Advances in Artificial Intelligence, Big Data and Algorithms
G. Grigoras and P. Lorenz (Eds.)

doi:10.3233/FAIA230898

The Broad Category Asset Investment Forecast Solutions Based on LSTM Algorithm and Macroeconomic Development

Hongtao ZHANG[a], Yuantong CHEN[a1], Wanying ZHANG[a], Yefeng LI[a], Qiting ZENG[b]
[a] *School of Electrical and Computer Engineering, Nanfang College, Guangzhou*
[b] *School of Accounting, Nanfang College, Guangzhou*

Abstract. Globally, category asset allocation as the core investment method has been the consensus of the industry. Macroeconomic index data and market data of various asset indexes reflect different economic states of each economic cycle and can calculate the risk and return characteristics of a certain type of asset, and construct passive investment portfolio. However, for China's investors and investment environment, we need a more suitable macroeconomic and broad asset category forecast model. In order to find high-frequency and effective macroeconomic indicators, we divided China's macroeconomic operation status from 2001 to 2021 into different economic states. In this paper, macroeconomic environment such as economic growth, inflation and interest rate of China in the next five years is simulated through macroeconomic models or other mathematical models. According to the above calculated GDP growth rate, the consumer price index (CPI) is used to calculate the inflation rate of money and commodities, and the monthly interest rate of demand deposit is calculated to simulate and forecast the macroeconomic development situation of China in the next five years. LSTM algorithm is used for sliding window forecast of economic growth, inflation and interest rates in the next five years. The risk-return characteristics, including expected return, standard deviation of return and Sharpe ratio, of major asset indexes in the data set under various economic conditions were calculated by using the four economic conditions divided by the Merrill Lynch clock frame. The indexes listed in the data set were arranged and combined, and the risk-return characteristics of the recent five years were selected for prediction, so as to obtain the investment portfolio with relatively suitable recovery stage and high return in the next five years.

Keywords. Merrill Clock Frame, LSTM algorithm, Pearson coefficient, GDP growth rate, M2 index

1. Introduction

The Long Short Term Memory (LSTM) model is essentially a Recurrent Neural Network (RNN) of a particular form [1]. LSTM model solves the problem of RNN short-term

[1] Corresponding author: Yuantong CHEN, School of Electrical and Computer Engineering, Nanfang College; e-mail: chenyt200022@stu.nfu.edu.cn

memory by adding Gates on the basis of RNN model [2], [3], so that the cyclic neural network can effectively utilize the long distance time sequence information [4], [5].

Major asset index allocation [6] can smooth portfolio risks for investors through investment diversification, and allocate investment funds among different asset classes according to investment needs. However, it is difficult for us to predict which type of asset will perform better in the future, so major asset allocation will inevitably face the problem of dynamic adjustment in the course of time. Therefore, the selection of appropriate macroeconomic and broad asset forecasting models and investment portfolios is conducive to the sound development of China's investment environment and investors' choice of appropriate portfolios to achieve higher returns.

Merrill Lynch clock theory [7] is the basic theory of investment timing. The Merrill Lynch clock framework divides macroeconomic operation conditions into four economic states: recession, recovery, overheating and stagflation. According to economic principles and historical experience, we can rotate capital allocation among major asset classes, and start to allocate this type of asset when the price of such asset is expected to rise [8]. Bonds do best in downturns, stocks do best in recoveries, commodities do best in overheating and cash-like assets do best in stagflation.

The Consumer Price Index (CPI) is a relative number reflecting the trend and degree of price changes of consumer goods and services purchased by urban and rural residents in a certain period. It is the result of comprehensive calculation of the consumer price index of urban residents and rural residents. Through the index can observe and analyze the retail price of consumer goods and service item price changes on the actual living expenses of urban and rural residents.

Sharpe Ratio, also known as Sharpe Index, a standardized index for fund performance evaluation [9], [10]. The research of Sharpe ratio in modern investment theory shows that the size of risk plays a fundamental role in determining portfolio performance. Risk-adjusted rate of return is a comprehensive index that can consider both returns and risks at the same time, which can eliminate the adverse impact of risk factors on performance evaluation in the long run.

According to the existing public data set of macroeconomic indicators and market data of major asset indexes, it can be seen that the macroeconomic indicators include finance, industry, national economic accounting, price index, prosperity index, employment and wages, interest rate and exchange rate, population, banking and currency, and other nine specific quantitative data of indicators.

According to the data collected above, solve the following problems:

Two or more high-frequency and effective indicators are found from the macroeconomic indicator data as specific indicators to classify different economic states under the domestic macroeconomic operation from 2001 to 2021. Establish macroeconomic models and other mathematical models to simulate China's macroeconomic environment in the next five years, including economic growth, inflation, interest rate and other indicators, and judge which economic state it is in according to the divided economic state. According to the market data of major asset indexes, four indexes are selected to represent four types of assets, including stocks, commodities, bonds, cash and their equivalents, and the risk-return characteristics, which including expected return, standard deviation of return rate, Sharpe ratio or others, under different economic conditions are predicted. At the same time, correlation analysis is carried out among various asset indexes.

2. Problem Analysis

2.1. Find Two or More Quantitative Macroeconomic Data to be Used as Indicators of Economic Development during the Past 20 Years from 2001 to 2021

At least two or more macroeconomic indicators are selected as the classification indicators of different economic states based on the trend and situation of macroeconomic development from 2001 to 2021. Among them, the inflation rate, money supply M2 and GDP growth rate in the Merrill Lynch clock framework are selected as the important indicators of domestic macroeconomic operation. The selection of the above three macroeconomic indicators can more objectively reflect the different economic states of domestic macroeconomic development from 2001 to 2021.

2.2. According to the Mathematical Model, the Macroeconomic Environment of China in the Next Five Years is Simulated and the Economic State that China will Face in the Next Five Years is Judged

Based on the known yearly GDP and monthly consumer price index CPI, according to the inflation rate of money supply M2 index, the CPI demand deposit interest rate in December of each year is selected to simulate and forecast the economic development trend of China in the next five years.

2.3. According to Various Economic Conditions of the Broad Asset index of Its Risk and Return Characteristics and Calculate the Correlation between the Broad Asset Index

Based on various economic states divided, which are recession, recovery, overheating and stagflation, appropriate broad asset indexes are selected for calculation. As can be seen from the quantitative values of broad asset indexes in Annex 2: The four categories of assets include stocks, bonds, commodities, cash and their equivalents, etc. Stocks include four indexes, namely Shanghai Stock Exchange 50 (SSE 50), Shanghai and Shenzhen 300 (CSI 300), China Stock Exchange 500 (CSI 500) and China Stock Exchange 1000 (CSI 1000), while commodities include two indexes, namely South China Commodity Index (NHCI) and S&P Gaosen Total Return Index. Bonds include three indexes: (1) China Bonds and composite wealth, (2) cash and its equivalents include money funds and (3) Sharpe ratio. The Sharpe ratio is the difference between comparing an investment to a benchmark representing the entire investment class and getting an excess return, that is, the higher the Sharpe ratio, The higher the excess rate of return per unit of risk. In correlation calculation, significance is used to calculate whether there is correlation between asset indexes of various categories. Pearson's coefficient is used to calculate the value of correlation coefficient to further compare the size of correlation.

2.4. Select an Appropriate Investment Portfolio and Predict Its Risk-Return Characteristics according to the Economic Conditions Predicted in the next Five Years

Based on the risk and return characteristics of various economic states and the correlation

between various asset indexes, the suitability of various asset categories for investment is determined according to the correlation coefficient between various indexes. The investment portfolio is mainly based on the broad categories of assets with high expected returns and moderate Sharpe ratio. According to the economic state of China in the next five years, the stock is given priority, followed by bonds. Considering the risk-return characteristics of each stock and bond, two appropriate stocks, bonds, monetary funds and commodity indexes are selected for portfolio investment.

3. Proposed Methods

3.1. Data Preprocessing

Data related to macroeconomic performance during the twenty years from 2001 to 2021 are selected as indicators to classify different economic states. Various indexes of macroeconomic performance are cut off in time slices of one unit in each year, and the CPI demand deposit interest rate in December of each year is selected as compared with the same period (per month).

The Dataframe was stored as a two-dimensional array, and then the difference transformation was carried out to convert the time series form data into a supervised learning set. Meanwhile, the data set was divided into a training set and a test set. The MinMaxScaler was used to scale the data to [0,1] to accelerate the convergence.

3.2. Model Preparation

3.2.1. Determine Quantitative Index

a) The GDP growth rate is calculated according to the GDP growth rate calculation formula, as follows:

$$\text{GDP growth rate} = \frac{\text{Current GDP} - \text{Previous Period GDP}}{\text{Previous Period GDP}} * 100\% \quad (1)$$

GDP growth rate is the real GDP growth rate calculated by the retail price index of commodities in 1978.

b) According to the M2 index of money supply in the data set, we can know:

$$\begin{aligned}&\text{M2}\\&= \text{M1} + \text{quasi currency}(\text{time deposit} + \text{resident time deposit} + \text{other deposit}\\&+ \text{deposit of securities company clients} + \text{housing provident fund}\\&+ \text{deposit of non}\\&- \text{depository financial institutions in depository financial institutions})\end{aligned}$$

c) According to the definition of LSTM neural network, we can know:

At time ***t***, the formula defined by LSTM neural network is as follows:

$$f_t = \text{sigmoid}(W_f * [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \text{sigmoid}(W_i * [h_{t-1}, x_t] + b_i) \quad (3)$$

$$o_t = \text{sigmoid}(W_o * [h_{t-1}, x_t] + b_o) \quad (4)$$

$$\tilde{c}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \quad (5)$$

$$c_t = f_t \times c_{t-1} + i_t * \tilde{c}_t \quad (6)$$

$$h_t = o_t \times \tanh(c_t) \quad (7)$$

3.2.2. Identify Classified Macroeconomic Indicators

a) Gross domestic product GDP growth rate

Gross Domestic Product (GDP) growth rate refers to the annual growth rate of GDP, which is calculated using gross domestic product at comparable prices. GDP growth rate is one of the four important macroeconomic indicators.

b) The growth rate of money supply M2

M2, a broad measure of money supply, includes all forms of money that may become real purchasing power and usually reflects the pressure of changes in social aggregate demand and future inflation. In recent years, many countries have taken M2 as the target of money supply regulation.

c) Expected return

The expected return can be obtained by subtracting the closing index of the previous day from the closing index of the next day.

$$\text{Standard deviation of return} = \sqrt{\frac{\sum_{i=1}^{n}(\text{Day's gain} - \text{average gain})^2}{\text{Days of the year} - 1}} \quad (8)$$

$$\text{Sharpe ratio} = \frac{E(R_p) - R_f}{\sigma_p} \quad (9)$$

Where, $E(R_p)$ denotes the expected rate of return of the portfolio, R_f denotes the risk-free rate of interest, and σ_p denotes the standard deviation of the portfolio.

d) Correlation coefficient (Pearson coefficient)

Covariance is an index that reflects the degree of correlation between two random variables. If one variable increases or decreases with the other variable, the covariance of the two variables is positive.

$$\text{cov}(X, Y) = \frac{\sum_{n}^{i=1}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \quad (10)$$

Although covariance can reflect the degree of correlation between two random variables, where covariance greater than zero means positive correlation, and less than zero means negative correlation, the size of covariance value cannot be a good measure of the degree of correlation between two random variables. Therefore, Pearson correlation coefficient was introduced to divide the standard deviation of the two random variables on the original basis.

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (11)$$

3.2.3. Combined with Quantitative Data of Macroeconomic Indicators in the Data Set, We Can Obtain as Fig.1 Shows:

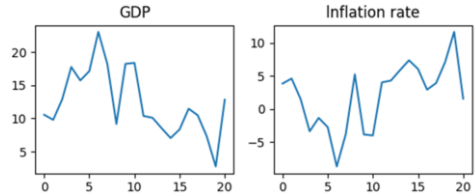

Figure 1. GDP growth and inflation line graph

3.3. Concrete Steps

Step1: The data set is sorted into the common data type in supervised learning. One behavior is one sample, the number of rows is the number of samples, and the number of columns is the total number of variables.

The preprocessed data is converted from DataFrame to two-dimensional array and the data type is converted to float64; The data was normalized and scaled to 0~1 using the MinMaxScaler.

Step2: Build a function to generate time series sliding window data set.

Because the model is used to predict the future data, time series sliding window method is needed. Continuously delete the first element of the input array, and continuously add the future prediction value to the last element of the input array, thus achieving rolling prediction. Single-step prediction of scrolling is achieved by continuously scrolling back through the input array.

Step3: Separate the training set from the test set.

Generate a scrollable training set and test set using the sliding window function.

Step4: Construct a LSTM model and add training set to train the model.

The training set generated in the above step is converted into a dimension shape acceptable to the LSTM model, which is added into the LSTM model for module training, and the trained prediction model is obtained.

Step5: The trained LSTM model and time sliding window test set are used to predict future data.

Step6: Pick four indexes that represent four asset classes, including stocks, bonds, commodity indexes, cash and its equivalents.

According to the Merrill Lynch clock frame, the four economic states are divided into recovery stage, overheating stage, recession stage and overheating stage, and the corresponding major asset indexes of the four stages are stocks, commodity indexes, bonds, cash and their equivalents.

Step7: Calculate the risk-return characteristics, including expected return, standard deviation of return, Sharpe ratio, etc. in each macroeconomic state.

[Assume Sharpe ratio risk-free ratio of 3.0]

Step8: Calculate correlations between broad asset indices

Step9: Based on the LSTM algorithm of the time series prediction model, the risk-return characteristics, whcih including expected return, standard deviation of rate of return, Sharpe ratio, of each year in the next five years are calculated.

Step10: According to a total of 24 permutations and combinations, the optimal investment portfolio is selected based on the correlation coefficient thermal map, that is,

the portfolio with high Sharpe ratio and high expected return is selected, and the risk-return characteristic prediction results for the next five years are obtained.

4. Experimental Results

Correlation coefficient and significance index among asset indexes of various categories were calculated according to Pearson coefficient, as shown in **Fig.2** below:

Figure 2: Correlation coefficient heat map

From the figure above: The highest correlation coefficient is 1.0, which are China debt-Composite Wealth (7-10 years) index and China debt-Composite wealth (3-5 years) index.Then is 0.997, which are China debt-composite wealth (3-5 years) index and China debt-composite wealth (≤1 year) index. Then is 0.996, which are China National Bond - Composite Wealth (7-10 years) index and China National Bond - Composite Wealth (≤1 year) index, etc.

China's macroeconomic environment of economic growth, inflation, interest rate and so on in the next five years can be calculated by GDP growth rate, monthly CPI inflation rate, demand deposit interest rate, etc. The calculation results are shown in the following table as **TABLE 1, TABLE 2, TABLE 3** shows:

TABLE 1: China's macroeconomic environment simulation forecast for the next five years

Year	GDP Growth rate	Inflation rate	Interest rate	CPI	M2 money in circulation
2022	13.0655	1.729649	0.3947955	3.893008	13.11112
2023	15.4162	0.660311	0.4037655	3.964447	14.01813
2024	15.2934	-1.874359	0.4080639	4.876406	15.28134
2025	16.8868	-2.376777	0.4129230	4.616366	15.86647
2026	16.6093	-2.186923	0.4189941	4.897334	16.29738

TABLE 2: South China Commodity Index forecast results for the next five years

Year	Standard deviation of return	Sharpe ratio	Expected return
2017	0.043374619	0.467239611	124.97
2018	0.022668053	-0.76199074	-84.39
2019	0.058590237	1.273462598	198.52
2020	0.034964991	0.353907851	113.26
2021	0.076560696	0.845784217	306.77
2022	0.03280576	-0.2849801	-49.476562
2023	0.03137872	-0.40299416	-26.07444
2024	0.03132049	-0.41605207	-26.609228
2025	0.0313181	-0.41749585	-26.59543
2026	0.031318	-0.41765544	-26.595781

TABLE 3: Prediction results of risk - return characteristics of monetary funds in the next five years

Year	Standard deviation of return	Sharpe ratio	Expected return
2017	0.015466712	5.014416118	54.1422
2018	0.015120144	4.498544144	54.4926
2019	0.010644697	-2.52078601	39.7483
2020	0.008642011	-7.26647575	32.6279
2021	0.009201418	-6.60841389	35.6539
2022	0.00966733	-5.505425	35.878704
2023	0.00978025	-5.303847	35.879242
2024	0.00980811	-5.26689	35.879242
2025	0.00981502	-5.260112	35.879242
2026	0.00981673	-5.258868	35.879242

[Calculation results of other asset categories are the same as shown in the table above]

5. Model Evaluation

5.1. Advantage

CNN and other models are not fully applicable to learning time series, so a variety of auxiliary processing is needed, and the effect may not be good. Faced with problems and tasks sensitive to time series, it is more appropriate to adopt the model of RNN class. RNN is used for sequence data and has certain memory effect.

As an excellent variant model of RNN, LSTM inherits most of the characteristics of RNN models, and solves the problem of ladder disappearance caused by gradual reduction in gradient backpropagation process, and can realize long-term data storage input.

5.2. Shortcoming

The LSTM model has disadvantages in parallel processing and is not as effective as some of the latest networks.

The gradient problem of RNN is solved in LSTM and its variants to a certain extent, but it is difficult to deal with a large number of sets of sequence data.

6. Conclusions

In this paper, LSTM algorithm is used for sliding window prediction of China's economic growth, inflation and interest rate in the next five years, and it is found that China's macroeconomic environment in the next five years is in the recovery stage. Based on this economic state, the risk-return characteristics of the investment portfolio are predicted. Among the major asset classes, stocks are preferred with high expected return and moderate sharpe ratio, followed by bonds. Therefore, a large number of stocks can be purchased, and then appropriate major asset indexes can be selected to construct the investment portfolio according to the relevant heat diagram of the obtained major asset indexes. This paper calculates the correlation coefficient and significance index among

various asset indexes according to Pearson coefficient, and finds that the suitable investment portfolios are: SSE 50, CSI 300, NHCI, China Bond - Comprehensive wealth (3-5 years), China bond - comprehensive wealth (7-10 years), money fund, etc.

Acknowledgment

This work is supported by the Ministry of Education industry-school cooperative education project (220505211275629).

References

[1] Xingjian Shi, Zhourong Chen, Hao Wang, and Dit-Yan Yeung, 'Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting'.Available: http://arxiv.org/abs/1506.04214
[2] I. Sutskever, O. Vinyals, and Q.V. Le, "Sequence to sequence learning with neural networks," In Proceedings of Advances in Neural Information Processing Systems, 2014, pp. 3104-3112.
[3] R. Pascanu, T. Mikolov and Y. Bengio, "On the difficulty of training recurrent neural networks," Computer Science, 2013, 52(3), pp.337-345.
[4] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Computation, 1997.
[5] T. Mikolov, "Statistical Language Models based on Netural Networks," ph.D.thesis, Brno University of Tecnology, 2012.
[6] Xueyong Zhang and Lin Zhang, 'A Review of Large Asset Class Allocation Theory Research'.Available: https://kns.cnki.net/kcms/detail
[7] Greetham T, Hartnett H. The Investment Clock Special Report# 1: Making Money From Macro[J]. Merrill Lynch, 2004.
[8] Yao Qiong. Validity test, Improvement and prediction of Merrill Lynch Clock for China's broad asset allocation [D]. University of International Business and Economics, 2019.
[9] William F. Sharpe, 'Mutual Fund Performance', Vol. 39, No. 1, Part 2: Supplement on Security Prices (Jan., 1966), pp. 119-138 (20 pages)
[10] Sharpe, William F. The Sharpe Ratio. The Journal of Portfolio Management. 1994, 21 (1): 49–58 [June 12, 2012]. doi:10.3905/jpm.1994.409501