1. Introduction

Economy is a complex system. Micro individuals can interact to determine the form of macro economy, and the dynamics of macro economy will affect the behavior of micro individuals [1]. Foreign research shows that genetic algorithm is an effective tool for complex systems. This method has stable measurement results of macroeconomic uncertainty, consistent with the actual economic operation, can effectively identify economic fluctuations and exogenous shocks, and can accurately reflect the impact degree difference of different shocks [2].

Macroeconomic aggregate, generally using regional gross national product (GDP) or stock market index and other controllable factors to describe macroeconomic state [3]. Commonly used macroeconomic state description data is: regional GDP, regional GDP per capita, monetary nominal purchasing power, iconic main grain price, precious metals and legal currency exchange price, commodity volume and transaction price, stock market index, freight index, etc., conventional macroeconomic state analysis, may be applied to dozens of commonly used macroeconomic state description data [4]. These data mostly expand statistics along the time line, in economics, using time-sharing data and K-line data for statistics. The K-line statistical method is the initial value, final value, highest value, and lowest value in the record statistical period. When the final value is higher than the initial value, it is recorded as the forward K-line, and the reverse K-line when the final value is lower than the initial value.

In this study, a series of macroeconomic models based on genetic algorithm are developed to analyze the stock market trading system, including: the linear programming
results are obtained by the least square method. In the analysis, the linear estimation can be obtained for the linear data \( x \), the macro data is a group of dimensional data, and the essence of the mutation algorithm is to obtain the movement direction of the mutation of the data nodes. And the distance between the algorithm obtained by the linear formula and the original data, calculate the average value in the original data, and then undergo binary t-correction, which refers to the binary t-correction results of the one-dimensional matrix before and after adding the final evaluation output factor. In this study, genetic algorithm is introduced as the core algorithm. In the algorithm efficiency verification test, the calculation model based on genetic algorithm is constructed in Matlab environment, and the data space construction mode and genetic variation mode of genetic algorithm are explored. Finally, a high-throughput macroeconomic timing prediction scheme based on genetic algorithm is designed.

In the macroeconomic analysis, it is an important analysis goal to use the previous K-line data to make an early warning or forecast of the K-line data in a certain period of the future [5]. Genetic algorithm was introduced in this study to use certain periodic data as initialization data to obtain more accurate macroeconomic prediction data by obtaining target data for maximum fitness. Using the macroeconomic indicators (GDP) and demographic data, based on the macro vulnerability model and the genetic algorithm, we can also effectively predict and evaluate the economic losses in other emergencies [6].

2. Design of the Macroeconomic Genetic Algorithms

The essence of genetic algorithm is to form the determinable original data into data space, take the target data as the target data, use the mutation algorithm to mutate the target data, and finally form the data with the highest evaluation result as the output data. Because the dynamics of complex adaptive systems are usually nonlinear, sometimes even chaotic, it is difficult to obtain accurate analysis results by using traditional linear and fixed-point methods. The algorithm is mainly used for data prediction and analysis with large chaotic effect or unclear data transmission logic. The general mode of genetic algorithm is shown in Figure 1.

For the macroeconomic analysis process, there is a large volume of Basic Data data (most of the original data cycles provided by the platform software are several years to decades), and the analysis and prediction cycle is short (generally need to push the data from 1 to 3 days before). Therefore, the influence of the different raw data selection cycles and the different relevant raw data volume on the accuracy of the final analysis results was explored. The available data structure of the macroeconomic economy is shown in Figure 2.

Combined with the calculation link in Figure 2 above, the following algorithms need to be elaborated.

2.1. Linear Planning Algorithm. Linear planning results were obtained by least squares, and in the analysis, for the linear data \( X \), linear estimates can be obtained:

\[
\hat{\beta} = (X^TX)^{-1}X^Ty,
\]

where \( X \) is the statistical array data; \( X^T \) is the statistics of the last data in the 1D array; and \( \hat{\beta} \) is the output data.

2.2. Variability Algorithm Based on Target Weights. The essence of the variation algorithm is to obtain the movement direction of data node variation, because the macroeconomic data are 1-dimensional array data, so the movement direction of freedom is only one dimension, obtaining the direction and forward or reverse movement, the algorithm in the previous formula (1), and the distance between the original data calculation average in the original data:

\[
\Delta x = \frac{\text{Avg}(X) - \hat{\beta}}{\max(X) - \min(X)},
\]

where \( \hat{\beta} \) is the linear planning results; \( \max(X) \) is the maximum value of the raw data \( X \); \( \min(X) \) is the minimum value of the raw data \( X \); \( \text{Avg}(X) \) is the arithmetic average value of the raw data calculated in formula (5); and \( \Delta x \) is the final variation factor of the data variation.
2.3. Decision Coefficient $R^2$ Value Algorithm. In the study design, the determination factor $R^2$ value and bivariate check $t$ value are used to form the evaluation factor, and the final $R^2$ value and $t$ value are considered to reach the output standard, where the $R^2$ value algorithm is

$$R^2 = \frac{\beta^T X^T y - (1/n)y^T uu^T y}{\sum_{i=1}^{n} x_i^2 - 1/m(\sum_{i=1}^{n} x_i)^2}$$

where $R^2$ is the decision coefficient evaluation result; other mathematical symbols are found in formula (1) and formula (2).

2.4. Bivariate $t$ Calibration Algorithm. The $t$ test, also known as the bivariate $t$ test, is a $t$-distribution theory used to infer the probability of differences occurring and thus compare whether the difference between the two mean values is significant. It is mainly suitable for a normal distribution with a small sample size, and the overall standard deviation is unknown. The difference of the data is compared by statistical bivariate $t$ calibration method, where the Value value is $t$ value. When $t < 10.000$, the smaller statistical difference. Log value is $P$ value, statistical reliability is considered when $P < 0.05$, and statistical significance is considered when $P < 0.01$.

The bivariate $t$ test designed in this study refers to the results of the bivariate $t$ test for adding the one-dimensional matrix before the final evaluation output factor and the one-dimensional matrix after the final evaluation output factor, as shown in

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{((n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2/n_1 + n_2 - 2)((1/n_1) + (1/n_2))}}$$

where $\overline{X}_1, \overline{X}_2$ is the average calculated value of column $X$ before and after the final evaluation of the output factor; and $n_1, n_2$ is the number of column $X$ elements before and after adding the final evaluation output factor.

The mean value algorithm is the formula (5):

$$\overline{X} = \sum_{i=1}^{n} x_i,$$

where the meaning of mathematical symbols is the same as the previous formulas;

The standard deviation rate is calculated as

$$\sigma = \frac{1}{n-1} \sqrt{\sum_{i=1}^{n} (x_i - \overline{X})^2},$$

The raw data acquisition mode and analysis results used in this study. Opening & Maximum & Minimum & Closing: initial value, maximum value, minimum value, final value in K-line data; result of $t$-bivariate $t$-verification, and evaluation value of determination coefficient in $R^2$ regression analysis.
where the meaning of mathematical symbols is the same as the previous formulas.

3. Validation of the Algorithm Efficiency and the Final Algorithm Optimization

In the algorithm efficiency verification test, the genetic algorithm-based computational model is constructed under Matlab, and the final validation results are obtained using the algorithm accuracy. The algorithm accuracy calculates the absolute value between the difference and the actual data, and then calculates the arithmetic average of the ratio, as shown in Equation (7):

$$ B = \frac{1}{n} \sum_{i=1}^{n} \frac{|x_i - \bar{x}|}{x_i}, $$

where $B$ is the results of algorithm accuracy evaluation; other mathematical symbols share the previous formulas.

3.1. Effect of the Raw Data Cycle on the Prediction Accuracy.

In this comparison model, the prediction results of the data of this column supported by a single column data are discussed. Comparing the data of 1 day, 3 days, 7 days and 15 days supported by 1 month to 12 months, respectively, the prediction data travel K-line and form 4 prediction data per day. The final evaluation results are shown in Table 1.

| Data prediction accuracy (%) | Raw data cycle |
|-----------------------------|---------------|
| 1d                          | 93.69         |
| 3d                          | 98.76         |
| 7d                          | 99.63         |
| 15d                         | 99.87         |
| 1m                          | 98.76         |
| 2m                          | 99.63         |
| 3m                          | 99.87         |
| 4m                          | 99.94         |
| 5m                          | 99.91         |
| 6m                          | 99.89         |
| 7m                          | 99.93         |
| 8m                          | 99.92         |
| 9m                          | 99.99         |
| 10m                         | 99.99         |
| 11m                         | 99.99         |
| 12m                         | 99.99         |

Table 2: Relationship between the number of relevant data columns and the prediction accuracy.

| Number of original data columns | Data prediction cycle |
|---------------------------------|-----------------------|
|                                 | 3 m-7d                |
|                                 | 3 m-15d               |
| 1                               | 70.13                 |
| 2                               | 88.61                 |
| 3                               | 93.58                 |
| 4                               | 97.42                 |
| 5                               | 98.85                 |
| 6                               | 99.34                 |
| 7                               | 99.63                 |
| 8                               | 99.74                 |
| 9                               | 99.86                 |

$d = \text{day}; m = \text{month}.$

To observe Table 1 data more visually, Table 1 data were visualized to obtain the results of Figure 3.

Observe Table 1 and Figure 3 data. The following data rules were found. The forecast target was 1 or 3 days. High data prediction accuracy can be obtained even with only 1 month of data support. According to the engineering industry consensus on AI data analysis, when the prediction accuracy exceeds 95%, the prediction software has an engineering value. So, under the push-forward prediction period of 1 day ago, select the correlation columns for 2 month data as the base data. Under the push-forward
prediction period of 3 days ago, select 3 month data as base data. But under the 7 days and 15 days, it has been impossible to simply increase the amount of relevant data using the algorithm to complete accurate data prediction above 95% accuracy.

3.2. The Effect of the Number of Reference Data Columns on the Prediction Accuracy. To further provide the system accuracy under the longer push forward push prediction period requirements, the study assumes hidden conduction logic relationships between other relevant economic data and before the predicted data. For example, there is a hidden transmission logic relationship between bulk material delivery and regional GDP. These relationships are difficult in economics to construct conduction coefficients under statistical models, but the broad association between economic indicators supports genetic algorithms when making forward predictions of certain economic indicator data. Additional data are introduced as the reference data. The raw data space was constructed using the corresponding data columns of the predicted data together with the other data columns, to explore the impact of this approach on the machine learning results of genetic algorithms. The results are shown in Table 2:

To observe Table 2 data more visually, Table 2 data were visualized to obtain the results of Figure 4:

![Figure 4: Relationship between the number of relevant data columns and the prediction accuracy.](image-url)

In Figure 4, whether the prediction data of 7 days ago or the forecast data, increase the reference of original data columns, can make the prediction data accuracy is greatly improved, when the reference data column reaches 4 columns, 7 days of prediction accuracy of more than 95%, when the reference data column reaches 6 columns, 5 days of data support prediction accuracy of more than 95%. However, the data in Figure 3 above show that under certain conditions, the accuracy of the macro data prediction and analysis

| Number of original data columns | Data prediction cycle(15d) |
|-------------------------------|--------------------------|
|                               | 1 m | 3 m | 5 m | 7 m | 9 m | 11 m |
| 1                             | 33.29 | 42.25 | 47.64 | 51.59 | 52.64 | 53.32 |
| 2                             | 54.83 | 63.18 | 67.24 | 74.19 | 82.43 | 84.29 |
| 3                             | 68.76 | 85.27 | 87.85 | 88.65 | 91.29 | 92.35 |
| 4                             | 86.42 | 92.36 | 93.12 | 94.58 | 94.36 | 95.57 |
| 5                             | 93.24 | 94.18 | 95.28 | 96.73 | 97.27 | 97.69 |
| 6                             | 94.15 | 95.35 | 97.43 | 97.92 | 98.61 | 98.94 |
| 7                             | 95.33 | 96.47 | 98.82 | 99.05 | 99.36 | 99.73 |
| 8                             | 96.25 | 97.29 | 99.53 | 99.57 | 99.89 | 99.99 |
| 9                             | 96.18 | 97.94 | 99.79 | 99.98 | 99.99 | 99.99 |

$d =$ day; $m =$ month.
results can exceed 99.9%. For the macroeconomic data prediction results of radical investment projects, the higher the prediction accuracy represents the stronger the risk control ability. Therefore, the study continues to assume that when the reference data column and the original data cycle increase simultaneously, the prediction data accuracy will be more significantly improved. Therefore, in the 15 days before the prediction, different reference data columns and reference data cycles, and the experimental results are obtained in Table 3:

To observe Table 3 data, Table 3 data were visualized to obtain the results of Figure 5:

Observe the data in Table 3 and Figure 5, it is found that increasing the number of reference data columns and the original data cycle can improve the accuracy of pushing data prediction 15 days ago. Taking the 99.9% accuracy threshold as an example, when the original data cycle reaches 7 months and the number of reference data columns reaches 9 columns, 99.9%, while the data cycle reaches 11 months, only 8 columns can reach 99.9%. Data prediction accuracy. It can be seen that increasing the number of reference data columns and the original data cycle can realize the accurate prediction of macroeconomic data.

4. The Final Evaluation Ability Test of Macroeconomics by Genetic Algorithms

The application of macroeconomic data prediction software basically focused on the trading platform (stocks, futures, precious metals, etc.) transaction price and comprehensive index prediction, the study uses the genetic algorithm to build the Matlab simulation environment, the original data cycle set to 12 months, related data columns set to 9 columns, push cycle evaluation 1 days, 7 days and 15 days, comparison software for paid version full function flush stock price prediction software (10jqka), from 2000, the algorithm simulation software and reference software, get Table 4

In Table 4, all data have $t < 10.000$, $t < 0.000$ and $P < 0.01$ with significant statistical differences. From the data ratio, the prediction accuracy of the 1-day push prediction period varied, but as the prediction period increased, the two sets of software gradually increased and decreased both $t$-value and $P$-value. We prove that the software has significant algorithm lead in the 10jqka period compared with the 10jqka software. The principle of the process is analyzed, because the 10jqka prediction method is to construct
multiple predictors based on all previous data, and after weighting each index, we form the prediction results. The software focuses more on using machine learning to search for data rules to achieve accurate prediction of data. Moreover, the algorithm can not only realize the prediction of stock price and stock index, but also realize that of other macroeconomic data, as long as the original data can be generated for a long enough time.

5. Summary

The development of computer technology makes it possible for us to intelligently simulate the behavior of a monomer in the whole environment. In recent years, the introduction of genetic algorithm in artificial intelligence has applied genetic algorithm to the research of economics. As the core model of macroeconomic forward time cycle prediction, it can achieve high data prediction accuracy and provide data support for macro-control, bulk trading and financial investment with the support of sufficient computing capacity and equipment. However, the algorithm requires high computing power [7]. When most institutions use the algorithm at the same time, the technology leading position of the algorithm will decline. Therefore, using this algorithm can improve the short-term risk control ability of financial institutions, but in the process of economic game, new algorithms still need to be developed.

Data Availability

The data underlying the results presented in the study are available within the manuscript.

Disclosure

We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

Conflicts of Interest

There is no potential conflict of interest in our paper.

Authors’ Contributions

All authors have seen the manuscript and approved to submit to your journal.

References

[1] F. Tong and S. Zhi, “Study on the design and optimization algorithm of a multivariate hybrid GARCH model based on LASSO,” Quantitative, Economic, Technical and Economic Research, vol. 38, no. 12, pp. 146–163, 2021.
[2] Y. Shuai, “Macroeconomic uncertainty measure study —— is based on the GA-SVR model,” Qinghai Finance, vol. 3, pp. 28–35, 2021.
[3] Y. Jingfeng, “Application of gene expression programming in macroeconomic prediction,” Journal of Xuchang College, vol. 37, no. 12, pp. 64–67, 2018.
[4] Zhang, “Prediction of grey neural networks based on genetic algorithms,” Statistics and Decision-Making, vol. 17, pp. 79–81, 2014.
[5] W. Jinsong and S. Wei, “Application of multiple group genetic neural networks in stock index prediction,” Statistics and Decision-Making, vol. 14, pp. 76–79, 2014.
[6] L. Jing and L. Li, “Rapid assessment of earthquake economic loss based on the macro-vulnerability model,” People’s Yangtze River, vol. 40, no. 24, pp. 79–81, 2009.
[7] C. Zhigao, “Application of genetic algorithms and BP neural networks in GDP prediction,” Computer and Digital Engineering, vol. 37, no. 9, pp. 172–175, 2009.