Establishment and research of combined forecasting model of carbon

Zuxu Zou, Jiaojiao Huang, and Chenlu Li

(Wuhan Polytechnic University, College of Civil Engineering and architecture, Hubei, Wuhan 430000)

Abstract. This paper analyzes the influencing factors of carbon dioxide emissions from four aspects: Population, economy, industrial structure and energy, then from the carbon emissions, economic development, industrial structure, energy consumption structure to show the status quo of carbon emissions in Hubei Province. Based on the analysis of the influencing factors, the main influencing factors of carbon emission are population, regional gross product and coal consumption. The multivariate linear regression model and the polynomial curve model are established and the error analysis is carried out. The combination weight coefficients of two single models are obtained through the linear programming model and the combination forecasting model is established, finally, the corresponding countermeasures to reduce carbon emissions are put forward.

1 Introduction

In March 2011, Hubei Province was selected as one of the country’s first seven carbon trading pilot provinces and cities. In August 2013, Hubei Province formulated and promulgated the “measures for the administration of carbon emissions trading in Hubei Province”, to complete the system construction including management system, quota allocation, trading platform, Verification Report, etc. The effective use of energy is closely related to the great economic development. Under the extensive economic development pattern of many years, Hubei Province has developed bottlenecks: insufficient industrial resources, low energy utilization efficiency, and worsening environmental problems, it can be seen that the development of intensive and low-carbon economy is indispensable. The study and comparative analysis of the current situation of carbon emissions in Hubei Province and the whole country can not only provide better decision-making recommendations for the environmental protection and low-carbon economic transformation in Hubei Province, it also provides the effective reference for the National Energy Utilization and the carbon emission control, and has certain practical significance. In this paper, the data of carbon emission in Hubei Province is analyzed and compared with the national carbon emission situation, combining the influence factors of population, regional GDP, coal, oil, natural gas and hydropower consumption, the main influencing factors are screened through the grey relational grade evaluation model, and the multiple linear regression and polynomial curve combination model are established to forecast the carbon emission of Hubei Province, based on the assessment of the achievement of the carbon emission reduction target and the analysis of the emission reduction potential in Hubei Province, the corresponding countermeasures for reducing carbon emissions are put forward.

2 The establishment of combination forecasting model

Combining the historical forecast data with the actual value, we can give an optimal weight combination of various forecast models, so that the total error of these historical forecast data is minimum after weighting. There are many ways to build a combination forecasting model, the typical ones are:

(1) standard deviation
Suppose there are n models to be combined. Suppose the standard deviation of Model i is \( \sigma_i \) (\( i = 1, 2, 3, ..., n \)) (the standard deviation is the square root of the average of the sum of squares of the differences between each data and its average),

\[
\sigma = \sum_{i=1}^{n} \sigma_i \quad (i=1,2,3,... n)
\]

To minimize \( \sigma \), then

\[
o_i = \frac{\sigma_i - \bar{\sigma}}{\sigma} \cdot \frac{1}{n-1} \quad (i=1,2,3,... n)
\]

This method first needs to calculate the standard deviation of n models.

(2) sum of squares of errors
Suppose that two single models need to be combined, let them be model a and model b. The errors of the two models are one \( e_a, e_b \) respectively, and let the error sum

\[
e = e_a + e_b
\]

\( e_a, e_b \) are the weight values of models a and b. Because there are positive errors and negative errors, in order to find the minimum value of e, the square of the error is

\[
Var(e) = Var(e_a) + Var(e_b) + 2\rho Var(e_a)e_b + 2\text{cov}(e_a,e_b),
\]
Minimizing \( w_t \) to \( e \) is worth finding
\[
\omega_t = \frac{\text{Var}(e_t) - \text{cov}(e_t, e_\theta)}{\text{Var}(e_t) + \text{Var}(e_\theta) - 2\text{cov}(e_t, e_\theta)}
\]  
(2)

The principle is that \( \text{Var}(e_t) \rightarrow x, \omega_t = 0; \text{Var}(e_\theta) \rightarrow x, \omega_\theta = 1 \) the more reliable model a is, the smaller the prediction error, the larger the model a’s weight coefficient, and vice versa.

The above two methods, the principle is similar, is based on the error minimum establishment function solution. Then we introduce the third method, the principle is the same as the second method, but we need to solve the weight value with the help of the established linear programming model, the calculation method is simpler and the results are more intuitive. By comparing the errors of the combination forecasting model and the single forecasting model, the rationality of the weight value and the superiority of the combination forecasting model can be tested.

Because there are large and small errors in the prediction model, and the single model can only contain local carbon emission information, in order to make full use of the effective information reflected by the single model and reduce the randomness of the single model prediction, consider building a combined forecasting model. The combination forecasting model can effectively use the information of multiple samples and determine the contribution value of single model to the combination model by the weight value, so the error can be minimized.

There are \( m \) kinds of methods to predict the same object \( m \geq 2 \), and. If the actual value of the t period is \( \hat{y}_t \), the forecast value is \( \text{wt} \), and the forecast error is \( \epsilon_t \), then for the \( m \) method, the corresponding t period actual value, the forecast value and the forecast error are respectively \( y_t, \text{wt}, \epsilon_t \), so
\[
\sum_{i=1}^{m} \omega_i = 1, \quad \sum_{i=1}^{m} \epsilon_i y_i = \sum_{i=1}^{m} \epsilon_i \text{wt}_i, \quad \sum_{i=1}^{m} \epsilon_i \epsilon_i = \epsilon^2
\]
The combination forecast Error Square sum is \( e^2_{ct} \), so
\[
e^2_{ct} = \sum_{i=1}^{m} \epsilon_i^2 = \sum_{i=1}^{m} \omega_i \epsilon_i + \sum_{j=1}^{n} \epsilon_j ...
\]
(3)

Based on the principle of the combination model, that is, the sum of the squares of the prediction error is the minimum, the linear programming equation is established to obtain the weight of each prediction method, and the linear programming model is as follows:

\[
\begin{align*}
\min & \quad e^2_{ct} \\
\text{subject to} & \quad \sum_{i=1}^{m} \omega_i = 1 \\
& \quad \omega_i > 0, i = 1, 2, ..., m
\end{align*}
\]
(4)

By using the error and variance between the predicted value and the actual value of the multivariate linear equation and the polynomial curve model, the optimal solution of the linear programming equation can be obtained, and then the weight value can be obtained.

3 Analysis of combinatorial forecasting models

Using the multiple linear regression equation and the quadratic curve model, the carbon dioxide emissions of Hubei Province from 2018 to 2022 can be measured. The actual value of carbon emissions from 2018 to 2022 can be calculated through the data found in the statistical yearbook. The value is compared with the actual value, and then the error percentage can be calculated, as shown in Table 1:

| Year | Actual value | Multivariate linear regression equation | POLYNOMIAL quadratic equation |
|------|--------------|-----------------------------------------|--------------------------------|
|      | Predicted value | Percentage error (%) | Predicted value | Percentage error (%) |
| 2018 | 29560 | 28071.9 | 4.39 | 28267 | 3.72 |
| 2019 | 28193 | 30101.7 | 6.77 | 29455.6 | 4.48 |
| 2020 | 30736 | 31086.9 | 1.14 | 30556 | 0.59 |
| 2021 | 30125 | 30469.7 | 1.14 | 31574.6 | 4.81 |
| 2022 | 33183 | 31932.3 | 3.77 | 32508.5 | 2.03 |

Combining the data analysis of Table 1 with the formula (1) ~ (2), we get \( \omega_1 = 0.096; \omega_2 = 0.904 \), so the combined forecasting model can be expressed as:
\[
Y = 0.096 \times (-197843.9 + 34.991x_1 + 1.676x_2) + 0.904 \times [-4144240 + 2079t - 42.4(t - 1999)]
\]
(5)

4 Test of combination forecasting model

4.1 the combination forecasting model is compared with the multiple LINEAR REGRESSION MODEL

Based on the consideration of population and raw coal consumption, the combined forecasting model also considers the Change Law of carbon emission with time (year) in Hubei Province, which is equivalent to giving more influence factors to the sequence, the dependence of the predicted results on the data of population and raw coal consumption is reduced. Using the indexes MAD, MSE and MAPE, the errors of the combined forecasting model and the multiple linear regression equation were compared and analyzed:

| Year | Actual value | Predicted value | Percentage error (%) | Predicted value | Percentage error (%) |
|------|--------------|----------------|----------------------|----------------|----------------------|
| 2018 | 29560 | 28071.9 | 4.39 | 28267 | 3.72 |
| 2019 | 28193 | 30101.7 | 6.77 | 29455.6 | 4.48 |
| 2020 | 30736 | 31086.9 | 1.14 | 30556 | 0.59 |
| 2021 | 30125 | 30469.7 | 1.14 | 31574.6 | 4.81 |
| 2022 | 33183 | 31932.3 | 3.77 | 32508.5 | 2.03 |

Table2. Comparison of errors between multiple linear regression model and combined forecasting model

| Indicators | Model | MULTIPLE LINEAR REGRESSION MODEL | Combined Forecasting Model |
|------------|-------|----------------------------------|-----------------------------|
| MAD        | 1028.62 | 927.75                           |
| MSE        | 1421707 | 1068985                          |
| MAPE       | 3.44   | 3.11                             |
error index values of the combined forecasting model are smaller than those of the multiple linear regression model, which proves that the combined forecasting model is superior to the single multiple linear regression model and can achieve the goal of reducing the error.

4.2 the combination forecasting model is compared with the polynomial curve model

The combined forecasting model retains the polynomial curve model and gives higher weight, but does not exclude the multivariate linear regression and gives up its use. It shows that both models are suitable for fitting and predicting CO2 emissions and that each of the two single models has its own focus. While affirming that the carbon emission in Hubei Province has obvious characteristics of changing with time, the effect of population and raw coal consumption on carbon emission can be incorporated into the combined forecasting model to reduce the fluctuation of data. Using the indexes MAD, MSE and MAPE, the errors of combined forecasting model and polynomial curve model are compared and analyzed as follows: Table 3:

| Indicators | Polynomial curve model | Combined forecasting model |
|------------|------------------------|---------------------------|
| Mad        | 931.94                 | 927.75                    |
| MSE        | 1075500                | 1068985                   |
| MAPE       | 3.13                   | 3.11                      |

Through comparative analysis, although the error index value of the polynomial curve model is close to that of the combination forecasting model, it is not difficult to see that the index value of the combination forecasting model is smaller, which proves that the combination forecasting model is better than the single polynomial curve model, the prediction accuracy is improved when the multivariate linear regression equation is used, so the combined prediction model can be used to predict the CO2 emission in Hubei Province.

5 Summary

On the basis of factor analysis, the model combination and prediction and analysis are carried out in this paper. The contents include:

1. the linear programming model is established on the principle of minimum variance, the combination weights of multivariate linear regression model and polynomial curve model are 0.096 and 0.904 respectively. The composite model can be expressed as:
   \[ y = 0.096(-197843.9 + 34.991x + 1.676z) + 0.904(-4144204 + 207920.79 - 42.4(t + 1999)) \]

2. using MAD, MSE and MAPE to test the superiority of the combination forecasting model, the combination forecasting model is superior to the single forecasting model.

3. the forecast value of CO2 emission in Hubei Province from 2018 to 2022 is calculated by using the combined forecasting model, and it is found that the growth rate is 1%-2% , and the carbon emission is increasing year by year.

4. to analyze the result of carbon emission prediction, and to calculate the emission reduction potential of Hubei Province by sensitivity analysis method. The population is more sensitive to carbon emission.

Fund

This paper is based on the Philosophy and Social Science Research Project of the Hubei Provincial Department of Education (19Y078) “Influencing factors of Hubei Province's carbon emissions under the new economic normal and the establishment and research of their combined forecasting model”

References

1. Ehrlich P, Holdren J. P. Impact of population [J]. Science, 171: 1212-1217(1971)
2. Thomas Dietz, Eugene a Rosa. Effects of population and influence on CO2 emissions [J]. Ecology, 94 : 175-179(1997)
3. Rattanar Pani, Ujjaini Mukhopadhay. Identifying The major players behind increasing global divergences emission decomposition analysis [J]. The Environmentalist, 30(2) : 183-205(2010)
4. Brant Liddle Sidney Lung. Age-structure, urbanization, and climate change in developed countries: revisiting STIRPAT for disaggregate population and p-consumption related Impacts. Popul Environ, (1)007-031.(2010)
5. Impact of Carbon Dioxide Emission on GNP Growth: Interpretation of Proposed Scenarios [R] . Presentation to the Energy and Industry Subgroup, Response Strategies Working Group, IPCC, Paris, (1989)
6. Hulten C.DEVISA Index numbers[J].Econometrical, 41(6) : 1017-1025(1973)
7. Howarth R. B. Manufacturing Energy use in eight OECD countries [J].Energy Economics,13(2) : 135-42(1991)
8. Ang B. W. , Zhang F. Energy, 23(6) : 489-495(1998)
9. Potential futures for road transportation CO2 emissions in the Asia Pacific [J]. Asia Pacific Viewpoint,48(3) : 355-377(2007)
10. CHANGE IN CO2 emissions and its transmission between Korea and using Japan International input-output analysis [J].Ecological Economics,,58(4) : 788-800(2006)
11. Bin Shui, Robert C. Harriss. The role of CO2 granditent in US-China Trade [J]. Energy Policy, 30(6) : 4062-4070,(2006)
12. Druckman A, Jackson T. The carbon footprint of UK households 1990-2004: Asocioeconomic ally
segregated, quasi-multi-regional input-output model
[J]. Ecological Economics, 30(8): 2066-2077(2009)