Optimal trading strategies based on time series analysis

Yifei Wang¹, a, 8, Xiaofeng Zhao¹, 8, Feng Zhang², Siyang Xie³ and Zhihan Liu⁴

¹ Evergrande School of Management, Wuhan University of Science and Technology, China;
² School of Finance, Jiangxi Normal University, China;
³ School of Information Science and Engineering, Wuhan University of Science and Technology, China;
⁴ Evergrande School of Management, Wuhan University of Science and Technology, China.

wyf1657892128@163.com

These authors contributed equally to this work

Abstract. Quantitative investment has been widely used in the field of foreign finance, especially the rapid development of international investment in the past decade. And financial activity is an important field of national economic activity. The frequency of financial transactions is an important indicator of the complexity of a country's economy, so it is of great significance to study the optimal investment strategy. This article uses daily price streams from past investments in gold, cash, and bitcoin to determine whether traders should buy, hold, or sell assets in their portfolios. The outlier data were processed by boxplot analysis, and the EM algorithm based on maximum likelihood estimation was used to visualize the case data. The ARIMA model and GARCH model are used to establish the portfolio optimization model and obtain the best portfolio scheme. The time series prediction model is used to conduct specific quantitative analysis on gold and Bitcoin and obtain the investment forecast of the initial $1000 in the future.

Keywords: Time series analysis; EM algorithm; ARIMA model; GARCH model.

1. Introduction

Trading strategy plays a very important role in financial asset trading. How to automatically select trading strategy in complex and dynamic financial markets is an important research direction of modern finance [1]. Market traders often buy and sell volatile assets with the aim of maximizing their total returns. There is usually a commission on every transaction. Two of those assets are gold and bitcoin. Investors usually use historical data for investment analysis [2]. According to the historical data, the prediction model of gold and bitcoin daily flow is established, and it is found that the daily fund flow is volatile whether it is gold or bitcoin.

In this paper, the outlier data are first processed by boxplot analysis, and then the data are visualized by EM algorithm based on maximum likelihood estimation, where the case data are sorted by date in a unified format. Then a stock is selected for qualitative analysis of relevant parameters, mainly analyzing the volatility trend and data distribution of gold and bitcoin. Visual trend analysis of gold bit volatility obtained by adding trend line.

Through ARIMA model and GARCH model, the portfolio optimization model is established to get the best portfolio scheme. Finally, based on the stock market price of USA from September 1, 2016 to September 1, 2021, the time series prediction model is used to conduct specific quantitative analysis and obtain the investment value of the initial $1000 in five years based on the best investment plan.

2. Model preparation

2.1 Time series model background
Time series analysis is a mature subject developed gradually on the basis of mathematical statistics, which is widely used in the economic field [3]. Time series, also known as dynamic series, refers to a numerical series that arranges the index values of a phenomenon in chronological order. Time series analysis can be roughly divided into three parts: describing the past, analyzing the law and predicting the future.

2.2 Visual processing of data

Because there are outliers in the case data and the data need to be cleaned, this paper mainly processed the outlier data through box graph analysis and completed the defective data. In order to fill the missing value data more accurately, this paper considered using the EM algorithm based on maximum likelihood estimation to fill the missing value. The proposed algorithm can avoid the possible dragon lattice phenomenon of high-degree interpolation, maintain the smoothness and continuity of the data well, and reduce the loss of information quantity.

2.3 EM algorithm steps for maximum likelihood estimation

Step 1: Missing value description, the analyst needs to take full account of the data, such as missing proportion, missing value distribution, missing value type (price type, usdpm), etc.

Step 2: Detect whether there is correlation between related variables. Missing value analysis is to use relevant variables to fill in missing values. For example, the padding of regression models in OLS estimation, and the correlation information between relevant variables (amos) in ML estimation.

Step 3: The correlation between variables can be linear or nonlinear. Then a scatter plot is used to describe the linear relationship between the variables. In general, regression model padding is recommended for linear relationships and EM for nonlinear methods.

Through the above data preprocessing steps, we model our portfolio using the preprocessed gold and bitcoin-related data.

2.4 The qualitative analysis.

Firstly, a stock in the US market is selected for qualitative analysis of relevant parameters, mainly to analyze the volatility trend of gold and bitcoin and the distribution of data. By adding trend lines, the trend analysis of gold and Bitcoin volatility can be visualized. As shown in the figure below:

![Fig. 1 Trend analysis diagram of the fluctuation visualization of gold and Bitcoin](image)

3. ARIMA Bitcoin Prediction model

(1) According to the case data, take Bitcoin as an example, model based on ARIMA model, and so on, each time series is modeled in this way. The time series diagram is as follows:
(2) Descriptive statistics were performed for the daily yield rate of \( r \)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r )</td>
<td>1,825</td>
<td>72.76342</td>
<td>246.9307</td>
<td>-98.6593</td>
<td>2123.474</td>
</tr>
</tbody>
</table>

(3) ADF checkout. Test whether the yield sequence \( r \) is the root per unit (the original hypothesis: the unit root sequence, the alternative hypothesis: the smooth sequence)

<table>
<thead>
<tr>
<th>Teststatistic</th>
<th>Dickey-Fuller-criticalvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z(t)</td>
<td>1%</td>
</tr>
<tr>
<td>-52.78</td>
<td>-3.43</td>
</tr>
</tbody>
</table>

MacKinnon approximate p-value for \( Z(t) = 0.0000 \). A p-value of 0 means rejecting the null hypothesis, so we consider the \( r \) sequence stationary.

(4) Observe the ACF map (The figure below is left) and the PACF (Right of the figure below) graph, and judge the order of the AMRA model.
Observe the ACF map (The figure below is left) and the PACF (Right of the figure below) graph, and judge the order of the AMRA model. The mathematical model is:

\[
(1 - \phi_1 \Delta - \ldots - \phi_p \Delta^p) y_t = c + (1 + \theta_1 \Delta) \epsilon_t,
\]

(1)

(6) They were simulated separately with four alternative ARIMA models.

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(1,0,6)</td>
<td>23328.09</td>
<td>23377.68</td>
</tr>
<tr>
<td>ARIMA(6,0,1)</td>
<td>20337.78</td>
<td>20387.36</td>
</tr>
<tr>
<td>ARIMA(6,0,1)</td>
<td>25192.67</td>
<td>25214.71</td>
</tr>
<tr>
<td>ARIMA(6,0,6)</td>
<td>20325.33</td>
<td>20402.46</td>
</tr>
</tbody>
</table>

In preliminary judgment, four alternative ARIMA models were used to fit the results. Since both AIC and BIC are minor selection principles, the average of AIC and BIC values of ARIMA (6,0,1) models were the smallest, so the ARIMA (6,0,1) model was selected for fit. Below we use this model for estimation to obtain the predictive value of the residuals and generate a histogram of the residue distribution as shown in Figure 6 below:

![Fig. 6 Histogram of the residue distribution](image)

(7) Q-test. Check whether the residuals are white noise sequence.

<table>
<thead>
<tr>
<th>Portmanteau(Q)statistic</th>
<th>Prob&gt;chi2(12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.5230</td>
<td>0.8875</td>
</tr>
</tbody>
</table>

The residual sequence was subjected by white noise test to generate the square of the residues, and the Q test to generate the residual square sequence ressq.

White noise test was performed on the residual squared sequence ressq:

<table>
<thead>
<tr>
<th>Portmanteau(Q)statistic</th>
<th>Prob&gt;chi2(12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4778</td>
<td>0.9999</td>
</tr>
</tbody>
</table>

(8) LM test. Test for ARCH error. The residual squared terms were regressedly fitted to their lag terms. The output LM statistic is 1.490736; the calculated p-value is 0.91413639.

(9) Select appropriate models using the AIC and BIC minimum principles for estimation.
4. GARCH model

4.1 Bitcoin Prediction model GARCH (2,2)

Set up a set of equations as follows:

\[ y_t = \mu + \varepsilon_t \]
\[ \varepsilon_t = \sigma_t^2 + (1 - \gamma_1 L - \gamma_2 L^2) \sigma_t^2 = \kappa + (\alpha_1 L + \alpha_2 L^2) \varepsilon_t^2 \]  

(2)

4.2 The ARMIA Gold Prediction Mode

The above ARIMA model is used to forecast gold. Firstly, each time series is modeled based on stock market related data. Then, descriptive statistics of the daily output of r were carried out. Then, the ADF test was conducted to obtain the MacKinnon approximate P-value with \( Z(t) = 0.0000 \). A p-value of 0 indicates rejection of the null hypothesis, so the R-series is considered stationary. By observing ACF and PACF images, a mathematical model was established:

\[ (1 - \phi_1 L - \ldots - \phi_p L^p)(1 - L) y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3) \varepsilon_t \]  

(3)

Four alternative ARIMA models were used for simulation:

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(3,0,8)</td>
<td>9298.079</td>
<td>9364.822</td>
</tr>
<tr>
<td>ARIMA(8,0,8)</td>
<td>9359.234</td>
<td>9425.977</td>
</tr>
<tr>
<td>ARIMA(3,0,3)</td>
<td>9384.29</td>
<td>9425.363</td>
</tr>
<tr>
<td>ARIMA(8,0,3)</td>
<td>9276.672</td>
<td>9369.086</td>
</tr>
</tbody>
</table>

The average AIC and BIC values of the ARIMA (8,0,3) model are the smallest, so the ARIMA (8,0,3) model is suitable. Then Q test. Check if the residuals are white noise sequences. Finally, the P-value of 7.409e-38 was obtained by LM test. The appropriate model was selected using the minimum principles of AIC and BIC, as shown below:
5. Results analysis of the single-target investment decision model

Establish mathematical models of ARIMA and GARCH according to time series analysis with the following regression equations:

\[ y = \frac{9485.9039}{(1 + 13051.9510 \times \exp(-0.1255 \times x))} \] (4)

The model regression is ideal and the obtained data have strong reliability. Next, we present quantitative value results for the initial $1000 investment harvest on September 1, 2021. The data was calculated using MATLAB, and the final result was calculated by adding the gold and Bitcoin predictions, which was about $7,645.

6. Conclusions

With the continuous expansion of the stock market, more and more people become stock investors. Quantitative investment method has gradually received the favor and pursuit of investors. Under the condition of ensuring asset returns, how to combine stock investment so as to reduce the overall risk has also become a hot research [4].

Based on the daily price flows of gold and Bitcoin over the past five years, this paper builds prediction models to provide the best trading strategies for each day. In this paper, the outlier data are processed by box plot analysis, and the EM algorithm is used to fill in the missing values. Then, ARIMA model and GARCH model are established to study the stock return rate. In the Bitcoin prediction model, it is found that ARIMA (6,0,1) model has the best fitting effect by combining the autocorrelation coefficient, the number of bias relations and AIC criterion. Among the prediction models for gold, ARIMA (8,0,3) model has the best fitting effect.

Finally, based on time series analysis, mathematical models of ARIMA and GARCH were established, and the data were substituted to obtain the investment value of the initial $1000 investment income and the final harvest was $7645. The portfolio investment strategy based on ARIMA model and GARCH model in this paper can effectively reduce the risk of portfolio investment, which has positive significance for investors to optimize financial decisions.

References