Analysis of Affecting Factors of Proportion of Tertiary Industry Based on Ridge Regression

Zhang Bin

School of Education, Hubei University for Nationalities

Yuan Shoucheng

School of Mathematics and Statistics, Pu`er University

Abstract—**Optimization of Industrial structure has played a significant role in the process of national economy's sustainable and stable growth. With six affecting factors including proportion of labor productivity, proportion of employed people, proportion of fixed-asset investment, proportion of the actual utilization of foreign capital, the proportion of total energy consumption, and proportion of resident population of Beijing from 2000 to 2015, this paper first uses the method of multiple linear regression model to study these variables' influence on proportion of tertiary industry, and finds that the model itself has multi-collinearity. Then, the method of Ridge Regression and variable selection model are utilized to quantify these variables' influence over proportion of tertiary industry. The study shows that these six factors had significantly affected the development of tertiary industry. The most influential one among of them is energy, then is population and investment. By using the method of ridge regression model to quantitatively analyse these factors, this paper proposes related advice about the development of the tertiary industry and deepening of industrial structure from three perspectives of energy, population and investment.**

Keywords—Tertiary industry; Ridge regression; Affecting factors

I. INTRODUCTION

Since the reform and opening up, China's economy has made great progress. Especially from the year 2000 to now, Gross Domestic Product (GDP) has been maintaining fast and long-term growth. The proportion of value of tertiary industry has exceeded that of primary industry and that of secondary industry, which has become the most important component of the national's economy. In today's society, the developmental level of tertiary industry has been an important measurement that can mark the developed level of social economy of one country or a region [1]. In recent years, the development of tertiary industry has been one of the hot issues of social and economical research. Li Jiangfan (2001) studied several related issues about tertiary industry and China's construction of modernization, and fitted the correlations between the tertiary industry and economical development with mathematical models [2]. Zhou Liangjun (2004) utilized Statistical tools to

analyze the developmental level of China's tertiary industry [3]. Zou Yuan has studied the influence of evolution of industrial structure in China over the quality of economic growth from the perspective of quantitative economics [3]. Besides, Zhou Yu (2007) and Li Jiangfan (2014) and so on respectively did quantitative analyses of developmental level of tertiary industry and their affecting factors in Jiangsu Province, Beijing and Shanghai and some other areas [5-6]. Liu Sisi (2008) and Wu Xiyan (2011) each has respectively studied the influence of population over the developmental level of tertiary industry, theoretically and empirically [7-8]. Gao Suying (2005) noted that the linear regression model were found to have the problem of multicollinearity when it was used to analyze the economic growth factors, and gave a suggestion to improve it by using the ridge regression [9].

According to the current research results of the development of tertiary industry and comprehensively considering the influence of theses six factors above over tertiary industry, this passage quantitatively analyzed the effects of these affecting factors by using multiple linear regression model, ridge regression model, and variable selection method and proposed some policy suggestions about how to promote the development of Beijing's tertiary industry.

II. DESCRIPTIVE ANALYSIS OF DATA

In order to quantize the relationships between tertiary industry and every affecting factor, the data from the year of 2000 to the year of 2015 in Beijing were adopted in the research (Data from Beijing Government Data Resource Network). To make it brief, we regarded the Y axis as the proportion of tertiary industry's increase in the whole city, X1 as the proportion of labor productivity, X2 as the proportion of year's end population of the employed, X3 as the proportion of fixed-asset investment, X4 as the proportion of the actual utilization of foreign capital, X5 as the proportion of total energy consumption, X6 as the proportion of permanent resident population (million) in Beijing. The time diagram of these variables is as follows:
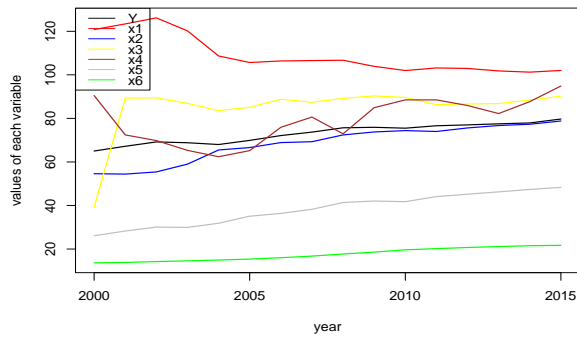
Fig. 1   Variation Trend of Every Index

Figure 1 can intuitively shows the variation trends of the proportion of Beijing's tertiary industry and its affecting factors, from which it can be obviously observed that the proportion of tertiary industry(Y) has kept a good and stable increasing trend since 2000. Besides, the proportion of permanent resident population(X6), the proportion of year's

end population of the employed(X2) and the proportion of total energy consumption(X5) have always shared a similar increasing trend, which is in accordance with that of the proportion of tertiary industry. What's more, the proportion of the actual utilization of foreign capital(X4) goes up like a wave, and the proportion of fixed-asset investment increases in a stable way but with a small average annual growth rate. The proportion of labor productivity(X1) once experienced a declining trend but tended to be stable after the year of 2005. In general, there are strong linear relationships between tertiary industry and all the affecting factors above.

III.   REGRESSION MODELS BETWEEN TERTIARY INDUSTRY AND ALL AFFECTING FACTORS

At first, what should do is to fit the data from 2000 to 2015 in Beijing by using the traditional multiple linear regression model.

TABLE I   MULTIPLE LINEAR REGRESSION COEFFICIENTS AND THEIR significance

	Intercept Term	X1	X2	X3	X4	X5	X6
Coefficient	-4.88082	0.27469	0.39872	0.03673	0.06601	0.57977	-0.55736
Standard Deviation	12.16766	0.06285	0.12432	0.01303	0.01822	0.12165	0.19397
Value of P	0.69768	0.00180 **	0.01071 *	0.02008 *	0.00554 **	0.00102 **	0.01837 *

Considering the value of every index, the adopted six affecting factors are all significant, and goodness of fit and adjusted goodness of fit are respectively 0.9964 and 0.9941. However, intercept term is negative and insignificant, and the index of resident population is also negative. It means that the proportion of tertiary industry will decrease when the population increase, which is not consistent with the reality because the proportion of tertiary industry will be less than zero when the value of every affecting factor is not very big. Consequently, there are some limitations when the model is adopted in the real situation.

Variance inflation factor (VIF) is used here to test the referred linear model's multicollinearity. If the value of variance inflation factor is 1, it indicates that there is no multicollinearity. However, the value is more than 5 or 10, it shows that multicollinearity exists among these variables. Through calculation, the VIF value of the proportion of labor productivity(X1) is 35.91, and the VIF value of the proportion of year's end population of the employed (X2) is 137.32, and the VIF value of the proportion of fixed-asset investment(X3) is 3.17, and the VIF value of the proportion of the actual utilization of foreign capital(X4) is 4.28, and the VIF value of the proportion of total energy consumption(X5) is 98.37, and the VIF value of the proportion of permanent resident population(X6) is 40.89. Apart from the proportion of fixed-asset investment and the proportion of the actual utilization of

foreign capital, the VIF values of the other 4 factors are all comparatively big. Especially, the VIF values of the proportion of year's end population of the employed and the proportion of total energy consumption are even more than others. In a result, it indicates that there is much multicolinearity in this linear regression model (Figure 2).
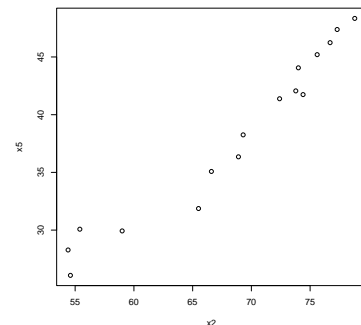

Fig. 2   The Colinearity of the proportion of total energy consumption(X5) and the proportion of year's end population of the employed(X2)

Figure 2 indicates that the scatter diagrams of the proportion of total energy consumption and the proportion of year's end population of the employed present the linearity. However, multicollinearity is inclined to result in this model's

instability and the increasing predictive errors [10]. One of the frequently-used methods to erase colinearity is ridge regression, the fundamental principle of which is to enhance stability of numerical value by loosening the requirements of unbiasedness on the basis of the method of least squares [11].

The next step is to fit these years' data by using the ridge regression model.

TABLE II   THE COEFFICIENT AND SIGNIFICANCE OF RIDGE REGRESSION

	Intercept Term	X1	X2	X3	X4	X5	X6
Coefficient	28.13740	0.07241	0.15944	0.06575	0.06575	0.29554	0.23174
Standard Deviation	—	1.09626	1.02948	0.83624	0.97622	1.18340	1.30463
Values of P	—	0.028365 *	3.74e-07 ***	0.000183 ***	0.007767 **	1.37e-12 ***	0.041519 *

Ridge parameter was automatically selected to be 0.03386481 which is based on the generalized cross validation principle. According to the value of P of every coefficient, it can be seen that the significance of these affecting factors, such as X2, X3, X4, X5 and so on, has been highly improved. Especially, the significance of X2, X3, X5 among these factors has been improved a lot and their intercept term is positive. So, this model is more in line with the reality.

Based on the ridge regression, it can be found that the proportion of fixed-asset investment and the proportion of the

actual utilization of foreign capital have the same weight, which means that both of them have the same influence on the proportion of tertiary industry. What's more, coefficient of resident population is also positive, which is closer to the actual and means that this model is better explicable than the multiple linear regression model. In order to measure the importance of every affecting factors more, variable selecting method called Lasso is used here to compare priority and importance of every variable.

TABLE III   STEPS OF CHOOSING VARIABLES

Steps	1	2	3	4	5	6
Sequence of variable	X5	X3	X4	X1	X2	X6
Value of Cp	79.734	56.012	28.858	22.642	18.395	7.000

The table above indicates that when we study the proportion of tertiary industry, the proportion of total energy consumption is at first pitched on and the second one is the proportion of fixed-asset investment and the last one is the proportion of permanent resident population, which shows that though Beijing's industrial structure has something to do with the increasing population, industrial structure optimization's achievements are not only made because of quantity changes, but more due to qualitative improvements. When the model

contains several independent variables, the independent variable's subset will be selected according to the value of Cp. The smaller the value of Cp is, the more accurate the model will be. The smallest value of Cp in the model above is 7, which tells that all these six factors have great influence on the proportion of tertiary industry and should be taken seriously. In conclusion, the ridge regression's model of tertiary industry is as follows:

$$Y = 28.13740 + 0.07241*X1 + 0.15944*X2 + 0.06575*X3 + 0.06575*X4 + 0.29554*X5 + 0.23174*X6.$$

The model above quantitatively describes influence of every affecting factors on the proportion of tertiary industry. And the proportion of total energy consumption has the greatest influence on tertiary industry, and the second most influential are the proportion of permanent resident population and population of the employed, which are followed by the proportion of fixed-asset investment and the proportion of the actual utilization of foreign capital.

IV. CONCLUSIONS AND POLICY SUGGESTIONS

This research analyzed six affecting factors of the proportion of Beijing's tertiary industry based by using ridge regression model, which indicated that all these six factors had influenced the proportion of Beijing's tertiary industry a lot. And the most influential factor is the proportion of total energy consumption. Then what takes second place is the proportion of permanent resident population and fixed-asset investment. So, actions should be taken from three aspects to promote the development of Beijing's tertiary industry. First of all, what should be done is to deepen market reform, increase energy efficiency and total consumption, such as making a reasonable price-and-market mechanism, providing more access to clean energy and so on [12]. Secondly, what should be done is to maintain long-term stability of population quantity and quality, raise employment rate, and improve the number and quality of the employed, such as striving to develop the emerging tertiary industry, and enhancing absorptive capacity of labor from service industry, leisure industry and industry of culture and sports. Lastly, what should be done is to consolidate and further expand the level of reform and opening up, promote fixed-asset investment, and accelerate urbanization, such as gradually expanding reform and opening up in the service-oriented fields of finance, insurance, business, and tourism, and participating actively in the competition of international service market and so on.

REFERENCES

[1] Li Jiangfan. Strategic Position and Development Direction of China's Tertiary Industry [J]. Finance & Trade Economics, 2004(1): 65-73.

[2] Li Jiangfan. Several Problems of China's Tertiary Industry and Modernization Construction [J]. Macroeconomics, 2001(10): 45-49.

[3] Zhou Liangjun. Statistical Analysis of the Development Level of China's Tertiary Industry [D]. Hunan University, 2005.

[4] Zhou Yuan. Study on the Influence of the Evolution of China's Industrial Structure on the Quality of Economic Growth. Chongqing University, 2016.

[5] Zhou Yu, Zhen Chuiyong. Grey Prediction of Tertiary Industry in Jiangsu Province [J]. Water Economics, 2007, 25(5): 49-51.

[6] Li Jiangfan, Lan Wenyan, Zhu Shengyong. Study on the Level, Structure and Influence of Tertiary Industry in Beijing and Shanghai [J]. Journal of Beijing Technology and Business University (Social Sciences Edition), 2014, 29(2): 7-17.

[7] Liu Sisi. Influence of Population Density on the Development Level of Tertiary Industry [D]. Zhongshan University, 2008.

[8] Wu Xiyan. The Influences of Population Density on the Developmental Level of Tertiary Industry-An Empirical Analysis from China [J]. Journal of Foshan University (Social Science Edition), 2011, 29(3): 37-40.

[9] Gao Suyin, Li Yanjun, Jin Hao. The Application of Ridge Regression in the Analysis of Affecting factors of Economic Growth [J]. Statistics and Decision, 2005(5x): 142-144.

[10] Zhao Dongbo. Study on the Issue of Multicollinearity in Linear Regression Model [D]. Bohai University, 2017.

[11] Zhang Bo. Study on Estimation of Ridge Regression Parameters in Econometric Model [D]. Tianjing University of Finance and Economics, 2016.

[12] Jiang Lei, Ji Minghe. On Relationship between Technological Progress, Industry Structure, Energy Consumption Structure and Energy Efficiency in China-An Empirical Study Based on Ridge Regression [J]. Contemporary Economy & Management, 2011, 33(5): 13-16.