

Research on Influencing Factors of Regional Taxation Based on ANN-RBF Algorithm

Xiangpeng Huang *

College of Liberal Arts and Sciences, University of Illinois Urbana-Champaign, 61820,
Urbana-Champaign, United States

* Corresponding Author Email: huangxp2022@163.com

Abstract. Taxation is an important part of national fiscal revenue and an important means for the government to regulate the macroeconomy and maintain stable economic growth. With the development of the economy, taxation is also increasing accordingly. Studying the influencing factors of taxation can better formulate tax policies and optimize the structure of economic development. This paper developed a new method, ANN-RBF (artificial neural network-radial basis function). This paper constructs 14 index systems affecting regional fiscal taxation from four aspects: population, national economy, international trade, and investment expenditure. R^2 (coefficient of determination) and MSE (mean square error) are set as algorithm test indicators. By calculating the contribution of independent variables to taxation, the core factors affecting taxation are analyzed. Finally, the Anhui Province of China is taken as the research object. The results show that: When the number of neurons is 14, the accuracy converges to the maximum. From the contribution analysis, it can be found that the total exports, the secondary industry GDP, the tertiary industry GDP, and the total investment in fixed assets are the core indicators affecting the tax revenue of Anhui Province. By analyzing the tax influencing factors in Anhui Province, this study can provide suggestions for the government to adjust tax revenue and promote economic development.

Keywords: ANN-RBF; influencing factors; Anhui Province; contribution analysis

1. Introduction

As an important means of macroeconomic regulation and control, fiscal revenue plays an important role in both countries and regions. Rich fiscal revenues provide the basis for the government to allocate and balance public resources and help countries and regions develop better. Especially for developing countries, as the economy grows, taxes also gradually increase, and taxes will affect the economy. Therefore, studying the influencing factors of taxation can provide suggestions for the formulation and improvement of tax policies, adjust the structure of economic development, and promote a more stable and rapid economic development. In view of the important influence of fiscal taxation on the development of countries and regions, many scholars have used various methods to analyze the tax-influencing factors of different countries and regions.

Ding et al. [1] obtained that employment and fixed asset investment have an important impact on China's tax revenue through multiple regression analysis. Cai [2] obtained by constructing a multiple regression model that total retail sales, agricultural output, and technology turnover have a greater positive impact on fiscal revenue. Sun [3] used the stepwise regression method to conclude that fixed asset investment has a significant role in promoting China's fiscal revenue. Huang et al. [4] analyzed the influence of economic growth, fiscal expenditure, and commodity retail price index on the tax revenue by establishing a multiple regression model. Ni [5] analyzed the influencing factors of taxation in Shandong Province through the adaptive Lasso method and concluded that the GDP of the secondary industry, the total retail sales of social consumer goods, and fiscal expenditure could all promote the increase of fiscal revenue in the province. Karran [6] pointed out that political factors such as tax policy, as well as macroeconomic factors such as GDP growth and inflation rate, have an important impact on British taxation.

In summary, previous studies have provided some reference for exploring the influencing factors of fiscal taxation, but most of the previous studies have focused on mathematical statistical analysis,

and the accuracy of regression is not high. The ANN-RBF used in this paper has strong nonlinear mapping ability. This algorithm is used to analyze the influencing factors of fiscal taxation in Anhui Province, China. Therefore, this paper will use the ANN-RBF model from four angles, including the population indicators, national economic indicators, international trade indicators, and investment spending indicators to analyze the influencing impact of taxation in Anhui Province, while using R^2 and MSE accuracy test, so as to provide suggestions for the Anhui provincial government to regulate tax revenue and promote economic development.

2. Construction of Index System of Influencing Factors of Regional Fiscal Taxation

2.1. Population indicators

Population indicators (X_1) can reflect the living situation and the level of consumption in the area. The rise in the total population is accompanied by an increase in the level of consumption, which can indirectly affect tax revenue [3]. Total retail sales of consumer goods is an important indicator to reflect the income of residents, living standards, prices, and industrial structure [4]. Per capita disposable income of urban residents, per capita disposable income of rural residents, per capita consumption expenditure of urban residents, and per capita consumption expenditure of rural residents are used to measure the level of economic development of a region and the extent to which taxpayers in the region bear the tax burden [7].

In sum, the population indicators should include total population (X_{11}), total retail sales of social consumer goods (X_{12}), per capita disposable income of urban residents (X_{13}), per capita disposable income of rural residents (X_{14}), urban per capita consumption expenditure (X_{15}), rural per capita consumption expenditure (X_{16}).

2.2. National economic

The national economic indicators (X_2) can intuitively reflect the economic development of a region and has a strong correlation with taxation in statistics. Among them, GDP is the core index to measure economic development status [8]. This index can be used to preliminarily determine the macroeconomic operation status of a region and measure the level of local people's living standards. With the increase in regional GDP, people's living standards will increase. Therefore, consumer spending and fiscal tax will also increase. There is a positive correlation between GDP and tax [9]. In addition, each industry's GDP can be a more detailed display of the impact of various industries on the tax [10]. Industrial structure optimization has a positive effect on tax revenue [11].

In sum, the national economic indicators should include gross primary industry output (X_{21}), gross secondary industry output (X_{22}), gross tertiary industry output (X_{23}), GDP (X_{24})

2.3. International trade

International trade indicators (X_3) mainly measure the openness of foreign trade in the region, and foreign trade is one of the troikas that drives economic growth. It can not only drive local economic growth but also have a positive impact on taxation. Generally, the total value of local exports and imports can be used to indicate the degree of openness of the region. Usually, the higher the degree of opening, the greater the volume of import and export trade. Import and export tariffs levied more taxes, indicating that the region's tax revenue will be more [7]. China's dependence on foreign trade, the ratio of total imports and exports to GDP, is about 50%, so total imports and exports are particularly critical in the study of taxes in China and its regions.

In sum, the international trade indicators should include total exports (X_{31}) and total imports (X_{32}).

2.4. Investment expenditure

Investment expenditure (X_{41}) reflects the main purpose of fiscal taxation. To be more specific, fixed asset investment is an important means of enterprise reproduction. The consumption-based VAT implemented in China allows the corresponding deduction of purchased fixed assets, which can stimulate the investment activities of enterprises, promote industrial restructuring, promote economic growth, and indirectly affect the tax base, thus affecting the total tax revenue [10]. Fiscal expenditure is the main purpose of tax revenue. The amount of tax revenue depends largely on the budget of the year. Therefore, the correlation between fiscal expenditure and tax revenue is highly related. Thus, fiscal expenditure is also an indispensable factor in studying the impact of tax revenue [10].

In sum, investment expenditure should include: fixed asset investment (X_{41}), fiscal expenditure (X_{42})

To sum up, the index system of influencing factors of regional tax revenue constructed in this paper is shown in Table 1

Table 1. Index system of influencing factors of regional tax revenue

Secondary Index	Tertiary Index	Index number	References
Population indicators (X_1)	Total population	X_{11}	[3]
	Total retail sales of social consumer goods	X_{12}	[4]
	Per capita disposable income of urban residents	X_{13}	[7]
	Per capita disposable income of rural residents	X_{14}	
	Urban per capita consumption expenditure	X_{15}	
	Rural per capita consumption expenditure	X_{16}	
National economic (X_2)	Gross primary industry output	X_{21}	[10][11]
	Gross secondary industry output	X_{22}	
	Gross tertiary industry output	X_{23}	
	GDP	X_{24}	[9][8]
International trade (X_3)	Total exports	X_{31}	[7]
	Total imports	X_{32}	
Investment expenditure (X_4)	Fixed asset investment	X_{41}	[10]
	Fiscal expenditure	X_{42}	

3. ANN Method

RBF neural network is a kind of feedforward network with a single hidden layer, which is composed of three layers: input layer, hidden layer and output layer. In 1985, Powell proposed a multivariate differential RBF method. ANN is a method of imitating animal nerves consisting of neurons and synapses. This method has a strong nonlinear fitting ability, can map any complex nonlinear relationship, and the learning rules are simple, which is convenient for computer implementation. At the same time, it is also the best method for function approximation, which can approximate any continuous function with arbitrary accuracy. In addition, due to its high nonlinear mapping ability and fast convergence speed, it has received extensive attention in time series prediction. The ANN diagram is shown in Figure 1. The steps of ANN are as follows:

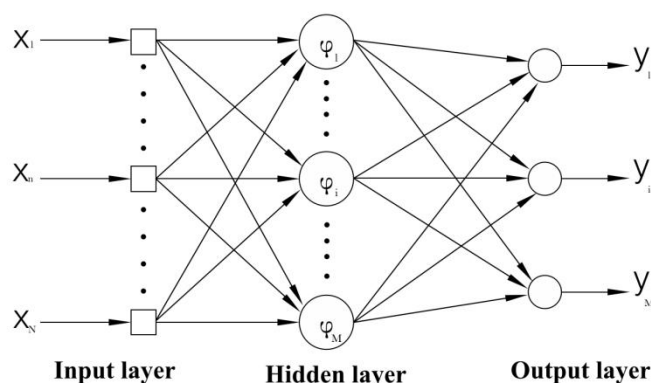


Figure 1. ANN Diagram

Each layer contains some neurons, and the adjacent two layers of neurons are connected in the same direction. As shown in the figure above, the input layer has N neurons, any one of which is represented as n ; the hidden layer has M neurons, any one of which is represented by i . The basis function $f(X, X_i)$ is the excitation output of the i th hidden layer neuron; The output layer has J neurons, any of which is represented by j . The connection weight of the output layer and the input layer is expressed as $w_{ij} (i = 1, 2, 3, \dots, M, j = 1, 2, 3, \dots, J)$. Assuming that a training sample set contains K training samples, where any sample stock is expressed as k . For example, the input vector is $X_k = [X_{k1}, X_{k2}, X_{k3}, \dots, X_{kn}, \dots, X_{kN}]^T (k = 1, 2, 3, \dots, K)$, corresponding to the actual output vector $Y_k = [Y_{k1}, Y_{k2}, Y_{k3}, \dots, Y_{kJ}]^T (k = 1, 2, 3, \dots, K)$, and the required vector is $d_k = [d_{k1}, d_{k2}, \dots, d_{kj}, \dots, d_{kJ}]^T (k = 1, 2, \dots, K)$. When X is the input of the j th output neuron, the actual output:

$$y_{kj}(X_k) = \sum_{i=1}^M \omega_{ij} \phi(X_k, X_i), j = 1, 2, 3, \dots, J \quad (1)$$

Green function is the basis function:

$$f(X_k, X_i) = G(X_k, X_i) \quad (2)$$

Gaussian is selected as the basis function, then:

$$f(r) = \exp\left(-\frac{(r-t)^2}{2s^2}\right) (s > 0, r \in R) \quad (3)$$

t is the key point of the Gaussian function and s is the variance of the Gaussian function, then:

$$G(X_k, X_i) = G(\|X_k - X_i\|) = \exp\left(-\frac{1}{2\sigma^2} \sum_{n=1}^N (X_{kn} - X_{in})^2\right) \quad (4)$$

$X_i = [X_{i1}, X_{i2}, \dots, X_{iN}]^T$ is the midpoint of the Gaussian function and s_i is the variance of the Gauss formula.

4. The setting of test indicators

The purpose of hypothesis testing is to determine whether there is a significant difference between the original hypothesis and the actual population. Using sample indicators to estimate the overall indicators, some of the conclusions are completely reliable, and some have only varying

degrees of reliability, which need to be further tested and confirmed. Through the test, we can judge whether there is a difference between the sample index and the overall index of the hypothesis so as to decide whether to accept the original hypothesis. Two groups of parameters, MSE and R^2 , are set to test the algorithm.

$$R^2 = \frac{\sum_{i=1}^n w_i (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n w_i (y_i - \bar{y}_i)^2} \quad (5)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

R^2 is generally used in regression models to assess the degree of conformity between predicted and actual values. The value range is between 0 and 1, and the domain is R. The closer R^2 is to 1, the better the model quality and the more accurate the prediction. In the formula, w_i is the weight; $\hat{y}_i$ is the predicted value; y_i is the real value; $\bar{y}_i$ is the average value.

MSE is the mean square error. n is the number of samples. y_i is the true value. $\hat{y}_i$ is the predicted value. The smaller the MSE, the better the model quality and the more accurate the prediction.

5. Empirical Analysis

5.1. Research Object Selection and Data Collection

The research object of this paper is Anhui Province, including population indicators (population, total retail sales of social consumption, per capita disposable income of urban residents, per capita disposable income of rural residents, per capita expenditure of urban residents, per capita expenditure of rural residents), national economic indicators (gross primary industry output, gross secondary industry output, gross tertiary industry output, gross domestic product), international trade indicators (total exports, total imports), investment expenditure indicators (fixed asset investment, government expenditure). The selection of these indicators is based on the indicators that have a greater impact on the factors affecting fiscal tax through research in previous scholars' papers. Due to the rapid economic growth in Anhui Province in recent years, the per capita income of social consumption retail sales, and other resident indicators have a huge growth, which have played a greater positive role in the growth of fiscal tax, so these are important factors to be analyzed. GDP and fiscal tax are highly correlated in statistics. In addition, the second and third industries in Anhui Province contribute 90 % to the GDP of Anhui Province, which is the core industry of Anhui Province. Therefore, they are important influencing factors. As a measure of local openness, total exports and total imports also have an important impact on fiscal revenue. Generally speaking, the higher the degree of openness, the higher the fiscal tax. Government expenditure and fixed asset investment as part of the expenditure can promote industrial restructuring, promote enterprise investment, and economic growth, so it also has an important impact on the tax revenue of the year.

The following data are from Anhui Statistical Yearbook, China Statistical Yearbook, and other databases. The data of Anhui Province from 2011 to 2020 are selected. The descriptive statistics of the indicators are shown in Table 2.

Table 2. Indicators Descriptive statistics

Indicators/units	Max	Min	Mid	Mean
X_{11} /million persons	6105.0	5957.0	6011.0	6024.2
X_{12} /100 million yuan	18334.0	4197.7	8908.0	9391.0
X_{13} /yuan	39442.1	15566.0	26936.0	27397.8
X_{14} /yuan	16620.2	5776.0	10821.0	10955.5
X_{15} /yuan	23781.5	11512.6	17233.5	17815.8
X_{16} /yuan	15024.0	4013.3	8975.0	9308.5
X_{21} /100 million yuan	3184.7	1729.0	2456.7	2455.4
X_{22} /100 million yuan	15671.7	6391.0	11077.7	11458.6
X_{23} /100 million yuan	19824.3	4193.7	8602.1	10081.2
X_{24} /100 million yuan	38681.0	12359.3	22006.0	23999.4
X_{31} /100 million dollar	455.8	109.2	306.0	297.7
X_{32} /100 million dollar	331.3	118.6	173.8	197.1
X_{41} /100 million yuan	32629.4	11849.4	21230.8	21230.8
X_{42} /100 million yuan	7473.6	2587.6	5239.0	5206.3

5.2. ANN Calculation and Parameter Comparison

Based on the data collected by 5.1, the SPSS platform is used to calculate the parameters. The independent variables of the neuron including total population, the total retail sales of social consumption, the per capita disposable income of urban residents, the per capita disposable income of rural residents, the per capita consumption expenditure of rural residents, the per capita consumption expenditure of urban residents, gross primary industry output, gross secondary industry output, gross tertiary industry output, GDP, the total export, the total import, the government investment, and the fixed asset investment are the independent variables of the neuron while the tax is the covariate. By adjusting the number of neurons, where the adjustment step is 1, the range is 10-25. The MSE and R^2 constructed in Section 4 are used to test the accuracy of the algorithm. The results are shown in Table 3.

Table 3. Algorithm accuracy comparison

The number of neurons	MSE	R^2
10	5703.995	0.980
11	6758.329	0.973
12	14004.106	0.958
13	4058.071	0.980
14	2045.077	0.991
15	4860.691	0.974
16	2821.329	0.986
17	5297.831	0.977
18	2382.398	0.988
19	3766.994	0.980
20	3139.368	0.985
21	6952.438	0.970
22	2996.903	0.985
23	6730.192	0.966
24	24568.350	0.951
25	3842.240	0.983

Through the iteration and operation of SPSS, it is found that when the number of neurons is 14, the accuracy of the algorithm converges at this time, and the MSE and R^2 are only 2045.077 and

0.991, indicating that the algorithm has high confidence at this time, and 5.3 contribution analysis can be conducted.

5.3. Contribution analysis

When the neuron is 14, the contribution of the independent variable to the dependent variable is analyzed. By measuring the synaptic weight of the independent variable to the dependent variable, the degree of influencing factors is quantified. The calculation results are shown in Figure 2.

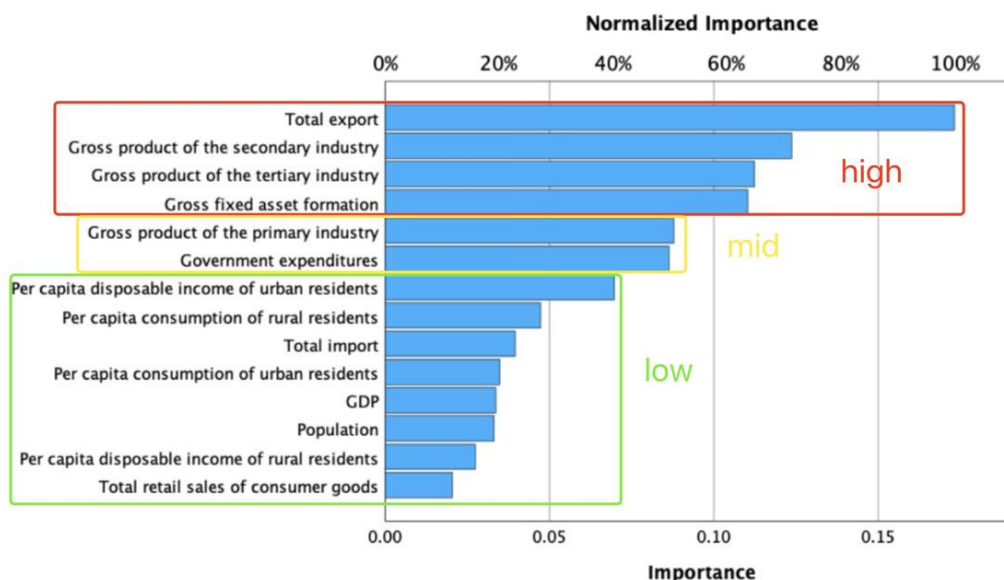


Figure 2. Contribution analysis

It can be seen from the graph that the total export, the gross secondary industry output, the gross tertiary industry output, and the investment in fixed assets have a greater contribution to the tax revenue of Anhui Province. The weights are: 0.147, 0.129, 0.111, 0.109. The gross primary industry output and government expenditure contribute less to tax revenue, with weights of 0.096 and 0.092 respectively. The per capita disposable income of urban residents, per capita consumption of rural residents, total imports, per capita consumption of urban residents, GDP, total population, per capita disposable income of rural residents, and total retail sales of social consumer goods contribute the least to tax revenue, with weights of 0.055, 0.053, 0.052, 0.051, 0.034, 0.029, 0.021 and 0.02 respectively.

The GDP of the secondary industry and the tertiary industry account for about 40 % and 50 % of Anhui's GDP respectively. They are the two most important industrial types in Anhui Province and have a huge impact on the economy and tax revenue. The impact of the secondary industry on tax revenue mainly comes from the manufacturing industry. With the increase in manufacturing output value, it will bring about an increase in taxes such as value-added tax and corporate income tax. These are the main taxes in Anhui Province, so they contribute greatly to the tax revenue of Anhui Province. Tertiary industry refers to the industry that provides services for production and people's livelihood, which has a significant impact on tax revenue [12]. The manufacturing industry depends on exports. 90 % of Anhui's export products are industrial products, and exports account for 8% -10% of GDP. Therefore, exports have a huge impact on the income of Anhui Province, especially the income of the secondary industry, so the export value contributes more to tax revenue. Fixed asset investment is an important means of enterprise reproduction, which can stimulate the investment activities of enterprises and also reflect the industrial activities of Anhui Province. Since from 2012 to 2018, the growth rate of fixed asset investment in Anhui Province has remained above 10%, reaching more than 20% in some years, driving the development of the secondary industry and indirectly promoting economic and tax growth. Although the primary industry accounts for less than the secondary and tertiary industries, it still contributes about 10% to GDP. Therefore, the increase in the GDP of the primary industry plays a certain role in the growth of taxation.

Government expenditure can play the role of GDP redistribution. Specifically, in Anhui Province, from 2010 to 2019, the average government expenditure on economic services in Anhui Province was 11.68%, mainly investment in infrastructures such as public transportation, communication, and electricity. Economic services have a significant role in promoting the high-quality development of Anhui's economy and indirectly affect the growth of tax revenue [13]. Because the per capita disposable income of urban residents, the per capita consumption expenditure of urban residents, the per capita disposable income of rural residents, and the per capita consumption expenditure of rural residents in Anhui Province are relatively low, the average value is only 27397,17816,10956,9308. During the same period, the average value of various indicators in the country is 31586,21299,11499,9089, except that the consumption expenditure of rural residents is slightly higher than the national average. The other indicators are far below the average. The total retail sales of social consumer goods are also low, so they contribute less to taxation in consumption. Total imports are small, at only about 6 % of GDP, and thus contribute less to the economy and taxes. Although the population size can directly affect the tax contribution, due to the low income and consumption capacity of the population in Anhui Province, except Hefei City, the contribution of population growth to tax revenue is limited.

6. Conclusion

In this paper, by using the ANN-RBF algorithm, 14 indicators, including the total population, retail sales of social consumer goods, per capita disposable income of urban residents, per capita disposable income of rural residents, consumption expenditure of rural residents, consumption expenditure of urban residents, total import and export, GDP of various industries, GDP, government expenditure, total investment in fixed assets are used as dependent variables, and the tax revenue of Anhui Province is used as a covariate. By adjusting the number of neurons, multiple sets of analysis results are obtained. Finally, R^2 and MSE are used as test indicators, and a set of data with the highest accuracy is determined as the experimental conclusion. The conclusions are as follows:

(1) When the number of neurons is 14, the accuracy converges and reaches the highest, indicating that the model constructed in this paper has a certain validity and practicality. This method can be extended to further research, such as the influencing factors of GDP, total import and export, and per capita disposable income of residents.

(2) Based on the ANN contribution analysis, it can be found that the total exports, the gross secondary industry output, the gross tertiary industry output, and the total investment in fixed assets have a great influence. They are the main impacts of the tax. Urban managers can focus on these points to make improvements.

References

- [1] Ding Hao, Long Ziyue. Analysis of the main factors affecting China 's fiscal revenue [J]. Modern Marketing (Chuangfu Information Edition), 2018 (09): 34.
- [2] Cai Simei. Analysis of influencing factors of fiscal revenue based on stepwise regression [J]. Accountants, 2019 (14): 17-18.
- [3] Li Jie. Empirical study on the main influencing factors of fiscal revenue in Tianjin [J]. Tianjin economy, 2015 (04): 12-14.
- [4] Zhang Lunjun. Quantitative relationship between taxation and major economic indicators such as GDP [J]. Tax Research, 2005 (07): 10-13.
- [5] Ni Jie. Influencing Factors and Prediction Analysis of Fiscal Revenue in Shandong Province [D]. Shandong Normal University, 2021.
- [6] Karran T. The determinants of taxation in Britain: An empirical test [J]. Journal of Public Policy, 1985, 5 (3): 365-386.

- [7] Guo Shuhua, Bao Weijie, Li Fang. An Empirical Study on the Influencing Factors of Tax Revenue in Yunnan Province - Based on Data from 1988 to 2016 [J]. *Economic Issues Exploration*, 2018 (03): 145-150.
- [8] Yang Han, Cultural Revolution. Analysis of influencing factors of fiscal revenue in Hebei Province [J]. *Cooperative Economy and Technology*, 2022 (21): 181-183.
- [9] Gao Tianhui, Zhang Yujing, Zhu Jiaming, Cui Wenkai. An Empirical Study on the Influencing Factors of Tax Revenue in Anhui Province Based on Principal Component Analysis [J]. *Journal of Natural Science*, Harbin Normal University, 2020, 36 (05): 12-17.
- [10] Zuo Yuting, Wu Li. A Study on the Influencing Factors of Tax Revenue in Anhui Province - An Empirical Analysis Based on Data from 2001 to 2018 [J]. *Journal of Chifeng University (Natural Science Edition)*, 2020, 36 (05): 66-71.
- [11] Li Na, Yang Huiru, Zhao Yijing. Research on the influencing factors of tax revenue in China based on panel quantile regression model [J]. *Statistical theory and practice*, 2022 (09): 41-45.
- [12] Zhao Shanshan, Liu Wei. An Empirical Analysis of Factors Affecting Tax Revenue in China [J]. *Taxation*, 2021, 15 (16): 23-24.
- [13] Mei Tingting. Research on the impact of fiscal expenditure structure on high-quality economic development in Anhui Province [D]. Anhui University, 2021.