Research Article
An Improved Data Mining Model for Predicting the Impact of Economic Fluctuations

Liting Zhou, Qiang Chen, and Tianyu Zhu

1Department of Finance, Huaibei Vocational & Technical College, Huaibei 235000, China
2Training Centre, Huaibei Vocational & Technical College, Huaibei 235000, China
3School of Information, Huaibei Normal University, Huaibei 235000, China

Correspondence should be addressed to Tianyu Zhu; zhuty@chnu.edu.cn

Received 8 January 2022; Revised 27 February 2022; Accepted 11 March 2022; Published 31 March 2022

In order to explore the influencing factors of economic fluctuations, this article combines data mining technology to mine the factors that affect economic fluctuations and introduces financial network theory to the transfer problem of investment strategies, which provides an effective research method for economic behavior research. Moreover, this article calculates the fund network parameters from a quantitative perspective by constructing the topological characteristic index of the financial network to illustrate the influence of the degree of network evolution and the time scale. In addition, this article constructs an economic fluctuation data mining model based on economic fluctuations and uses data to verify the effects of the method proposed in this article. The experimental results show that the data mining method proposed in this article can play an important role in the analysis of factors affecting economic fluctuations and accurately mine the relevant factors that affect the economy.

1. Introduction

With the continuous advancement of economic globalization and world economic integration, countries in the world have formed interconnected and interdependent organisms to a certain extent through economic activities such as trade and investment [1].

In order to promote the rapid recovery of their own economies after the financial crisis, governments around the world have implemented economic policies more actively. However, the frequent promulgation of economic policies has increased the complexity of economic ties between countries, reduced the stability of the country’s economic system, and aggravated the uncertainty of economic policies, thereby adversely affecting macroeconomic fundamentals and slowing the speed of economic recovery. There is currently no authoritative definition of the concept of economic policy uncertainty. The mainstream view is that the inconsistency between the current and expected views of economic participants on the impact of economic policies on the economy has caused economic policy uncertainty, and economic uncertainty is closely related to economic operations and economic policies. The variability of the market makes possible mismatches and time lag in the transmission of economic policies, resulting in failure to produce the expected effect of resource allocation by the government, resulting in uncertainty in economic policies. Especially during the financial crisis, while the government implemented a series of fiscal or monetary policies to stimulate the economy, it also greatly increased the uncertainty of economic policies [2].

The current methods for measuring the uncertainty of economic policies can be summarized into three categories according to the measurement ideas: proxy index substitution method, subjective expected deviation method, and state quantity estimation method. The more common in the proxy index substitution method is the frequency of keywords such as economic uncertainty in statistical
newspapers and periodicals. This method believes that the formulation and influence of economic policies and people’s views on the economy must be reflected in the media. By counting the frequency of occurrence of keywords related to uncertainty, economic uncertainty can be reflected, to a certain extent. However, the choice of newspapers and media depends entirely on the subjective wishes of statisticians, and the views of newspapers and media do not represent the subjective perceptions of all members of the economy. Therefore, it is difficult to fully reflect the macroeconomic conditions.

The occurrence of major events has varying degrees of impact on the politics, culture, and economy of various countries. As a core content of world development, the financial market is also greatly affected. The financial market has a core position in the social economy, and the most important part of the financial market is the securities market, so the securities market can represent the financial market to a certain extent. The occurrence of major events will cause volatility in the stock market and affect the correlation between stock markets.

At present, some scholars only use the univariate GARCH model to discuss the volatility of a certain stock market without considering the correlation between the markets; other scholars consider the correlation between the markets and use the multivariate GARCH model to analyze the stock market; and relatively little research has been done on how the occurrence of major events affects the stock market. The influence of the securities market on the development of the country and people’s social life cannot be ignored. Therefore, it is of practical significance to study the impact of the occurrence of major events on the correlation between the stock markets, whether it is for the development of society or the investment of investors.

The organizational structure of this article: the first part is the introduction, which mainly analyzes the research background, research status, research motivation, and research significance. The second part is the literature review part, which summarizes and analyzes the quantitative methods of economic fluctuations. The part is to improve the big data algorithm, which provides the basis for the intelligent method of the fourth part. The fourth part is the model construction part of this article, and the fifth part is the system experiment part of this article. The conclusion part is the research content of this article. The summary and outlook for future work are then presented.

The main contributions of this article are as follows: (1) big data methods are used combined with financial network theory to explain the phenomenon of economic fluctuations and dynamically identify the changing process of fund investment strategies. (2) It provides an effective method for the prediction and analysis of economic fluctuations in the era of big data.

This article analyzes the economic fluctuations with big data technology and obtains the main factors that affect economic fluctuations through data mining and the economic fluctuations are analyzed. Moreover, this article combines experimental research to evaluate the effect of economic fluctuation forecasting to verify the reliability of the method.

2. Related Work

Academia usually adopts VAR and GARCH family models to measure spillover effects between markets. Zhou et al. [4] used the recursive VAR method to examine the relationship between quantitative easing in the United States and the volatility spillover effects of major international financial markets. It is found that the unconventional monetary policy of the United States has a significant impact on volatility spillovers and potential global systemic risks. Bhattacharya et al. [5] measured the time-varying spillover effects between my country’s real economy and the stock market and bond market by constructing a mixed-frequency VAR model. The study found that the time-varying characteristics of the spillover effect between the real economy and the two markets are significant, and the spillover effect increased during the financial crisis, and then the spillover effect decreased.

At present, some scholars only use the univariate GARCH model to discuss the volatility of a certain stock market without considering the correlation between the markets; other scholars consider the correlation between the markets and use the multivariate GARCH model to analyze the stock market; and relatively little research has been done on how the occurrence of major events affects the stock market. The influence of the securities market on the development of the country and people’s social life cannot be ignored. Therefore, it is of practical significance to study the impact of the occurrence of major events on the correlation between the stock markets, whether it is for the development of society or the investment of investors.

The organizational structure of this article: the first part is the introduction, which mainly analyzes the research background, research status, research motivation, and research significance. The second part is the literature review part, which summarizes and analyzes the quantitative methods of economic fluctuations. The part is to improve the big data algorithm, which provides the basis for the intelligent method of the fourth part. The fourth part is the model construction part of this article, and the fifth part is the system experiment part of this article. The conclusion part is the research content of this article. The summary and outlook for future work are then presented.

The main contributions of this article are as follows: (1) big data methods are used combined with financial network theory to explain the phenomenon of economic fluctuations and dynamically identify the changing process of fund investment strategies. (2) It provides an effective method for the prediction and analysis of economic fluctuations in the era of big data.

This article analyzes the economic fluctuations with big data technology and obtains the main factors that affect economic fluctuations through data mining and the economic fluctuations are analyzed. Moreover, this article combines experimental research to evaluate the effect of economic fluctuation forecasting to verify the reliability of the method.

2. Related Work

Academia usually adopts VAR and GARCH family models to measure spillover effects between markets. Zhou et al. [4] used the recursive VAR method to examine the relationship between quantitative easing in the United States and the volatility spillover effects of major international financial markets. It is found that the unconventional monetary policy of the United States has a significant impact on volatility spillovers and potential global systemic risks. Bhattacharya et al. [5] measured the time-varying spillover effects between my country’s real economy and the stock market and bond market by constructing a mixed-frequency VAR model. The study found that the time-varying characteristics of the spillover effect between the real economy and the two markets are significant, and the spillover effect increased during the financial crisis, and then the spillover effect decreased.

At present, some scholars only use the univariate GARCH model to discuss the volatility of a certain stock market without considering the correlation between the markets; other scholars consider the correlation between the markets and use the multivariate GARCH model to analyze the stock market; and relatively little research has been done on how the occurrence of major events affects the stock market. The influence of the securities market on the development of the country and people’s social life cannot be ignored. Therefore, it is of practical significance to study the impact of the occurrence of major events on the correlation between the stock markets, whether it is for the development of society or the investment of investors.

The organizational structure of this article: the first part is the introduction, which mainly analyzes the research background, research status, research motivation, and research significance. The second part is the literature review part, which summarizes and analyzes the quantitative methods of economic fluctuations. The part is to improve the big data algorithm, which provides the basis for the intelligent method of the fourth part. The fourth part is the model construction part of this article, and the fifth part is the system experiment part of this article. The conclusion part is the research content of this article. The summary and outlook for future work are then presented.

The main contributions of this article are as follows: (1) big data methods are used combined with financial network theory to explain the phenomenon of economic fluctuations and dynamically identify the changing process of fund investment strategies. (2) It provides an effective method for the prediction and analysis of economic fluctuations in the era of big data.

This article analyzes the economic fluctuations with big data technology and obtains the main factors that affect economic fluctuations through data mining and the economic fluctuations are analyzed. Moreover, this article combines experimental research to evaluate the effect of economic fluctuation forecasting to verify the reliability of the method.

2. Related Work

Academia usually adopts VAR and GARCH family models to measure spillover effects between markets. Zhou et al. [4] used the recursive VAR method to examine the relationship between quantitative easing in the United States and the volatility spillover effects of major international financial markets. It is found that the unconventional monetary policy of the United States has a significant impact on volatility spillovers and potential global systemic risks. Bhattacharya et al. [5] measured the time-varying spillover effects between my country’s real economy and the stock market and bond market by constructing a mixed-frequency VAR model. The study found that the time-varying characteristics of the spillover effect between the real economy and the two markets are significant, and the spillover effect increased during the financial crisis, and then the spillover effect decreased.

At present, some scholars only use the univariate GARCH model to discuss the volatility of a certain stock market without considering the correlation between the markets; other scholars consider the correlation between the markets and use the multivariate GARCH model to analyze the stock market; and relatively little research has been done on how the occurrence of major events affects the stock market. The influence of the securities market on the development of the country and people’s social life cannot be ignored. Therefore, it is of practical significance to study the impact of the occurrence of major events on the correlation between the stock markets, whether it is for the development of society or the investment of investors.

The organizational structure of this article: the first part is the introduction, which mainly analyzes the research background, research status, research motivation, and research significance. The second part is the literature review part, which summarizes and analyzes the quantitative methods of economic fluctuations. The part is to improve the big data algorithm, which provides the basis for the intelligent method of the fourth part. The fourth part is the model construction part of this article, and the fifth part is the system experiment part of this article. The conclusion part is the research content of this article. The summary and outlook for future work are then presented.

The main contributions of this article are as follows: (1) big data methods are used combined with financial network theory to explain the phenomenon of economic fluctuations and dynamically identify the changing process of fund investment strategies. (2) It provides an effective method for the prediction and analysis of economic fluctuations in the era of big data.

This article analyzes the economic fluctuations with big data technology and obtains the main factors that affect economic fluctuations through data mining and the economic fluctuations are analyzed. Moreover, this article combines experimental research to evaluate the effect of economic fluctuation forecasting to verify the reliability of the method.
3. Principles of Financial Networks Based on Big Data Algorithms

3.1. Analysis of Financial Parameters. \( r_{ij}^M(t) \) is the percentage change of daily net value \( Z \) at time \( t \) and time scale (span) \( \Delta t \). In order to eliminate the error caused by the difference in the net value of different funds and maintain the stability of the data series, a first-order logarithmic difference is carried out on the net value of the fund:

\[
\Delta r_{ij}^M(t) = \ln Z(t + \Delta t) - \ln Z(t). \tag{1}
\]

Based on the logarithmic rate of return \( \Delta r_{ij}^M(t) \), the Pearson correlation coefficient \( \rho_{ij}^M \) between fund \( i \) and fund \( j \) on the time scale \( \Delta t \) can be calculated by the following formula:

\[
\rho_{ij}^M = \frac{\langle \Delta r_{ij}^M \Delta r_{ij}^M \rangle - \langle \Delta r_{ij}^M \rangle \langle \Delta r_{ij}^M \rangle}{\sqrt{\langle \Delta r_{ij}^M \Delta r_{ij}^M \rangle - \langle \Delta r_{ij}^M \rangle \langle \Delta r_{ij}^M \rangle} \sqrt{\langle \Delta r_{ij}^M \Delta r_{ij}^M \rangle - \langle \Delta r_{ij}^M \rangle \langle \Delta r_{ij}^M \rangle}}. \tag{2}
\]

\( \rho_{ij}^M \) represents the correlation between the net value of fund \( i \) and fund \( j \) on a time scale of Mr. Among them, \( \langle \cdots \rangle \) represents the expectation operator. \( \rho_{ij} \in [-1, 1] \) represents the correlation coefficient, which measures the degree of correlation between the actual strategies of funds. When \( \rho_{ij} = 0 \), it means that there is no correlation between the actual strategies of the two funds. When \( \rho_{ij} > 0 \), the actual strategies of funds \( i \) and \( j \) are positively correlated, and the larger \( \rho_{ij} \), the greater the degree of correlation between the actual strategies of the two funds. When \( \rho_{ij} = 1 \), the two funds adopt the same actual strategy. When \( \rho_{ij} < 0 \), the actual strategies of funds \( i \) and \( j \) are negatively correlated, and the larger \( \rho_{ij} \), the smaller the correlation between the actual strategies of the two funds. When \( \rho_{ij} = -1 \), the two funds adopt opposite actual strategies. Due to the finiteness of the time series of fund net value, there may be false correlations between funds. The correlation coefficient of fund samples is used to infer whether the actual strategies of the two fund variables in the fund population are related. The significance test of the null hypothesis that the overall correlation coefficient is 0 can be carried out using the t-distribution statistics of the overall degree of freedom \( df = n - 2 \) (\( n \) is the number of fund samples) for the overall correlation coefficient of the fund (formula (5)). This article will take a hypothesis test on the correlation coefficient at a 95% confidence level. If the t test is significant, formula (3) holds, which means that there is no correlation between the actual strategies of the two tested funds, namely, \( \rho_{ij} = 0 \). If the t test is not significant, formula (4) holds, indicating that there is a correlation between the actual strategies of the two funds, and the correlation coefficient is calculated by formula (2).

The hypothesis test is as follows [18]:

\[
H_0: \rho = 0, \tag{3}
\]

\[
H_1: \rho \neq 0, \tag{4}
\]

\[
t = \frac{n - 2}{1 - \rho_{ij}^2} \sim t(n - 2). \tag{5}
\]

The correlation coefficient matrix between funds cannot be directly represented by a network diagram. The distance between funds is calculated by the Euclidean distance formula and the correlation coefficient matrix is converted into a distance matrix to construct an undirected weighted financial network to describe the fund market.

We set the following:

\[
\bar{r}_i = \frac{r_i - \langle r_i \rangle}{\sqrt{\langle r_i^2 \rangle - \langle r_i \rangle^2}}. \tag{6}
\]

Among them, \( \bar{r}_i \) is the time series of the net value of fund \( i \). The \( n \) records of the vector \( \bar{r}_i \) in the same time interval are taken as the distance \( \bar{r}_{ik} \) between the points of the \( n \)-dimensional vector \( \bar{r}_i \). Then, \( d_{ij} \) of the two funds can be obtained from the Pythagorean relationship:

\[
d_{ij}^2 = \sum_{k=1}^{n} (\bar{r}_{ik} - \bar{r}_{jk})^2. \tag{7}
\]

From the definition of \( \bar{r}_i \), the length of the vector \( \bar{r}_i \) is 1; that is,

\[
\sum_{k=1}^{n} \bar{r}_{ik}^2 = 1. \tag{8}
\]

Therefore, formula (7) can be rewritten as follows:

\[
d_{ij}^2 = \sum_{k=1}^{n} \left( \bar{r}_{ik}^2 - 2\bar{r}_{ik}\bar{r}_{jk} + \bar{r}_{jk}^2 \right) = 2 - 2\sum_{k=1}^{n} \bar{r}_{ik}\bar{r}_{jk} = 2 - 2\rho_{ij}. \tag{9}
\]

From formula (8), we can get the following:

\[
d_{ij} = \sqrt{2(1 - \rho_{ij})}. \tag{10}
\]

Among them, \( \bar{r}_i \) is equivalent to the Euclidean distance of the net value time series vectors of any two funds \( i \) and \( j \), \( d_{ij} \in [0, 2] \). It is satisfied with the three properties that must be satisfied by the Euclidean distance, namely (1), \( d_{ij} = 0 \iff i = j \) (2); \( d_{ij} = d_{ji} (3) \); \( d_{ij} \leq d_{ik} + d_{kj} \). The properties (1) and (2) are easy to verify because \( \rho_{ij} = \rho_{ji} \Rightarrow d_{ij} = d_{ji} \). The triangle inequality of property (3) can be proved by the equivalence of (7) and (10). The economic meaning of \( \bar{r}_i \) is the correlation coefficient between funds, and the smaller the
distance, the more similar the actual strategies between funds, and vice versa [19].

3.2. Analysis of Financial Network Algorithms. The establishment of a fund network of N funds can be calculated by calculating the correlation coefficients between N funds. Calculate the distance between funds by formula (8). If the hypothesis test of the correlation coefficient between the funds is not considered, N funds will construct a distance matrix of $N \times N$, and a complete graph representing the actual strategy network of the funds can be constructed through the distance matrix between the funds. A complete graph is a simple graph in which each pair of different vertices is connected by an edge. Therefore, the fund strategy network of the complete graph has a large amount of information and is not easy to handle. Therefore, it is necessary to use a specific method to filter some redundant edges while retaining the valuable actual strategy edges in the fund network to form the final fund actual strategy financial network. Next, starting from the fund distance matrix, the final fund actual strategy financial network is generated based on the minimum spanning tree method and the planar maximum filter graph method.

There are two basic algorithms for minimum spanning trees: Kruskal’s algorithm and Prim’s algorithm. The principle of the Kruskal algorithm mainly starts from the connection edges and gradually determines the connection edges that meet the conditions to obtain the final minimum spanning tree network. Prim’s algorithm expands around nodes and determines the nodes and connecting edges in the network one by one to get the final minimum spanning tree network. There is no essential difference between the two algorithms, and the final result is the same. In this article, Kruskal’s algorithm is used to screen the edges of the distance matrix step by step to obtain the minimum spanning tree network. The specific methods of Kruskal’s algorithm are as follows:

(1) The algorithm arranges the weights of all connected edges (that is, the distance of the fund) in ascending order and selects the edge with the smallest weight:

(2) In each step, the edge with the smallest weight is selected from the unselected edges so that it does not form a circle with the selected edges, until the $N - 1$ edges are selected.

The Kruskal algorithm of the minimum spanning tree is programmed by Matlab7.0.

We assume that the space of N funds is a hypermetric space. This hypothesis is based on the “posterior” motivation; that is, the research results obtained based on this hypothesis are meaningful from an economic point of view. The supermetric space refers to the space in which the distance between objects is the supermetric distance. The supermetric distance satisfies the first two properties of distance, namely, (1) $d_{ij} = 0 \iff i = j$ and (2) $d_{ij} = d_{ji}$. However, the property (3) of distance The triangle inequality is replaced by the hypermetric inequality, namely, $d_{ij} \leq \max\{d_{ik}, d_{kj}\}$. Rammal et al.’s article introduced the concepts related to hypermetrics in detail. Several supermetric spaces can be obtained by segmenting a set of N objects with a certain metric distance relationship. Among all possible hypermetric structures corresponding to distance $d_{ij}$, the subhypermetric space is the simplest and has good properties. In the metric space where N objects are associated together, the subsupermetric space can be obtained by determining the minimum spanning tree associated with the N objects. The minimum spanning tree of the subsupermetric space corresponds to a unique exponential hierarchical tree (hierarchical tree), which can directly determine the supermetric distance matrix $d_{ij}$. Each element in the matrix $d_{ij}$ is equal to the maximum distance between any two adjacent targets when moving along the shortest MST path connecting the starting object to the ending object. Compared with the matrix $d_{ij}$, there are no more than $N^2 - 1$ different elements in the supermetric distance matrix $d_{ij}$ [20].

The exponential hierarchical structure tree is based on the working principle of the hierarchical clustering method to cluster the N objects to be clustered and the $N \times N$ distance matrix hierarchically. When studying the relationship between investment strategies of securities market funds, the article believes that funds at the same level have significant common attributes; that is, their actual strategies are the same or similar. The basic steps of the hierarchical clustering method are as follows:

(1) First, the algorithm classifies each object of the system into one category separately and obtains N categories in total. The distance between the categories is the distance between the objects they contain.

(2) The algorithm finds the two categories with the closest distance in the system and merges them into a new category to reduce the number of system categories.

(3) The algorithm recalculates the distance between the new class and all old classes.

(4) The algorithm loops the second and third steps until all the objects in the system are finally merged into one category (this category contains N objects).

The hierarchical clustering method can be divided into single linkage cluster analysis, average linkage cluster analysis, and complete linkage cluster analysis according to the different calculation methods of the distance between clusters in step 3 of the clustering principle. Among them, the distance between clusters in the single linkage cluster method is equal to the minimum distance between two types of objects. The distance between clusters in the average linkage cluster method is equal to the average of the pairwise distances between two types of objects. The distance between classes in the complete linkage cluster method is equal to the maximum distance between the two classes of data. The index hierarchical structure tree in the paper uses the single-link clustering method to reflect the hierarchical relationship of the actual strategies between funds, and the distance obtained by the single-link clustering method is equal to
each element of the overmetric distance matrix $d_{ij}$. Figure 1 is the generation process of the minimum spanning tree and the nonunique minimum spanning tree corresponds to the unique exponential hierarchical structure tree.

The Planar Maximaly Filtered Graph adds more edges to the minimum spanning tree to ensure that more effective information is retained in the financial network and avoid serious filtering of information on the fund’s actual strategy network. The plane maximum filter graph is also based on a complete graph constructed by the distance matrix, and the sum of the distances is made as small as possible under the condition that it is a plane graph. The method of constructing the planar maximum filter graph is similar to the method of minimum spanning tree, with the main differences as follows:

1. The constraints on the edges are different. MST requires that the edges on the spanning tree cannot appear loops. PMFG relaxes this constraint and requires the edges on the final network graph to be on a plane graph; that is, all edges can be drawn on a plane without crossing out.

2. The number of edges is different. For a network graph composed of $N$ nodes, there are $N – 1$ edges on MST and $3N – 6$ edges on PMFG. According to the Kuratowski theorem of plan graphs, PMFG only allows three factions and four factions. The so-called $N$ faction refers to the complete graph containing $N$ nodes in the subgraph of the network graph.

The structure of PMFG is more complex than that of MST; it retains more effective information about the actual fund strategy and is a better supplement to simple MST. Due to the nonuniqueness of the MST method and the seriousness of MST’s filtering of fund network information, it is passed. The analysis of topological characteristics of PMFG and MST in time evolution and time scale can verify the effectiveness of the MST method. In addition, the three factions and four factions in PMFG can dig out faction characteristics related to the fund’s actual strategy that are not in MST.

The actual fund strategy network represents the relationship between the actual strategies of all funds in the fund market composed of 94 selected funds. From the actual strategy network, the clustering characteristics of funds with the same strategy on the network diagram can reflect the actual fund, whether the strategy is consistent with the declared strategy. When examining the influence of time evolution on the actual strategy of the fund and the change of the time scale on the fund network structure, the position of the fund node in the actual strategy network of the fund will change, and the clustering characteristics of the fund will also be affected. However, this time evolution has an impact on the fund. The degree of influence of the actual strategy and the degree of change of the fund network structure on the time scale are not enough to show from the network diagram. The article calculates the fund network parameters from a quantitative perspective by constructing the financial network topological characteristic index to illustrate the degree of network evolution and the influence of time scale.

![Figure 1: Generation of minimum spanning tree and corresponding index hierarchical structure tree.](image)

The financial network topological characteristic index is the correlation coefficient of the final spanning tree (MST, PMFG), average value, standard tree length, average network path length, average network aggregation coefficient, and central node.

### 3.3 Analysis of Fund Strategy Network Nodes

The position relationship of nodes on the fund’s actual strategy network indicates the degree of correlation between the actual strategies of the funds. The connection of two nodes in the fund’s actual strategy network indicates that the actual strategies of the two funds are the same or similar. However, there are many unconnected nodes in the actual strategy network of the fund, and the degree of relevance of their actual strategies cannot be expressed in the actual strategy network of the fund. The article examines the correlation degree of the actual strategy in the network by calculating the mean value $\rho$ of the correlation coefficient between the funds in the actual strategy network of the fund. The larger the mean value of the correlation coefficient $\rho$, the greater the correlation between the nodes on the fund’s actual strategy network, and the higher the actual strategy similarity between funds Equations (11) and (12) are used to calculate the mean values of the correlation coefficients of the minimum spanning tree and the plane maximum filter graph, respectively.

\[
\rho_1(\Delta t, T) = \frac{1}{N-1} \sum_{i \neq j}^{N} \rho_{ij} \tag{11}
\]

\[
\rho_2(\Delta t, T) = \frac{1}{3N-6} \sum_{i \neq j}^{N} \rho_{ij} \tag{12}
\]
Among them, $\rho_1(\Delta t, T)$ and $\rho_2(\Delta t, T)$, respectively, represent the mean value of the correlation coefficients of the fund’s actual strategy network on the time scale $M$ and the time period $T$ under the minimum spanning tree and planar maximum filter graph methods, and $N$ is the number of fund nodes, and $R_{C,T}^\Delta$ is the correlation coefficient matrix under the condition of the significant correlation coefficient.

In addition, the article uses the correlation coefficient variance $r_1(\Delta t, T)$ and $r_2(\Delta t, T)$ under the minimum spanning tree and the flat maximum filter graph to investigate the actual strategy stability of the fund’s actual strategy network. A greater variance of the correlation coefficient means a smaller stability of the fund’s actual strategy, and the greater the diversity of the actual strategy:

$$r_1(\Delta t, T) = \frac{1}{N^2} \sum_{\rho_{ij} \in R_{C,T}^\Delta} (\rho_{ij} - \rho_1(\Delta t, T))^2,$$

$$r_2(\Delta t, T) = \frac{1}{3N-6} \sum_{\rho_{ij} \in R_{C,T}^\Delta} (\rho_{ij} - \rho_2(\Delta t, T))^2.$$  

The normalized tree length (NTL) is usually used to measure the tightness of the network in financial network theory. The smaller the standard tree length, the tighter the actual strategy network of the fund, and the more similar the actual strategy network between funds. Equations (15) and (16) are the calculation formulas for the minimum spanning tree under the time scale $\Delta t$ and the time period $T$ and the standard tree length under the flat filter graph:

$$L_{NTL_1}(\Delta t, T) = \frac{1}{N-1} \sum d_{ij},$$

$$L_{NTL_2}(\Delta t, T) = \frac{1}{3N-6} \sum d_{ij}.$$  

Among them, $d_{ij}$ is the Euclidean distance between nodes $i$ and $j$, and $D_{C,T}^\Delta$ is the correlation coefficient matrix $R_{C,T}^\Delta$ calculated by the distance formula.

The average path length (APL) refers to the average value of the path length between any two nodes. It represents the average number of intermediary edges connecting any two nodes in the network. From the perspective of network theory, it shows that the actual strategies of any two funds in the fund market need to be related at least as many intermediary funds on average.

$$L_{APL} = \frac{2}{N(N-1)} \sum I_{ij}.$$  

Among them, $I_{ij}$, the path length of any two points, is defined as the number of edges on the shortest path.
connecting two nodes. The smaller the average path length of the fund network is, the more compact the fund network is, and the possibility that the actual strategy similarity between funds will increase increases.

The clustering coefficient $C_i$ of a node is the ratio of the number of edges that are connected to each other in the network of the node set formed by all the nodes connected to the node and the number of edges that constitute a complete graph. The network average clustering coefficient $AC$ is the average value of the clustering coefficients of all nodes on the network.

$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

$$AC = \frac{1}{N} \sum_{i=1}^{N} C_i.$$  \hspace{1cm} (18)

Among them, $k_i$ is the degree value of node $i$, and $E_i$ is the number of edges, where $k_i$ nodes exist in the network.

The degree distribution of each node on the network graph is not uniform. Some nodes have a large degree value, while other nodes have a small degree value. This feature is obvious on the interpersonal network. People usually rank the node with the highest degree value, defined as a key node. Because the degree of key nodes is very large and the correlation with other nodes is great, key nodes can often reflect the characteristics of the network. In the article, the nodes with the highest node degree in the network are considered key nodes.

The actual strategy of nodes on the fund’s actual strategy network is constantly changing over time. Fund managers will change their actual strategies according to the market situation and their own technical means. Regarding the fund market as a network, there are correlations between nodes. Key nodes represent nodes with a greater degree of correlation between the fund and the actual strategies of other funds, representing the most common actual strategies in the fund market, and it can reflect the overall fund market and the change trend of the actual strategy over time. It is worth noting that it is not the biggest influencer of the actual strategy of the fund’s actual strategy network. It cannot affect the changes of the actual strategy of other funds with the same or similar strategies but the actual strategy of the actual strategy network and the best representative of time evolution.

---

**Figure 3:** Schematic diagram of the algorithm.

**Figure 4:** Flowchart of the algorithm.
4. Analysis Model of Influencing Factors of Economic Fluctuation Based on Big Data

4.1. System Model Building. The model is mainly divided into three stages in the forecasting process. The first stage is the decomposition stage. Its main task is to find out the potential factors that cause time series fluctuations, to effectively separate independent source signals through independent component analysis algorithms, and to analyze economic meanings simply in combination with their fluctuation patterns. The second stage is the prediction stage. Its main task is to find the optimal ARIMA model to predict each independent component based on its own data characteristics. The third stage is the reconstruction stage. Its main task is to reconstruct the predicted value of each independent component through a certain mechanism to form the predicted value of the original time series, and this mechanism is the inverse operation of independent component analysis. The framework of the model is shown in Figure 2.

Figure 3 describes the principle of the Infomax algorithm and draws a flowchart. The core idea of the algorithm is to measure the independence between various variables through information entropy and apply it to the time series observation data by selecting a nonlinear function so as to realize the separation of the independent components of the time series.

The calculation process of the Infomax algorithm is drawn as the following flowchart in Figure 4.

The modeling flowchart of each independent component ARIMA prediction model in the prediction stage of the ICA-ARIMA model proposed in this article is shown in Figure 5.

In order to avoid the impact of different sample data sets on the experimental results, the experiment will be repeated.
4.2. Analysis and Discussion. In order to verify the effectiveness of the model proposed in this article in the economic market, this article selects the PMI index to form the relevant experimental data set to carry out the experiment and conducts the simulation test in Matlab.

On the basis of the above analysis, this article combines the PMI index to carry out the application of big data mining technology in the analysis of the influencing factors of economic fluctuations, calculates the PMI index in recent years, and draws it into a statistical chart, as shown in Table 1 and Figure 7:

On the basis of the above analysis, the analysis results of economic fluctuation factors are counted, and the prediction accuracy of the factors affecting economic fluctuations is calculated. The system in this

Table 1: Statistical table of economic volatility index.

| Index                   | Jan-20 | Feb-20 | Mar-20 | Apr-20 | May-20 | Jun-20 | Jul-20 | Aug-20 | Sep-20 | Oct-20 |
|------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Comprehensive PMI index| 53     | 28.9   | 53     | 53.4   | 53.4   | 54.2   | 54.1   | 54.5   | 55.1   | 55.3   |
| Manufacturing PMI      | 50     | 35.7   | 52     | 50.8   | 50.6   | 50.9   | 51.1   | 51     | 51.5   | 51.4   |
| Nonmanufacturing PMI   | 54.1   | 29.6   | 52.3   | 53.2   | 53.6   | 54.4   | 54.2   | 55.2   | 55.9   | 56.2   |
| Index                  | Nov-20 | Dec-20 | Jan-21 | Feb-21 | Mar-21 | Apr-21 | May-21 | Jun-21 | Jul-21 | Aug-21 |
| Comprehensive PMI index| 53.7   | 55.1   | 52.8   | 51.6   | 55.3   | 53.8   | 54.2   | 52.9   | 52.4   | 48.9   |
| Manufacturing PMI      | 52.1   | 51.9   | 51.3   | 50.6   | 51.9   | 51.1   | 51.0   | 50.9   | 50.4   | 50.1   |
| Nonmanufacturing PMI   | 56.4   | 55.7   | 52.4   | 51.4   | 56.3   | 54.9   | 55.2   | 53.5   | 53.3   | 47.5   |

until all the stocks in the bank stock pool are selected at least once. The final result will also be displayed by taking the average of the times of all data sets. The specific experimental process is shown in Figure 6.

Figure 6: The comparative experimental process of the ICA-ARIMA model and ARIMA model.

Figure 7: Statistical diagram of economic fluctuations.
The method of this paper is compared with the literature [17] through the simulation platform, and the statistical effect of the statistical economic fluctuation index and the analysis effect of the influencing factors of economic fluctuation are compared, and the results are shown in Table 2 and Figure 8.

It can be seen from the above research that the data mining iteration proposed in this article can play an important role in the analysis of the influencing factors of economic fluctuations and accurately dig out the relevant factors that affect the economy.

5. Conclusion

Data mining often selects historical data for the statistics of keyword frequency, and it is difficult to estimate the economic uncertainty of the current or even forecast the future. According to the efficient market hypothesis, all kinds of information will be absorbed by all markets at the same time in the shortest time, so there is no spillover effect between financial markets. In fact, with the deepening of research, there is a widespread spillover effect in financial markets. According to existing research, the transmission of information through yield or volatility channels is called mean spillover and volatility spillover, respectively. Mean spillover refers to changes in asset prices or returns in one market that affect changes in asset prices or returns in other markets. Volatility spillover refers to the volatility of asset prices in one market affecting other market volatility, which is generally measured by the conditional variance of prices or returns. This article proposes an analysis model of the

| Number | The method of this paper | The method of [17] | Number | The method of this paper | The method of [17] |
|--------|-------------------------|------------------|--------|-------------------------|------------------|
| 1      | 89.66                   | 87.05            | 31     | 83.69                   | 80.08            |
| 2      | 92.48                   | 83.61            | 32     | 83.47                   | 77.48            |
| 3      | 89.71                   | 85.22            | 33     | 88.03                   | 83.14            |
| 4      | 86.64                   | 84.31            | 34     | 82.91                   | 80.02            |
| 5      | 80.98                   | 77.03            | 35     | 83.94                   | 83.78            |
| 6      | 80.25                   | 75.88            | 36     | 91.07                   | 88.90            |
| 7      | 84.30                   | 77.45            | 37     | 87.84                   | 82.05            |
| 8      | 82.51                   | 74.94            | 38     | 87.35                   | 82.25            |
| 9      | 84.84                   | 79.35            | 39     | 84.07                   | 80.11            |
| 10     | 83.29                   | 76.37            | 40     | 85.95                   | 79.29            |
| 11     | 87.96                   | 80.05            | 41     | 81.94                   | 75.63            |
| 12     | 92.42                   | 91.63            | 42     | 88.51                   | 84.17            |
| 13     | 93.24                   | 89.00            | 43     | 88.55                   | 84.83            |
| 14     | 81.27                   | 73.97            | 44     | 84.15                   | 82.70            |
| 15     | 92.22                   | 92.03            | 45     | 83.44                   | 80.24            |
| 16     | 80.24                   | 74.49            | 46     | 92.20                   | 83.02            |
| 17     | 91.11                   | 89.42            | 47     | 84.51                   | 77.75            |
| 18     | 90.80                   | 89.33            | 48     | 80.66                   | 76.64            |
| 19     | 92.35                   | 87.86            | 49     | 88.07                   | 84.80            |
| 20     | 93.68                   | 91.64            | 50     | 82.62                   | 77.61            |
| 21     | 84.39                   | 82.12            | 51     | 85.56                   | 85.47            |
| 22     | 87.79                   | 80.46            | 52     | 87.66                   | 87.39            |
| 23     | 81.03                   | 77.52            | 53     | 80.86                   | 75.74            |
| 24     | 85.17                   | 77.67            | 54     | 92.48                   | 89.79            |
| 25     | 83.65                   | 75.31            | 55     | 83.99                   | 83.12            |
| 26     | 92.45                   | 83.90            | 56     | 86.62                   | 81.19            |
| 27     | 84.03                   | 78.63            | 57     | 88.86                   | 80.87            |
| 28     | 88.75                   | 88.07            | 58     | 92.79                   | 85.57            |
| 29     | 82.93                   | 81.74            | 59     | 85.21                   | 79.71            |
| 30     | 80.82                   | 72.97            | 60     | 92.99                   | 87.30            |

Figure 8: Analysis of the influencing factors of economic fluctuations in the era of big data.
influencing factors of economic fluctuations based on data mining and combines actual data to conduct research to calculate economic fluctuations. Through verification, it can be seen that the data mining method proposed in this article can play an important role in the analysis of economic fluctuations and accurately mine the relevant factors that affect the economy.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the (1) Teaching Innovation Team of Teachers from E-Commerce, Teaching Innovation Team of Provincial Teachers from Anhui Province, Project Number: 2019cxtd036; (2) Research on E-Commerce Issues of Huaibei Agricultural Products under Digital Village Strategy, Humanities Key Project of Huaibei Vocational and Technical College, Project Number: 2020-B-6.

References

[1] K. Ataka, “Prediction of election result and economic indicator,” Resuscitation, vol. 96, no. 6, p. 84, 2014.
[2] S. Barde, “Back to the future: economic self-organisation and maximum entropy prediction,” Computational Economics, vol. 45, no. 2, pp. 337–358, 2015.
[3] A. Ferramosca, D. Limon, and E. F. Camacho, “Economic MPC for a changing economic criterion for linear systems,” IEEE Transactions on Automatic Control, vol. 59, no. 10, pp. 2657–2667, 2014.
[4] L. Zhou, K. K Lai, and J. Yen, “Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation,” International Journal of Systems Science, vol. 45, no. 1-3, pp. 241–253, 2014.
[5] D. Bhattacharya, J. Mukhoti, and A. Konar, “Learning regularity in an economic time-series for structure prediction,” Applied Soft Computing, vol. 76, no. 2, pp. 31–44, 2019.
[6] Y. Geng, Z. Wei, and H Zhang, “Maimaituerxun M. Analysis and prediction of the coupling coordination relationship between tourism and air environment: yangtze river economic zone in China as example,” Discrete Dynamics in Nature and Society, vol. (10), pp. 1–15, 2020.
[7] H. L. Vu, K. T. W. Ng, and D. Bolingbroke, “Time-lagged effects of weekly climatic and socio-economic factors on ANN municipal yard waste prediction models,” Waste Management, vol. 84, no. 2, pp. 129–140, 2019.
[8] C. Teljeur, M Neill, P Moran et al., “Using prediction intervals from random-effects meta-analyses IN an economic model,” International Journal of Technology Assessment in Health Care, vol. 30, no. 01, pp. 44–49, 2014.
[9] P Rajsic, A Weersink, A Navabi, and K. P. Pauls, “Economics of genomic selection: the role of prediction accuracy and relative genotyping costs,” Euphytica, vol. 210, no. 2, pp. 1–18, 2016.