

Research Article

Big Data-Driven Macroeconomic Forecasting Model and Psychological Decision Behavior Analysis for Industry 4.0

Jie Liu 

School of Psychology, Fujian Normal University, Fuzhou, Fujian 350117, China

Correspondence should be addressed to Jie Liu; 117012018045@student.fjnu.edu.cn

Received 19 August 2021; Accepted 20 September 2021; Published 10 November 2021

Academic Editor: Muhammad Javaid

Copyright © 2021 Jie Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the advent of Industry 4.0, economic development has become a rapid information age. The content of macroeconomic forecast is very extensive, and the existence of big data technology can provide the government with multilevel, diversified, and complete information and comprehensively process, integrate, summarize, and classify these pieces of information. This paper forecasts the CPI value in the next 12 months according to the CPI in China in the recent 20 years. Compared with the traditional forecasting methods, the forecasting results have higher accuracy and timeliness. At the same time, the trend of growth rate of industrial value-added is analyzed, and the experiments on MAE and RMSE show that the method proposed in this paper has obvious advantages. It also analyzes the disadvantages of traditional psychological decision-making behavior analysis, introduces the development status and advantages of big data-driven psychological decision-making behavior analysis, and opens up new research ideas for psychological decision-making analysis.

1. Introduction

The method of big data technology makes macroeconomic forecasting more accurate and effective. Big data technology can provide data resources for national development strategies and can also help judge the development of the whole economic situation to a certain extent. Up to now, psychology has made great achievements. However, in the process of development, due to the limitation of objective factors, many problems gradually appear. The traditional research model that promoted the development of psychology in the past has now become a stumbling block to the development of psychology. The emergence of big data provides a new research route for the development of psychology and creates favorable conditions for the development of psychology.

Macroeconomic data plays an important role in supporting the decision-making of various departments of the country, but the timeliness of this data is relatively poor [1], and it is often quarterly data, so how to improve its timeliness is particularly urgent. The mixed-frequency data method is used to analyze and forecast the macroeconomic

data, and the high-frequency data and mixed-frequency data are sampled and then analyzed. The research shows that the mixed-frequency model is superior to the autoregressive integral moving average model as our benchmark model in most cases. With the rapid development of economy, a large number of data sets have been produced in production. It is necessary to analyze these data sets scientifically and then explore new insights, which requires new modeling techniques to analyze data [2]. After discussion, researchers and managers from different backgrounds and fields summarized the latest applications of big data technology in economic and financial fields and future research aspects and places. Traditional MIS cannot meet the decision analysis of different users for different needs, because the decision needs are divided into different levels or have special needs [3]. A new type of decision support system (DSS) is a modular design decision support system based on data and online processing and analysis system. Researchers set up multi-dimensional data sets for OLAP analysis and proposed a support vector machine acceleration algorithm for boundary sample selection. Tests show that data mining has a good effect on economic forecasting and is worth popularizing. At

present, the development speed of cruise tourism in China is slowing down slowly, and the corresponding investment in cruise ships and ports also has great risks, so predicting investment demand in advance is the premise of decision-making. The least squares support vector regression model based on gravity search algorithm (LSSVR-GSA) is based on big data such as data and economic indicators. Experiments show that the framework is feasible and big data can be used as a forecasting tool [4]. Word embedding and deep learning models have attracted attention in many fields in recent years. It can combine richer information sets including news with the most advanced machine learning models, so as to establish solutions and improve the accuracy of predicting financial time series [5]. Affected by system changes, low signal-to-noise ratio, and so on, people have tried to use standard methods to evaluate, but the effect is not good, and the performance outside the sample is very poor. Based on the data of entrepreneurial enterprises, we create an intermediary model that links entrepreneurial networks with entrepreneurial opportunities through decision-making and use it to study the relationship between decision-making in entrepreneurial networks and entrepreneurial opportunities. Combined with the needs of practical research and what we know, entrepreneurial network is very important for determining entrepreneurial opportunities and development in the context of social media. This model shows our understanding of the formation mechanism of emerging entrepreneurial opportunities and also provides reference opinions for managers and decision makers in various fields [6]. On the one hand, the rapid development of big data has changed the basis of traditional financial forecasting; on the other hand, the performance of its forecasting model is also a challenge [7]. Big data-driven economic and financial forecasting plays an important role in the future social development, which is helpful for economic system feature mining, forecasting research, economic and financial system management mechanism formation, and so on. Predicting the decision-making style of actors, perception is helpful to explore social perception, because when we interact with others, we often rely on our own characteristics of partners to judge the level of cooperation between them [8]. Through the research, we find that the perception of the actor's decision-making style is positively correlated with the inference of the actor's competence and conversely correlated with the intention of cooperation between the perceiver and the actor, which shows that if the intention of strengthening cooperation, it is necessary to infer from the other party's rational decision-making style instead of intuitive decision-making style. With the increase of data scale, macroeconomic forecasting has important social science significance, while the traditional method is basic small-scale data, and the forecasting accuracy is low by mathematical model.

2. Macroeconomic Forecasting Model

2.1. Big Data Overview. The amount of data information processed by big data is extremely large. Compared with traditional data analysis, the amount of information collected by big data is at GB, TB, and even PB and EB

levels. For this reason, traditional computers are unable to deal with such problems. The information includes pictures, voice, network search, and text and log information. These pieces of information have diversified characteristics, and big data itself has high-value characteristics [9]. However, compared with traditional data, the value of big data needs to be reflected through a lot of data analysis.

2.2. Big Data Drives the Development Trend of Macroeconomic Forecast. Survey statistics is the main measure of data collection in traditional macroeconomic forecasting, which usually uses the data released by official statistical departments to ensure accuracy, but this will make the data lose timeliness [10]. If we deliberately pursue timeliness and adopt the latest data, it will increase the possibility of macroeconomic forecasting errors, and the actual results are usually unsatisfactory. The development of big data brings hope to solve this dilemma. It makes people less dependent on statistical resources, and various social media data and retrieval data can become data resources. The channels and types of data sources have greatly increased, and there is no human interference factor, so the data is more authentic and reliable.

At present, traditional statistical data is the main data type supporting the existing forecasting models in China, so the current macroeconomic forecasting will take a long time, and most of them are monthly, quarterly, and annual forecasting models. However, timeliness and accuracy are very important for the country to formulate policies and strategies. At present, traditional macroeconomic forecasting is difficult to meet this requirement. Foreign countries have achieved good results in macroeconomic forecasting, and the data they use are accurate, and the results are also very reliable and accurate. Because of this, we should seize this opportunity, conform to the trend, and apply big data to China's macroeconomic forecasting. Gradually change from medium- and long-term forecasting analysis to immediate forecasting analysis.

Traditional macroeconomic forecasting methods have been developed for many years and have formed a relatively comprehensive and complete research system, which can conduct comprehensive and in-depth analysis of data structures under different circumstances, and then give results. However, due to the problem of data statistics, the prediction effect of its model is difficult to be further improved. Big data technology solves the problems of data acquisition, preservation, and analysis. Because it is a new technology, compared with traditional data sets, it is less used in macroeconomic forecasting, and its theory is not perfect enough. The advantage of big data technology is that it can make macroeconomic forecasting obtain valuable in a short time and then solve economic problems on the basis of traditional models and improve the accuracy of forecasting. The immediacy of big data itself can change the lag of traditional statistical data process. For example, sensors can be used to collect transaction data, and the inflation rate of the current month can be calculated immediately according

to the data. In this way, the data can be kept in the latest state, and various problems caused by the Lucas Critique can be solved [11].

Before the emergence of big data technology, it was difficult to achieve the accuracy and timeliness of prediction at the same time. However, with the development of technology, firstly, the processing power of computer itself is strengthened. Secondly, with the application of big data gradually penetrating into the field of macroeconomic management, the accuracy is gradually improved on the premise of ensuring immediacy. Immediate forecasting or short-term forecasting came into being, and now, instant forecasting has gradually moved from theory to practice.

2.3. Macroeconomic Forecast Model Driven by Big Data. At present, there are four basic methods for real-time prediction: bridging equation, mixed-data sampling, mixed-frequency vector autoregressive (MF-BVAR), and dynamic factor model.

2.3.1. Theoretical Model

Bridge Model. In the case of mixed-frequency data, the earliest econometric method is to use the bridging equation. Linear regression method is a bridging equation-dependent method, which estimates the dynamic equation by using the index after time aggregation of low-frequency index and high-frequency index. For example, the forecast of quarterly GDP growth rate is divided into two steps: the first step is to predict monthly indicators in the remaining period of the quarter based on the univariate time series model and summarize them to obtain the corresponding quarterly values. In the second step, the aggregated quarterly values are used as regression variables for real-time prediction.

Mixed-Data Sampling. Mixed-data sampling is an econometric regression or filtering method developed. MIDAS model is a direct prediction tool, which can directly use high-frequency index and its lag index to predict and analyze low-frequency series immediately through simple regression method. Mixed-data sampling model includes four weight function forms: exponential Almon polynomial, Almon polynomial, Beta polynomial, and step function polynomial. In addition, there are many derivative models for predictive analysis of mixed-data sampling model, including autoregressive AR-mixed-data sampling model, unconstrained U-mixed-data sampling model, nonlinear N-MIDAS model, asymmetric A-MIDAS model, and smooth transfer ST-mixed-data sampling model, and Markov region MS-mixed-data sampling model.

MF-VAR Model. The idea for this model comes from a model proposed by Zadrozny that can be used directly to estimate sampling at different frequencies. In this model, all sequences are generated at the highest frequency, and some of them are considered unobservable. Therefore, those variables that can be observed only at low frequencies are

considered to be periodically missing. At present, there are three estimation methods for MF-VAR, including classical state space method, Bayesian framework, and stacked vector method. An alternative framework for MF-VAR model estimation is a Bayesian technique, and a Gibbs sampling method has been developed to estimate VAR of mixed and irregular sampled data. Different from the classical parameter-driven model, the stacked vector method is actually an observation-driven model. Stacking vector method is set by observable data, and there is no potential process and impact, so as to avoid the setting of measurement equation and filtering equation. MF-VAR is regarded as a multivariate extension of MIDAS analysis.

Factor Model. There are two main methods for real-time prediction analysis using factor model. One is the MF-GRS model. MF-GRS model realizes the real-time forecast analysis of American GDP for the first time, but this model generally selects one quarterly index and several monthly indexes for analysis. Since then, MF-GRS model has been used in real-time prediction analysis of macroeconomic indicators in Britain and France. MF-GRS model is generally limited to the real-time prediction of monthly and quarterly two-frequency mixing index data. The other is the MF-DFM model proposed by Aruoba et al. It not only expands the constraint that the frequency of mixing index is limited to monthly and quarterly, but also improves the estimation method and realizes the dynamic estimation of Kalman filter and mixing data. MF-DFM model is mainly used by some local Federal Reserve in the United States, such as the “immediate forecast” model of New York Federal Reserve, the GDP now model of Atlanta Federal Reserve, and the ENI index of St. Louis Federal Reserve. Generally speaking, the application of real-time forecasting model is mainly concentrated in the Federal Reserve of various regions in the United States. In addition, the real-time forecast website (<https://www.now-casting.com>) also provides real-time forecast analysis of the eurozone, G7, BRICS countries, etc. [12].

2.3.2. Characteristics of Real-Time Prediction Model Data Sets. Maximizing the available information set is the key to real-time forecasting. The data set used in real-time forecasting is quite different from the data set used in traditional economic forecasting, and it is a real-time data set. Croushore and Stark confirmed that real-time data sets are helpful to improve prediction accuracy, but their modeling is very difficult. It takes a lot of work to analyze and process data, mainly reflected in the following four aspects:

High Dimensions. Wide coverage and refined components are the current characteristics of macro-economy, and there are many kinds of corresponding indicators, so the analysis of macroeconomy is actually a big data problem.

Mixing. Because of the difference in obtaining difficulty, the statistical frequency of different indicators is quite different. For different econometric models, it is necessary to unify the

indicators with different frequencies before they can be used in the same model.

Sawtooth. Different types of data are published in an asynchronous manner, and different types of data have different lag periods, which makes the latest available date in different positions between sequences, resulting in possible data missing at the end of the information set. The existence of “jagged” edges makes data integrated into unbalanced panels. In addition, the middle part of a partial sequence may also contain a large amount of missing data.

Amendments. Macroeconomic forecasts usually use official statistics, and official data are usually based on preliminary estimates at the initial stage of release and will be adjusted at a later stage, which makes the data values usually change gradually. More importantly, data revision has a substantial impact on structural modeling, policy analysis, and forecasting [13].

2.3.3. Advantages of Real-Time Macroeconomic Forecasting Models. Because of the difference of the basic theory, nowcasting model has the advantage that other models cannot replace. Since Giannone et al. began to build nowcasting model based on dynamic factor model and used a large number of monthly economic indicators to predict the quarterly GDP growth rate of the United States, many scholars started the exploratory research on nowcasting model. The basic idea is that if there are highly correlated common movements among macroeconomic indicators, most of the dynamics of the sequence can be captured by a few potential dynamic factors, and these factors will be driven by some power.

$$\begin{aligned} x_t &= \Lambda F_t + \xi_t, \quad \xi_t \sim N(0, R), \\ F_t &= \sum_{s=1}^p A_s F_{t-s} + \mu_t, \quad \mu_t \sim N(0, Q). \end{aligned} \quad (1)$$

Here, F_t represents a common component composed of R factors, x_t is a data set composed of N variables ($R < N$), and ξ_t is a heterogeneous component. It is assumed that the heterogeneous component is only related to specific variables and follows the white noise process and is orthogonal to the common component.

To improve the accuracy of macroeconomic forecasting, it is necessary to have high-dimensional data input, but because of the increase of the number of indicators, more parameters need to be determined, and the problem of “dimension curse” also appears. Based on the motion theory of high correlation between macroeconomic indicators, Giannone et al. proposed “potential factors” that can capture most macroeconomic dynamics and then predicted the change of GDP growth rate through the change of “potential factors.” All macroeconomic indicators are composed of two parts: one is the public part, that is, affecting all macroeconomic indicators, that is, “potential factors”; the other part is heterogeneous; that is, it only affects a certain index.

$$x_{i,t|v_j} = u_i + \lambda_{i1} f_{1,t} + \cdots + \lambda_{ir} f_{r,t} + \xi_{i,t|v_j}, \quad i = 1, \dots, n, \quad (2)$$

where u_i is a constant term and $x_{i,t|v_j}$ represents the sequence of macroeconomic indicators after processing. That is to say, the changes of all macroeconomic index data come from the changes of common component $x_{it} = \lambda_{i1} f_{1,t} + \cdots + \lambda_{ir} f_{r,t}$ and heterogeneous component ξ_{it} . The matrix expression is

$$x_t = \mu + \Lambda f_t + \xi_t, \quad (3)$$

where f_t represents the dynamic factor vector of $r \times 1$ dimension ($r < n$) and λ is the factor load matrix.

It is essentially different from the traditional prediction model. When the new data is released, the corresponding new economic information is obtained, and then the corresponding “potential factor” changes, and then the GDP growth rate is updated immediately. In this way, a small number of “potential factors” can describe huge macroeconomic indicators, thus realizing dimension reduction.

Traditional prediction models can only use the same frequency index to predict. Giannone et al. convert the corresponding monthly index series into quarterly index series, extract quarterly dynamic factors, and realize the prediction of quarterly GDP index. There are three main transformation methods:

$$\begin{cases} x_{it} = (1 - L^3)(1 + L + L^2)X_{it} \\ x_{it} = (1 - L^3)(1 + L + L^2)\log X_{it} \times 100 \\ x_{it} = (1 - L^3)(1 + L + L^2)(1 - L^{12})\log X_{it} \times 100 \end{cases} \quad (4)$$

X_{it} represents the original monthly indicator series, x_{it} represents the corresponding quarterly observed series at the end of the quarter, and L is the lag operator.

Giannone et al. can also deal with mixing problems, but it is required to convert monthly indicators into corresponding quarterly values first, and there is a risk that some information may be lost. Giannone et al. assume that low-frequency variables are potentially unobservable high-frequency variables, so it is not necessary to transform high-frequency indicators into corresponding low-frequency values and directly model low-frequency variables. Taking quarterly GDP as an example, firstly, the quarterly variables are modeled, and the potential observation value of quarterly variables in monthly frequency is constructed.

$$\text{GDP}_t^Q = \text{GDP}_t^M + \text{GDP}_{t-1}^M + \text{GDP}_{t-2}^M, \quad t = 3, 6, 9, \quad (5)$$

where GDP_t^Q represents the quarterly level value of GDP corresponding to time t and GDP_t^M represents the corresponding monthly level value. For convenience of expression, the monthly growth rate of potential unobservable GDP corresponding to $y_t^Q = 100x \log(\text{GDP}_t^Q)$, $Y_t^M = 100x \log(\text{GDP}_t^M)$ can be expressed as $y_t = \Delta Y_t^M$, and the expression of quarterly GDP growth rate is as follows:

$$y_t^Q = \begin{cases} Y_t^Q - Y_{t-3}^Q, & t = 3, 6, 9 \dots \\ \text{unobservable}, & \text{others} \end{cases} \quad (6)$$

Since quarterly GDP growth rates are finally predicted immediately, it is necessary to link quarterly growth rates with monthly growth rates. Using the approximation proposed by Mariano and Murasawa, the quarterly growth rate can be decomposed into the potential monthly growth rate, assuming that the geometric mean is approximately equal to the arithmetic mean.

$$\begin{aligned} y_t^Q &= Y_t^Q - Y_{t-1}^Q \approx (Y_t^M + Y_{t-1}^M + Y_{t-2}^M) - (Y_{t-3}^M + Y_{t-4}^M + Y_{t-5}^M) \\ &= y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4}. \end{aligned} \quad (7)$$

By establishing the above connection, the potential unobservable monthly growth rate can be expressed by the same dynamic factor model as the actual monthly variable:

$$y_t = \mu_Q + \Lambda_Q f_t + \varepsilon_t^Q. \quad (8)$$

Therefore,

$$\begin{aligned} y_t^Q &= \mu_Q + \Lambda_Q f_t + 2\Lambda_Q f_{t-1} + 3\Lambda_Q f_{t-2} + 2\Lambda_Q f_{t-3} + \Lambda_Q f_{t-4} \\ &\quad + \varepsilon_t^Q + 2\varepsilon_{t-1}^Q + 3\varepsilon_{t-2}^Q + 2\varepsilon_{t-3}^Q + \varepsilon_{t-4}^Q, \end{aligned} \quad (9)$$

maximizes the use of mixing index sequences by modeling low-frequency variables.

The release time of different index data is different. In order to use the information contained in the data in time when obtaining new data, it is necessary to use the data set with “sawtooth” characteristics to estimate the factors. Based on the dynamic characteristics of “potential factor,” combined with the process of factor extraction, we can conveniently use Kalman filtering technology. Set the heterogeneous component variance at the data that has not been published yet to a very large number.

$$E(\xi_{it|vj}^2) = \psi_i = \begin{cases} \psi_i x_{i,t|vj} & \text{available} \\ \infty x_{i,t|vj} & \text{not available} \end{cases}. \quad (10)$$

Therefore, when using Kalman filtering technology to estimate the factor, the weight obtained at the data that has not been published temporarily at time t is almost 0, and its expression is as follows:

$$\begin{cases} \widehat{f}_t = A_t \widehat{f}_{t-1} + \mu_t \\ \overline{V}_t A_t \overline{V}_{t-1} A_t^T + Q \\ K = \overline{V}_t \Lambda_t^T (\Lambda_t \overline{V}_t \Lambda_t^T + R)^{-1}, \\ \widehat{f}_t = \overline{f} + K(x_t - \Lambda_t \overline{f}_t) \\ \widehat{V}_t = (I - K \Lambda_t) \overline{V}_t \end{cases}, \quad (11)$$

where R represents the covariance matrix of heterogeneous components. When the data is temporarily unavailable, the value at this point is very large so that the Kalman gain K at this point is very small, and the estimated dynamic factor \widehat{f}_t and covariance matrix \widehat{V}_t are almost unchanged. This can not only effectively deal with the “sawtooth” feature at the end of the data set, but also

be suitable for the situation that there are missing data in the interpart of the data set.

As long as new data are available, the growth rate of GDP will be reestimated. The nowcasting model uses real-time data sets, which is different from other models. Whenever new data is released, the “news” contained in the new data will be used to revise the forecast value of GDP growth rate. In this process, we need to reestimate the dynamic factors every time. When the data is revised, the dynamic factors at that time will change accordingly, thus revising the estimated GDP growth rate.

The biggest difference from other real-time forecasting models is that the marginal impact of “news” released by each economic indicator series on GDP growth rate can be obtained by nowcasting model. According to the new data release, nowcasting model can immediately calculate the marginal impact value of this information on GDP growth rate and obtain the impact weight value of this information on GDP, so as to accurately obtain the impact intensity of this index change on GDP.

$$E[y_t^Q | I_{v+1}] = b_{j,v+1} I_{v+1}. \quad (12)$$

Among them, I_{v+1} indicates the “news” contained in data release, and $b_{j,v+1}$ indicates the influence weight of “news” on GDP growth rate. According to the weight, it is easy to judge the impact of indicators on GDP growth rate. The price $E[y_t^Q | I_{v+1}]$ indicates the impact of “news” on GDP growth rate.

That is to say, because the model can easily predict other macroeconomic variables when predicting GDP growth rate in real time, the “news” (unpredicted part) included in the data can be obtained. That is to say, when a new data is released, the difference with the predicted value will have an impact on the predicted value of GDP growth rate.

$$I_{v+1,j} = x_{j,T_{j,v+1}} - E[x_{j,T_{j,v+1}} | \Omega_v]. \quad (13)$$

$I_{v+1,j}$ means “news”; that is, it is not the release of new data itself that has an impact on the predicted value of GDP, but its unpredicted part (news). The predicted value of GDP after the release of new data (new predicted value) can be composed of two parts: one is the last predicted value of GDP (old predicted value), and the other is the predicted value based on the “news” part. Its expression is as follows:

$$E[y_t^Q | \Omega_{v+1}] = E[y_t^Q | \Omega_v] + E[y_t^Q | I_{v+1}]. \quad (14)$$

y_t^Q represents GDP growth rate (variable to be predicted), $E[y_t^Q | \Omega_{v+1}]$ represents new forecast value, $E[y_t^Q | \Omega_v]$ represents old forecast value, and $E[y_t^Q | I_{v+1}]$ represents forecast value based on “news” section. The latter part can be further decomposed:

$$\begin{aligned} E[y_t^Q | I_{v+1}] &= E[y_t^Q I_{v+1}^T] E[I_{v+1} I_{v+1}^T]^{-1} I_{v+1} \\ &= \sum b_{j,t,v+1} (x_{j,T_{j,v+1}} - E[x_{j,T_{j,v+1}} | \Omega_v]). \end{aligned} \quad (15)$$

Therefore, by solving $E[y_t^Q I_{v+1}^T]$ and $E[I_{v+1} I_{v+1}^T]^{-1}$, the influence weight of the “news” part corresponding to each newly

released data on the GDP forecast value can be calculated, and the influence value of the new data release on the GDP growth rate can be obtained by multiplying the weight with the “news” so as to update the GDP growth rate forecast value.

The CPI data predicted in this paper comes from China Economic Database (CEIC). From the historical data chart of China’s CPI and industrial added value year-on-year growth rate in Figures 1 and 2, it can be seen that the economic data in some special periods (such as 1998, 2008, and 2020) are obviously different from other periods, and the growth rate is either too high or too low. Therefore, we introduce a dummy variable to deal with it to ensure the rationality of the data; that is, the data in the special period is set to 1, otherwise it is 0. At the same time, in order to ensure the reliability of the research results, this paper also carries out the following processing on the data: we take January 1998 as the base period and use the price index to deflate the output value to get the true value. At the same time, the missing values in the data set are filled with mean values.

Usually, the growth rate of industrial added value will not be around the same value for a long time, so we need to make a first-order difference. For the variables related to China’s macroeconomy, we use factor model to predict. The data set includes 10 categories, and each category contains 50 variables, taking the time from January 1998 to March 2020 as the time dimension. These factors mainly include the following aspects:

- (1) Price category, including CPI, PPI, and RPI
- (2) Investment category, including fixed assets investment in various industries
- (3) Export category, including the year-on-year growth rate of imports and exports
- (4) Financial category, including money supply M_0 , M_1 , M_2 and weighted average interest rate, exchange rate, foreign exchange reserves, and bond issuance of interbank lending
- (5) Prosperity index, including manufacturing PMI and consumer confidence index
- (6) International economic category, including the import and export volume indicators of major economies such as the United States, Europe and Japan and China. These data come from China Economic Database (CEIC) and National Bureau of Statistics. All data have been stabilized and seasonally adjusted.

Then, we divide the sample into two parts: regression sample interval and out-of-sample prediction interval in Figure 3. The time window U of regression sample interval is from January 1998 to March 2020, with a total of 267 observed values; the length of the scroll window U is 120. The forecast range is 1, 3, 6, and 12 months. In the experiment, we first calculate the model parameters by using the sub-samples in the rolling window and then predict the test set samples by logistic regression equation.

Finally, we compare the prediction accuracy by two quantitative criteria, which are relative root mean square prediction error standard (RMSE) and absolute average prediction error (MAE). In the experiment, we use simple

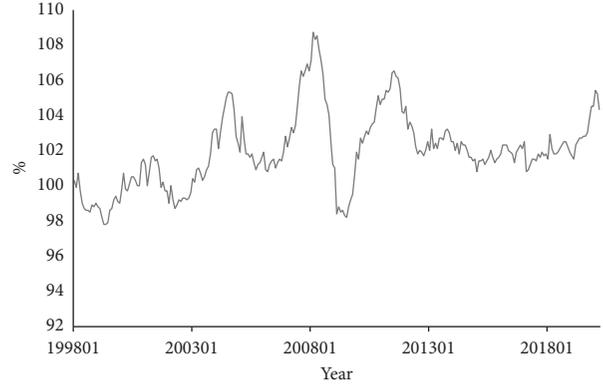


FIGURE 1: Historical data map of CPI in China.

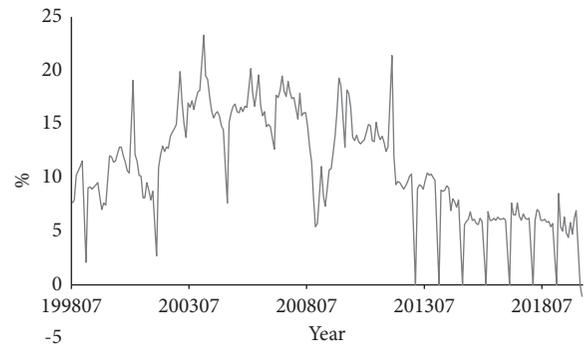


FIGURE 2: Historical data chart of China’s industrial value-added growth rate.

mean model, autoregressive moving average model (ARMA), VAR model, and BVAR model as the benchmark model of this experiment.

Flow window is used to test the training set. If only one or one of 120 data sets is tested, the prediction effect of the whole data area may be poor. Therefore, the proposed flow window training set can make the whole data set show good prediction performance. Extra-sample prediction refers to the prediction of samples around the training set, which generally has a good effect. Using flow window, extra-sample prediction can obtain the training effect of flow window in real time, thus improving the prediction accuracy of the whole algorithm.

The RMSE and MAE formulas are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y'_i - y_i)^2}, \quad (16)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y'_i - y_i|.$$

3. Comparison and Analysis of Prediction Results

According to RMSE standard, in the prediction of CPI and industrial added value (IP) growth rate in China, the

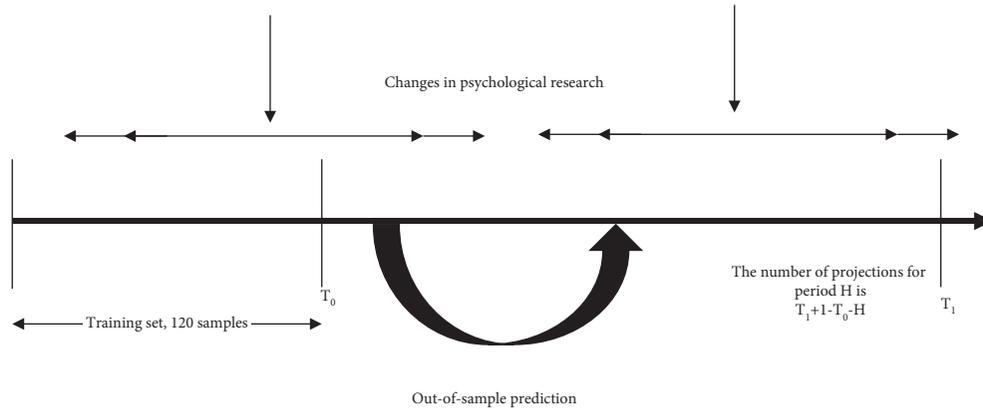


FIGURE 3: Schematic diagram of out-of-sample prediction model.

prediction error of factor model is obviously lower than that of several benchmark models in almost all prediction periods, and the factor model has obvious advantages over several benchmark models when the prediction period is longer in Table 1.

According to MAE standard, in the prediction of CPI, the prediction error of factor model is obviously lower than that of several benchmark models in almost all prediction periods. Therefore, the factor model has obvious advantages over several benchmark models. In the prediction of the growth rate of industrial added value, the prediction accuracy of medium- and long-term BVAR model and factor model is obviously improved compared with the traditional model, and only the moving average model has better prediction effect. The reason of better benchmark results in short-term forecast may be related to the smaller sample size of macro data in China and the types and quantities of selection factors in Table 2.

Generally speaking, the direct prediction accuracy of factor model is higher than that of general model. According to the comparison results of RMSE and MAE quantitative evaluation criteria, on the whole, in the prediction accuracy evaluation, the factor model is consistently superior to other benchmark models in the medium and long term, which shows that the factor model can effectively solve the problems of too many parameters and overfitting existing in the traditional model, and embodies the remarkable superiority of applying the factor model to predict macroeconomic variables.

Through the estimation method of the number of factor models mentioned above, it can be seen that only two factors are needed to approximate the running characteristics of CPI data and industrial added value growth rate in China. These two factors explain 71% and 68% of the data set, respectively. Further study shows that the optimal number of factors of Bai's factor model obtained according to the optimal IC criterion is $K=7$. According to Fond et al.'s (2009) explanation of the economic meaning of predictors, predictors represent domestic macroeconomic indicators that have an impact on their own macroeconomics. The actual results show that compared with the predictors of international economy, China's own macroeconomic factors

have a more obvious impact on the changes of CPI and industrial added value in Figures 4 and 5.

Figures 4 and 5, respectively, give the forecast results of China's CPI and industrial added value growth rate from May 2020 to May 2021 based on the number of optimal factors calculated by the above method (the red part is the confidence interval of 95% of the forecast value).

The macroeconomic forecasting model driven by big data has important practical significance. As the most mainstream model at present, real-time forecasting can provide reference for making public policies and enterprise decisions by effectively using available information to forecast macroeconomy [14].

4. Analysis of Psychological Decision-Making Behavior Driven by Big Data

4.1. Characteristics and Advantages of Data under the Background of Big Data. In the context of big data, data is behavior. A lot of information is recorded in big data, and this information is actually the behavior of universal people. It has the characteristics of voluntariness and universality.

First of all, most of the behaviors in big data are naturally and voluntarily generated by people. For example, messages from chat tools such as Weibo, although most of the time, have a certain purpose of display. But these data are voluntary and not forced. Experienced researchers will find that most psychological tests are course subjects, and the subjects who carry out the tests have the purpose of obtaining credits or bonuses, even including online data like Amazon MTurk, and the subjects also take psychological tests because of money. The results produced in this case are indeed controversial as representatives, because the experimenter's motivation will add unnecessary variables to the psychological test. Regardless of the subjects for credit and money, the subjects were mostly women, in which case they were different in some characteristics. However, the behavioral data in the context of big data is naturally generated without coercion. Moreover, data acquisition is also very convenient, such as wearable devices, which can make them easily obtain behavioral data. This is a big advantage of big data behavior data.

TABLE 1: RMSE value of CPI and industrial value added forecast in China.

CPI	$h=1$	$h=3$	$h=6$	$h=12$	IP	$h=1$	$h=3$	$h=6$	$h=12$
Mean	2.196	2.264	2.658	3.097	Mean	4.147	4.239	4.416	4.776
MA	1.602	1.964	2.785	3.203	MA	3.534	3.822	4.419	4.835
AR(1)	1.588	1.875	2.571	3.161	AR(1)	3.904	4.014	4.495	5.062
AR(p)	1.564	1.770	2.616	3.264	AR(p)	3.783	3.893	4.374	4.941
ARM(1,1)	1.574	1.712	2.842	3.215	ARM(1,1)	3.630	3.764	4.341	4.898
VAR(p)	1.608	2.299	2.783	3.431	VAR(p)	4.466	3.303	4.515	6.314
BVAR	1.593	1.988	2.578	2.602	BVAR	3.729	3.804	4.303	4.443
Factor(2)	1.678	1.821	2.289	2.386	Factor(2)	3.582	3.799	4.136	4,366
Factor(k)	1.582	1.729	2.339	2.512	Factor(k)	3.530	3.884	4.190	4.398

TABLE 2: MAE values predicted by CPI and industrial added value in China.

CPI	$h=1$	$h=3$	$h=6$	$h=12$	IP	$h=1$	$h=3$	$h=6$	$h=12$
Mean	1.622	1.675	1.750	2.189	Mean	3.667	3.764	3.906	4.253
MA	0.656	0.997	1.476	2.195	MA	1.741	1.992	2.490	2.869
AR(1)	0.423	0.826	1.352	2.082	AR(1)	2.062	2.250	2.779	3.390
AR(p)	0.457	0.858	1.447	2.304	AR(p)	2.214	2.094	2.592	3.115
ARM(1,1)	0.424	0.786	1.296	1.910	ARM(1,1)	2.163	2.313	2.927	3.766
VAR(p)	0.480	0.975	1.711	2.870	VAR(p)	2.227	2.383	3.158	4.293
BVAR	0.441	0.838	1.561	2.028	BVAR	2.184	2.272	2.722	3.747
Factor(2)	0.497	0.868	1.301	1.767	Factor(2)	2.011	2.142	2.603	3.060
Factor(k)	0.472	0.896	1.321	1.831	Factor(k)	2.024	2.177	2.615	3.087

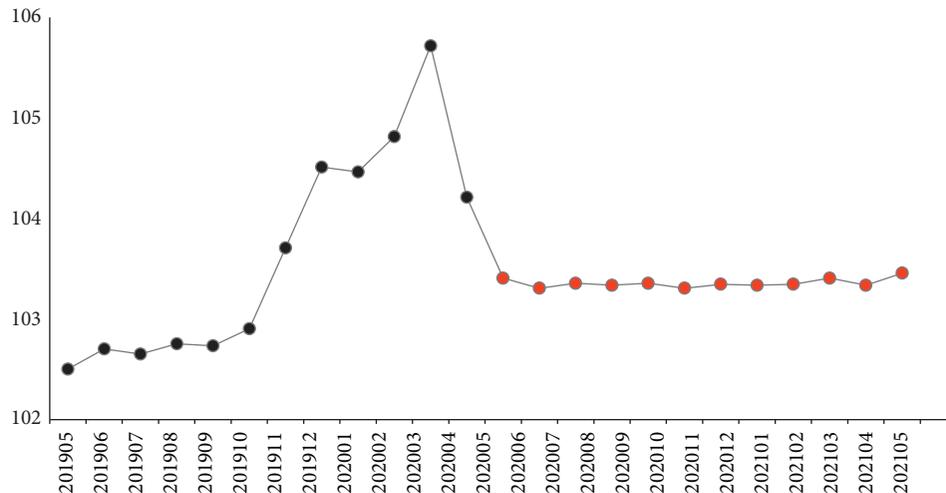


FIGURE 4: CPI forecast value for 12 months.

Secondly, the behavioral data of big data is generalized and comparable. Because of the inability to observe psychological phenomena directly, there are various psychological behaviors that can be observed in psychological research. In psychological research, because different research methods are more easily favored by magazine editors, researchers usually like to observe the same psychological behavior with different research methods. This makes the system of disciplines very large but disorganized, and many similar concepts are not comparable. For example, to study aggressive behavior, researchers can observe whether people bump into a doll, whether people click on a person who makes a mistake, and even the extent to which a person will

tear up other people's photos. Although it is the same behavior, the definition of each observation and operation is different, which makes it difficult to compare. The emergence of big data makes a turning point in solving this problem. When data is collected, researchers can mostly mine behaviors from existing data. At this point, the definition of behavior and operation are unified so that the results produced by big data can be compared [15].

4.2. Analysis of Traditional Psychological Decision-Making Behavior and Its Disadvantages. Scientific epistemology and methodology are the guiding principles of psychology since

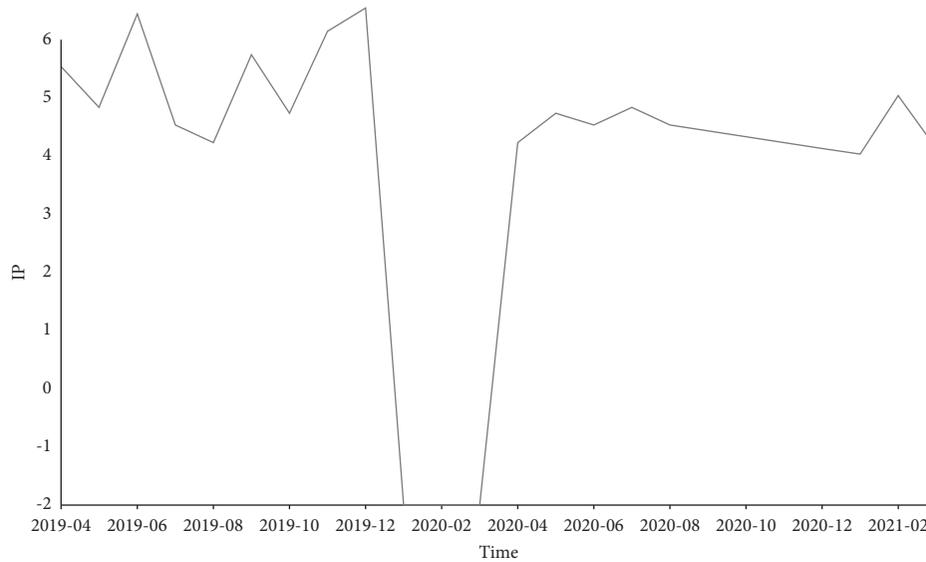


FIGURE 5: Forecast value of industrial added value growth rate in 12 months.

its birth. Psychology has been exploring and improving its research methods, striving to form a rigorous and mature discipline research paradigm. However, for a long time, the scientific nature of psychological research has been questioned both inside and outside the discipline. After Watson's criticism, psychology began to turn its attention to behavioral data. This makes psychological research change its most important research object from the structure and function of consciousness to the study of behavior. However, traditional psychology still uses a large number of behavioral experiments to explore psychological phenomena. Typical experimental research is usually carried out in the setting situation of the laboratory, which is highly abstract and obviously divorced from real life, which undoubtedly reduces the external ecological effectiveness of psychological research and restricts the further development of psychology [16].

4.3. Research on Psychological Decision-Making under the Background of Big Data

4.3.1. Changes Brought by Big Data to Psychological Research. The popularity of the Internet is accompanied by the explosive growth of data. In order to analyze the data and get useful information, it is necessary to adopt scientific collection methods and strong technical support. The arrival of big data era has created a good environment for psychological research. Psychological research includes three parts: reasonable description, accurate prediction, and effective control, which are interrelated and influence each other. For example, through the analysis of individual personality types, consumption preferences, education level, and other dimensions, the research object can be positioned. On this basis, the psychological decision-making behavior analysis of the research object can achieve accurate prediction to a certain extent. Therefore, big data has a far-reaching impact on psychological research.

4.3.2. Research on Psychological Decision-Making in the Age of Big Data. In the era of big data, network platform plays an increasingly important role in people's social interaction. As long as people search and browse on the Internet, they will leave traces, which also reflect people's psychology and behavior habits to a certain extent. Using these browsing traces of users on the network platform, we can explore the psychological characteristics and behavior habits of users more deeply. Nowadays, people's daily life and work have already been integrated with big data. The analysis and integration of various data can effectively enhance the value of data and, on this basis, promote the progress of psychological research.

At the same time, due to the popularity of the Internet, online shopping, a new shopping model, has become the mainstream and developed rapidly. Under this popular trend, merchants need to collect customers' psychology in order to better cater to customers' preferences and get better development. Users will leave purchase records when purchasing related products and gradually form a large amount of data. Through the analysis of the data, we can understand the customer's purchasing psychology, so as to effectively screen and recommend similar products to customers. Therefore, in the era of big data, merchants can obtain consumers' purchase psychoanalysis, which provides convenience for meeting customers' needs, while users can enjoy more comprehensive services and get better consumption experience. Relying on big data technology can realize the accuracy of psychological decision analysis, at the same time, it also enhances the value of data and gives full play to the utility of data.

Generally, the amount of data recorded by individual data is less, and its role is limited. However, if a large number of individuals and groups join, the amount of data generated will be difficult to estimate, and the value of data will be greatly improved. Research and analyze big data, and master the cognitive and psychological situation of the public. Then, when dealing with group events, we can understand the

changes of group emotions through a certain amount of data analysis, realize effective control of emotions, grasp the trend of events, play a harmonious role, and maximize the occurrence of vicious events.

In the era of big data, we should give full play to the role of psychological interference and combine existing research and reports to deeply understand the role of psychological intervention in psychological research. The traditional psychological research mode cannot expand the psychological intervention function to the maximum extent. The emergence of big data technology not only helps psychological research solve this problem, but also expands in other aspects, and the speed of information acquisition is guaranteed. Psychological intervention has been widely popularized, which promotes the process of psychological research and lays a solid foundation for its progress [17].

In the Internet information age, science and technology are still making continuous progress, which is closely linked with people's lives. Psychological research is related to big data, which is the inevitable result of the development of the times. In the era of big data, emerging technologies are mainly related to psychology, and corresponding information can be learned through data analysis in different situations and states, thus promoting the progress of psychoanalysis and solving the problems in society.

Researchers' research driven by big data can dig deeper into the changes in consumers' minds. Especially when the macroeconomic development is favorable, the proportion of consumers' psychological investment increases, and the economy can be further increased. On the contrary, when the economic development is not good, the proportion of consumers' investment decreases and the economic growth is slow. Big data background can analyze the changes of consumers' economic activities, control the changes of consumers' hearts, and benefit the needs of economic development.

5. Conclusion

This paper expounds the advantages of big data technology in macroeconomic forecasting and analysis and introduces the principle on this basis. Finally, through an application example, it is found that big data technology can significantly improve the accuracy and effectiveness of macroeconomic forecasting. The change of big data not only enables psychologists to find a new research method and data source but also enables us to rethink about the research object, purpose, and theoretical system of psychology. Psychological research is related to big data, which is the inevitable result of the development of the times. The advanced nature of big data technology can help psychological research break the tradition and adopt a more emerging model for research.

Data Availability

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

Conflicts of Interest

The author declared that there are no conflicts of interest regarding the publication of this work.

References

- [1] C. Duarte, P. M. M. Rodrigues, and A. Rua, "A mixed frequency approach to the forecasting of private consumption with ATM/POS data," *International Journal of Forecasting*, vol. 33, no. 1, pp. 61–75, 2017.
- [2] V. Murdock, C. L. A. Clarke, J. Kamps, and J. Karlgren, "Report on the workshop on search and exploration of x-rated information (SEXI 2013)," *Acm Sigir Forum*, vol. 47, no. 1, pp. 31–37, 2013.
- [3] M. Zhang and R. Qi, "Data mining and economic forecasting in DW-based economical decision support system[J]," *International Journal of Reasoning-Based Intelligent Systems*, vol. 11, no. 4, pp. 300–307, 2019.
- [4] G. Xie, Y. Qian, and S. Wang, "Forecasting Chinese cruise tourism demand with big data: an optimized machine learning approach," *Tourism Management*, vol. 82, Article ID 104208, 2021.
- [5] X. Wei, W. Chen, and X. Li, "Exploring the financial indicators to improve the pattern recognition of economic data based on machine learning," *Neural Computing and Applications*, vol. 33, no. 2, pp. 723–737, 2020.
- [6] W. Li, C. Gong, and J. Li, "Research on some problems before decision making of party committee of state-owned enterprises based on data model," *Journal of Physics: Conference Series*, vol. 1744, no. 4, Article ID 042014, 2021.
- [7] W. Alschner, J. Pauwelyn, and S. Puig, "The data-driven future of international economic law," *Journal of International Economic Law*, vol. 20, no. 2, pp. 217–231, 2017.
- [8] T. Ming-Hong, V. M. Nadhilla and B. H. Verlin, "The effects of perceived decision-making styles on evaluations of openness and competence that elicit collaboration," *Personality and Social Psychology Bulletin*, vol. 46, no. 1, pp. 124–139, 2019.
- [9] J. Chen, "Data analysis and knowledge discovery in web recruitment—based on big data related jobs," in *Proceedings of the 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, pp. 142–146, Taiyuan, China, November 2019.
- [10] P. M. E. Garboden, "Sources and types of big data for macroeconomic forecasting," *Macroeconomic Forecasting in the Era of Big Data*, vol. 52, pp. 3–23, 2020.
- [11] M. Neupauerová and J. Vravec, "Monetary strategies from the perspective of intermediate objectives," *Panoeconomicus*, vol. 54, no. 2, pp. 219–233, 2007.
- [12] M. Camacho, G. Perez-Quiros, and P. Poncela, "Green shoots in the euro area: a real time measure," *Ssrn Electronic Journal*, no. 1026, pp. 9–41, 2010.
- [13] G. T. Wodtke, "Are smart people less racist? Verbal ability, anti-black prejudice, and the principle-policy paradox," *Social Problems*, vol. 63, no. 1, 2016.
- [14] B. S. Paye, "Déjà vol': predictive regressions for aggregate stock market volatility using macroeconomic variables," *Journal of Financial Economics*, vol. 106, no. 3, pp. 527–546, 2012.
- [15] S. E. Tolwinski-Ward, M. P. Tingley, M. K. Evans, M. K. Hughes, and D. W. Nychka, "Probabilistic reconstructions of local temperature and soil moisture from tree-ring data with potentially time-varying climatic response," *Climate Dynamics*, vol. 44, no. 3–4, pp. 791–806, 2015.

- [16] L. . Harding, "Children's quality of life assessments: a review of generic and health related quality of life measures completed by children and adolescents," *Clinical Psychology and Psychotherapy*, vol. 8, no. 2, pp. 79–96, 2010.
- [17] E. Ferrer and J. J. Mcardle, "Longitudinal modeling of developmental changes in psychological research," *Current Directions in Psychological Science*, vol. 19, no. 3, pp. 149–154, 2010.