Carbon Emissions Prediction of Jiangsu Province Based on Lasso-BP Neural Network Combined Model

Xiaodie Liu, Xiangrui Meng* and Xiangqian Wang

School of Economics and Management, Anhui University of Science and Technology, Huainan, China

*Corresponding author e-mail: xrmeng@aust.edu.cn

Abstract. Under the policy of achieving carbon peak before 2030, the paper predicts whether Jiangsu province can achieve the peak of carbon emissions before 2030. Based on the data of Jiangsu province from 2001 to 2018, the paper uses Lasso regression model to screen out 8 significant factors affecting carbon emissions, sets the values of each influencing factors during 2019-2030, and uses the BP neural network model to predict the carbon emissions of Jiangsu province during 2019-2030. The prediction results show that Jiangsu province will achieve the carbon peak in 2023 under the hypothetical scenario, with the peak carbon emissions of 335.1755 million tons.

Keywords: Lasso regression; BP neural network; scenario simulation; carbon emissions prediction.

1. Introduction

The issue of global warming has attracted much attention, and it has become the consensus of all mankind to deal with climate change. According to the assessment report released by the United Nations (IPCC), 90% of global warming is caused by greenhouse gas emissions [1], especially carbon dioxide emissions. Whether carbon emissions can be controlled effectively is related to the future fate and development of mankind. In response to the climate crisis, China has shown its responsibility to tackle climate change. At the seventy-fifth United Nations General Assembly in September 2020, Xi Jinping announced that China will increase its own emissions reduction contributions independently, and strive to achieve the carbon peak before 2030 and achieve carbon neutrality before 2060.

Jiangsu province is one of the most economically developed coastal areas in China, and the processing industry there plays an important role in our country. Therefore, identifying the influencing factors of its carbon emissions and predicting the medium and long-term carbon emissions are of great significance to China's goal of achieving carbon peak and carbon neutrality. This paper takes the data of Jiangsu province from 2001 to 2018 as the study object, uses lasso regression to screen out the influencing factors of carbon emissions, utilizes the screened variables as the input of the BP neural network, then simulates the variable values from 2019 to 2030 and predicts the carbon emissions value of Jiangsu province from 2019 to 2030.
2. Lasso regression model

2.1. Lasso regression

Lasso is a biased estimation method [2] proposed by Tibshirani in 1996. It compacts the coefficients of the model through penalty function, compressing some coefficients to zero and screening out significant variables. Moreover, Lasso model can deal with the multi-collinearity of data on its own [3].

Assuming that \( X \) as the independent variable and \( Y \) as the dependent variable, the normalized value of the observed data obtained from \( n \) times of sampling is the matrix \((X, Y)\), where \( X \) is the matrix of order \( n \times p \) \((n > p)\), \( Y \) is the matrix of order \( n \times 1 \), and the observed values of \( X \) is \( X_i = (X_{i1}, X_{i2}, \cdots, X_{ip})^T \), \( i \in [1, 2, \cdots, n] \) and the observed values are independent of each other, \( Y = (Y_1, Y_2, \cdots, Y_n)^T \). The regression model of \( Y \) and \( X \) is expressed as follows:

\[
y_i = \alpha + \sum \beta_j x_{ij} + \epsilon_i \tag{1}
\]

In formula (1), \( \epsilon_i \sim N(0, \sigma^2) \), according to the definition, \( \bar{\alpha} = \bar{y} \), standardized data \( \bar{y} = 0 \), so tidy formula (1) get formula (2).

\[
Y = \beta X + \epsilon \tag{2}
\]

In formula (2), \( \epsilon \sim N(0, \sigma^2) \), to screen out the significant variables, we need to add a condition to formula (2), and the constraint expression is:

\[
\arg \min_{\{\beta_1, \beta_2, \cdots, \beta_n\}} \|y - \beta X\|^2 \\
\text{s.t.} \sum_{j} |\beta_j| \leq s \tag{3}
\]

In formula (3), \( s = \frac{t}{\sum \beta_j} \), the value range of \( s \) is \([0, 1]\), \( t \) is a harmonic parameter, \( t \geq 0 \), Lasso regression is through constantly adjust values \( t \), reduce model integral regression coefficient, no significant compression variable coefficient, until to 0. Lasso regression is to adjust the value of \( t \) continuously to reduce the overall regression coefficient of the model, and compresses the coefficient of the insignificant variable continuously until it is compressed to 0.

2.2. Analysis of influencing factors of carbon emissions

All data are taken from "Jiangsu Statistical Yearbook", "China Statistical Yearbook" and "China Energy Statistical Yearbook".

2.2.1. Calculation of carbon emissions

For the measurement of carbon emissions, the consumption of raw coal, coke, crude oil, gasoline, kerosene, diesel, fuel oil and natural gas is selected to count the carbon emissions of Jiangsu province [4]. The calculation formula is as follows:

\[
C = \sum_{i=1}^{8} k_i \times f_i \times e_i \tag{4}
\]

In Formula (4), \( k_i \) represents the consumption of energy \( i \), \( f_i \) represents the folding coefficient of energy \( i \), and \( e_i \) represents the carbon emission coefficient of energy \( i \). Carbon emissions can be calculated according to the energy conversion coefficients and carbon emission factors in the guidelines of national greenhouse gas inventories published by OECD in 2002 and IPCC in 2006 [5].
Table 1. Energy conversion coefficients and carbon emission factors

| Index                      | Raw coal | Coke  | Crude oil | Gasoline | Kerosene | Diesel | Fuel oil | Natural gas |
|----------------------------|----------|-------|-----------|----------|----------|--------|---------|-------------|
| Energy conversion coefficients | 0.714    | 0.97  | 1.429     | 1.471    | 1.471    | 1.457  | 1.429   | 13.3        |
| Carbon emission factors     | 0.748    | 0.11  | 0.585     | 0.553    | 0.342    | 0.591  | 0.618   | 0.448       |

2.2.2. Analysis of influencing factors. This paper preliminarily selects 10 factors that may affect carbon emissions from the perspectives of economy, society, energy and environment, and the specific index system is shown in Table 2.

Table 2. Influencing factors of carbon emission in Jiangsu province

| Number | Variables                  | Variable’s description                        |
|--------|----------------------------|-----------------------------------------------|
| A1     | Economic level             | GDP (100 million yuan)                        |
| A2     | Industrial structure       | Percentage of secondary industry in GDP (%)   |
| A3     | Population size            | Total population (10 thousand)                |
| A4     | Urbanization level         | Ratio of Urban Population to Total Population (%)|
| A5     | Energy consumption         | Total energy consumption (tons of standard coal) |
| A6     | Energy structure           | Ratio of coal consumption in total energy consumption (%) |
| A7     | Energy intensity           | Energy consumptions per GDP (tons of standard coal/100 million yuan) |
| A8     | Electricity consumption    | Total electricity consumption (billion kilowatt-hours) |
| A9     | Transportation development | Number of private cars (10 thousand)           |
| A10    | Forest coverage rate       | Proportion of forest area to total land area (%) |

First of all, the data of carbon emissions and influencing factors are normalized to eliminate the possible non-linear relationship between the data. Then, the processed data are imported into the software R and the influencing factors of carbon emissions are analyzed by Lasso regression analysis. K-fold CV cross validation is used to determine value s. The result of CV cross validation is shown in Figure 1. The abscissa is λ, the ordinate is the root-mean-square error in the model, namely value CV. With the increase of λ, the root-mean-square error of the model also increases. The number above the dotted line on the left of the figure indicates that s is at step 8.

![Figure 1. CV residual plot](image-url)
Figure 2 describes the change trajectory of each independent variable. With the increase of the punishment intensity on the horizontal axis, almost all the coefficients of the variables decrease continuously and finally reduce to 0.

Select the optimal value of $\lambda$ and bring it into lasso model to fit the coefficients of lasso regression model. As shown in Table 3, the coefficients of A4 and A7 are compressed to 0 and eliminated by the model. Eight significant variables are selected, which are economic level (A1), industrial structure (A2), population size (A3), total energy consumption (A4), energy structure (A5), electricity consumption (A6), transportation development (A7) and forest coverage rate (A8), respectively.

| Variables | A1     | A2     | A3     | A4    | A5     | A6     | A7    | A8     | A9    | A10   |
|-----------|--------|--------|--------|-------|--------|--------|-------|--------|-------|-------|
| Coefficients $\beta$ | 0.263  | 0.0204 | 0.106  | 0     | 0.176  | 0.034  | 0     | 0.493  | 0.030 | -0.026|

3. BP neural network prediction model

3.1. BP neural network model construction

BP neural network is a multilayer feedforward network trained by error propagation algorithm. It can approximate any continuous function and has a strong ability of nonlinear mapping. As shown in Figure 3, the structure of BP model includes input layer, hidden layer and output layer [6]. Its main idea is to input learning samples, adjust and train the deviation and weight of the network repeatedly through the back propagation algorithm, and make the output value as close as possible to the expected value.

Utilizing the result from Lasso regression, takes the eight influencing factors that economic level, industrial structure, population size, total energy consumption, energy structure, electricity consumption, transportation development and forest coverage rate as input variables, and takes the carbon emissions of Jiangsu province as output variables, then uses the BP neural network model to predict.
3.2. Scenario simulation of influencing factors

According to the macro development planning of China and the regional development planning of Jiangsu province, sets reasonable growth rates for each variable, so as to be able to describe the future development of Jiangsu Province more accurately.

3.2.1. Economic level. The “13th Five-Year Plan” of Jiangsu province pointed out that the average annual GDP growth rate of Jiangsu province is expected to be 7.5%. This paper assumes the GDP growth rate from 2019-2020 is 7.5%, from 2021-2025 is 6.5%, and from 2026 to 2030 is 5.5%.

3.2.2. Industrial structure. The proportion of the secondary industry in Jiangsu province has shrunk by 9% from 2009 to 2018, and the degree of reduction has decreased year by year. According to the national green development strategy, the proportion of the tertiary industry will continue to increase in the future. Therefore, it is assumed that the share of the secondary industry will shrink by 0.3% per year from 2019 to 2020, 0.2% per year from 2021 to 2025, and 0.1% per year from 2026 to 2030.

3.2.3. Population size. The average annual population growth rate of Jiangsu Province was 5‰ from 2001 to 2018, 3‰ from 2009 to 2018, and 2‰ from 2017 to 2018. According to Xu Jiaming's prediction [7], Jiangsu's population is expected to reach the peak in 2022 and then decline year by year. The population change is set as shown in Table 4.

| Table 4. Natural population growth rate of Jiangsu province from 2019 to 2030 |
|-----------------|----|----|----|----|----|----|----|----|
| Year            | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2026-2030 |
| Natural growth rate (%) | 1.5  | 1    | 0.5  | 0    | -0.5 | -1   | -1.5 | -2        |

3.2.4. Total energy consumption. The “13th Five-Year Energy Development Plan” of Jiangsu province pointed out that we will strive to control the total energy consumption at 337 million tons of standard coal (with an average annual growth of 2.2% and an elasticity coefficient of 0.29) by 2020, creating conditions for the total energy consumption to peak around 2025 and the total carbon emission to reach peak before 2030. Therefore, it is assumed that the annual growth rate of energy consumption will be 2.2% in 2019-2020, 1.1% per year from 2021 to 2025, and -0.5% per year from 2026 to 2030.

3.2.5. Energy structure. The “13th Five-Year Energy Development Plan” of Jiangsu province pointed out that we should continue to control and reduce coal consumption, and the proportion of coal consumption in total energy consumption will continue to maintain “negative growth”. It is assumed that the proportion of coal consumption will decrease by 1.5% per year from 2019 to 2020, 1% per year from 2021 to 2025, and 0.5% per year from 2026 to 2030.

3.2.6. Electricity consumption. The average annual growth rate of electricity consumption in Jiangsu province is 4.8% from 2013 to 2018, and 5.5% from 2017 to 2018. Therefore, the growth rate of electricity consumption in Jiangsu province is assumed to be 5.5% from 2019 to 2020, 6% from 2021 to 2025, and 6.5% from 2026 to 2030.

3.2.7. Transportation development. According to the growth trend of private cars in Jiangsu province, the number of private cars will continue to increase in the future. It is assumed that the annual growth rate of private cars from 2019 to 2020 is 9%, from 2021 to 2025 is 7%, and from 2026 to 2030 is 5%.

3.2.8. Forest coverage rate. The national government report mentioned that our country will build a natural protection system with national parks as the main body, the forest coverage rate will reach 24.1%. At present, the forest coverage rate in Jiangsu province has stabilized at 15.2%, and the forest coverage rate for 2019-2020 is assumed to be 15.2%, the forest coverage rate from 2021 to 2025 is 15.5%, and from 2026 to 2030 is 16%.
3.3. Carbon emissions prediction
Select the data from 2001-2014 as the training samples, and the data from 2015-2018 as the test samples. Through the training of the training samples for several times, the network structure is 8-6-1, the training times is assumed to be 1000, the learning rate is assumed to be 0.01, and the minimum error of the training target is assumed to be 0.00001. Use the trained model to predict the test samples, and obtain the error graph between the predicted values and the actual values from 2015 to 2018. As shown in Figure 4, the average absolute error is 0.016315, indicating that the BP neural network model has a high degree of fit for the carbon emission prediction of Jiangsu province under the current data.

![Test results of BP neural network](image)

Figure 4. Test results of BP neural network

Taking the simulation data of influencing factors from 2019 to 2030 into the trained BP neural network model, the carbon emission prediction value of Jiangsu province from 2019 to 2030 is obtained, as shown in Table 5. From the prediction results, it can be seen that under the simulation scenario, the carbon emission of Jiangsu province will continue to rise from 2019 to 2023 and reach the peak in 2023, and then will decrease year by year.

| Year | Predictive value (10 thousand tons) | Year | Predictive value (10 thousand tons) |
|------|-----------------------------------|------|-----------------------------------|
| 2019 | 31731.46                          | 2025 | 33349.70                          |
| 2020 | 32487.27                          | 2026 | 33253.05                          |
| 2021 | 33106.62                          | 2027 | 33146.57                          |
| 2022 | 33403.76                          | 2028 | 33089.06                          |
| 2023 | 33517.55                          | 2029 | 33064.47                          |
| 2024 | 33470.28                          | 2030 | 33056.99                          |

4. Conclusions
To predict the carbon emissions of Jiangsu province from 2019 to 2030 through the establishment of the Lasso-BP neural network combined model, the main conclusions are as follows:

1) Using Lasso regression to screen out the factors that have significant influence on carbon emissions in Jiangsu Province, including economic level, industrial structure, population size, total energy consumption, energy structure, electricity consumption, transportation development and forest coverage rate.

2) Using BP neural network model to train, test and predict, and the prediction error is 0.016315. Under the simulation scenario, the carbon emissions of Jiangsu province will continue to rise from 2019 to 2023 and reach the peak in 2023, and it will decrease year by year after 2023.
Acknowledgments
This work was supported by the National Natural Science Foundation of China under Grant No.51874003 and No. 72074003.

References
[1] IPCC, Climate Change 2006: The scientific basis. Third assessment report of intergovernmental panel on climate change. Cambridge: Cambridge University Press, 2006.
[2] L. Cavalcante, R. J. Bessa, M. Reis, and J. Browell, “LASSO vector autoregression structures for very short-term wind power forecasting,” Wind Energy, vol. 20, no. 4, pp. 657–675, 2017, doi: 10.1002/we.2029.
[3] R. Tibshirani, “Regression Shrinkage and Selection Via the Lasso,” J. R. Stat. Soc. Ser. B, vol. 58, no. 1, pp. 267–288, 1996.
[4] C. Huiqiang and C. Bao, “Dynamic analysis of economic growth and carbon emission in Anhui Province based on decoupling theory,” West. Forum, vol. 23, no. 04, pp. 91–97, 2013.
[5] L. Dong, “The empirical analysis for the impact of consumption structure upgrade on carbon dioxide emission,” Commer. Econ. Res., no. 05, pp. 38–41, 2020.
[6] SKAPUR A D, Building Neural Networks. New York: Addision-Wesley, 1999.
[7] X. Jiaming, “Prediction on the trend of population change after the implementation of the two-child policy in Jiangsu province,” Stat. Manag., no. 09, pp. 3–8, 2019.