World Scientific Research Journal Volume 6 Issue 8, 2020
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0013

Research and Investment Model of Stock Index Fluctuation based on Entropy Method

Ting Liao[1, a]

[1]Faculty of Accounting, Jiangxi University of Finance and Economics, Nanchang, China.

[a]ltingha@163.com

Abstract

Stock index fluctuation is a complex nonlinear process, which should be improved by considering other factors besides the advancement of it with the uniform time process. In this paper, a stock turnover prediction model based on time series analysis will be constructed based on the regularity research of stock in the stock market. Through factor analysis dimension reduction, the influential variables are found out and the investment index model is constructed. The entropy method is used to assess whether there are differences in parameters. If there are differences, the optimization model is needed to be optimized. If the parameters are relatively perfect, the model does not need to be optimized, so as to evaluate the scale of the model. The parameters of the stock investment index model are modified by an entropy method to achieve the model optimization effect. Finally, through the stock yield sequence analysis of the trend of stock volatility effect, build the ARIMA model forecast analysis for the coming year index fluctuation, can be used in the index of the current fluctuations and next year's index analysis and prediction of investment for the future of its positive volatility will benefit, the opposite is not conducive to future investment, summed up the short-term investment direction, the section chief and investment advice and strategies.

Keywords

Entropy method, investment index model, ARIMA model.

1. INTRODUCTION

With the development of the economy and the increase of participation in securities investment in recent years, more funds pour into the securities market. At the same time, our country continues to strengthen the market trading system, the information disclosure system and the system of securities laws and regulations such as the improvement of the regulations, but although the price of the securities market volatility has certain regularity, so to study the law of its fluctuation and provides the factors which influence the fluctuation of feasibility, and the study of the people for the analysis of risk and return, maximize shareholder rights and regulatory effective regulation has a vital role.

2. ASSUMPTIONS OF THE MODEL

The stock price with non-normal absence on the normal trading day rises or falls by the daily limit on the day of the defect.

The stock market will be closed on holidays and weekends.

The current closing price of each share minus the opening price is the earnings of that share

Long-term investment is assumed to be greater than or equal to one year in the investment cycle.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0013

The trend of stock volatility can be analyzed according to the volatility of stock returns

Investment risk threshold =0.5.

3. ESTABLISHMENT AND SOLUTION OF THE MODEL

The training process established by the model follows the following steps: download data, data processing, factor analysis dimension reduction, data normalization, regression analysis training, an investment index equation, entropy fine-tuning parameters, and ARIMA model prediction.

3.1. Download Data

Using the Tushare package that comes with Python, download a list of 10 stocks for 2019 that includes timing data such as the opening price, high, low, volume, adjusted closing, and closing prices.

3.2. Data Processing

Considering that the data of the original volume sequence may be non-stationary, the volume change rate sequence is obtained after the first-order difference of the original volume sequence. ADF unit root test was carried out:

```
Augment Dickey-Fuller Test
Date: ddat
Dickey-Fuller = -9.5663, lag order = 6, p-value = 0.01
Alternative hypothesis: statistionary
```

Fig 1. Results of unit root test

The results show that through the significance test, it can be known that the change rate sequence of volume is stable. According to its ACF autocorrelation function diagram and PACF partial autocorrelation function diagram, it can be seen that the lag order 3 is appropriate in ACF and the lag order 4 is appropriate in PACF, so the ARIMA (4, 1, 3) is established.

```
Box-L jung test
Date: fit$residuals
X-squared = 0.00092855, df = 1, p-value = 0.9757
     Point Forecast    Lo 80   Hi 80   Lo 95   Hi 95
280      103805586 53074230 154536942 26218672 181392500
```

Fig 2. Residual stationarity test results

3.3. Factor Analysis Dimension Reduction

As a rational person, the main purpose of investing in the securities market is to seek profit maximization. In this paper, several factors that evaluate the stock selection effect are selected for factor analysis and dimension reduction to find the maximum factor. The selected factors are date, opening price, closing price, maximum price, minimum price, volume and yield. Among them, yield is defined as (closing price - opening price), which can be used to analyze the investment value of constituent stocks and provide portfolio Suggestions.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0013

(1) KMO and Bartlett tests were used to consider whether variables were suitable for factor analysis. It can be seen from the test that the observed value of Bartlett sphericity test statistics is 4299.829, and the probability P value is close to 0. Since the probability P value is less than the significance level, the null hypothesis should be rejected and the correlation coefficient matrix should be considered significantly different from the identity matrix. Meanwhile, the KMO value is 0.726. According to the KMO metric provided by Kaiser, the original variable is suitable for factor analysis.

(2) According to the correlation coefficient matrix of the original variables, principal component analysis was used to extract the factor and select the characteristic root value greater than 1. According to the explained total variance table, the variance contribution of the first factor is 3.988, which explains 79.76 of the total variance of the original five variables. The variance contribution of the second factor is 0.997, which accounts for 19.943% of the total variance of the original five variables, and the cumulative variance contribution rate is 99.703%. According to the lithotriptic diagram, the eigenvalue of the first factor is very high and contributes the most to the interpretation of the original variables. After the second factor, the eigenvalue of the factor is small, and the contribution to the interpretation of the original variable is small, so it can be ignored. So it's appropriate to extract two factors.

(3) In this paper, the maximum variance method is used to implement an orthogonal rotation of the factor loading a matrix to make the factor have naming interpretation. It can be seen from the rotation component matrix that the opening, lowest, highest and closing values have a higher load on the first factor, which mainly explains these variables and is named as "price factor". Volume has a higher load on the second factor, which mainly explains these variables and is named the "sales factor".

(4) Calculate factor scores

The regression method was used to estimate the factor score coefficient and output the factor score coefficient. According to the component score coefficient matrix table, the following factor score function can be written. The interpretation degree of variance is respectively 79.564% and 20.139%. The ratio of the interpretation degree of variance to the total interpretation variance is used as the weight of the factor to calculate the comprehensive score, namely the investment index.

3.4. Regression Analysis

Two factor score of the above equation can be calculated each factor score of samples, to test the two factors influence on purchased stock returns, to establish the return of two factors on the share price, the regression equation obtained through F test, F value to 9.731, p values < 0.000 0.05, suggesting that there is no multicollinearity, two factors of the share price and show high goodness of fit.

Therefore, the regression equation is:

$$Y = 0.018 - 0.007F_1 + 0.062F_2$$

According to the regression equation, the regression coefficient of the sales coefficient in the independent variable is the largest, indicating that the stock trading volume has the greatest impact on the return of a stock.

3.5. Investment Index Equation

According to the investment index equation:

$$F = 0.197 opening\ price + 0.212 The\ highest + 0.195 The\ minimum + 0.206 closing\ price + 0.229 volume$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0013

Therefore, the corresponding investment index of each stock can be obtained by substituting the corresponding data about ten stocks.

Firstly, the stock data is normalized and the stock investment index can be obtained by substituting it into the investment index equation.

If α_0=0.5 is the cut-off point of the stock investment index. Therefore, when, we believe that the stock investment index is too low, the risk is large, do not recommend investment; Instead, it is recommended to invest in the stock.

3.6. Entropy Fine-Tuning Parameters

Entropy method is used to optimize, construct and solve the model. Form the original index matrix $\boldsymbol{X}=(\boldsymbol{x}_{ij})_{10\times5}$, where $\boldsymbol{X}_{ij}$ is the value of the $\boldsymbol{j}$ index of the $\boldsymbol{i}$ stock.

Since entropy method is used to calculate the ratio of a certain index in each scheme to the same index, there is no dimensional influence and standardization is not required. So I'm going to shift the data:

For the larger the better index (the lowest), do the following treatment:

$$\boldsymbol{X}_{ij}^{'}=\frac{\boldsymbol{X}_{ij}-\min(\boldsymbol{X}_{1j},\boldsymbol{X}_{2j},...,\boldsymbol{X}_{nj})}{\max(\boldsymbol{X}_{1j},\boldsymbol{X}_{2j},...,\boldsymbol{X}_{nj})-\min(\boldsymbol{X}_{1j},\boldsymbol{X}_{2j},...,\boldsymbol{X}_{nj})}+1 \quad (\boldsymbol{i}=1,2,...,10;\boldsymbol{j}=1,2,...5)$$

For the smaller the better indicators (opening, highest, closing, volume), do the following:

$$\boldsymbol{X}_{ij}^{'}=\frac{\max(\boldsymbol{X}_{1j},\boldsymbol{X}_{2j},...,\boldsymbol{X}_{nj})-\boldsymbol{X}_{ij}}{\max(\boldsymbol{X}_{1j},\boldsymbol{X}_{2j},...,\boldsymbol{X}_{nj})-\min(\boldsymbol{X}_{1j},\boldsymbol{X}_{2j},...,\boldsymbol{X}_{nj})}+1(\boldsymbol{i}=1,2,...,10;\boldsymbol{j}=1,2,...5)$$

We get a new data matrix, and here we still remember the processed data as $\boldsymbol{X}_{ij}$. Then calculate the proportion of the $\boldsymbol{i}$ stock in the $\boldsymbol{j}$ index and get the new data matrix $\boldsymbol{P}_{ij}=\frac{\boldsymbol{X}_{ij}}{\sum_{\boldsymbol{i}=1}^{10}\boldsymbol{X}_{ij}}(\boldsymbol{j}=1,2,...,5)$. And then $\boldsymbol{e}_j=-\boldsymbol{k}*\sum_{\boldsymbol{i}=1}^{10}\boldsymbol{P}_{ij}\log(\boldsymbol{P}_{ij})$, where $\boldsymbol{k}$ is a constant and is associated with $\boldsymbol{m}$, let's say $\boldsymbol{k}=\frac{1}{\ln \boldsymbol{m}}$, calculate the $\boldsymbol{j}$ entropy of the term. For the $\boldsymbol{j}$ index, greater the difference in the index value $\boldsymbol{X}_{ij}$, the greater the effect on the scheme evaluation, and the smaller the entropy value. Let the difference coefficient of the $\boldsymbol{j}$ index be $\boldsymbol{g}_j=1-\boldsymbol{e}_j$. Then $\begin{bmatrix}\boldsymbol{g}_1 & \boldsymbol{g}_2 & \boldsymbol{g}_3 & \boldsymbol{g}_4 & \boldsymbol{g}_5\end{bmatrix}$ = Then the weight of the $\boldsymbol{j}$ index $\boldsymbol{W}_j=\frac{\boldsymbol{g}_j}{\sum_{\boldsymbol{j}=1}^{5}\boldsymbol{g}_j},\boldsymbol{j}=1,2,3,4,5$ finally calculates the comprehensive score $\boldsymbol{S}_i=\sum_{\boldsymbol{j}=1}^{5}\boldsymbol{W}_j*\boldsymbol{P}_{ij}(\boldsymbol{i}=1,2...,10)$ of each stock to obtain the comprehensive evaluation result of entropy value method.

Through analysis, we can see that the weights of five indicators are respectively $0.199424871,0.199424458,0.201130291,0.199424766,0.200595615$; Compared with the results($0.197,0.212,0.195,0.206,0.229$) of the factor analysis method, the highest, with an increase in volume weight coefficient in more objective real cases, the highest price and volume to an investor's investment guide effect more obvious, and the lowest coefficient of weight drops, suggests that investors in the stock market investment, for the sensitivity of the stock price is not high, the stock is expected low-cost purchase psychology, extremely easy to miss the best buying opportunities. This further demonstrates that the investment index model itself has

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0013

certain deficiencies. But the entropy value method is based on the variation degree of indicators to determine the index weight, is a kind of objective method, to avoid the artificial factor differences, and factor analysis subjectively put some related variables comes down to some comprehensive factor, so the entropy value method is more objective results, and further optimization of the model.

3.7. ARIMA Model Prediction

In the reasonable modeling of the current index fluctuation and the index fluctuation in the next year, the volatility of the index needs to calculate the volatility of the stock itself. Therefore, the stock return rate can be calculated and the future volatility trend can be predicted and analyzed by using the ARIMA model according to the volatility of the return rate.

In the next year (March 27, 2019 solstice March 26, 2019), we remove non-trading days such as weekends and holidays to get 243 trading days.

Considering that the data of the original return sequence$R_I$may be non-stationary, the stock return sequence $R_I$is obtained after the first-order difference of the original return sequence $\{R_I\}$. ADF unit root test is carried out and the results show that the significance test is passed. Therefore, it can be known that the stock return sequence is stable. According to figure the ACF autocorrelation function and partial autocorrelation PACF function diagram, lag of 2 order is appropriate in ACF and lag in PACF 11 order is appropriate. Thus establishing the model of ARIMA (11,1,2) by a residual autocorrelation coefficient test can be found that the model residuals didn't pass the significance test. Namely the residual autocorrelation coefficient is zero, ARIMA (11,1,2) can be well fitted 001 stock trading volume.

4. CONCLUSION

By using the ARIMA (11,1,2) model to predict the volume data of the next 243 periods, it can be seen that 001 stock tends to have negative returns in the future volatility chart of the stock, so it has no investment value in the short term or the long term.

5. MODEL EVALUATION

(1) Advantages of the model

Entropy method is an objective weighting method, and the index weight value obtained is more credible and accurate than the subjective weighting method.

The ARIMA model is very simple and only requires endogenous variables without recourse to other exogenous variables.

In the stock investment index model, a factor analysis method is used to minimize information loss, condense many original variables into a few factor variables, and has strong explanatory ability and high credibility, which can effectively explain problems. And the new variables are independent of each other, eliminating multicollinearity. Can identify common factors affecting variables.

(2) Shortcomings of the model

There is a dispute about whether each indicator is positive or negative by an entropy method;

The entropy method lacks the horizontal comparison among various indexes;

The weight of each indicator of entropy method changes with the change of samples.

The ARIMA model requires that time sequence data be stationary or stable after differencing, which is tricky for data preprocessing.

In essence, ARIMA model can only capture the linear relationship, but not the nonlinear relationship, which limits the promotion of the model.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202008_6(8).0013

In the stock investment index model, different factor analysis methods are used to calculate the factor score, which results in different coefficient results.

REFERENCES

[1] Hu di, huang wei. Stock price prediction based on ap-svm combination model [J]. Journal of wuhan university of technology,2019,41(03):296-302.

[2] Peng yan, liu yuhong, zhang rongfen. Modeling and analysis of stock price prediction based on LSTM [J]. Computer engineering and applications,2019,55(11):209-212.

[3] Gao Qinjiao, ZHANG Shenggang, JIA Xiaowei. Research and Application of Stock Price Prediction Model Based on Support Vector Machine [J]. Curriculum Education Research,2016(28):227.

[4] Su Zhenghao. Disposal Effect in Stock Investment Decision and its Countermeasures [J]. Finance and Economics (Academic Edition),2016(15):100+144.