Research Journal of Applied Sciences, Engineering and Technology 5(6): 2073-2077, 2013
DOI:10.19026/rjaset.5.4752
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: July 27, 2012 Accepted: September 03, 2012 Published: February 21, 2013

Research Article

Prediction Models of Energy Consumption Structure of Shandong Province of China

[1]Jiekun Song and [2]Hailing Wang
[1]School of Economics and Management, China University of Petroleum, Qingdao, 266580, China
[2]School of Management, University of Science and Technology, Hefei, 230026, China

Abstract: In order to predict the energy consumption structure of Shandong province of China, linear regression model, gray model and ARIMA model are constructed respectively. On the basis of the single predicted results, the optimal weighted combination model is constructed for combination prediction of Shandong province's energy consumption structure. The empirical test shows that combination prediction model can effectively increase the prediction accuracy, providing a new method for the energy consumption structure prediction.

Keywords: Combination prediction, energy consumption structure, Shandong province, single prediction

INTRODUCTION

Energy has very important effect on economy and environment. On one hand, energy is the core impetus and strength source of economic development and a lot of energy need to be consumed in the process of the economic development; on the other hand, energy consumption will bring about some environmental problems and produce waste gas, waste water and solid waste. Energy consumption structure refers to the structure and the ratio relationship of all kinds of energy in total energy consumption. In 2010, the primary energy consumption structure of Shandong province of China is that, coal 76.21%, oil 21.98%, power 0.09% and other 1.81%. The coal-based energy consumption structure determines that the economic development of Shandong province is based on the high pollution, which is not harmonious with the low carbon target of today's world. Therefore, scientific prediction of energy consumption structure of Shandong province is conducive to make effective policies on optimization of energy consumption structure and achieve energy-economy-environment system's coordinated development. At present, prediction models of energy consumption structure include statistical prediction (Wei *et al*., 2006), grey prediction (Chen *et al*., 2007), neural network prediction (Li *et al*., 2009) and time series prediction (Huang *et al*., 2004). Consideration the simplicity, applicability of prediction, this study uses regression prediction, gray prediction and ARIMA prediction models to predict energy consumption structure of Shandong province and makes combination prediction on the basis of three prediction models.

SINGLE PREDICTION MODELS OF ENERGY CONSUMPTION STRUCTURE

Regression prediction model: Assuming that there is a time series $X = \{x_1, x_2, \ldots, x_n\}$, where x_i means the energy consumption of the i^{th} year, we use the time serial number as variable and make linear regression on energy consumption. The steps are as follows:

Step 1: Input all samples into the least squares estimation equation and get the least squares estimation of parameters $\hat{a}$ and $\hat{b}$:

$$\begin{cases} \hat{b} = \dfrac{\sum_{t=1}^{n} tx_t - n\bar{t} \bullet \bar{x}_t}{\sum_{t=1}^{n} t^2 - n\bar{t}^2} \\ \hat{a} = \bar{x}_t - \hat{b}\bar{t} \end{cases} \quad (1)$$

where, $\bar{t}$ and $\bar{x}_t$ stand for the mean values of t and x_t, respectively.

Step 2: Input $\hat{a}$ and $\hat{b}$ into regression equation:

$$\hat{x}_t = \hat{a} + \hat{b}t \quad (2)$$

And then compute prediction values of all the years.

Step 3: Input x_t and $\hat{x}_t$ into F statistics:

$$F = \frac{S_1^2}{S_2^2/(n-2)} \quad (3)$$

Corresponding Author: Jiekun Song, School of Economics and Management, China University of Petroleum, Qingdao, 266580, China

where,

$$S_1^2 = \sum_{t=1}^{n}\left(\hat{x}_t - \bar{\hat{x}}_t\right)^2, \quad S_2^2 = \sum_{t=1}^{n}\left(x_t - \hat{x}_t\right)^2$$

Give significance level α (usually 0.05) and if $F>F_{\alpha}(1, n\text{-}2)$, the regression equation can be considered significant, i.e., there is a linear relation between t and x_t and it can be used for prediction. Otherwise, the regression equation has no meaning and cannot be used for prediction.

Step 4: Input $t = n + 1, n + 2 \ldots$ into Eq. (2) and get the prediction values.

Grey prediction model: The basic Grey prediction Model is GM (1, 1). It constructs a linear one-order differential equation model for grey exponential time sequences and the corresponding time response function is exponential function. Grey prediction process can generally be divided into grey generation, parameters calculation and model tests. The steps are as follows (Deng, 1990):

Step 1: Grey generation: Usually, the original sequence is unlikely to have grey exponential law. So we need to make accumulation operation and weaken the influence of bad data in the original series. The cumulated series is recorded as $Y = \{y_1, y_2, \ldots, y_n\}$, where:

$$y_t = \sum_{i=1}^{t} x_i \tag{4}$$

Based on the cumulated series, the grey Generation Model GM (1, 1) is:

$$\frac{dy}{dt} + ay = u \tag{5}$$

where,
a = Called development grey number
u = Called endogenous control grey number

Step 2: Parameters calculation: Record $\hat{a}$ as parameters vector $\hat{a} = [a\ u]^T$, where:

$$\boldsymbol{B} = \begin{bmatrix} -0.5(y_1+y_2) & 1 \\ -0.5(y_2+y_3) & 1 \\ \vdots & \vdots \\ -0.5(y_{n-1}+y_n) & 1 \end{bmatrix}, \quad C_N = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

We can use least squares method to calculate parameters a and u:

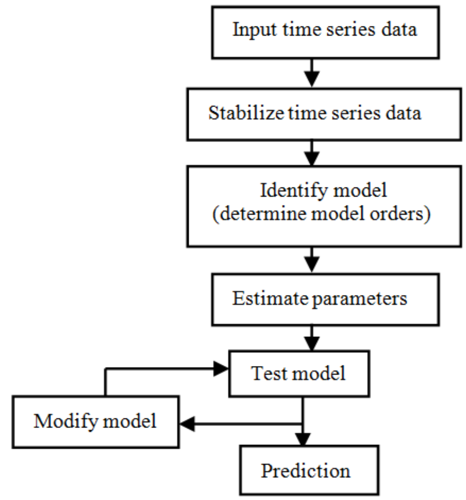

Fig. 1: Process of ARIMA construction

$$\hat{a} = \begin{bmatrix} a \\ u \end{bmatrix} = (B^T B)^{-1} B^T C_N \tag{6}$$

The corresponding time response function is:

$$\hat{y}_{t+1} = (x_1 - \frac{u}{a})e^{-at} + \frac{u}{a},\ t = 0,1,\cdots,n \tag{7}$$

Which is GM (1, 1) prediction model.

Make subtraction operation and we can get the prediction values of raw sequence:

$$\hat{x}_{t+1} = \hat{y}_{t+1} - \hat{y}_t,\ \ t = 0,1,\cdots,n \tag{8}$$

Step 3: Model tests: We need to make residual test, correlation test and after test to determine whether GM (1, 1) prediction model meets the precision requirements. Residual test generally requires that mean relative error $e\leq 20\%$, preferably $e\leq 10\%$; correlation test requires correlation degree $r>0.6$; after test requires posterior error $C<0.5$, preferably $C<0.35$, small error frequency $p>0.8$, preferably $p>0.95$.

Step 4: If GM (1, 1) prediction model does not pass the above tests, we should make residual correction, then modify grey model. It has passed tests; we can apply it to predict energy consumption in the future years.

ARIMA prediction model: ARIMA (p, d and q) model was put forward in 1970s by Box and Jenkins, which is a time series prediction method (Gao, 2006). AR means auto-regressive, MA means moving average, p is auto-regressive item, q is moving average item and d is the difference frequency made to stabilize time series. The process of ARIMA construction can be seen as Fig. 1.

The steps of ARIMA are as followed:

Step 1: Smooth and white noise test of series: Check the autocorrelation function of original sequence ***X*** to determine whether it is a stable sequence or not. If ***X*** is unstable, we need make difference operation. Repeat this step *d* times to ensure that $Z = \Delta^d X$ is a stable sequence.

Step 2: Model identification: Calculate the autocorrelation function and partial autocorrelation function of *Z* to determine the parameters *p* and *q*. Where, according to the *q* steps truncation of correlation function to determine the moving average item *q*; according to the *p* steps truncation of partial autocorrelation function to determine autoregressive item *p*. If the autocorrelation function and partial autocorrelation function decay according to exponential law, but they are both not censored, then we should fit ARMA model order from low order to high order.

Step 3: Parameters estimation: Calculate parameters $\varphi_1, \varphi_2, \ldots, \varphi_p$ and $\theta_1, \theta_2, \ldots, \theta_q$ and get the ARIMA model:

$$\hat{y}_t = -\phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{9}$$

Step 4: Model diagnosis: We can use AIC information criterion, correlation coefficient or fitting error to make model tests. Generally, AIC should be small, the absolute value of correlation coefficient should be close to 1 as possible and the fitting error should be small. Usually, step 1 to step 4 are performed at the same time.

Step 5: Prediction: If ARIMA prediction model does not pass model tests, we should return step 2. Otherwise, if it has passed model tests, we can apply it to predict energy consumption in the future years.

OPTIMAL WEIGHTED COMBINATION PREDICTION OF ENERGY CONSUMPTION STRUCTURE

Combination prediction will combine different prediction models together. It utilizes different information provided by different prediction methods comprehensively and gives combination prediction models in appropriate weighted form (Chen, 2008). Optimal weighted combination method is to construct the objective function based on some optimal criterion and in certain conditions; it will minimize objective function to acquire weighted coefficients of different prediction methods.

Assume that w_1, w_2, w_3 are the weights of regression prediction, grey prediction and ARIMA prediction methods respectively and $\hat{x}_{1t}, \hat{x}_{2t}, \hat{x}_{3t}$ are the prediction values of the t^{th} year by the three methods, we can get the combination prediction value of the t^{th} year is:

$$\hat{x}_t = \sum_{i=1}^{3} w_i \hat{x}_{it} \tag{10}$$

Use minimization the sum of squared prediction error as the objective function and we can construct optimal weighted combination prediction model as follows:

$$\min \sum_{t=1}^{n} (\hat{x}_t - x_t)^2 \qquad s.t. \sum_{i=1}^{3} w_i = 1,\ w_i \geq 0, i = 1,2,3 \tag{11}$$

Solve the above mathematical programming model and we can get the optimal solutions of w_1, w_2, w_3. Input them into model (10) and then we can use it to predict energy consumption in the future years.

In order to estimate the prediction effects of all the prediction models, we use the followed error indexes:

- **Mean square error:**

$$MSE = \frac{1}{n}\sqrt{\sum_{t=1}^{n}(x_t - \hat{x}_t)^2} \tag{12}$$

- **Mean absolute error:**

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|x_t - \hat{x}_t| \tag{13}$$

- **Mean absolute percentage error:**

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \tag{14}$$

- **Mean square percentage error:**

$$MSPE = \frac{1}{n}\sqrt{\sum_{t=1}^{n}[(x_t - \hat{x}_t)/x_t]^2} \tag{15}$$

ENERGY CONSUMPTION STRUCTURE PREDICTION OF SHANDONG PROVINCE

Single prediction: The statistics of primary energy consumption structure of Shandong province in 1996-2010 is as shown in Table 1. To make energy consumption structure prediction, we can predict the total primary energy consumption, coal consumption and oil consumption and then get the future trend of energy consumption structure. This study will only

Table 1: Energy consumption of Shandong province

Year	Total energy	Coal	Oil
1996	10117.67	7940.35	2036.69
1997	10128.81	7852.87	2183.77
1998	10028.80	7875.62	2069.94
1999	10104.56	7779.50	2251.30
2000	9977.11	7392.04	2530.20
2001	11649.88	8955.26	2539.67
2002	13121.91	10738.97	2326.51
2003	15974.50	12694.94	3162.95
2004	19606.14	14896.75	4566.27
2005	25687.50	20744.27	4714.89
2006	28786.10	22972.14	5540.50
2007	31194.99	25101.87	5822.50
2008	32116.22	25044.54	6610.10
2009	34535.66	26637.66	7347.15
2010	36357.25	27707.90	7990.73

Table 2: Prediction results of energy consumption by regression prediction method

Year	Total energy	Coal	Oil
2011	37556.38	29356.36	7691.88
2012	39756.04	31073.12	8139.26
2013	41955.69	32789.87	8586.64
2014	44155.35	34506.63	9034.01
2015	46355	36223.39	9481.39

Table 3: Prediction results of energy consumption by grey prediction method

Year	Total energy	Coal	Oil
2011	45119.66	35139.45	9236.72
2012	50707.85	39419.57	10403.56
2013	56988.14	44221.02	11717.81
2014	64046.27	49607.31	13198.08
2015	71978.56	55649.67	14865.34

Table 4: ADF test of $\Delta^2 \ln X$ series

		t-statistic	Prob.*
Augmented dickey-fuller test statistic		-4.293225	0.0005
Test critical values	1% level	-2.771926	
	5% level	-1.974028	
	10% level	-1.602922	

describe the modeling process of total energy consumption and the processes of coal consumption and oil consumption prediction are the same with it. It can be seen from Table 1 that the total energy consumption of Shandong province has raised and the trend of it is linear basically.

Use linear regression prediction and we get the regression equation is:

$$\hat{x}_t = 2199.66t + 2361.90$$

At the same time, we get $F = 133.41 > F_{0.05}(1, 13) = 4.67$, which means that the equation is significant and can be used for prediction. Input t = 16, 17, …, 20 and we get the prediction value of total energy consumption in 2011-2015 as shown in Table 2. The prediction values of coal consumption and oil consumption are also shown in Table 2.

Use grey prediction and we get a = -0.12, u = 7113.97. GM (1, 1) model is:

$$\hat{y}_{t+1} = 71044.44e^{0.12t} - 60926.77$$

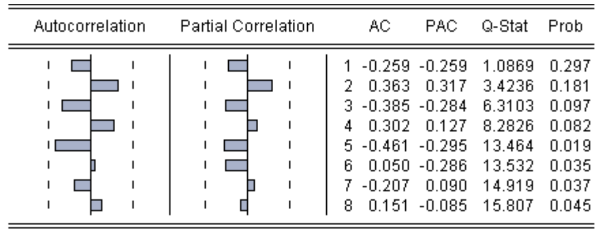

Fig. 2: Autocorrelation function and partial autocorrelation function of $\Delta^2 \ln X$ series

Make subtraction operation and we get the prediction model of raw sequence is:

$$\hat{x}_{t+1} = 7829.35e^{0.12t}$$

Meanwhile, we get the mean relative error e = 10.18%, correlation degree r = 0.99>0.6, posterior error c = 0.2281<0.35 and small error frequency p = 1.0>0.95. So the grey prediction model has passed tests. Input t = 15, 16, …, 16 and we get the prediction values in 2011-2015 as shown in Table 3.

Then we use ARIMA method to predict energy consumption. Firstly, we take the natural logarithm of original time series: ln$X$ = log (X). Secondly, we make unit root test and find that after two orders difference, the time series become stable, namely, the difference order d = 2. ADF test result is as shown in Table 4. We can see that t statistic value is less than critical value under confidence level 1%, so it refuses the hypothesis of having unit root.

After two orders difference, we get autocorrelation function and partial autocorrelation function as shown in Fig. 2.

By using Eviews 6.0 to calculate iteratively and comparing AIC and SC, we consider ARIMA (1, 2 and 1) model is the best model. ARIMA model is as follows:

$$\Delta^2 \ln X_t = -0.934621 \Delta^2 \ln X_{t-1} + \varepsilon_t + 0.997364 \varepsilon_{t-1}$$
$$(-17.33003) \qquad\qquad (5.293560)$$

where,

$$\Delta^2 \ln X_t = d(\ln X, 2) = \ln X_t - 2\ln X_{t-1} + \ln X_{t-2}$$

The prediction model is:

$$\hat{X}_t = \exp(2\ln X_{t-1} - \ln X_{t-2} - 0.934621 \Delta^2 \ln X_{t-1} + \varepsilon_t + 0.997364 \varepsilon_{t-1})$$

And AIC = -2.5685. All the parameters are significant in 5% level and have passed significance test. Residual correlation coefficients and partial correlation coefficients are all in the confidence interval, so they are white noise series. The prediction values of energy consumption in 2011-2015 are shown in Table 5.

Table 5: Prediction results of energy consumption by ARIMA prediction method

Year	Total energy	Coal	Oil
2011	45119.66	35139.45	9236.72
2012	50707.85	39419.57	10403.56
2013	56988.14	44221.02	11717.81
2014	64046.27	49607.31	13198.08
2015	71978.56	55649.67	14865.34

Table 6: Prediction results of energy consumption by combination prediction method

Year	Total energy	Coal	Oil
2011	40251.61	30396.23	9370.77
2012	43154.44	32213.92	10478.76
2013	46804.61	34115.07	11776.13
2014	50257.29	36285.98	13236.13
2015	54506.46	38665.39	14879.95

Table 7: Results of error indexes of four prediction models

Model	MSE	MAE	MAPE	MSPE
Regression model	784.7	2421.3	0.15	0.060
Grey model	705.0	2147.5	0.12	0.040
ARIMA model	387.5	940.3	0.04	0.020
Combination model	216.6	637.1	0.03	0.012

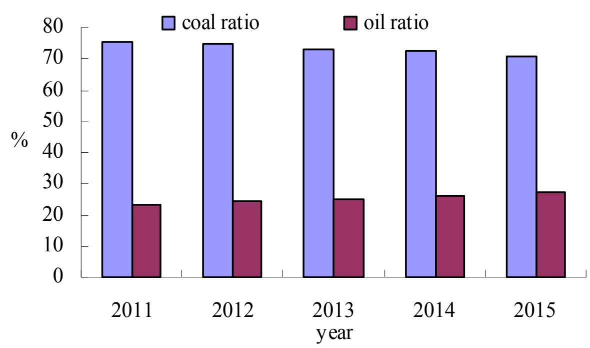

Fig. 3: Predicted coal and oil ratios in energy consumption

Combination prediction: Calculate the prediction errors of total energy consumption of three methods and input them into the optimal weighted combination prediction model, then we get the optimal solution: w_1 = 0.1106, w_2 = 0.2242 and w_3 = 0.6652. The prediction values of energy consumption in 2011-2015 by combination prediction are shown in Table 6.

The energy consumption ratios of coal and oil are shown in Fig. 3. We can see that although coal ratio will decrease, but it will still exceed 70% and the average value of it in 2011-2015 will be 73.24%. Oil ratio in energy consumption will rise gradually and the average value of it in 2011-2015 will be 25.27%. Coal and oil ratio in the energy consumption will still exceed 98% and the average value of it in 2011-2015 will be 98.52%.

The results of error indexes of three single prediction models and optimal weighted combination prediction model are as shown in Table 7. We can see that the results of four error indexes are the smallest among four prediction models, which indicate that the combination prediction model is the best prediction model.

CONCLUSION

This study constructs linear regression model, grey prediction model and ARIMA model to predict energy consumption structure of Shandong province of China. Based on the three single prediction models, it constructs the optimal weighted combination prediction model. The prediction results show that the combination prediction model has higher accuracy than the three single prediction models, which can achieve very good prediction effect and provide a new method for prediction of energy consumption structure.

ACKNOWLEDGMENT

This study is supported by Humanities and Social Sciences Project of Ministry of Education in China (Grant No. 10YJC630207), Shandong Provincial Natural Science Foundation (Grant No. ZR2011GQ004), University Scientific Research Development Program of Shandong Province (Grant No. J10WG94), the Fundamental Research Funds for the Central Universities (Grant No. 10CX04012B and 11CX04034B).

REFERENCES

Chen, H., 2008. Effective Theory of Combination Prediction Methods its Application. Science Press, China.

Chen, H. and Z. Dequn, 2007. Study on prediction of China's energy consumption based on GM (1, 1). Mining Res. Dev., 7: 77-79.

Deng, J., 1990. Grey System Theory. Wuhan University Press, China.

Gao, T., 2006. Methods and Models of Econometric Analysis. Qinghua University Press, China.

Huang, J., Z. Meng and W. Junhai, 2004. Application of ARMA model in prediction of China's energy consumption. Stat. Decision, 12: 49-50.

Li, J., 2009. China's future energy demand forecast and potential crisis. Res. Financial Econ. Issues, 2: 6-21.

Wei, Y. and L. Qiaomei, 2006. China's regional energy demand prediction in 2010-2020. Prediction report of Chinese Academy of Sciences.