Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5025656, 10 pages
https://doi.org/10.1155/2022/5025656

Research Article

Prediction Algorithm of Digital Economy Development Trend Based on Big Data

Pan Xiangyan

Business School, Shunde Polytechnic, Foshan 528300, Guangdong, China

Correspondence should be addressed to Pan Xiangyan; xiangyanpan@sdpt.edu.cn

Received 5 January 2022; Revised 28 March 2022; Accepted 7 April 2022; Published 23 April 2022

Academic Editor: Dost Muhammad Khan

With the development of economic globalization, people have higher requirements for economic analysis and prediction, while the prediction accuracy of the traditional economic analysis and prediction model is quite low. Therefore, with the adaptation of the times, the “big data and artificial intelligence” technology is presented for researchers and data analysts. Based on the “big data and artificial intelligence” technology, economic analysis and prediction have unique advantages in many aspects. This study analyses the impact of “big data and artificial intelligence” technology on economic analysis and prediction. Based on the mean square error criterion, the variable selection method for constructing the interval data model is given. The interval financial time-series data of stock market, fund market, futures market, and money market are used to predict and analyze the macroeconomy, and the macroeconomic interval prediction method is given, which is different from the traditional point data model. The empirical results show that the Shenzhen Component Index, Shanghai fund index, futures market transaction amount, and narrow money supply in the interval financial data have little fitting error to the macroeconomic interval prediction model. Through the first mock examination, we get the macroeconomic forecasting model based on interval financial time-series data. We also use the single model structure and the combined model structure to give the macroeconomic change interval of 2020–2023 years in China. The experimental research results show that the proposed model has good performance in the prediction of economic development trends and that it can be used for forecasting subsequent economic development forecasts.

1. Introduction

Big data are an emerging research field highly concerned by academia and industry in recent years. In 1995, Fayyad and others first proposed the concept of big data at the Knowledge Discovery Conference (KDD) [1], the top summit in the field of international data mining, and summarized its basic objectives into two aspects: descriptive and predictive. The former is to mine various association patterns hidden behind massive data. Victor Mayer Schoenberg and others believe that because big data break through the data-scale limitations of traditional sample collection methods, it can largely use full-sample massive data for analysis. Therefore, it can make extensive use of correlation mining methods to find the clue information hidden behind massive data. The latter is predicted according to the potential patterns in the data. Different from the traditional statistical prediction method mainly based on limited statistical sample data, big data can help people break through the limitations of sample collection methods and realize the collection of all-sample, all-weather, all-scene, and all-round data, to improve the prediction ability of economic and social operation subjects.

In recent years, with the rise of big data, the use of various nontraditional statistical data to carry out macroeconomic analysis and judgment has not only become the focus of economists, but also attracted the common attention of researchers from many disciplines, including information science, biology, physics, and sociology [2]. According to the author’s incomplete statistics, there have been more than 150 articles directly themed on big data economics in domestic and foreign journals in the past decade. This study comprehensively reviews the attention of researchers from different disciplines such as economics and information science to macroeconomic big data analysis in recent ten years, combs the application research context of big data in macroeconomic analysis, and looks forward to the next research trend and development direction.

2. Related Work

Macroeconomic big data analysis is a typical interdisciplinary research field. From the perspective of the discipline background of the more active research team in this field, economists are an important force, but they are by no means the only discipline group paying attention to this field. Researchers from information science, life science, physics, complex science, and other fields also began to pay a lot of attention to big data analysis in macroeconomic operation, which makes macroeconomic big data research become a comprehensive field with the multidisciplinary intersection. A simple example is that the WOS database is searched directly with the title "big data + economics/economy" [3–5]. In the search results, the literature in the fields related to economics (economics, business, management, finance, etc.) accounts for 42%, while the literature in the fields related to information science (computer, communication, artificial intelligence, library informatics, automation, etc.) accounts for 43.5%, which has exceeded economics. Scholars in other fields such as life science, physics, sociology, engineering science, and urban planning also began to pay attention to macroeconomic big data analysis. There are still many challenges for big data technology to assist macroeconomic monitoring and analysis:

Based on the monitoring, prediction, and analysis of macroeconomy and data technology, a large amount of behavior information of economic subjects, and information of many other economic variables, it has high requirements for the real-time data and the diversity of key indicators. The data of Guangxi macro-government departments mainly come from statistical departments, with a relatively single source and poor timeliness, and some indicators often have a one-month lag, which makes it impossible to understand the macroeconomic situation timelier, comprehensively and accurately [6]. At the same time, the reliability of data will also affect the subsequent macroeconomic situation analysis. At present, Guangxi has not established a department and bureau-level macro-big data sharing and linkage mechanism, and the macro-big data analysis and judgment mainly depend on Internet data. The reliability of the data remains to be discussed, the value density is low, the data noise is large, and the continuous support of credible data from various departments and bureaus is lacking [7].

The vast majority of macroeconomic big data belong to unformatted data such as text, pictures, and videos, which has great uncertainty, mainly in terms of high-dimensional, changeable, and strong randomness [8]. Computers can not directly recognize and understand the information contained in unstructured data, and it is necessary to use the interdisciplinary knowledge including mathematics, economics, sociology, computer science, and management science to clean and process the data for analysis and research. The macro-big data analysis system in Guangxi is not perfect, and the ability of data cleaning and processing and feature extraction is not enough, which affects the data collection and summary.

At present, Guangxi has established macroeconomic monitoring index system and macroeconomic prediction index system according to domestic and foreign research and has preliminarily realized the function of macroeconomic analysis and monitoring, but the construction of prediction model algorithm and index system is not achieved overnight. At present, the prediction effect of macro-big data system is unstable, manual judgment and calibration are often needed, and the system algorithm and index system need to be adjusted and upgraded in time according to the new macroeconomic situation [9]. At the same time, the system function is still imperfect and many index system data are missing. The system still needs to be further improved and extended to municipal and county development and reform departments. The stock trend prediction has obtained an important and significant attention from data analysts in recent times and machine learning. Hao et al. [10] has proposed a machine learning-based stock trend prediction system with a focus on minimizing data sparseness in the acquired datasets. The results obtained show the effectiveness of the proposed system, when compared with baseline studies.

The core of big data is data capture and analysis, and the analysis link is inseparable from manually setting variables and establishing models [11]. The improvement and operation and maintenance of the system also need the support of big data technical talents. However, there are few talents in Guangxi who master both professional knowledge in the economic field and computer big data analysis technology and can skillfully use it, there are few comprehensive talents who can use big data technology to carry out macroeconomic analysis and prediction at the same time, and the shortage of big data analysis talents is very prominent.

2.1. The Machine Learning. Machine learning not only provides new analysis tools but also tries to solve problems that cannot be solved by traditional measurement methods. Machine learning can be roughly divided into four categories: supervised machine learning, unsupervised machine learning, enhanced machine learning, and semi-supervised machine learning.

(i) Supervised Machine Learning: supervised machine learning mainly focuses on the prediction problem: it is necessary to use the labeled sample data, including feature (x) and corresponding tag value (y), and build an optimal model by learning the corresponding relationship between feature and tag value [12]. In this way, when facing the new feature x, we can use this optimal model to predict its label value y. This process of finding the best estimation parameters through known data is similar to the regression analysis in econometrics. However, supervised machine learning does not need to set the relationship between Y and X in advance, and a more flexible function form can be selected according to the data itself. In the process of empirical analysis based on the classical econometric model, people often pay attention to whether the estimation and statistical inference of parameters before x variable are accurate. While the algorithm of machine learning focuses on the prediction effect of X variable on y, we do not care much about the accuracy of parameter estimation. This is why the parameters estimated by machine learning often do not meet the consistency. At present, there are corresponding software packages in common data analysis software (such as python and R), which can easily implement such algorithms [13].

(ii) Unsupervised Machine Learning: in reality, it often happens that there is a lack of sufficient prior knowledge, or the cost of manually labeling data is too high, so only data samples containing feature (x) without corresponding tag value (y) can be obtained [14]. Because there is no tag value, we cannot find the optimal prediction model according to the error (prediction value and label value). Therefore, unsupervised machine learning needs to solve another problem: when facing the unmarked video, image, and text data, we hope the machine can find the potential laws between features from a large sample set. Some representative features and aggregate individuals are identified with similar features. It should be noted that the machine hardly adds human judgment to the classification, to maximize objectivity. The most typical example of unsupervised machine learning is clustering [14]. The authors of [15] used machine learning technology to classify the news when studying the impact of Google's closure of Google News in Spain on its news report types. After dimensionality reduction in the data set, the impact of closing events on all different types of news reports is studied at the same time. This kind of technology is also widely used in life. For example, some popular recognition software (shape and color, Microsoft flower recognition, etc.) can accurately identify the type of flowers according to a picture of flowers taken by users. In addition, before traditional regression analysis, we can also first use the clustering algorithm to test the correlation between variables and preprocess the data to reduce the amount of calculation [16, 17].

3. Prediction Algorithm of Digital Economy Development Trend

3.1. Application of Big Data in Economic Forecasting. Economic forecasting is one of the most important functions of economics. It is also the reference and basis for government, enterprises, and individuals to make decisions. However, for a long time, people have questioned the prediction theory of economics. Hayek once argued that human economic behavior is unpredictable. He pointed out that "Despite our knowledge of the workings of human reason, the main fact remains that it is impossible for us to explain all the specific facts that lead individuals to do a specific thing at a specific time. For us, individual personality is still a very unique and difficult phenomenon to calculate, and we are expected to develop some practices from experience For example, praise and blame exert influence on it in a desirable direction, but we cannot predict or control its specific behavior because we cannot understand all the specific facts that determine it."

With the advent of the big data era, people's prediction of economics has changed from the prediction theory paradigm based on causal inference to the prediction theory paradigm based on probability theory. Cai Yuezhou believes that the impact of big data on economic forecasting involves almost every link such as data sources, forecasting methods, and forecasting results, which has changed the basic paradigm followed by conventional economic forecasting to some extent. From the database of economic forecast, big data economic forecasting is systematically compared with traditional economic forecasting in terms of index generation methods, prediction methods, and tools. With the increasing ability of big data to widely obtain various behavioral data of economic and social subjects, it will provide a new path for measuring the expectations of economic and social subjects and quantifying the emotions of subjects and is expected to gradually open the process of the formation of expectations of economic subjects. "Black box" n greatly improves the ability of prediction and analysis. Li Huajie and others believe that the prediction method based on big data is not a subversion of traditional economic research methods in many aspects, but a supplement to traditional research methods. From the existing research results, the application of big data in economic prediction research can be divided into several categories: constructing leading indicators, improving traditional prediction models, and constructing prediction models and complexity prediction, which will be discussed in detail one by one later. It should be pointed out that many literature studies also include the current prediction research into the scope of big data prediction in economics, but the author believes that its essence is still the real-time analysis of the current situation, which is only advanced to a certain extent using the "time difference" in the relative lag of statistical methods. It has been introduced in the previous section and will not be repeated here.

3.2. Big Data Forecast Based on Leading Indicators. The prediction of leading indicators can be divided into three categories:

(i) From the perspective of human economic behavior chain, the behavior of the next link is predicted through the analysis of the previous link of the behavior chain. For example, Liu et al. used PLSA to mine user views and emotions from blog content data to predict film box office sales. Schneider has launched an Amazon Com-based model for predicting laptop word packs from user reviews. Khadivi et al. used data to predict Hawaii's tourism demand by analyzing Wikipedia. For another example, a large number of researchers judge people's psychological expectations of the stock market based on the public opinion sentiment 2 of our media such as microblog and Twitter, as well as financial search data, to predict the operation trend of the stock market.

(ii) Based on various economic theories, this study constructs leading indicators that can reflect the trend of economic operation in advance. For example, the author has constructed the "lubricating oil purchase index" based on the tax invoice data and believes that manufacturing enterprises generally estimate the amount of lubricating oil required for commencement according to the order reserve, so the change in lubricating oil purchase can reflect the actual operation rate of manufacturing industry to a certain extent. The research shows that, compared with the purchasing managers' index (PMI), the trend of the index can be ahead of others for about three months. However, due to the limited data sources, such studies are rare. More typical ones, such as akitas, established the toll station index with the monthly heavy truck cross-border data recorded by the German toll station as the leading indicator of the GNP index and demonstrated that the index can effectively reflect the production index officially released by the German statistics office; Qian Binhua constructs the leading indicator 2 of tax development in Ningbo based on the change rate of the number of employees, the change rate of total assets, and the change rate of owner's equity and other indicators; Cui Chenxin and others predicted the trend of regional GDP based on the treasury revenue index of Hebei Province. Zhang Qiuyan and others predicted the trend of economic prosperity based on power consumption; Qu Yanling proposed to predict the macroeconomic trend based on the key indicators of deposits and loans and the changes in social financing structure; and Bok et al. predicted GDP growth *n* Zhen based on real-time data and dynamic factor model of New York Federal Reserve Bank.

At present, interval data widely exist in economic, financial, and social life. The variation range of variables can be described by the lower and upper limits of variables in a specific time period. It has the advantage of rich data information, and the interval prediction results obtained by the interval data model have more reference value. In addition, the existence of uncertain factors makes it difficult to obtain accurate point value prediction in the research of real economic, financial, and system engineering problems, and interval prediction is becoming more and more important. As the Beijing News reported on January 20, 2019, China's multi-provincial government work report sets the CDP growth target as an interval value. So, the main innovation of this study is to extend the research object of macroeconomic forecasting model from traditional point data to interval data (the value result of random variables is interval data), to increase the use of data information and improve the efficiency of statistical inference and prediction accuracy. At the same time, for the interval time-series data in the financial market, a method of variable selection of interval data model based on the mean square error criterion is proposed, to construct the interval prediction model of macroeconomic and analyze its interval prediction effect, and then provides decision-making reference for the formulation of China's macroeconomic monitoring and early warning policies and measures in the future.

3.3. Interval Data Model and Its Variable Selection Method

3.3.1. Interval Data. Let $K_c(R)$ be a set of nonempty compact intervals in R, any $A, B \in K_c(R)$ is taken, and interval addition operation $A + B$ and interval number multiplication operation λA on $K_c(R)$, and the Hukuhara difference operation of interval data $C = A - B_H$ are defined, which satisfies the following:

$$A = B = C,\ C \in K_c(R). \tag{1}$$

Based on the above interval data algorithm, the general structure of linear interval time-series data model is as follows:

$$Y_t = \alpha_0 + \beta_0 I_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} + \sum_{l=0}^{s} \delta_l' X_{t-1} + u_t, \tag{2}$$

where $\alpha_0, \beta_0, \beta_j\,(j = 1, \ldots, p)$ and $\delta_l = (\delta_{l,1}, \delta_{l,2}, \ldots, \delta_{l,h})'\,(l = 0, 1, \ldots, s)$ are the parameters for estimating, $I_0\,[-1/2.1/2]$ is the constant unit interval, $\alpha_0 + \beta_0 I_0 = [\alpha_0 - \beta_0/2, \alpha_0 + \beta_0/2]$ is the intercept term of the interval, $Y_t = [Y_{L,t}, Y_{R,t}]$ is a stationary interval time-series process, and its stationarity can be realized by the Hukuhara difference, $X_t = (X_{1t}, X_{2t}, \ldots, X_{mt})'$ is a stationary vector interval data process, and $X_{it} = [X_{L,it}, X_{R,it}]$, $i = 1, \ldots, h$, $u_t = [u_{Lt}, u_{Ht}]$ is an interval martingale difference sequence process and satisfies $E(I_{t-1}) = [0, 0]$.

The important point value data model can be obtained from the interval data model through interval operation. For example, the following interval midpoint model and interval range model can be obtained using the operation of the left and right boundaries of the interval:

$$\begin{aligned} Y_t^m &= \alpha_0 + \sum_{j=1}^{p} \beta_j Y_{t-j}^m + \sum_{l=0}^{s} \delta' X_{t-j}^m + u_t^m, \\ Y_t^r &= \beta_0 + \sum_{j=1}^{p} \beta_j Y_{t-j}^T + \sum_{l=0}^{s} \delta' X_{t-j}^r + u_t^r, \end{aligned} \tag{3}$$

where Y_t^m and Y_t^r represent the midpoint and range of interval process $Y_t = [Y_{L,t}, Y_{R,t}]$, and X_t^m and X_t^r represent the midpoint vector and range vector of vector interval process X_t. The sum is an additive new interest item and satisfies $E(I_{t-1}) = 0$ and $E(I_{t-1}) = 0$. Note that the parameters β_0 and α_0 cannot be identified in equations (3) and (4), because equations (3) and (4) only use the midpoint information and range information of interval prices, respectively, and lose some other information.

Table 1: Results of stationarity test of interval time-series variables.

Indicator symbol	Data sample	Right boundary ADF test statistics	Right boundary p value	Left boundary ADF test statistics	Left boundary p value	Inspection results
GDP	Original interval sequence	−1.608	0.4794	−2.147	0.2261	Unstable
	Hukuhara difference sequence	−5.565	0.0000	−5.522	0.0000	Stable
SHCI	Original interval sequence Hukuhara difference sequence	−1.584	0.4919	−2.492	0.1175	Unstable
	Hukuhara difference sequence	−6.287	0.0000	−6.549	0.0000	Stable
SZCI	Original interval sequence Hukuhara difference sequence	−1.606	0.4804	−1.784	0.3885	Unstable
	Hukuhara difference sequence	−4.959	0.0000	−5.747	0.0000	Stable
CCI	Original interval sequence Hukuhara difference sequence	−1.996	0.2882	−2.405	0.1404	Unstable
	Hukuhara difference sequence	−5.662	0.0000	−5.827	0.0000	Stable
SHFI	Original interval sequence Hukuhara difference sequence	−1.151	0.6942	−0.526	0.8869	Unstable
	Hukuhara difference sequence	−3.151	0.0230	−3.544	0.0069	Stable
FV	Original interval sequence	−3.908	0.0020	−4.774	0.0001	Stable
FM	Original interval sequence	−4.496	0.0002	−5.588	0.0000	Stable
M1	Original interval sequence	−3.228	0.0184	−3.674	0.0045	Stable
RMBEXR	Original interval sequence Hukuhara difference sequence	−2.191	0.2096	−1.966	0.3017	Unstable
		−4.979	0.0000	−4.470	0.0002	Stable

3.3.2. Variable Selection of Interval Data Model Based on Mean Square Error. For the parameter estimation of the above interval data model, $D_k(.,.)$ is the interval distance measurement form, which is defined as for any $A, B \in K_c(R)$, then:

$$D_K(R,B) = \sqrt{\iint [s_A(u) - s_B(u)][s_A(v) - s_B(v)]dK(u,v)}, \tag{4}$$

where K is a kernel function and satisfies the following:

$$\begin{aligned} s_A(u) &= \langle u, a\rangle, \\ u \in R^l.S^0 &= \left\{u \in R^1, |u| = 1\right\} \\ &= \{1,-1\}. \end{aligned} \tag{5}$$

According to the interval data model in (2), the parameter vector corresponding to the model is set to be $\phi = [\alpha_0, \beta_0, \beta_1, \ldots, \beta_p, \delta_0', \delta_1', \ldots, \delta_s']^{'}$, so the estimated parameter ϕ based on D_k distance is as follows:

$$\begin{aligned} \widehat{\phi} &= \arg\min_{\phi\in\Phi} \widehat{Q}_T(\phi), \\ \widehat{Q}_T(\phi) &= \frac{1}{T}\sum_{i-1}^{T} q_t(\phi), \\ q_t(\phi) &= \|\widehat{u}(\phi)\|_k^2, \\ &= \left\|Y_t - Z_t'(\phi)\phi\right\|_k^2, \\ &= D_K^2\left[Y_t, Z_t'(\phi)\phi\right], \end{aligned} \tag{6}$$

where $Z_t(\phi)$ stands for the parameter vector and $q_t(\phi)$ is the residual item. Assuming that the number of alternative interval explanatory variables in interval data model (2) is $n = p + 2 + (s+1)h$, there are C_n^k regression subsets with k independent variables, which are recorded as $A(k)$, and then, the number of independent variables k should be selected so that the mean square error $s^2(A) = T\widehat{Q}_T/T - k - 1$ is minimized.

3.3.3. Data Acquisition and Analysis. This study selects the quarterly data of cumulative GDP as the measurement index of macroeconomy. At the same time, according to the research results of economic and financial theory and economic prosperity analysis, in the empirical research of macroeconomic interval prediction model, the required interval financial data indexes come from stock market, fund market, futures market, and money market, respectively. These financial data indicators affect household consumption expenditure, the implementation of fiscal policy, and import and export trade and have an impact on China's macroeconomic aggregate. According to different data frequencies, this study selects the maximum and minimum values of financial data indicators in a certain year to construct the interval time series of corresponding indicators. All samples are from the WAND database. The specific indicators are shown in Table 1. It should be noted that in the process of selecting fund market indicators, since the Shenzhen fund index was terminated after June 2017 and replaced by Lok Fu index, to ensure the consistency of data, this study only selects the Shanghai fund index.

To ensure the stationarity of interval financial time-series prediction model variables, ADF stationarity test is carried out for each interval data sample series. The specific test results are shown in Table 1.

From the stability test results of financial time-series data in Table 1, it can be seen that the original interval series of interval

Table 2: Basic statistical analysis results of interval time-series samples.

Indicator symbol	Interval attribute variable	Mean value	Median	Max. value	Min. value	Standard deviation	Skewness	Kurtosis
	ΔGDPR	−0.1167	−0.1000	2.8000	−2.9000	1.5062	−0.0065	2.5565
GDP	ΔGDPL	−0.1056	−0.1500	4.2000	−4.1000	1.8154	−0.1169	4.1498
	GDPT	0.7278	0.5000	3.0000	0.0000	0.7865	1.4905	4.8474
	ΔSHCIR	0.0291	−0.0379	0.8194	−0.4624	0.3468	0.9387	3.2403
SHCI	ΔSHCIL	0.0326	−0.0307	0.7827	−0.4230	0.2618	1.1252	5.0818
	SHCI	0.4567	0.3587	1.1991	0.1344	0.2762	1.3508	4.0570
	ΔSZCIR	0.0462	−0.0155	1.0753	−0.3637	0.3658	1.4780	4.7883
SZCI	ΔSZCIL	0.0414	−0.0211	0.8293	−0.2916	0.2517	1.7387	6.4107
	SZCI	0.5221	0.4398	1.2373	0.2114	0.2901	1.2652	3.5472
	ΔCCIR	0.5650	0.1500	14.5300	−7.1000	4.2502	1.7698	8.0088
CCI	ΔCCIL	0.4961	0.2000	9.7000	−11.0000	5.0414	−0.3168	3.1199
	CCI	4.6250	4.4200	11.7000	0.5000	3.4247	0.7030	3.7793
	ΔSHF1R	0.0976	0.4340	0.8940	−0.2578	0.3295	1.4434	4.3642
SHF1	ΔSHF1L	0.0961	0.0546	0.8908	−0.1692	0.2474	1.9765	7.0087
	SHF1	0.3781	0.2782	0.9179	0.0852	0.2784	1.1303	2.8469
	FVR	85.5483	61.3300	236.5800	−26.1500	80.5813	0.5935	2.3313
FV	FVL	9.0952	0.4132	87.2374	−43.7900	39.6490	0.3948	2.1410
	FV	76.4532	67.0900	233.9400	13.7200	57.4186	1.1411	3.7206
	FMR	100.8111	84.5450	305.6700	−61.6900	95.4186	0.4923	2.5563
FM	FML	19.3034	2.0550	124.5500	−73.2100	55.3134	0.5528	2.2463
	FM	81.5077	74.8600	181.1200	11.5200	49.9537	0.4737	4.0449
	M1R	19.5083	18.4900	38.9600	7.3000	7.7424	0.9976	2.1615
M1	M1L	10.0128	9.7000	20.8700	1.2000	5.9927	0.2643	5.6518
	M1	9.4956	7.6150	27.9500	2.2400	6.1243	1.6273	3.0511
	ΔRMBEXRR	−0.0612	−0.0179	0.4906	−0.5420	0.2468	0.2201	3.3881
RMBEXR	ΔRMBEXRL	−0.0947	−0.0079	0.3619	−0.5390	0.2095	−0.2941	2.3564
	RMBEXR	0.2096	0.1883	0.6376	0.0006	0.1918	0.5344	2.5563

data indicators such as gross domestic product (GDP), Shanghai Composite Index (SHCI), Shenzhen Component Index (SZCI), consumer confidence index (CCI), Shanghai fund index (SHFI), and people's currency exchange rate (RMB XR) is unstable, but the interval series after the Hukuhara difference is stable. Therefore, for these data indicators, the Hukuhara difference operation is used in the prediction model to set the corresponding interval explanatory variables and interval explained variables. The original interval series of futures market turnover (FV), futures market turnover (FM), and narrow money supply (M1) is stable. Therefore, in the later interval prediction model, these variables do not need to be transformed into data stability through the Hukuhara difference operation. Based on the interval variables determined by the stationarity test, the basic statistical analysis of the sample is given, and the results are shown in Table 2.

The statistical results of each point value attribute of interval variables in Table 2 show that the characteristics of standard deviation, kurtosis, skewness, and other aspects of interval right boundary (also known as interval upper bound), interval left boundary (also known as interval lower bound), and interval range of each interval data process are very different [17]. Although these intervals' valued attribute variables are extracted from the same interval process, the information contained is very different. Therefore, this study attempts to study the macroeconomic prediction based on interval financial time-series data from the perspective of interval data overall modeling [18].

4. Experimental Results

Based on the basic analysis of interval financial data indicators, an interval financial data indicator set:

$$A = \{\Delta SHCI_{t-1}, \Delta SZCI_{t-1}, \Delta CCI_{t-1}, \Delta SHFI_{t-1}, FV_{t-1}, FM_{t-1}, M1_{t-1}, \Delta RMBEXR_{t-1}\}, \tag{7}$$

which can be used to predict macroeconomy is determined. Therefore, the general structure of the macroeconomic interval prediction model used in this study is as follows:

$$\Delta GDP_t = \alpha_0 + \beta_0 I_0 + \lambda \Delta GDP_{t-1} + \sum_{t=1}^{k} \delta_l X_{t-1}^l + u_t. \tag{8}$$

Among them, the interval variable X_{t-1}^l is selected from the interval independent variable index set A. In this study, the interval variable X_{t-1}^l will be selected from the stock market, fund market, futures market, and money market, respectively, and the interval explanatory variables selected in equation (8) will be gradually determined according to the minimum mean square error criterion. The detailed results are shown in Tables 3–7, in which the symbol ✓ in each table represents the corresponding variables selected in the corresponding interval prediction model, and each serial number represents different interval variable combination structures.

In Table 3, by comparing the mean square error of the interval model corresponding to the combination of financial variables in different stock markets, the Hukuhara

Table 3: Results of variable selection of macroeconomic interval prediction model based on stock market financial data index.

Order	α_0	β_0	ΔGDP_$(t-1)$	ΔSHCI_$(t-1)$	ΔSZCI_$(t-1)$	ΔCCI_$(t-1)$	MSE2
1	✓	✓					68.5660
2	✓	✓	✓	✓			70.9486
3	✓	✓	✓		✓		67.3342
4	✓	✓	✓	✓		✓	69.9477
5	✓	✓	✓	✓	✓		71.0457
6	✓	✓	✓			✓	73.9438
7	✓	✓	✓	✓	✓	✓	71.1464
8	✓	✓	✓		✓	✓	74.8738
9			✓				61.1446
10				✓			66.9370
11					✓		66.2403
12						✓	64.8534
13			✓	✓			65.3981
14			✓		✓		61.8689
15			✓			✓	64.4728
16				✓	✓		67.7883
17				✓		✓	68.5598
18					✓	✓	67.7145
19			✓	✓	✓		65.1464
20			✓	✓		✓	67.8495
21			✓		✓	✓	65.1464

difference interval variable ΔGDP_{t-1} of GDP and the Hukuhara difference and the Shenzhen component index $\Delta SZCI_{t-1}$ are preliminarily determined as the candidate prediction variables of the macroeconomic interval prediction model, which are further compared with the fund market, and the range financial variables of futures market and money market are analyzed. In this study, the macroeconomic interval prediction model corresponding to the optimal result ΔGDP_{t-1} and $\Delta SZCI_{t-1}$ variables in Table 3 is abbreviated as model I. Variable selection results of macroeconomic interval prediction model based on financial data index of fund market are shown in Table 4.

According to the mean square error comparison results of the interval model of the financial data indicators of the fund market in Table 4, the Hukuhara difference in the Shanghai Stock Exchange Fund Index is preliminarily determined as the candidate prediction variable of the macroeconomic interval prediction model and further combined with the regional financial variables of the stock market, futures market, and money market. In this study, the macroeconomic interval prediction model corresponding to the optimal result variable in Table 4 is abbreviated as model II.

In Table 5, by comparing the mean square error of the interval model corresponding to the combination of financial variables in different futures markets, the transaction amount in the futures market is preliminarily determined as the candidate prediction variable of the macroeconomic interval prediction model and further combined with the interval financial variables in the stock market, fund market, and money market. In this study, the macroeconomic interval prediction model corresponding to the optimal result variable in Table 5 is abbreviated as model III.

In Table 6, by comparing the mean square error of interval models corresponding to different combinations of

Table 4: Variable selection results of macroeconomic interval prediction model based on financial data index of fund market.

Order	α_0	β_0	ΔGDP_$(t-1)$	ΔSHCI_$(t-1)$	MSE2
1	✓	✓			67.1991
2				✓	59.2855
3			✓	✓	63.4567

Table 5: Results of variable selection of macroeconomic interval prediction model based on financial data index in futures market.

Order	α_0	β_0	ΔGDP_$(t-1)$	FV_$(t-1)$	FM_$(t-1)$	MSE2
1	✓	✓				67.1723
2				✓		63.1001
3					✓	61.0152
4			✓	✓		67.2987
5			✓		✓	65.1679
6				✓	✓	61.3855
7			✓	✓	✓	66.0816

money market financial variables, the narrow money supply is preliminarily determined as the candidate prediction variable of $M1_{t-1}$ macroeconomic interval prediction model and further combined with the interval financial variables of stock market, fund market, and futures market. In this study, the macroeconomic interval prediction model corresponding to the optimal result $M1_{t-1}$ variable in Table 5 is abbreviated as model IV.

Based on the candidate prediction variables $\Delta SZCI_{t-1}, \Delta SHFI_{t-1}, \Delta FM_{t-1}$, and $M1_{t-1}$ of the macroeconomic interval prediction model preliminarily selected from the stock market, fund market, futures market, and money market in Tables 3 to 6, the combination analysis of interval financial variables is further carried out to determine the optimal macroeconomic interval prediction model. See Table 7 for the specific results.

Table 6: Variable selection results of macroeconomic interval prediction model based on money market financial data indicators.

Order	α_0	β_0	ΔGDP_(t − 1)	M1_(t − 1)	ΔRMBEXR_(t − 1)	MSE2
1	✓	✓				51.5518
2				✓		48.9569
3					✓	50.6920
4			✓	✓		50.9221
5			✓		✓	52.7987
6				✓	✓	51.0391
7			✓	✓	✓	53.1168

Table 7: Variable selection results of macroeconomic interval prediction model based on financial data indicators.

Order	α_0	β_0	ΔGDP_(t − 1)	ΔSZCI_(t − 1)	ΔSHFI_(t − 1)	FM_(t − 1)	M1_(t − 1)	Errors2
1	✓	✓	✓	✓	✓	✓	✓	91.6767
2			✓	✓	✓			68.1735
3			✓	✓		✓		65.8909
4			✓	✓			✓	67.4801
5			✓		✓	✓		66.6546
6			✓		✓		✓	68.0650
7			✓			✓	✓	71.3898
8				✓	✓	✓		66.4587
9				✓	✓		✓	65.8597
10				✓		✓	✓	73.4322
11					✓	✓	✓	71.6713
12			✓	✓	✓		✓	71.2655
13			✓	✓	✓	✓		70.7993
14				✓	✓	✓	✓	71.1229
15			✓	✓		✓	✓	77.3397
16			✓		✓	✓	✓	67.4567
17			✓	✓	✓	✓	✓	75.1236

For different combinations of candidate prediction variables $\Delta G\,DP_{t-1}$, $\Delta SZCI_{t-1}$, $\Delta SHFI_{t-1}$, FM_{t-1}, and $M1_{t-1}$, Table 7 comprehensively compares and analyses the mean square error corresponding to the macroeconomic interval prediction model under various circumstances. The data results show that the mean square error of the combination of interval financial variables $\Delta SZCI_{t-1}$, FM_{t-1}, and $M1_{t-1}$ selected from the stock market, futures market, and money market and the combination of interval financial variables $\Delta SHFI_{t-1}$, FM_{t-1}, and $M1_{t-1}$ selected from the fund market, futures market, and money market is small, Therefore, this study preliminarily selects $\Delta SZCI_{t-1}$ (or $\Delta SZFI_{t-1}$), FM_{t-1}, and $M1_{t-1}$ as interval prediction variables of macroeconomy. Because the GDP prediction effects of the two groups are very similar, this study only lists the prediction results of one group of interval data models. At this time, the macroeconomic prediction model corresponding to the optimal result $\Delta SZFI_{t-1}$, FM_{t-1}, and $M1_{t-1}$ variables in Table 7 is abbreviated as model V.

Based on the comparison results of using the financial data of stock market, fund market, futures market, and money market to predict macroeconomy in Tables 3 to 7, this study obtains five better macroeconomic prediction models (model I to model V). Further, based on the sample fitting error MSE of interval time series of different prediction models, model I to model V are weighted by model combination, and three cases are considered in the weight structure of each model combination: equal weight structure, relative performance weight, and rank weight. The weight calculation formula is as follows:

$$\begin{aligned}\omega_i^{MSE} &= \frac{\left(MSE^{-1}\right)^{\gamma}}{\sum_{j=1}^{5}\left(MSE_j^{-1}\right)^{\gamma}},\\ \omega_i^{\mathrm{Rank}} &= \frac{R_i^{-1}}{\sum_{j=1}^{5} R_j^{-1}}.\end{aligned} \tag{9}$$

Among them, ω^{MSE} in $\gamma = 0$ and $\gamma = 1$ represents equal weight structure and relative performance weight based on MSE and ω_i^{MSE} represents MSE and model weight of the ith model; ω^{Rank} represents the rank weight according to the predicted performance order; and R_i and ω_i^{Rank} represent the prediction performance order and model rank weight of the ith model. Using the parameter estimation results of model I to model V, the MSE of each model can be calculated, respectively, and then, the combined weight of different models can be obtained.

Therefore, from the interval prediction results of China's macroeconomic GDP based on the combination model in Table 7, it can be seen that the fluctuation range of China's macroeconomic growth from 2020 to 2023 is 6.3426~6.6152, 6.3387~6.5622, 6.2291~6.5451, and 6.2177~6.4998, respectively. From the forecast results, China's macroeconomic operation continues to remain within a reasonable range from 2020 to 2023 and continues the overall stable, stable, and slow development trend. However, at present, China still urgently needs to solve problems such as technological innovation and upgrading and industrial structure adjustment. At the same time, the global economic situation is still complex and severe, external instability and uncertainties are increasing, and foreign trade frictions and geopolitical risks still exist. Therefore, China's macroeconomy as a whole will continue to face new downward pressure.

5. Conclusion

Reasonable prediction of macroeconomic environment is one of the key points of current macroeconomic and policy research. Accurately grasping the fluctuation range of future economic operation can provide important decision-making basis for the formulation of relevant policies in China. Based on the stock market, fund market, futures market, and money market, this study discusses the interval prediction ability of relevant interval financial data indicators to the macroeconomy, gives the variable selection method of constructing the interval data model based on the mean square error criterion, carries out the interval prediction analysis of the macroeconomy, and gives the macroeconomic prediction method, which is different from the traditional point value data model. The empirical results show that the Shenzhen Component Index, Shanghai fund index, futures market transaction amount, and narrow money supply in the interval financial data have a significant interval prediction ability for the macroeconomy, and the prediction results show that China's macroeconomy will continue to develop steadily and slowly in the future.

The macroeconomic interval prediction method proposed in this study expands the methodology of the traditional point value macroeconomic prediction model, and the interval data model can make full use of the data information than the traditional point value data model to make its advantages in statistical inference and interval prediction modeling, to give more accurate interval prediction results. This has important theoretical value and practical significance for mastering the future development trend of China's macroeconomy. From the macroeconomic range prediction results, China's future economic growth will be in an overall controllable slowdown and soft-landing process, which will also provide a huge development space for China to fully deal with the adverse changes in the external environment, the adjustment and upgrading of industrial structure, and the exploration of new economic growth points. Therefore, to better maintain stable economic growth, the government should continue to steadily promote the development measures of opening up, promoting reform and improving efficiency through competition, further stabilize market confidence, stimulate market vitality, and improve the quality of development. Secondly, with the support of active fiscal and tax policies, policy initiatives such as financial market service innovation are strengthened, enterprise financing environment is improved, enterprise competitiveness is enhanced, and industrial transformation and upgradation are promoted. Thirdly, monetary policy needs to provide reasonable and sufficient liquidity, monitor the flow channels of funds while providing stable and effective investment, and prevent the inflated development of some industries. Finally, the marketization of basic assets is improved, especially in the fields of oil, electricity, and natural gas closely related to macroeconomic development, the market transaction volume and liquidity quality are stimulated, and the leading response ability of the financial market to the macroeconomy is further improved [19].

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] T. B. Gotz and T. A. Knetsch, "Google data in bridge equation models for German GDP," *International Journal of Forecasting*, vol. 35, pp. 45–66, 2019.

[2] P. Kuosmanen and J. Vataja, "Time-varying predictive content of financial variables in forecasting GDP growth in the G-7 countries," *The Quarterly Review of Economics and Finance*, vol. 71, pp. 211–222, 2019.

[3] E. Ghysels, "Macroeconomics and the reality of mixed frequency data[J]," *Journal of Econometrics*, vol. 193, pp. 294–314, 2016.

[4] Y. Jiang, Y. Guo, and Y. Zhang, "Forecasting China's GDP growth using dynamic factors and mixed-frequency data," *Economic Modelling*, vol. 66, pp. 132–138, 2017.

[5] R. E. Moore, *Interval Analysis*, Prentice-Hall, Englewood Cliffs N J, USA, 1966.

[6] E. d. A. Lima Neto and F. d. A. T. De Carvalho, "An exponential-type kernel robust regression model for interval-valued variables," *Information Sciences*, vol. 454-455, pp. 419–442, 2018.

[7] L. C. Souza, R. M. C. R. Souza, G. J. A. Amaral, and T. M. Silva Filho, "A parametrized approach for linear regression of interval data," *Knowledge-Based Systems*, vol. 131, no. 131, pp. 149–159, 2017.

[8] L. T. He, C. Hu, and K. M. Casey, "Prediction of variability in mortgage rates: interval computing solutions," *The Journal of Risk Finance*, vol. 10, no. 2, pp. 142–154, 2009.

[9] C. Hu, "A note on probabilistic confidence of the stock market ILS interval forecasts," *The Journal of Risk Finance*, vol. 11, no. 4, pp. 410–415, 2010.

[10] P. Hao and J. Guo, "Constrained center and range joint model for interval-valued symbolic data regression," *Computational Statistics & Data Analysis*, vol. 116, pp. 106–138, 2017.

[11] W. Yang, A. Han, Y. Hong, and S. Wang, "Analysis of crisis impact on crude oil prices: a new approach with interval time series modelling," *Quantitative Finance*, vol. 16, no. 12, pp. 1917–1928, 2016.

[12] Y. Sun, A. Han, Y. Hong, and S. Wang, "Threshold autoregressive models for interval-valued time series data," *Journal of Econometrics*, vol. 206, no. 2, pp. 414–446, 2018.

[13] A. L. Maia and F. A. De Carvalho, "Holt' s exponential smoothing and neural network models for forecasting interval-valued time series[J]," *International Journal of Fore-Casting*, vol. 27, pp. 740–759, 2011.

[14] M. Z. Asghar, A. Lajis, M. M. Alam et al., "A deep neural network model for the detection and classification of emotions from textual content," *Complexity*, vol. 2022, Article ID 8221121, 2022.
[15] Z. Yang, D. K. J. Lin, and A. Zhang, "Interval-valued data prediction via regularized artificial neural network," *Neurocomputing*, vol. 331, pp. 336–345, 2019.
[16] M. Z. Asghar, F. Subhan, H. Ahmad et al., "Senti-eSystem : a sentiment-based eSystem -using hybridized fuzzy and deep neural network for measuring customer satisfaction," *Software: Practice and Experience*, vol. 51, no. 3, pp. 571–594, 2021.
[17] Q. Xu, Q. Guo, and C. X. Wang, "Network differentiation: a computational method of pathogenesis diagnosis in traditional Chinese medicine based on systems science," *Artificial Intelligence in Medicine*, vol. 7724, Article ID 102134, 2021.
[18] Li. Minmin, H. Jiang, H. Yule et al., "A systematic review on botany, processing, application, phytochemistry and pharmacological action of Radix Rehmnniae," *Journal of Ethnopharmacology*, vol. 285, Article ID 114820, 2021.
[19] T. Xiong, C. Li, and Y. Bao, "Interval-valued time series forecasting using a novel hybrid HoltI and MSVR model," *Economic Modelling*, vol. 60, pp. 11–23, 2017.