http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (18): 4529-4533, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.4529.4533

Application of Correlation Coefficients Theory for the Prediction of Chinese Carbon Emissions from Primary Energy

Zhou Jianguo and Zhang Xigang
School of Economics and Management, North China Electric Power University, China

Abstract: The text is to establish a combined forecasting model based on the correlation coefficients theory, in consideration of the Chinese Carbon emissions from primary energy:uncertainty, imperfection and small sample properties. The Multiple regression prediction model not only consider the time factor but also takes into account the causal relationship between variables; the BP neural network is suitable for small samples, poor information system prediction;the gray system model can filter random amount mixed in raw data and find out the hidden rules in a volatile time series. We research combination forecasting method, from the point of view of the relevant indicators, used their advantage and theoretical correlation coefficient.This combined forecasting model is better than the traditional forecasting methods, It is able to improve the accuracy of the single prediction model. Finally, we predict the same period Chinese carbon emissions to verify the effectiveness of combination forecasting model, based on the data of the Chinese carbon emissions from 2002 to 2011 as well as the population, GDP and total energy consumption in the same period.

Key word: Correlation coefficient, carbon emissions, BP neural network, multiple regression, grey system

INTRODUCTION

Chinese energy consumption is given priority to the fossil energy, with the development of our national economy, the energy demand is increasing year by year, Chinese CO2 emissions increased significantly. According to related statistics, in 2007, Chinese CO2 emissions from fossil fuel consumption has surpass the United States, as the world's first big CO2 emitter. In 2009 (Jiang, 2011), the Chinese primary energy consumption reached 2.177 billion tons of oil equivalent, was 8.7% higher than in 2008, 19.5%r of the world primary energy consumption, almost the same with the (Cao *et al*., 2011) United States consumption. Along with our country economic growth and rigid demand of energy, carbon emissions will surely be a difficult task. In the domestic energy supply and demand contradiction and growing global climate warming trend is more and more serious today, quantitative analysis of the CO2 emissions, to our country for energy conservation and emission reduction of the work, has a strong practical significance (Ding *et al*., 2010).

For the prediction of CO2 emissions, there have been many scholars and research institutions that forecasted Chinese future carbon emissions from different point of view. But on the one hand, these studies mostly with energy-economic as the theme, seldom consider economic development level and other related factors in this complicated system of population, On the other hand, in the research of CO2 emissions problem, the method is given priority to of linear prediction, this makes the models have certain limitations in the prediction research (Wang *et al*., 2010).

Chinese CO2 emission is nonlinear system, influenced by many factors and emission rule complex. In addition, Chinese CO2 emissions prediction problem is a typical "small sample" analysis system. Taking into account these two factors, the paper uses Multiple regression prediction model, the BP neural network and Support Vector Machine to predict the CO2 emissions, in order to improve the prediction accuracy, using the correlation coefficients theoretical methods to overcome the shortcomings of traditional single forecasting methods.

ESTABLISH AND EVALUATE THE CORRELATION COEFFICIENT COMBINATION FORECASTING MODE

The combination forecasting model building process as follows: First, we will establish multiple regression prediction model and secondly, establish BP neural network model and then we establish gray prediction model and finally, according to the correlation coefficient principle, determine the weight coefficient of combination forecasting model and establish the combination forecasting model (Chen *et al*., 2012).

Corresponding Author: Zhou Jianguo, School of Economics and Management, North China Electric Power University, China

Select a sample of carbon emissions data: Since there is no direct announcement of carbon dioxide emissions, it must be estimated using the relevant method. Reference to the IPCC (2006) and the National Coordination Committee on Climate Change and the National Energy Research Institute (2007) method and the combination of Cha Donglan, Zhou Qun calculation method, we calculate China's annual carbon emissions. In this paper we use data, from 2002 to 2011 as a sample.

Establish multiple regression (MR) prediction model and its application: Through the relevant macroeconomic variables cointegration and causality test, the results showed that: it exists the obvious causal relationship between carbon emissions, population, Gross Domestic Product (GDP), the energy consumption. Therefore, we consider these economic indicators as explanatory variables and our carbon emissions as the dependent variable to establish multiple linear regression model.

$$E = b_0+b_1P+b_2GDP+b_3TEC$$

where, e is carbon emissions; P is Population; TEC is energy consumption.

According to the literature Yang (2008) proposed a method, using EXCEL table to achieve and test the multiple linear regression model, we input the date in Table 1 to EXEL table, we get: $b_0 = 1529637$ $b_1 = -2.3304$ $b_2 = 0.099036$ $b_3 = 0.99096$ and Multiple R = 0.999045, This shows that the variables P, GDP, TEL and the dependent variable E have a high positive correlation. Adjusted $R^2 = 0.997136$, This proves that the independent variables can explain the dependent variable E of 99.713%. It can be that we build a valid ternary linear regression model. As follows:

$$E = 1529637-12.3304P+0.099036GDP+0.99096TEC$$

Establish the BP neural network model: w_{ji} and θ_j are Initializated by a random number (typically 0-1).Wherein w_{ji} is the neuron-i to neuron-j connection weights and θ_j is the threshold of the neuron j (the hidden layer and output layer); Set pretreated training sample set $\{x_{pt}\}$ and the corresponding desired output set $\{y_{pt}\}$, where in p, t denote the sample class and the number of input vectors; calculate Q (the layers output of neurons; For the output layer neurons, its input and output are same, $Q_{pi} = x_{pi}$, Wherein x_{pi} is the i value of the section p samples. For the hidden layer and output layer, neurons's the output operation is as follows (Li, 2012) :

$$Q_{pj} = f[\sum_i w_{ij} Q_{pi} - \theta_j]$$

Calculate neurons's error signal of the respective layers.

Output layer:

$$\delta_{pj} = (y_{pj}-Q_{pj})Q_{pj}(1-Q_{pj})$$

Hidden layer:

$$\delta = Q_{pi}(1-Q_{pi})\sum_i \delta_{pj} w_{pj}$$

The backpropagation, correction weight:

$$w_{ij}(1/1) = \frac{w_{ij}(l)}{\alpha \delta_{pj} Q_{pj}}$$

In the above equation, α represents the learning speed of the algorithm.

Finally, calculate errors:

$$E_r = \sum_p \sum_k \frac{(Q_{pk} - y_{pi})^2}{2}$$

When E_r is less than a given error, the network training is finished.

Establish the grey system GM (1, 1) prediction model: According to Table 1 the original data of Chinese carbon emissions from 2002 to 2011 year.

We establish gray series:

$$x^{(0)} = [x^{(0)}(1), x^{(0)}(2), \ldots\ldots, x^{(0)}(t)]$$

$$(T = 1, 2, \ldots 10)$$

Wherein, $x^{(0)}$ represents the China's carbon emissions in the t period (Li, 2003).

Table 1: Original sample data

Years	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
E	103474	120661	139781	154877	169930	183952	189051	198696	207523	218311
P	128453	129227	129988	130756	131448	132129	132802	133450	134091	134101
GDP	120333	135823	159878	184937	216314	265810	314045	340903	401202	410231
TEC	147793	171846	199154	219949	241345	261433	269007	282729	296994	301102

In order to decrease the volatility of the raw data, the original series was conducted to generate a cumulative process, get a new series (Xie, 2008):

$$x^{(1)} = [x^{(1)}(1), x^{(1)}(2), \ldots\ldots, x^{(1)}(t)]$$

and:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$$

Because the sequence of $x^{(1)}(k)$ with the law of exponential growth and the first order differential equation is exponential growth form of solution, so we can think of $x^{(1)}$ sequences satisfy the first-order linear differential equation model:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u$$

For A and U values, we put it on the discretization to:

$$x^{(0)}(k+1) + \frac{1}{2}a[x^{(1)}(k+1) + x^{(1)}(k)] = u$$

Take on different k value can be obtained by the following equation:

$$\begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(2)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(n-1) + x^{(1)}(n)] & 1 \end{pmatrix} \begin{pmatrix} a \\ u \end{pmatrix}$$

Abbreviated as: $Y_n = Ba$.
Using matrix derivation equation can be obtained:

$$\hat{A} = (B^T B)^{-1} B^T Y_n = \begin{pmatrix} \hat{a} \\ \hat{u} \end{pmatrix}$$

Gray prediction model as:

$$\hat{x}^{(0)}(k+1) = (e^{-\hat{a}} - 1)(x^{(0)}(1) - \frac{\hat{u}}{\hat{a}})e^{-\hat{a}k} \quad (k = 0, 1, 2\ldots)$$

Input the the original data in the Table 1 to the MATLAB software can calculate, We can get a result, $\hat{a} = -0.06464$, $\hat{u} = 123055.6$. For it take the posterior variance test, posteriori error ratio C = 0.16717, small error probability p = 1. According to the Table 2, the evaluation of grey prediction model, the accuracy of the model is best.

Table 2: Comprehensive assessment of predictive models: The small error probability (P) and posterior error ratio (C)

	Best	Good	Qualified	Unqualified
P	>0.95	>0.8	>0.7	≤0.70
C	<0.35	0.35 = C<0.5	0.5≤C<0.65	≤0.65

Table 3: Result of singer model

Years	Actual value	MR	BP	GM
2002	103474	104129	89811	125639
2003	120660	119955	113583	134028
2004	139781	140016	133904	142977
2005	154876	153635	154563	152524
2006	169930	169412	169343	162708
2007	183952	185824	191896	173572
2008	189051	189807	192160	185162
2009	198696	198075	202137	197526
2010	207522	210279	212255	210714
2011	218310	215121	225728	224784

Establish combined forecasting model based on correlation coefficient (CFMC): Assuming our actual carbon emissions (Ding, *et al*, 2010) is $\{x_t, t = 1, 2, \ldots, n\}$, there are m types of single forecasting methods, $\{x_{it}, t = 1, 2, \ldots, n\}$ as forecasting value series of i-th method, $i = 1, 2, \ldots m$. Suppose A is the predictive value of the combination model, $l_1, l_2, \ldots, l_m$ is the weighting factor and:

$$\sum_{i=1}^{m} l_i = 1 \quad (l_i \geq 0,\ i=1,2,\cdots m)$$

So, for time t satisfy combination forecasting deviation:

$$\hat{e}_t = \sum_{i=1}^{m} l_i e_{it}$$

Among: $e_{it} = x_{it} - \bar{x}_i$, $e_t = x_t - \bar{x}$, $\hat{e}_t = \hat{x}_t - \bar{\hat{x}}$ and:

$$\bar{x}_i = \frac{1}{n}\sum_{t=1}^{n} x_{it},\ \bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t,\ \bar{\hat{x}} = \frac{1}{n}\sum_{t=1}^{n} \hat{x}_t$$

Therefore, the correlation coefficient is R_i between the predictive value and actual value for i-th method (Yan, 2003):

$$R_i = \frac{\sum_{t=1}^{n} e_t e_{it}}{\sqrt{\sum_{t=1}^{n} e_t^2}\sqrt{\sum_{t=1}^{n} e_{it}^2}}, i = 1, 2, \cdots, m$$

Three kinds of single forecast model predictions in the following Table 3 (Jia *et al*., 2006). We take the data in Table 3 into the combination forecasting model and using MATLAB optimization toolbox software calculate the

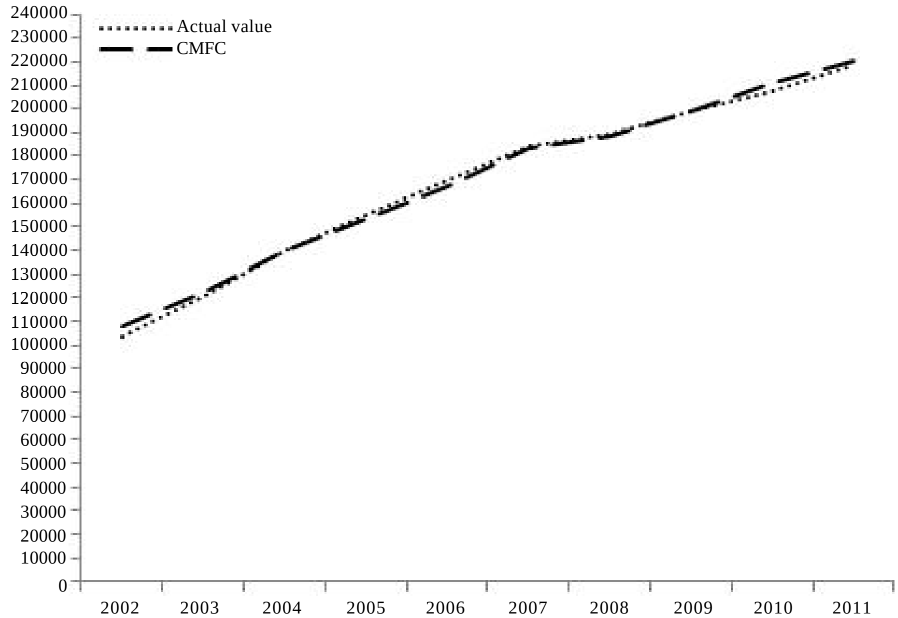

Fig. 1: Analysis result of fitting degree

Table 4: Result of the CMFC

Years	Result	Years	Result
2002	107719	2007	183363
2003	122903	2008	188884
2004	139682	2009	198722
2005	153487	2010	210805
2006	167387	2011	220141

weighting coefficients (Zhang *et al*., 2009). $l_E = 0.5$ $l_{BP} = 0.2$ $l_{GM} = 0.3$, Linear combination forecasting model, as follow (Chen, 2006):

$$\hat{x}_t = \sum_{i=1}^{3} l_i x_{it}$$

FEASIBILITY ANALYSIS OF COMBINATION FORECASTING MODELS

We put the data in Table 3 into the combination forecasting model, the following results can be obtained (Li and Jiang, 2006).

In order to verify the effectiveness of combination forecasting model, the combination forecast model results are compared with the actual value, the results shown in Fig. 1 (Yang, 2008).

From Fig. 1 we can see that the predictions of the combination forecasting model based on correlation coefficient and the actual valuehave a high degree coincidence, indicating that this combination forecasting model not only can well predict the trend of carbon emissions, more accurate prediction of the amount of carbon emissions, this model for prediction of carbon emissions is an effective.

CONCLUSION

In this article, we proceed from the two aspects of the uncertainty and complexity of Chinese CO2 emissions, we use the combination forecasting model based on correlation coefficient to predict, this approach overcomes the subjectivity of traditional methods and also avoids the nume rical calculation of nonlinear extremal problems. We can see this method is superior to the advanced support vector machines and BP neural network forecasting method for our CO2 emissions prediction.

REFERENCES

Cao, J., X.Z. Yuan, H.N. Zhu, H.W. Jiang, G.M. Zeng and H.J. Huang, 2011. To seek to Based energy structure IGT model CO-2 emissions study. Environ. Sci. Manage., 8: 50-54.

Chen, H., 2006. Based correlation coefficient superior combination forecasting model. Syst. Eng., 4: 353-360.

Chen, Y., T. Jing and J. Lu, 2012. Traffic accident multiple regression prediction model. Highway Trans. Res. (Applied Technol. Edn.), 1: 175-179.

Ding, Z.Q., Y. Wang, G. Zhou and H. Chen, 2010. Based correlation coefficient IOWGA operator combination forecasting model. Operations Res. Manage., 4: 45-50.

IPCC., 2006. Guidelines for national greenhouse gas inventories volume 2: Energy. Intergovernmental Panel on Climate Change. http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol2.html

Jia, C., C.H. Zhang, F. Jin and W.S. Cheng, 2006. Reliability and failure mode of the random variable correlation coefficient of sensitivity analysis and its application. Eng. Mech., 4: 12-16.

Jiang, G.L., 2011. Chinese carbon emissions estimates and its influencing factors. Resource Sci., 4: 597-604.

Li, C., 2012. Based on BP neural network prediction of stock index futures prices. Qingdao University, China.

Li, X. and W.H. Jiang, 2006. Correlation coefficient and correlation measure. Math. Practice Theory, 12: 188-192.

Li, Z.J., 2003. 1992-2001 Application of grey system theory research. Syst. Eng., 5: 8-12.

Wang, F., L.H. Wu and C. Yang, 2010. China's economic development in the drivers of growth in carbon emissions study. Econ. Res., 2: 123-136.

Xie, N.M., 2008. Grey system modeling research. Nanjing University of Aeronautics and Astronautics, China.

Yan, L.K., 2003. Correlation coefficient and partial correlation coefficient correlation analysis application. Econ. Univ. J., 3: 78-80.

Yang, G., 2008. China's dynamic economic fluctuations stylized facts-based on dynamic conditional correlation of theoretical and empirical. Nankai Econ. Stud., 3: 118-130.

Zhang, S.Q., J. Lu, Z. Jiang and L. Zhang, 2009. Discussion on correlation coefficients. Math. Practice Theory, 19: 102-107.