Series: Earth and Environmental Science*, 2022, 967(1).
- [9] Chyon Fuad Ahmed, Suman Md Nazmul Hasan, Fahim Md Rafiul Islam, Ahmmed Md Sazol. Time Series Analysis and Predicting COVID-19 Affected Patients by ARIMA Model Using Machine Learning. [J]. *Journal of virological methods*, 2021, 301.
- [10] Kim Sooin, Choi ChiYoung, Shahandashti Mohsen, Ryu Kyeong Rok. Improving Accuracy in Predicting City-Level Construction Cost Indices by Combining Linear ARIMA and Nonlinear ANNs [J]. *Journal of Management in Engineering*, 2022, 38(2).