Demand Analysis of Science and Technology Talents Based on Time Series - BP Neural Network Model

Jing Luo, Jingwen Qu*

School of Economics and Management Xi'an Shiyou University Xi'an, China
School of Economics and Management Xi'an Shiyou University Xi'an, China
**JingwenQu1122@163.com*

Abstract

Using Eviews7 and SPSS25, Granger causality test and stepwise regression analysis were carried out on the statistical data of *China Statistical Yearbook*, *Shaanxi Statistical Yearbook* and *Xi'an Statistical Yearbook* from 2010 to 2020. On this basis, a time series-BP neural network combined prediction model was constructed, and MATLAB software was used to train BP neural network for relevant data. Accordingly, the demand for scientific and technological talents in Shaanxi Province from 2021 to 2025 was predicted. The following conclusions were drawn: the total output value of industrial enterprises in Shaanxi Province can effectively predict the demand for scientific and technological talents; compared with the GM(1,1) model, the time series model has higher prediction accuracy for the gross industrial output value of industrial enterprises on the specification; the demand for science and technology talents in Shaanxi Province is estimated to increase exponentially.

Keywords- *Science and Technology Talents; Demand Analysis; Granger Causality Test; GM(1,1) Model; Time Series Model; BP Neural Network Model*

1. INTRODUCTION

In the "14th Five-Year" plan proposal adopted by the Fifth Plenary Session of the 19th Central Committee, it is stated that the intellectual resources of a country or a competitive entity in the process of accelerated catch-up, especially the innovative intellectual capital formed by the high level of top talents, the overall quality of talents and the vitality of talent development, have an efficiency multiplier effect on its strategic catch-up. "Recommendation" points out that there are many problems in the talent development planning of some local and sectoral areas, such as low strategy, weak leadership, insufficient systematicness, and insufficient precision[1].Taking Shaanxi Province as an example, although there are abundant talent reserves and natural resources, there are still problems such as unreasonable scientific and technological talent structure and unbalanced regional economic development. According to the relevant data of Shaanxi Statistical Yearbook and Xi'an Statistical Yearbook (2017–2020), the stock of R&D personnel of industrial enterprises in Shaanxi Province is 70832, 70156 and 56926. Among them, the proportion of Xi'an is higher than 50%, up to 52.05%, 56.95% and 60.62%. The proportion of R&D expenditure of industrial enterprises in Xi'an is also up to about 55%, and there is a phenomenon of "one city dominates" in Xi'an.

Based on the current situation of the development of scientific and technological talents in Shaanxi Province, Eviews7 and SPSS25 are used to screen out the prediction indexes of the demand for scientific and technological talents through Granger test and stepwise regression analysis. The time series-BP neural network combination model is constructed, and MATLAB is used to train the BP neural network for the relevant data. Then, the demand for scientific and technological talents in Shaanxi Province from 2021 to 2025 is predicted, which provides new ideas for the scientific method of the demand prediction of scientific and technological talents.

2. A SURVEY OF TALENT DEMAND FORECASTING

Since 2012, the research methods of "talent demand forecasting" published in high-level journals can be divided into two types: single model forecasting and combined model forecasting. The single prediction

G. Vilas Bhau et al. (Eds.): MSEA 2022, ACSR 101, pp. 248–254, 2023.
https://doi.org/10.2991/978-94-6463-042-8_37

model mainly includes GM(1,1) model[2], differential autoregressive moving average(ARIMA) model[3], trend extrapolation prediction[4] and BP neural network model[5]. Combination models include correlation analysis-regression prediction model[6], ARIMA-LSSVM model[7], GM(1,1)-time series combination model[8] and multiple regression-grey prediction combination model[9].

The research shows that talent demand is affected by economy, city, policy and so on. It is not just a simple linear change. The single prediction model may have shortcomings. The exponential smoothing method of quadratic curve is one of the time series models. The accuracy of the predicted value of talent demand change only considering the time factor needs to be discussed. The grey forecasting model is only suitable for exponential growth forecasting and cannot meet the unstable change of talent demand caused by policy changes. Therefore, in order to fully combine the advantages of various single prediction models, this study explores the feasibility and effectiveness of the combined prediction model composed of grey GM (1,1) model, time series model and BP neural network for talent demand prediction on the basis of Granger test and stepwise regression analysis. Data sources and research methods

2.1. Data sources

The historical data of R&D personnel stock (people) in Shaanxi Province in *China Statistical Yearbook*, *Shaanxi Statistical Yearbook* and *Xi'an Statistical Yearbook* from 2010 to 2020 are selected to predict the demand for scientific and technological talents in Shaanxi Province. Because the economy, people's lives, social development and scientific and technological activities and other factors will have an impact on the stock of scientific and technological talents, so select the GDP of Shaanxi Province(billion yuan), the added value of the secondary industry(billion yuan), R&D investment(billion yuan) and technology market turnover (million yuan) and other indicators as alternative indicators for the demand forecast of scientific and technological talents in Shaanxi Province.

2.2. Research Method

Firstly, using Eviews7 and SPSS25 software to do Granger causality test and stepwise regression analysis of the data, the scientific and technological talent demand forecasting indicators are screened. Then, GM(1,1) model and time series model are used to predict the trend of "science and technology talent demand forecasting index" in five years, and the model with better prediction effect is selected according to the RMSE of the two models. Finally, MATLAB is used to train the relevant data by BP neural network to obtain the predicted value of the demand for scientific and technological talents in Shaanxi Province from 2021 to 2025.

3. Selection of Forecasting Indexes for Science and Technology Talent Demand

The indicators to be screened include eleven items, as shown in Table 1.

TABLE 1. INDICATORS TO BE SCREENED

Category	Index
Economic Development	X0-GDP of Shaanxi Province(billion yuan), X1-added value of secondary industry of Shaanxi Province(billion yuan), X2-added value of tertiary industry of Shaanxi Province(billion yuan)
People Living	X3-Per capita disposable income of urban residents in Shaanxi Province(Yuan)
Social Development	X4-Investment in fixed assets of the whole society in Shaanxi Province(billion yuan), X5-Gross industrial output of industrial enterprises in Shaanxi Province(billion yuan), X6-Gross construction output of Shaanxi Province(million yuan)
Sci-tech Activities	Y-Shaanxi Province R&D personnel stock (people), X7-Shaanxi Province R&D investment(billion yuan), X8-Shaanxi Province technology market turnover(million yuan), X9-Shaanxi Province R&D expenditure accounted for the proportion of GDP(%)

3.1. Unit Root Test

Granger causality test requires that the time series have stability, ADF stability test is needed, Eviews7 is used to treat the screening index for unit root test. Table 2 shows that the P of eleven indicators are all less than 0.1, where X5 and X9 reject the null hypothesis at the level of 10%, X0 –X3 reject the null hypothesis at the level of 5%, and their indexes reject the null hypothesis at the level of 1%, all the index sequences are stable without unit root.

TABLE 2. UNIT ROOT TEST RESULTS

Index	ADF Test		
	T Statistical	*p*	*Result*
Y	-3.550477	0.0045	stable
X0	-2.339874	0.0295	stable
X1	-2.219460	0.0346	stable
X2	-3.540076	0.0424	stable
X3	-2.449725	0.0229	stable
X4	-3.316011	0.0062	stable
X5	-3.679551	0.0920	stable
X6	-8.009899	0.0060	stable
X7	-3.001495	0.0089	stable
X8	-12.62910	0.0005	stable

Index	ADF Test		
	T Statistical	*p*	*Result*
X9	-4.139165	0.0568	stable

3.2. Granger Causality Test

Granger causality test was performed on the stable data. The results (Table 3) showed that X0, X2 and X5 were unidirectional Granger causes of Y. Y is the one-way Granger cause of X4 and X9; X1, X3, X6, X7 and X8 do not constitute a causal relationship with Y. The GDP of Shaanxi Province, the added value of the tertiary industry and the total industrial output of industrial enterprises will have an impact on the stock of R&D personnel.

TABLE 3. GRANGER CAUSALITY TEST RESULTS

Null Hypothesis	*F*	*P*	*Whether to reject the original hypothesis*
X0 is not Granger cause of Y	6.89262	0.0756	reject
Y is not Granger cause of X0	0.39619	0.7036	accept
X1 is not Granger cause of Y	3.65372	0.1570	accept
Y is not Granger cause of X1	0.00767	0.9924	accept
X2 is not Granger cause of Y	3.86453	0.0969	reject
Y is not Granger cause of X2	0.37447	0.5630	accept
X3 is not Granger cause of Y	1.68672	0.3229	accept
Y is not Granger cause of X3	1.25603	0.4015	accept
X4 is not Granger cause of Y	1.72888	0.3166	accept
Y is not Granger cause of X4	7.60697	0.0668	reject
X5 is not Granger cause of Y	8.16626	0.0611	reject
Y is not Granger cause of X5	0.89207	0.4966	accept
X6 is not Granger cause of Y	2.17158	0.2611	accept
Y is not Granger cause of X6	1.43411	0.3655	accept
X7 is not Granger cause of Y	2.59139	0.2220	accept
Y is not Granger cause of X7	1.36114	0.3796	accept
X8 is not Granger cause of Y	3.03017	0.1905	accept
Y is not Granger cause of X8	3.26935	0.1764	accept
X9 is not Granger cause of Y	1.60575	0.3357	accept
Y is not Granger cause of X9	8.94262	0.0544	accept

3.3. Regression Analysis

Taking the annual Y in Shaanxi Province from 2010 to 2019 as the dependent variable, according to the above Granger test, X0, X2 and X5 in Shaanxi Province were selected as independent variables, and SPSS25 was used to carry out stepwise and linear regression analysis.

3.3.1. Stepwise Regression Analysis

Stepwise regression analysis is to use the regression model to eliminate the variables that have no obvious influence on the dependent variable and are highly correlated with the dependent variable, which can reduce the multicollinearity as much as possible. Table 4 shows the results of stepwise regression model fitting, R^2 is 0.907, indicating that the model has high explanatory power. That is to say, the total X5 has 90.7% explanatory power for the change of Y in Shaanxi Province.

TABLE 4. MODEL SUMMARY

Model	*R*	*R^2*	*Adjusted R^2*	Errors in standard estimation
1	0.952^a	0.907	0.895	6981.671

According to Table 5, the VIF values of X0 and X2 are greater than 10, that is, there is a collinearity, which is eliminated in the regression model.

TABLE 5. EXCLUDED VARIABLES

	Enter Beta	*t*	Sig.
X0	-0.390^b	-0.506	0.629
X2	-0.149^b	-0.340	0.743

	Partial Correlation	Collinearity Statistics		
		Tolerance	*VIF*	*Minimum Tolerance*
X0	-0.188	0.022	46.222	0.022
X2	-0.128	0.068	14.648	0.068

3.3.2. Linear Regression Analysis

According to the results of stepwise regression analysis, Y is selected as the dependent variable, and X5 is selected as the independent variable for linear regression analysis(Table 6). The following regression equation is obtained :

$$Y=47832.364+4.255*X5 \quad (1)$$

The model R^2 is 0.907, X5 can explain the change of 90.7% of Y. Table 6 shows that the model can pass the F test(F=77.808, p=0.000<0.05), and the regression coefficient of X5 is 4.255(t=8.821, p=0.000<0.05), all X5 will have a significant positive impact on Y.

TABLE 6. RESULTS OF LINEAR REGRESSION ANALYSIS(N=10)

	Enter Beta		Standardized Coefficient		t	p
	B	*STDERR*	*Beta*			
constant	47832.364	9860.109	-		4.851	0.001
X5	5.255	0.482	0.952		8.821	0.000
	F		R^2	Adjusted R^2	VIF	
constant	F（1,8）=77.805，p=0.000		0.907	0.895	-	
X5					1.000	

4. SCIENCE AND TECHNOLOGY TALENT DEMAND FORECAST

This study constructs a combination model to predict the demand for scientific and technological talents in Shaanxi Province. Firstly, GM(1,1) model and time series model are used to predict X5 from 2021 to 2025, and the optimal prediction model is selected according to RMSE. Then, after the training of BP neural network model, the predicted value of Y from 2021 to 2025 is obtained.

4.1. Prediction X5

4.1.1. GM(1,1) Model

GM(1,1) model is suitable for short-term and medium-term prediction of time series data with less data such as X5, and does not consider the distribution law or change trend.

4.1.1.1. Level ratio test:

The level ratio test is a preliminary test on whether the original data sequence is suitable for the model construction. The test results are shown in Table 7. According to the formula (e^(−2/(n+1)),e^(2/(n+1))), the standard range of the level ratio of the model is calculated, so as to determine that the original data do not pass the level ratio test. On the basis of the original value, the translation conversion value 26623.00 is added to carry out, and the level ratio test values of the data are all within the standard range [0.834,1.199]. This data is suitable for the construction of GM(1,1) model.

TABLE 7. GM(1,1) MODEL LEVEL RATIO

NO.	1	2	3	4	5
level ratio λ	-	0.925	0.939	0.955	0.978
level ratio after conversion λ	-	0.925	0.939	0.955	0.978
NO.	**6**	**7**	**8**	**9**	**10**
level ratio λ	0.993	0.969	0.961	0.974	0.973
level ratio after conversion λ	0.993	0.969	0.961	0.974	0.973

4.1.1.2. Posterior difference ratio test:

Table 8 shows that the model calculation development coefficient a=−0.0299 is less than 0.3, which is suitable for medium-and long-term prediction, and the posterior error ratio C calculated by the model is 0.0147<0.35, so the model accuracy level is very good.

TABLE 8. MODEL CONSTRUCTION RESULTS

a	b	C
-0.0299	40263.6423	0.0147

4.1.1.3. Model fitting and prediction:

After building the model, the fitting value of the model and the predicted value of the latest 12 periods are obtained. Since the grey prediction model is more suitable for medium and short-term prediction, the predicted value of the last 6 periods(Table 9) is selected, that is, the predicted value of X5 in 2021-2025 by GM(1,1) model is 30070.747, 31794.186, 33570.016, 35399.830 and 37285.269 billion yuan. The fitting figure(Fig.1) shows that the fitting effect of the original data and the predicted values is good.

TABLE 9. PREDICTIONS

NO.	Original Value	Predictions
1	11199.840	11199.840
2	14283.480	15399.358
3	16926.490	16676.800
4	18982.470	17992.075
5	20015.880	19349.364
6	20333.980	20746.883
7	21837.610	22186.885
8	23825.180	23670.662
9	25192.360	25199.545
10	26623.110	26774.904
backward 1	-	28398.153
backward 2	-	30070.747
backward 3	-	31794.186
backward 4	-	33570.016
backward 5	-	35399.830
backward 6	-	37285.269

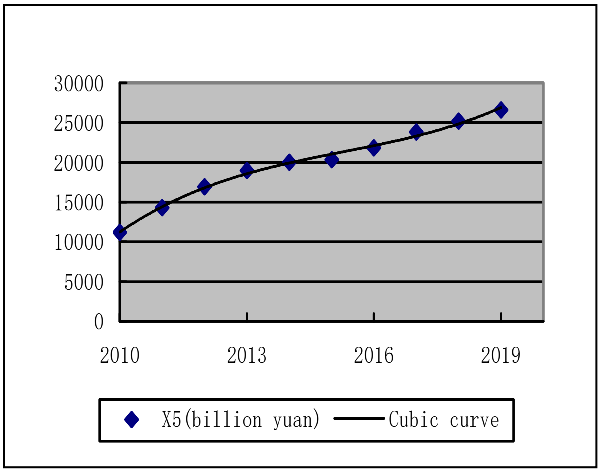

Figure 1. MODEL FITTING DIAGRAM

4.1.1.4. Model residual test:

The relative error value is calculated by the ratio of the absolute value of the residual value to the original value. When the relative error value is less than 20%, the fitting is proved to be good. The maximum relative error value of the model is 7.812%<20%, indicating that the fitting effect meets the high requirements. In general, when the level ratio deviation is less than 0.2, it meets the requirement. The maximum level ratio deviation of the model is 0.192⩽0.2, which meets the requirement of the model fitting effect. The model is tested by model residuals, as shown in Table 10.

TABLE 10. GM (1,1) MODEL TEST

NO.	Residual Error	Relative Error	Class Ratio Dispersion
1	0.000	0.000%	-
2	-1115.878	7.812%	0.192
3	249.690	1.475%	0.130
4	989.395	5.212%	0.081
5	666.516	3.330%	0.023
6	-412.903	2.031%	-0.014
7	-349.275	1.599%	0.041
8	154.518	0.649%	0.056
9	-7.185	0.029%	0.026
10	-151.794	0.570%	0.025

4.1.2. Time Sequence Model

The scatter plot of X5 data from 2010 to 2019(Fig.2) shows an overall upward trend in the decade since 2010.

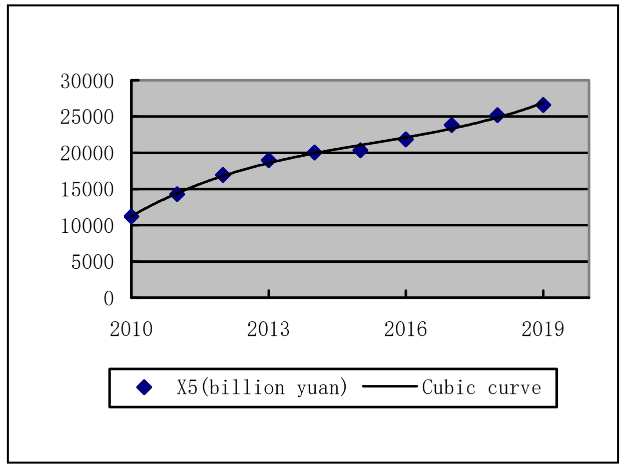

Figure 2. 2010-2019 X5

SPSS25 was used to fit the change trend of each point, and the maximum fitting value R^2 was 0.994, indicating that the time series was more in line with the cubic curve growth trend (Fig.2), which was suitable for time series model analysis. The data used in this paper were annual data, without seasonal analysis, and the traditional time series prediction model was constructed. The expert modeler was used to automatically select the optimal prediction model as the Brown model. The prediction results are shown in Table 11.

TABLE 11. 2021-2025 X5 FORECAST

Date	2021	2022	2023	2024	2025
X5	29484.600	30915.345	32346.090	33776.835	35207.580

4.1.3. Optimal Model Selection

In this paper, GM(1,1) model and time series model are used to predict X5 from 2021 to 2025. In order to ensure the optimal prediction effect, the RMSE of the two prediction models are compared. The time series model's RMSE=676.99 less than that of the GM(1,1) model=1752.177. According to the principle that the smaller the RMSE is, the better the prediction model is, the prediction results of the time series model are selected.

4.2. BP Neural Network Model Prediction Analysis of Science and Technology Talent Demand

BP neural network model is a common mathematical model. Information processing is carried out by constructing a synaptic connection structure similar to the brain nerve, and the processing process includes the input layer, the hidden layer and the output layer. In the BP neural network model constructed in this paper, the input layer is X5, and the output layer is Y. The training results show that when the hidden layer is 11, the network training MSE=0.047128, and the prediction effect is the best. The model is shown in Fig.3.

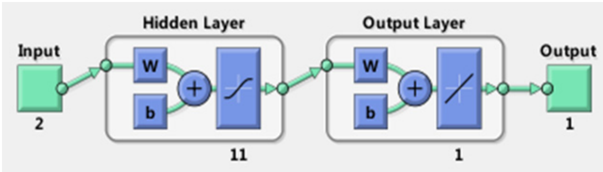

Figure 3. BP NEURAL NETWORK MODEL DIAGRAM

The BP neural network cannot restore 100% of the same results. Therefore, it is necessary to train the input data for many times and take the average of multiple training results as the final prediction value. Ten better training results are selected and the average is taken as the final result of the Y prediction value from 2021 to 2025, as shown in Table 12.

TABLE 12. TRAINING RESULTS AND AVERAGE VALUES

Date	2021	2022	2023	2024	2025
1	193820	196230	197870	198790	199010
2	193740	200660	206840	212610	217390
3	178610	188920	201210	214690	228290
4	204100	206830	207300	207400	207440
5	173950	175230	177220	180020	183660
6	196580	212940	228020	244320	262250
7	219270	223020	223890	224100	224170
8	202060	208540	220750	247520	279760
9	173790	175530	178830	183140	187410
10	178180	178820	179100	179370	179750
Average	191410	196672	202103	209196	216913

5. CONCLUSION

The total output value of industrial enterprises(X5) can effectively predict the demand for scientific and technological talents in Shaanxi Province. Through the Granger test, it is concluded that the Granger cause of the stock of R&D personnel in Shaanxi Province is GDP(X0), the added value of the tertiary industry(X2) and the total output value of industrial enterprises(X5). By stepwise regression analysis, the GDP and the added value of the tertiary industry in Shaanxi Province are excluded, and the total output value of industrial enterprises is better predicted.

Compared with the GM(1,1) model, the time series model has higher prediction accuracy for the total industrial output value(X5) of industrial enterprises on the specification. After screening the correlation between economic indicators such as GDP and the demand for scientific and technological talents in Shaanxi Province, X5 can explain 90.7% of the change in scientific and technological talents. GM(1,1) model and time series model were used to predict X5, respectively. According to the RMSE of the model, the time series model was selected to effectively improve the prediction accuracy.

REFERENCES

[1] Fifth Plenary Session of the Nineteenth Central Committee. “Fourteenth Five-Year” Talent Development Plan : Layout of a New System for High-quality Talent Development [EB/OL].[2021-03-22]

[2] Yao Juan, Liu Hongyuan, Liu Jianxian. Research on Regional Demand Trend of Scientific and Technological Innovation Talents - Forecast and Comparative Analysis of Sichuan, Shaanxi and Shanghai[J]. Scientific and Technological Progress and Countermeasures, 2019,36 (14) : 46-52.

[3] Li Xiao, Zhang Qinghua. Demand Forecast of Creative Industry Talents in Heilongjiang Based on ARIMA Model[J]. Science and Technology Management Research, 2016,36(18):122-125.

[4] Zhao Dongxu. Demand Forecast and Training Countermeasures of English Tourism Talents in Jilin Province Based on Trend Extrapolation[J]. Vocational and Technical Education, 2015,36(20): 13-16.

[5] Yang Junsheng, Xue Yongjun. The prediction of talent demand trend in ASEAN Free Trade Zone based on BP artificial neural network model—and the countermeasures of Yunnan Province[J]. Academic Exploration, 2014(04):83–87.

[6] Li Xiyuan, Wang Yanjiao. Research on overseas high-level talent demand of key industries-Taking automobile and steel industry in Hubei Province as an example[J]. Science and technology progress and countermeasures, 2014,31(08):135-139.

[7] Yang Huanhai. Modeling and Simulation Analysis of Talent Demand Combination Forecast[J]. Computer Simulation, 2013,30(10):253-256.

[8] Qu Qunzhen, Wang Jiaji, Tang Mengxue, Niu Ping. The demand forecast of Chinese scientific and technological talents during the 14 th Five-Year Plan based on the combination model[J]. Science and technology management research, 2021,41(21):129-135.

[9] Tan Kai, Wang Wensheng, Zhang Li, Liang Yi. Demand Forecast of Talents in Coal Industry Based on Multiple Regression-Grey Forecast Combination Method[J]. Coal Engineering, 2019,51(03):151-154.