

## Prediction of the Development of Science and Technology Service Industry Based on Multiple Linear Regression Analysis

Enyuan Zhu<sup>1</sup>, Fanhui Yan<sup>2</sup> and Keliang Jia<sup>2\*</sup>

<sup>1</sup> Shandong Research Center of Science and Technology Statistics and Analysis, Shandong, China

<sup>2</sup> School of Management Science and Engineering, Shandong University of Finance and Economics, Shandong, China

enyuanzhu@sina.com, 944617092@qq.com, jiakeliang@sina.com

**Keywords:** Science and Technology Service Industry; Development Level; Forecast Model; Shandong Province

**Abstract.** In order to enhance the development of science and technology service industry in Shandong Province, the paper established a statistical index system of science and technology service industry based on the relevant literature and the actual data analysis. And then the paper used multiple linear regression model to predict the development of science & technology service industry in Shandong province. Finally the paper collected the data set and analyzed them through SPSS and putted forward to some policy suggestions to promote its development.

### Introduction

As an indispensable part of national economy, science and technology service industry plays an increasingly important role in realizing economic growth. Therefore, it is of theoretical and practical significance to forecast the development of science and technology service industry. So many scholars focused on the prediction of the development of science and technology service industry.

Mo Li-ping, Lu Jun, Mei Yun, Zhu Jiang-jie summed up the development trend of Science and Technology Intermediary Service Industry in Guiyang, and effectively predicted its regional layout [1]. Dai Teng-hui, Tan Yuan-peng analyzed the relevant data of the present situation of the Production Service Industry and constructed a random rights network model based on the random rights network, and then predicted the indexes of the development of producer services in China in the next five years [2]. Zhou Hong, Song Chen selected sixteen specific indexes which reflected the development level of Tianjin Science and Technology Service Industry, and established a forecasting model of the development level of Tianjin Science and Technology Service Industry finally [3].

Many scholars used multiple linear regression analysis model to carry out prediction research. Gu Nai-hua obtained the regression model of influencing factors,  $VSGDP = \beta_0 + \beta_1 Y_0 + \beta_2 Location + \beta_3 Policy + \epsilon$ , and analyzed the differences of service industry development among provinces in China [4]. Zhang Ai-xia, Zhu Ming, Zhao Liang used the regression analysis method to forecast the trend of the casualty accident [5]. On the basis of analyzing the current situation of economic development, Jing Bing-jie established a mathematical model to reflect its relationship, and then calculated the trend of forecasting variables, so as to predict the future economic development trend [6]. Tang Shu-an, Zheng Ze-zhong, Zhu Xue-bo elaborated the principles and ideas of the multiple regression analysis method, and predicted the productivity of the oilfield by this method. [7].

Therefore, the paper aimed to use the time series data of Shandong Province to establish multiple regression model, and used regression analysis method to forecast the development of science and technology service industry, so as to promote the development of science and technology service industry.

**Research Methods**

Multivariate regression analysis is a mathematical statistical method for dealing with the statistical correlation between a dependent variable and several independent variables. The basic idea is to find the mathematical expression that best represents the relationship between the independent variable and the dependent variable. The main steps of the regression analysis are as follows:

**Model Establishment.** When considering the relationship between multiple independent variables and dependent variables, a multiple linear regression model can be established as  $y = b_0 + b_1X_1 + b_2X_2 + \dots + b_iX_i + \dots + b_kX_k$ .

**Reliability Test and Correlation Test.** Cronbach coefficient  $\alpha = K / (K - 1) (1 - \sum_{i=1}^K \sigma_{y_i}^2 / \sigma_x^2)$ , Pearson correlation coefficient  $r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$ .

**Calculate Regression Coefficient.** Calculate the coefficient  $b_1, b_2, \dots, b_i, \dots, b_k$  and  $b_0$ .

**Model Test.** Fitting degree test  $R^2 = RSS / TSS = 1 - ESS / TSS$ , Adjusted fitting degree test  $R^2 = 1 - \frac{ESS / (n - k - 1)}{TSS / (n - 1)}$ . F test  $F = S_1^2 / S_2^2$ , T test  $t = (\bar{X} - \mu) / \frac{\sigma_x}{\sqrt{n - 1}}$ .

**Economic Forecast.** If the regression equation established is significant, it can be used to predict what you are studying.

**Data Collection**

Based on the design principles of comprehensive, scientific, systematic, hierarchical and dynamic, the following 16 indexes are selected, as shown in Table 1:

Table 1 The Index System of Science and Technology Service Industry in Shandong Province

Variable Type	Variable	Variable Definitions	Unit
Independent Variable	X <sub>1</sub>	Number of Employees in Science and Technology Service Industry	Million People
	X <sub>2</sub>	R & D Personnel (Equivalent) Full Time Equivalent	People
	X <sub>3</sub>	Number of Domestic Patent Applications	Item
	X <sub>4</sub>	Investment in Fixed Assets of Science and Technology Service Industry	Billion
	X <sub>5</sub>	Investment in Fixed Assets of High - tech Industry	Billion
	X <sub>6</sub>	The Average Wage of Workers in the Service Industry	Yuan
	X <sub>7</sub>	The Proportion of Regional Science and Technology Financial Allocations to Fiscal Expenditure	%
	X <sub>8</sub>	Research and Development Funds (R & D) Expenditure as a Proportion of GDP	%
	X <sub>9</sub>	The Number of Contracts for the Technology Market	Item
	X <sub>10</sub>	Value Added of Technology Services	Billion
	X <sub>11</sub>	Per Capita GDP	Yuan
	X <sub>12</sub>	Local Financial Science and Technology Funding Expenditure	Billion
	X <sub>13</sub>	High - tech Industry Sales of New Products	Million
	X <sub>14</sub>	High - tech Industry Output Value	Billion
	X <sub>15</sub>	Number of Scientific and Technological Achievements	Item
Dependent Variable	Y	Annual Technical Market Turnover	Billion

The paper obtained the relevant sample data of Science and Technology Service Industry in Shandong province from 2005 to 2013. The data was from "Statistical Yearbook of Shandong Province" (2005-2013), "Shandong Science and Technology Statistics Handbook" (2005-2013), "China Labor Statistical Yearbook" (2005-2013), "Shandong Industrial Statistical Yearbook"(2005-2013).

## Empirical Analysis and Evaluation of Results

**Reliability Test.** Cronbach's alpha or Cronbach's  $\alpha$  is the most commonly used statistics of reliability measurement. Normally, the Cronbach's alpha value is required to be 0.7 or more. In the paper, by selecting analysis - measurement - reliability analysis, the Kronbach coefficient based on the normalization term was 0.972, there were 16 terms. It was believed that the consistency reliability was good and the reliability of the test was relatively high.

**Relevance Test.** The correlation test was performed on the variables. A correlation coefficient greater than 0.8 is highly correlated. Correlation coefficients between 0.5 and 0.8 are moderate correlation. Correlation coefficients less than 0.5 are low correlation. The results were shown in Table 2 below:

Table 2 The Results of Data Correlation Test

	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	X <sub>14</sub>	X <sub>15</sub>
Pearson Correlation	1	0.979* *	0.983 **	0.984 **	0.970 **	0.979 **	0.988 **	0.562	0.985 **	-0.172	0.986 **	0.991 **	0.988 **	0.987 **	0.880 **	0.051
Significant (Bilateral)		0.000	0.000	0.000	0.000	0.000	0.000	0.115	0.000	0.658	0.000	0.000	0.000	0.000	0.002	0.896

As shown in Table 2, the correlation coefficient of the index which greater than 0.8 are: X<sub>1</sub>(number of employees in science and technology service industry), X<sub>2</sub> (R & D personnel (equivalent) full time equivalent), X<sub>3</sub> (number of domestic patent applications), X<sub>4</sub>(investment in fixed assets of Science and Technology Service Industry), X<sub>5</sub>(investment in fixed assets of High - tech Industry), X<sub>6</sub> (the average wage of workers in the Service Industry), X<sub>8</sub>(research and development funds (R & D) expenditure as a proportion of GDP), X<sub>10</sub> (value added of Technology Services), X<sub>11</sub>(per capita GDP), X<sub>12</sub>(local financial science and technology funding expenditure), X<sub>13</sub>(High - tech Industry sales of new products), X<sub>14</sub>(High - tech Industry output value), indicating that these factors and the annual technical market turnover was significantly positive correlation.

**Multiple Regression Model → Stepwise Method.** First, opened the SPSS software, inserted the sample of science and technology service industry of Shandong Province. Clicked Analyze - Regression - Linear, and selected the stepwise method. And then selected the estimates in the statistics, the model fit, the Durbin-Watson, and the case diagnosis. Then selected the normal probability graph in the drawing options. And then in the save option, selected the non-standardized, Cook distance, mean, including covariance matrix. Finally selected the probability of F in the options, the constants in the equation, and the number excluded by the list [8].

In the paper, the probability of F-to-enter was less than 0.050, the probability of F-to-remove was more than 0.100. The first variable which entered the linear regression model was X<sub>11</sub>, followed by X<sub>5</sub> X<sub>6</sub>, and then X<sub>11</sub> was removed, and then X<sub>4</sub> and X<sub>7</sub> entered. Finally the optimal regression model was  $Y = -879.273 - 2.816X_4 + 3.215X_5 + 0.093X_6 - 197.272X_7$ , more details in document [9].

The overall model parameters were shown in Table 3:

**Table 3 Model Overall Parameters**

Model Summary					
Model	R	R Square	Adjusted R Square	Standard Estimation Error	Durbin-Watson
1	0.991a	.983	.980	342.48410	
2	.998b	.997	.996	159.74493	
3	.999c	.999	.998	103.05487	
4	.999d	.999	.998	100.20643	
5	1.000e	1.000	1.000	46.11678	
6	1.000f	1.000	1.000	29.78863	2.739
a. Predictive Variables: (Constant), VAR00011.					
b. Predictive Variables: (Constant), VAR00011, VAR00005.					
c. Predictive Variables: (Constant), VAR00011, VAR00005, VAR00006.					
d. Predictive Variables: (Constant), VAR00005, VAR00006.					
e. Predictive Variables: (Constant), VAR00005, VAR00006, VAR00004.					
f. Predictive Variables: (Constant), VAR00005, VAR00006, VAR00004, VAR00007.					
g. Dependent Variable: VAR00014					

It could be seen from Table 3, the adjusted goodness coefficient was 1.000, it meant that the equation had a strong ability to explain, the degree of fitting was very good. The Durbin-Watson statistic was 2.739, which showed that the residuals were normal distribution[10].

The regression variance was shown in table 4:

**Table 4 Regression Analysis of Variance**

Coefficient <sup>a</sup>						
Model		Non - normalized Coefficients		Standard Coefficient		
		B	Standard Error	Trial Version	t	Sig.
1	(Constant)	-1683.713	347.469		-4.846	0.002
	VAR00011	0.234	0.012	0.991	20.047	0.000
2	(Constant)	-680.647	254.372		-2.676	0.037
	VAR00011	0.148	0.018	0.629	8.436	0.000
	VAR00005	2.452	0.479	0.381	5.116	0.002
3	(Constant)	-1002.954	194.835		-5.148	0.004
	VAR00011	-0.055	0.067	-0.234	-0.820	0.449
	VAR00005	3.018	0.360	0.469	8.382	0.000
	VAR00006	0.112	0.037	0.781	3.069	0.028
4	(Constant)	-938.888	173.563		-5.410	0.002
	VAR00005	2.828	0.268	0.440	10.548	0.000
	VAR00006	0.083	0.006	0.575	13.788	0.000
5	(Constant)	-1324.629	112.954		-11.727	0.000
	VAR00005	3.209	0.146	0.499	21.914	0.000
	VAR00006	0.096	0.004	0.668	24.619	0.000
	VAR00004	-3.545	0.734	-0.153	-4.830	0.005
6	(Constant)	-879.273	173.686		-5.062	0.007
	VAR00005	3.215	0.095	0.500	33.980	0.000
	VAR00006	0.093	0.003	0.646	33.769	0.000
	VAR00004	-2.816	0.540	-0.121	-5.218	0.006
	VAR00007	-197.272	69.818	-0.017	-2.826	0.048
A. Dependent Variable: VAR00014						

Table 4 showed that the sig values of the four independent variables ( $X_4$ ,  $X_5$ ,  $X_6$ ,  $X_7$ ) were 0.006, 0, 0, 0.048, respectively, which were less than 0.05, indicated that the regression equation was significantly established on the whole and the regression coefficient T test value was very significant.

## Conclusion

The results of the correlation test showed that there was a significant positive correlation between the independent variable, such as  $X_1$ (number of employees in science and technology service industry),  $X_2$ (R&D personnel (equivalent) full time equivalent),  $X_3$ (number of domestic patent applications),  $X_4$ (investment in fixed assets of Science and Technology Service Industry),  $X_5$ (investment in fixed assets of High - tech Industry),  $X_6$ (the average wage of workers in the Service Industry),  $X_8$ (research and development funds (R&D) expenditure as a proportion of GDP),  $X_{10}$ (value added of Technology Services),  $X_{11}$ (per capita GDP),  $X_{12}$ (local financial science and technology funding expenditure),  $X_{13}$ (High - tech Industry sales of new products),  $X_{14}$ (High - tech Industry output value) and the dependent variable.

Multiple regression analysis methods showed that the regression equation fitted the best, which contained several variables such as  $X_4$ (fixed asset investment in science and technology services),  $X_5$ (high tech industry investment in fixed assets),  $X_6$ (the average wage of workers in the service industry of science and Technology), and  $X_7$ (financial allocation technology area accounts for the proportion of fiscal expenditure). So the paper suggested that the four factors should be strengthened in order to raise the annual technical market turnover of science and technology service industry in Shandong Province. Future work is to get the output value of science and technology service industry in Shandong Province and to construct the optimal regression model in order to raise the development of science and technology service industry.

## Acknowledgements

This work was supported by Humanity and Social Science foundation of Education Ministry of China (No. 14YJC860011), Shandong province key R & D projects (No. 2015GGB01518).

## References

- [1] Mo Li-ping, Lu Jun, Mei Yun, Zhu Jiang-jie. Development Trend and Layout Prediction of Science and Technology Intermediary Service Industry in Guiyang City[J]. Popular Science and Technology, 2009,(4): 189-190
- [2] Dai Teng-hui, Tan Yuan-peng. Research on China's Productive Service Industry Based on Random Weight Network Model and Policy Suggestions[J]. Economic Issues Exploration, 2016,(6): 94-100
- [3] Zhou Hong, Song Chen. Empirical Research on Evaluation of Developmental Level to Promote Science and Technology Service Industry——Take Tianjin as an Example[J]. Journal of Guizhou University of Finance and Economics, 2013, (4): 92-97
- [4] Gu Naihua. An Empirical Analysis of Regional Service Industries Development Difference and Its Influencing Factors in China: 1980-2000[J]. Finance and Trade Economics, 2004, (9): 84-88
- [5] Zhang Ai-xia, Zhu Ming, Zhao Liang. The Application of Regression Forecasting on Accident Forecasting[J]. Journal of Hebei Polytechnic University, 2007,29 (4): 11-13
- [6] Jing Bingjie. On the Application of Regression Analysis in Economic Prediction[J]. Journal of Shanxi Institute of Economic Management, 2004,12 (3): 32-34
- [7] Tang Shu'an, Zheng Zezhong, Zhu Xuebo. Application of Multiple Regression Analysis in oil and gas Productivity Prediction [J]. Science and Technology Innovation Herald, 2008, (7): 169-171
- [8] Bi Jianwu, Jia Jinzhang, Liu Dan. Forecast of the Gas Emission Quantity of the Working Face Based on the SPSS Multiple Regression Analysis[J]. Journal of Safety and Environment, 2013,13 (5): 183-186

- [9] Luo Fengming, Qiu Jinbiao, Li Minghua, Xiao Bingkun. How to do Regression Analysis by Statistical Software SPSS[J]. *Computer Knowledge And Technology*, 2008 ,(2): 293-304
- [10] Wei Zhijing, Liu Xiyu, Zhao Qingzhen, Jing Yuan. The Analysis Based on Statistic Software SPSS and Multiple Linear Regression Aanalysis[J]. *Information Technology and Informatization*, 2006(2):107-109