ISSN 2616-5872 Vol.3, Issue 4: 37-43, DOI: 10.25236/AJEE.2021.030408

Study on energy consumption characteristics based on grey multivariate model ——with an example of Beijing in China

Ruobing Zhang[1], Yufeng Wu[2]

[1]Faculty of Science, Tibet University, Lhasa, 850000, China
[2]School of Mathematical Sciences, Chongqing Normal University, Chongqing, 405400, China

Abstract: *With the economic growth, people's living standards have increased and energy has been consumed excessively, energy consumption is an important part of the sustainable development of the economy. In this paper, the three major industries of Beijing's energy consumption with the total energy consumption by grey correlation analysis, it is concluded that the industrial structure and energy consumption characteristics. In view of the theory of consumption demand of economics, select factors that influence the consumption of total energy, establish a multivariate linear regression model, and determine the relationship between each factor and total energy consumption. By examining significant differences, eliminating secondary factors and arriving at multiple linear regression equations. Finally the change of total energy consumption in Beijing is predicted.*

Keywords: *Energy consumption, Gray correlation, Multiple linear regression, Significant differences test*

1. Introduction

Since the past, energy is a prerequisite for civilization development, and the blood of urban. All activities of human activities are inseparable from energy. The rapid development of the economy has greatly improved people's material living standards, and it also caused excessive consumption of energy[1,2]. Under "Peak carbon dioxide emissions before 2030 and carbon neutrality before 2060" strategic objectives, we must accelerate China's energy green and low-carbon development[3], and reasonably control our country's energy consumption. This paper selects Beijing as an research object of energy consumption characteristics[4].

First determine the definition of energy consumption characteristics[5], then establish the grey relational analysis model. Gray correlation analysis between the respective energy consumption and total energy consumption of the three major industries in Beijing, and finally combined with sample data, the respective gray correlation is drawn, thereby analyzing the industrial structure and energy consumption characteristics of Beijing[6, 7].On the basis of the industrial structure and energy consumption characteristics of Beijing, determine other main influencing factors outside the industrial structure according to the economic consumption demand theory, establish the total energy consumption of Beijing with its various linear regression equations that affect the factors[7]. Finally, the regression equation and the regression coefficient perform a corresponding significance test, and the previous regression equation can be improved, then the energy consumption of Beijing can analyze forecasts. This paper analyzes the consumption structure based on the 2017 Beijing Statistical Yearbook. Beijing is a city with rapid development, but the recent research on the consumption structure is in 2009. The reference data in this paper is relatively new, which has a certain reference value for scholars who need to study the recent consumption characteristics of Beijing. The sample data in this paper is from "Beijing Statistical Yearbook 2017".

2. Gray correlation model

2.1. Establishment of gray correlation model

2.1.1. Reference series and comparison series

$$\begin{cases} x_0 = \{x_0(k) | k = 1,2, \dots, \mathrm{n}\} \\ x_i = \{x_i(k) | k = 1,2, \dots, \mathrm{n}\} \end{cases}$$

Where $x_0(k)$ represents the k-th value of total energy consumption, $x_i(k)$ represents the k-th value of the i-th factor affecting total energy consumption.

2.1.2. Gray correlation coefficient

Gray correlation coefficient of reference series $x_0(k)$ and comparison series $x_i(k)$ on the element k:

$$\xi_i(k) = \frac{\min\limits_i \min\limits_k |x_0(k) - x_i(k)| + \rho \cdot \max\limits_i \max\limits_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \cdot \max\limits_i \max\limits_k |x_0(k) - x_i(k)|}$$

Where ρ is the resolution, and 0<ρ<1, usually takes 0.5; $\min\limits_i \min\limits_k |x_0(k) - x_i(k)|$ is the double minimum difference; $\max\limits_i \max\limits_k |x_0(k) - x_i(k)|$ is the double maximum difference.

2.1.3. Gray correlation

Gray correlation between comparison series x_i and reference series x_0:

$$\gamma_i = \frac{1}{n} \sum_{k=1}^{n} \xi_i(k)$$

2.1.4. Correlation degree sort

After calculation, the correlation between the comparison series and the reference series γ_i is obtained. The greater the degree of relevance γ_i show that the closer the relationship between the comparison series and the reference series, their data and trend changes and more similar. The smaller γ_i, indicating that the relationship between them is more alienated, and their data and trend changes are different.

3. Multivariate linear regression model

3.1. Establishment of multiple regression model

3.1.1. General form of model

Assuming that there is a linear relationship between dependent variable Y and p independent variables, the regression equation is established as follows:

$$\begin{cases} Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

Where $\beta_0, \beta_1, \dots, \beta_p$ is the regression coefficient, indicating the marginal influence of the independent variables on the dependent variable Y; ε is the random error, indicating the part that cannot be determined by the independent variable.

3.2. Parameter estimation of model

3.2.1. Least square method

The least square method is a method to minimize the sum of squares of errors and obtain a more suitable functional relationship with the original sample data.

The specific operation process is relatively simple, that is, take $b_0, b_1, \dots, b_p$ as the estimated value of regression coefficient $\beta_0, \beta_1, \dots, \beta_p$, and the sample value $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$,(i=1,2,…,n) of dependent

ISSN 2616-5872 Vol.3, Issue 4: 37-43, DOI: 10.25236/AJEE.2021.030408

variable Y can be expressed as:

$$y_i = b_0 + b_1x_{i1} + \cdots + b_px_{ip} + e_i, (i = 1,2,\ldots,n)$$

Where e_i is the estimated value of ε_i, i.e. residual; let $\widehat{y_i}$ be an estimate of y_i:

$$\widehat{y_i} = b_0 + b_1x_{i1} + \cdots + b_px_{ip}$$

$$e_i = y_i — \widehat{y_i} = y_i — b_0 — b_1x_{i1} — \ldots — b_px_{ip}$$

Since $\beta_0, \beta_1, \ldots, \beta_p$ is taken to minimize the sum of squares of residuals $Q = \sum_i^n e_i^2$, the following p+1 equations are obtained by making each partial derivative of Q to $\beta_0, \beta_1, \ldots, \beta_p$ 0 through the minimum principle of calculus:

$$\begin{cases} \dfrac{\partial Q}{\mathrm{a}b_j} = -2\sum_{i=1}^{n}(y_i — \widehat{y_i})x_{ij} = 0, (j = 1,\ldots,p) \\ \dfrac{\partial Q}{\mathrm{a}b_0} = -2\sum_{i=1}^{n}(y_i — \widehat{y_i}) = 0 \end{cases}$$

By solving the above equations, the estimated value $\beta_0, \beta_1, \ldots, \beta_p$ of regression coefficient is solved, and the multiple regression equation is obtained:

$$\widehat{y_i} = b_0 + b_1x_1 + \cdots + b_px_p$$

3.3. Test of multiple regression model

By calculating the fitting degree between the sample data and the multiple regression equation, the overall fitting degree of the multiple regression equation can be determined, that is, the overall relationship between the dependent variable and all independent variables. Among them, the goodness of fit is expressed by the ratio of the sum of squares of multiple regression model to the sum of squares of total deviations.

Goodness of fit:

$$R^2 = \frac{S_R}{S_{yy}} = 1 — \frac{S_e}{S_{yy}} = 1 — \sum_{i}^{n}(y_i — \widehat{y_i})^2 / \sum_{i}^{n}(y_i — \hat{y})^2$$

Where $S_R/\sigma^2 \sim \chi^2(p)$ is the sum of squares of regression, $S_e/\sigma^2 \sim \chi^2(n-p-1)$ is the sum of squares of residuals, and $S_{yy}/\sigma^2 \sim \chi^2(n-1)$ is the sum of squares of total deviations. R^2 represents the degree of interpretation of the independent variables to the dependent variable. The closer its value is to 1, the better the fitting degree of the model is.

4. Instance analysis

4.1. Data processing

This paper uses Beijing as a research object, firstly the gray correlation model is used to calculate gray correlation between each of the three major industries of energy consumption and total energy consumption in Beijing. From this analysis to obtain the industrial structure and energy consumption characteristics. In addition, by reviewing the information and the consumer demand theory[13], it is determined that the seven major factors affecting the energy consumption of Beijing, then establish a multi-regression analysis model to analysis and predict the changes in the total energy consumption.

The data indicators required for the two models used in this paper is needed: the total energy consumption, the primary, secondary and tertiary industries in energy consumption and its seven major influencing factors, Beijing's gross domestic product (GDP level), the number of Beijing's permanent population, energy intensity, industrial structure, Beijing’s level of urbanization, and energy consumption structure. Sample data referred to herein are derived from the "Beijing Statistical Yearbook 2017" (1990 -2016 years).

To comprehensive evaluation, sample data should be normalized. Z normalized (standard deviation normalized) used to the standardize sample data:

ISSN 2616-5872 Vol.3, Issue 4: 37-43, DOI: 10.25236/AJEE.2021.030408

$$x'_{ij} = \frac{x_{ij} - \overline{x_j}}{s_j}, (i = 1,2, \dots ,7)$$

Where $\overline{x_j}$ is

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}\, , s_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})^2}$$

x'_{ij} is a variable value after Z standardization in the SPSS software.

4.2. Gray correlation between the three major industries and total energy consumption

4.2.1. Establish gray correlation analysis model

1) As the total energy consumption of Beijing as the reference series, the three major industries can be used as a comparison series.

2) Calculate the gray correlation coefficient $\xi_i(k)$ and gray correlations γ_i.

Where $\xi_i(k)$ is the association coefficient between the kth element of x_i and the kth element of x_0.

4.2.2. Model results

Based on the MATLAB software, the energy consumption of Beijing and the three major industries is obtained as shown in Figure 1.

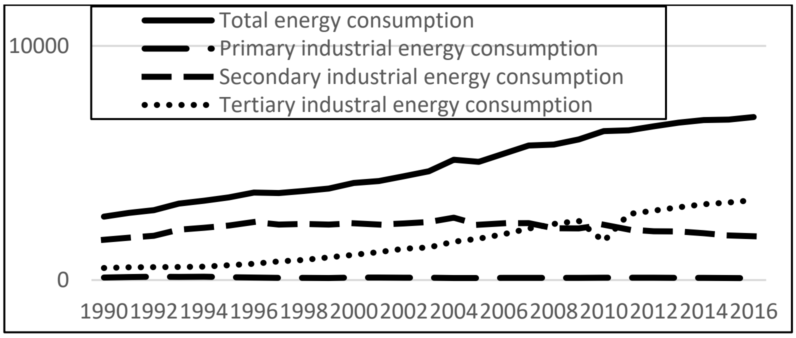

Figure 1: Energy consumption map of various industries in Beijing

4.2.3. Model results analysis

According to the calculation formula of gray correlation model, the results is performed using MATLAB software are as follows:

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0.6650 \\ 0.7298 \\ 0.8292 \end{pmatrix}$$

According to the above results, it can be seen that $\beta_3 > \beta_2 > \beta_1$, note that the energy consumption of Beijing's tertiary industry has the greatest impact on total energy consumption, and has the highest correlation, the secondary industry is lower than the tertiary, and the first is minimal. It can be seen from Figure 1 that the growth trend of total energy consumption and the energy consumption of the tertiary industry are closer. In conjunction with the current development of Beijing, its tertiary industry has developed rapidly and has been awarded the first position in the country for many years; the development of the secondary industry is slow, and there is a tendency to decline; and the first industry has been stable and slowly changing.

4.3. Multi-factor analysis model for total energy consumption

4.3.1. Influencing factors

This paper refers to the theory of consumer demand for Western economics[13], and combined with the current situation, the 7 major influencing factors of the total energy consumption of Beijing are obtained: energy consumption price, GDP level, industrial structure, urbanization level, total population,

ISSN 2616-5872 Vol.3, Issue 4: 37-43, DOI: 10.25236/AJEE.2021.030408

energy intensity and energy consumption structure, which are used respectively by $x_1, x_2, x_3, x_4, x_5, x_6, x_7$, the total amount of energy consumption in Beijing uses Y.

4.3.2. Establish multi-factor analysis model

Establish a multivariate regression equation between the total amount of energy consumption Y and 7 influencing factors $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ in Beijing:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \varepsilon$$

4.3.3. Minimum multiplier estimation of regression coefficient

The minimum multiplier estimate of the regression equation is performed by means of the SPSS software, and the specific result data is shown in Table 1 below.

Table 1: Minimum multiplier estimation of regression coefficient

variable	non-standard coefficient	t value	sig value
constant term	4859.274	216.535	0.000
x_1	4708.859	2.297	0.033
x_2	469.138	1.618	0.122
x_3	-82.260	-0.870	0.395
x_4	-190.914	-2.994	0.007
x_5	-4220.924	-2.221	0.039
x_6	368.771	3.185	0.005
x_7	17.538	0.094	0.926

4.3.4. Model results

According to the regression coefficient of the regression equation derived in Table 3, the multivariate regression equation:

$$\hat{Y} = 4859.274 + 4708.859x_1 + 469.138x_2 — 82.260x_3 — 190.914x_4 — 4220.924x_5 + 368.771x_6 + 17.538x_7$$

4.4. Multi-model test

4.4.1. Fitness test of regression equation

The specific test results data are obtained by means of SPSS software as follows Table 2.

Table 2: Goodness of fit

R	R^2	$\overline{R^2}$	standard error estimation
0.997	0.995	0.993	116.6070

The results show that the goodness of fit R^2=0.995 is high and close to 1; additionally, the adjusted goodness of fit $\overline{R^2}$=0.993 is also close to 1, indicating that the degree of fitting of the multivariate regression model is better and passed the test.

4.4.2. Significant test of regression equation

By the SPSS software, the F=532.173, sig=0<0.05 are calculated, thus rejected H_0, indicating that the regression equation is remarkable and passed by a significant test. This shows that the seven influencing factors selected in the model have a significant linear relationship with the total amount of energy consumption in Beijing.

4.4.3. Significant test of regression coefficient

Through the SPSS software, the significance test of the multi-regression equations is realized, and the test results data are shown in Table 3.

Analysis of test results: sig value of each variable in Table 3 as a result of judging whether the impact on the dependent variable y is significant. Based on the significance level of this paper is 0.05, if the sig value of a certain independent variable is less than 0.05, then the influence of this independent variable on dependent variable is significant, otherwise it will not be significant; therefore, it is necessary to eliminate the independent variables that have weak impact on dependent variable. Significant testing of regression coefficients each time should eliminate independent variables with maximum sig value. In Fig. 1, since sig(x_7)=0.926>0.05 is the maximum, eliminate independent variable x_7, re-establish regression equation for the remaining independent variable, the regression coefficients were subjected to

ISSN 2616-5872 Vol.3, Issue 4: 37-43, DOI: 10.25236/AJEE.2021.030408

a significant test, and the results were as follows Table 3.

Table 3: Minimum multiplier estimation of regression coefficient

variable	non-standard coefficient	t value	sig value
constant term	4859.274	221.109	0.000
x_1	4740.859	2.405	0.026
x_2	454.462	1.910	0.071
x_3	-80.551	-0.891	0.384
x_4	-194.499	-3.909	0.001
x_5	-4259.854	-2.356	0.029
x_6	368.803	3.267	0.004

Analysis of test results: after eliminating x_7, get Table 3 via SPSS software, R^2=0.995, $\overline{R^2}$=0.993, F=653.266, the error of the standard estimate is 113.6807. It can be seen from Table 3 after removing x_7, significance of the remaining independent variables changed, at this time sig(x_3)=0.384>0.05 maximum, thus eliminate x_3, carry out the regression analysis again, the results are as follows Table 4.

Table 4: Minimum multiplier estimation of regression coefficient

variable	non-standard coefficient	t value	sig value
constant term	4859.274	223.212	0.000
x_1	4424.323	2.293	0.032
x_2	421.603	1.802	0.086
x_4	-181.625	-3.834	0.001
x_5	-3945.580	-2.236	0.036
x_6	339.300	3.160	0.005

Analysis of test results: after eliminating x_3, get Table 4 via SPSS software, R^2=0.995, $\overline{R^2}$=0.993, F=791.561, the error of the standard estimate is 113.1192. It can be seen from Table 4 that after removing x_3, sig(x_3)=0.384>0.05 maximum, there is not much difference with 0.05, therefore, two cases are needed to be analysis.

Situation 1: do not eliminate x_2, the result is the same as Table 4.

Situation 2: eliminate x_2, carry out the regression analysis again, the results are as follows Table 5.

Table 5: Minimum multiplier estimation of regression coefficient

variable	non-standard coefficient	t value	sig value
constant term	4859.274	212.609	0.000
x_1	6352.866	3.770	0.001
x_4	-199.605	-4.106	0.000
x_5	-5527.486	-3.439	0.002
x_6	397.342	3.695	0.001

Analysis of test results: after eliminating x_2, get Table 5 via SPSS software, R^2=0.995, $\overline{R^2}$=0.993, F=896.950, the error of the standard estimate is 118.7602, at this time, the sig value of all remaining independent variables is less than 0.05.

4.5. Multi-model conclusion and explanation

By analyzing the results of the models obtained by the case one and two, the error of the standard estimation of the case one is smaller, thus the independent variable x_2 is retained. That is to say, the five independent variables x_1, x_2, x_4, x_5, x_6 that the equation last reserved have a significant impact on the dependent variable Y. At the same time, the fitness test and the significance test of the regression equation passed.

Therefore, the predictive regression equation of the total amount of energy consumption in Beijing is finally obtained:

$$\hat{Y} = 4859.274 + 4424.323x_1 + 421.603x_2 — 181.625x_4 — 3945.58x_5 + 339.3x_6$$

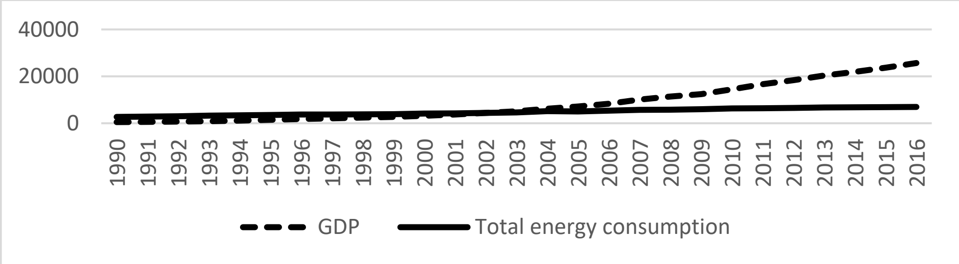

Figure 2: Comparison of changes in GDP and total energy consumption

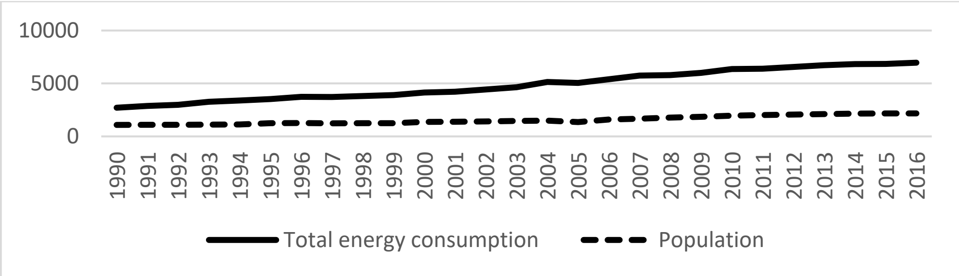

Figure 3: Comparison of changes in population and total energy consumption

5. Conclusion

This paper analyzes the industrial structure and energy consumption characteristics of Beijing by gray correlation analysis model. In addition, the multiple linear regression model is used to study the functional relationship between the total energy consumption and its seven influencing factors in Beijing[17,18], then carry on analysis and estimate. The grey correlation degree is used to analyze the correlation between the energy consumption of the primary, secondary and tertiary industries and the total energy consumption in Beijing from 1990 to 2016, so as to obtain the industrial structure and energy consumption characteristics of Beijing: that is, Beijing is a "three, two and one" industrial pattern dominated by the tertiary industry and supplemented by the secondary industry.

References

[1] Zhang M H, Li Y F. On the relationship between energy and economic growth in China [J]. Industrial technology and economy, 2004, 23 (4): 77-80.
[2] Wang H G. Research on the relationship between China's economic growth and energy consumption [D]. Jiangxi University of Finance and economics, 2006.
[3] Xiong X P. Accelerating China's green and low-carbon energy development [n]. China economic times, 2021-11-04 (004).
[4] Beijing Municipal Development and Reform Commission. Beijing energy development report [M]. Beijing: Beijing Municipal Development and Reform Commission, 2012.
[5] Jiang W, Gao W D. Analysis on the growth characteristics and influencing factors of China's energy consumption [J]. World Geography Research, 2013, 22 (3): 160-168.
[6] Yin C H, Gu P L. Grey correlation analysis of China's industrial structure adjustment and energy consumption [J]. Science of Tianjin University Bulletin, 2003, 36 (1): 104-107.
[7] Huang Y S, Wang Y W, Gai S.The Application and Research of a New Combinatorial Analysis and Forecasting Method in Real Estate Area based on Grey System Theory and Multivariate Linear Regression [J]. Procedia Engineering, 2011, 15(15): 32-37.