ATLANTIS PRESS
Advances in Engineering, volume 126
5th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2017)

Study on the Influential Factors of Carbon Emissions from Thermal Power Industrial in China Using Partial Least Square Regressive Model

Mei Liu [a], Jing-Yan Li [b]
School of North China Electric Power University BaoDing, Hebei 071000, China
[a]liumei_w@163.com, [b]15932022562@163.com

Keywords: carbon emissions, thermal power, the Partial Least Square Regressive model, influential factors.

Abstract. This paper firstly analyzed the current thermal power situation in China, and then determined several influential factors by referring to the Kaya identity. Finally, the Partial Least Square Regressive model is applied on this issue. To validate the model, this paper calculated the actual Carbon emissions from thermal power industrial by means of the formula formulating by the Intergovernmental Panel on Climate Change's 2006 guide from 1991-2014 across China. Under the model, we find that the amount of population, consumption of the coal and the ritio of the thermal power are the main factors influencing the carbon emissions from thermal power industrial.

Introduction

Due to the rapid grow of the economy of China, the consumption of primary energy and the emission of greenhouse gases has increased rapidly, meanwhile China has become the most greenhouse gases emission country in 2007. With the development of China's industrialization and urbanization, the scale of thermal power is becoming larger and larger. The installed capacity of China has up to 1.36 billion kilowatts by the end of 2014, which is ten times of the 0.138 billion kilowatts in 1990s, improving by 10% average annual growth. Among it, the installed capacity of non-fossil energy has up to 0.45 billion kilowatts, with the proportion surpassing 25 percent at the first time [1]. Therefore, to investigate how to control the carbon emissions basing the unceasing development of thermal power situation makes sense.

At present, related scholars have done a serious of researches centering on the calculation and prediction of carbon emissions. Bo-Qiang Lin adopted LMDI and STIRPA models to analyze the main factors of carbon emission per capita of China [2]. Xiao-Yan Shen adopted Grey model to predict the carbon emissions and conclude that we must adjust the energy consumption structure if the population policy will not change [3]. Chang-Kai Wang also adopted LMDI model to investigate the influential factors of electricity industrial, which found that the economy scale is the most important factor [4].

But by far, there has rarely paper to investigate the carbon emission from the thermal power industrial using the Partial Least Square Regressive model, so this paper plays a very significant role in controlling the carbon emission in China.

Establishment of the model

Selection and analysis of the indicators.

Kaya identity is created by Kaya Yoic, a Japanese scholar, and is widely used to calculate carbon emission and analyze the influential factors [5-7]. The model and its specific definitions of all variables are expressed as follows:

$$C = \frac{C}{E} \cdot \frac{E}{GDP} \cdot \frac{GDP}{P} \cdot P \qquad (1)$$

Where C represents the carbon emission, E represents the consumption of energy, GDP is the abbreviation of Gross Domestic Product, and P represents the amount of population. Based on this, and combing the characteristic of the thermal power industrial, this paper will select indicators from the generation, distribution, and usage of the power three aspects. Thus, this paper selected the ritio of thermal power, the consumption of the coal, the distribution loss of the power, the population and the GDP as the investigated indicators.

Introduction of the model.

In view of the issue that the variables we studied may have multiple linear correlation, this paper decide to adopt the Partial Least Square Regressive model to solve the problem.

The Partial Least Square Regressive method integrate principle component analysis, canonical correlation analysis and linear-regression analysis which can acquire the maximum to improve the precision of correlation analysis of the model [8]. The specific model process is as follows:

Firstly, we should acquire the observation sample of independent variables $\{x_1,x_2,\ldots,x_p\}$and dependent variables$\{y_1,y_2,\ldots,y_q\}$ and then get the data table $\mathbf{X}=[x_1,x_2,\ldots,x_p]$, $\mathbf{Y}=[y_1,y_2,\ldots,y_q]$. The main line of this model is: Respectively extract the ingredients $\mathbf{t_1}$ and $\mathbf{u_1}$, where $\mathbf{t_1}$ is the linear combination of the independent variables $\mathbf{x_1,\ldots,x_p}$, and $\mathbf{u_1}$ is the linear combination of the dependent variables $\mathbf{y_1,\ldots,y_q}$. In addition, $\mathbf{t_1}$ and $\mathbf{u_1}$ need to be satisfied that they should carry the most variability information and be up to the maximum correlation. The specific steps of algorithm are as follows:

(1) Standardizing the data of **X**, **Y** and respectively obtaining the standard array $\mathbf{E_0}$ and $\mathbf{F_0}$.

(2) Calculating the regularization eigenvector $\boldsymbol{w_1}$ corresponding to the maximum eigenvalue of matrix $\mathbf{E_0^T F_0 F_0^T E_0}$, and acquiring the first component $\mathbf{t_1=E_0 w_1}$.

(3) Calculating respectively the regression equation $E_0=t_1p_{1T}+E_1$ and $F_0=t_1r_{1T}+F_1$, thereinto $\mathbf{E_1}$ and $\mathbf{F_1}$ is the residual matrix of $\mathbf{E_0}$ and $\mathbf{F_0}$, and $\mathbf{p_1}$ and $\mathbf{r_1}$ can expressed as the follows:

$$p_1 = \frac{\mathbf{E}_0^T t_1}{\|t_1\|^2} \qquad (2)$$

$$r_1 = \frac{\mathbf{F}_0^T t_1}{\|t_1\|^2} \qquad (3)$$

(4) Testing the convergence. In this section, cross validation test should be applied to validate the precision: if it dissatisfies the equation $\mathbf{E_0=E_1}$, $\mathbf{F_0=F_1}$, it should repeat the above steps up to it meets the requirements. Finally, we can get the following equation:

$$F_0 = t_1 r_1^T + t_2 r_2^T + \cdots + t_h r_h^T \qquad (4)$$

(5) Establishing the regression model of $\mathbf{E_0}$ and $\mathbf{F_0}$, we can get $\mathbf{F_0=E_0B+F_A}$.

$$\mathbf{B} = \sum_{j=1}^{h} [\prod_{i=1}^{j-1} (\mathbf{I} - w_i p_i^T) w_j] r_j^T \qquad (5)$$

Where **I** is unit matrix and $\mathbf{F_A}$ is residual matrix.

(6) Establishing the original regression equation of independent variable **X** and dependent variable **Y** by de-normalizing $\mathbf{E_0}$ and $\mathbf{F_0}$, we can get the following equation:

$$\mathbf{Y} = \mathbf{X}\mathbf{B}_g + \mathbf{C} \qquad (6)$$

Where

$$\mathbf{B}_g = \mathbf{D}_x^{-1} \mathbf{B} \mathbf{D}_y \qquad (7)$$

$$\mathbf{C} = m_y - m_x \mathbf{B}_g \qquad (8)$$

In the above equation, $\mathbf{D_x}$ and $\mathbf{D_y}$ respectively represent the diagonal matrix consisting the variance of the list of **X** and **Y** , and $\boldsymbol{m_x}$ and $\boldsymbol{m_y}$ respectively represent the column vector consisting the lists of **X** and **Y**.

Model for evaluating carbon emissions of the thermal power industrial

The original data.

The original data are shown as the Tab. 1.

Tab. 1 Index data

year	the ritio of thermal power/%	the consumption of the coal/ hundred million tons	the distribution loss of the power/billoin kilowatt hour	Population/ hundred million people	GDP / hundred billion yuan
1991	81.53789388	3.01191	48.75	11.5823	22.0056
1992	82.42963631	3.34594	54.6	11.7171	27.1945
1993	81.58327965	3.68317	61.91	11.8517	35.6732
1994	80.37238169	4.005311	59.61	11.985	48.6375
1995	79.81502982	4.444015	74.45	12.1121	61.3399
1996	81.26793948	4.880859	76.95	12.2389	71.8136
1997	81.45389477	4.89792	79.84	12.3629	79.715
1998	80.95523924	4.948926	79.09	12.481	85.1955
1999	82.34809973	5.116352	86.19	12.5909	90.5644
2000	82.19164945	5.58112	93.67	12.6743	100.2801
2001	79.91828741	5.979786	103.35	12.7627	110.8631
2002	80.90326481	6.860003	116.87	12.8453	121.7174
2003	82.71624324	8.196584	126.07	12.9227	137.422
2004	81.49511417	9.196156	142.06	12.9988	161.8402
2005	81.88508395	10.366294	170.65	13.0756	187.3189
2006	82.68748277	11.876391	185.88	13.1448	219.4385
2007	82.97694687	12.791725	206.17	13.2129	270.2323
2008	80.47812442	13.265237	213.79	13.2802	319.5155
2009	80.29774003	14.396729	225.82	13.345	349.0814
2010	79.19665523	15.374248	256.82	13.4091	413.0303
2011	81.34274839	17.557853	270.07	13.4735	489.3006
2012	78.05054586	18.353099	289.62	13.5404	540.3674
2013	78.1901967	19.517738	314.07	13.6072	595.2444
2014	75.55694406	18.45253	309.99	13.6782	643.974

Empirical studying.

This paper investigates the carbon emissions of thermal power industrial based on the five factors mentioned above, that is the ratio of thermal power, the consumption of the coal, the distribution loss of the power, the amount of population, and GDP. They are respectively expressed as x_1, x_2, x_3, x_4, and x_5.

Using the Partial Least Square Regressive method mentioned above, we find that the first two gradients can account for 97.55 percent for the independent variable; thus, we only choose two ingredients. After implementing the steps mentioned above, the final regression equation can be expressed as follows:

$$y = -1596.856 + 4.627x_1 + 14.853x_2 + 0.888x_3 + 107.28x_4 + 0.357x_5 \qquad (9)$$

Using the model to calculate the carbon emission of the thermal power industrial, the results are shown as Tab. 2.

Model testing.

To validate the model, this paper collected some data about the amount of coal, petroleum, and natural gas by using the formula of carbon emissions formulating by the Intergovernmental Panel on Climate Change's 2006 guide from 1991-2014 across China. [9-11] The formula is as follows:

$$E = \sum_{i=1}^{3} EC_i \times EF_i \times f_i \qquad (10)$$

Where E represents the production of carbon emissions, EC_i represents the consumption of the *i*th energy, EF_i represents the coefficient of carbon emissions of different kinds of energy, f_i represents the value of the standard coal coefficient of different kinds of energy. In this paper, EF_{coal}=0.7266,

$EF_{petroleum}$=0.5588, $EF_{natural\ gas}$=0.4234, and f_{coal}=0.7143kg standard coal/kg, $f_{petroleum}$=1.4286kg standard coal/kg, $f_{natural\ gas}$=1.33kg standard coal/kg. The original data and the results are shown as Tab. 3:

Tab. 2 The prediction values of carbon emissions

year	The predicted values of carbon emissions/million tons
1991	118.810730586702
1992	149.402467319142
1993	174.447677861852
1994	190.512676121787
1995	225.787677740382
1996	258.556097798407
1997	278.355931260834
1998	290.765132202339
1999	319.703314003876
2000	344.934217554887
2001	362.187237345474
2002	404.551267481748
2003	454.862240149525
2004	495.124591231617
2005	557.011640548425
2006	615.551197195959
2007	673.917196737078
2008	700.949564420969
2009	745.095457473019
2010	811.720268186729
2011	899.954845194162
2012	939.277343971175
2013	1005.6623447536
2014	999.032937460309

Tab. 3 The original data and the results

year	Coal/ten thousand tons	Petroleum/ ten thousand tons	Natural gas/ hundred million cubic meters	Actual carbon emissions/million tons
1991	30119.1	1270.5	6.4	166.8240763
1992	33459.4	1299.7	6.5	184.3993153
1993	36831.7	1601.16	8.23	204.4058827
1994	40053.11	1178.4	8.31	217.7549398
1995	44440.15	1358.46	7.9	241.9384668
1996	48808.59	1238.58	7.45	263.6287793
1997	48979.2	1662.11	21.28	268.6741077
1998	49489.26	1304.76	16.05	268.1741282
1999	51163.52	1228.56	12.68	276.0656334
2000	55811.2	1178.15	15.21	299.927621
2001	59797.86	1213.55	13	320.7769491
2002	68600.03	1275.57	11.05	366.8464229
2003	81965.84	1394.4	13.24	437.2883098
2004	91961.56	1653.49	19.03	491.5615016
2005	103662.94	1306.42	30.12	550.1467151
2006	118763.91	931.25	57.56	627.0725352
2007	127917.25	718.55	80.68	674.1832703
2008	132652.37	504.08	81.97	697.1195598
2009	143967.29	393.17	134.24	757.9032113
2010	153742.48	385.33	161.78	810.1257132
2011	175578.53	319.76	197.83	924.9636877
2012	183530.99	292.39	204.41	966.3898197
2013	195177.38	265.08	218.65	1027.419662
2014	184525.3	254.09	228.19	972.5837464

*The data are all derived from China Energy Statistical Yearbook (physical quantity)

After calculation, we get the average relative residual (relative residual equals to the predicted carbon emissions subtracts out actual carbon emissions and then divided by the predicted value) of the actual values and predicted values, which reads 0.0709187. That means the model imitation has a good result. The algorithm of average relative residual is that: firstly, using the predicted values subtract out the actual values and then dividing by the actual values, finally averaging these numbers.

By observing the regression equation, we can easily find that the population, consumption of coal and the proportion of the thermal power are the main factors influencing the carbon emissions from thermal power industrial.

Conclusions

This paper firstly analyzed the current thermal power situation in China, and then determined several influential factors by referring the Kaya identity. Finally, the Partial Least Square Regressive model was applied on this issue. To validate the model, this paper calculated the actual Carbon emissions from thermal power industrial by means of the formula formulating by the Intergovernmental Panel on Climate Change's 2006 guide from 1991-2014 across China. Through calculation, the average relative residual is 0.0709187. Under the model, we find that the population, consumption of coal and the proportion of the thermal power are the main factors influencing the carbon emissions from thermal power industrial.

In addition, the structure of China's electric power is mainly the thermal power and the energy structure "rich coal, lack oil and miss gap" determined the thermal power industrial mainly used coal. This phenomenon results in China becoming the biggest carbon emission country in the world. To reduce the carbon emission, China should go on optimizing the power structure.

Based on the results of the model, and with the implement of two-children policy in China, the population will continue to increase and the demand for power will increasingly crease. Thus, how to contain the emission of carbon emission is becoming a critical problem for China. For the moment, optimizing energy structure in the thermal power industrial and advocating wind power generation and solar power generation are effective ways.

References

[1] China Electricity Council, Present situation and expectation of China's electricity industrial(2014)[R/OL], Information on http://www.weld.labs.gov.cn

[2] Bo-Qiang Lin, Zhu-Jun Jiang, Environment Kuznets curve prediction and influential factors analysis of China's carbon emission, Management World, 2009(4):27-36

[3] Xiao-Yan Shen, The analysis and prediction of China's carbon emission, Statistic and Decision, 2010(19):90-92

[4] Chang-Kai Wang, Hong-Zuo Xie, Analysis on Dynamic Characteristics and Influencing Factors

of Carbon Emissions from Electricity in China, China Population, Resources And Environment, 2015(4)

[5] Dalia Streimikiere, Kaya identity for analysis of the main drivers of GHG emissions and feasibility to implement EU "20–20–20" targets in the Baltic States, Renewable and Sustainable Energy Reviews, 1108-1113.

[6] Yu-Zhe Wu ,The impact of urbanization on carbon emissions in developing countries: a Chinese study based on the U-Kaya method, Journal of Cleaner Production, 589–603

[7] Georgios Mavromatidis, A strategy for reducing CO2 emissions from buildings with the Kaya identity – A Swiss energy system analysis and a case study, Energy Policy, 343-354

[8] Lin Yao, Quality monitoring method of strip hot-dip galvanizing based on partial least squares regression, Journal of University of Science and Technology Beijing, 28-33

[9] Hua-Ding Shi, Yong-Qing Qi, The Research of Greenhouse Gas Accounting System for Power Industry, Advances In Climate Change Research, 1673-1719 (2010) 01-0040-07

[10]Zhuo Li, Ning-Zhang, Studied on the accounting method of carbon emissions from power generation industrial, Artical of Science and Technology, 12-13

[11]Shen-Shen Su, Jing-Yang Zhao, Greenhouse Gas Emissions from Power Sector in China from 1990-2050.