

Prediction of Electricity Consumption Based on Multiple Liner Regression

He-rui Cui

Department of Economic Management
North China Electric Power University
Baoding, China

Ting-ting Wu

Department of Economic Management
North China Electric Power University
Baoding, China

Abstract—The prediction of power load is important content, prerequisite and basis of the planning system and running of power grid. The healthy development of power industry is based on scientific and accurate prediction. Multivariate linear regression analysis is a method to analyze the linear relationship between one dependent variable and multiple independent variables. The accuracy of load forecasting directly affects the operation of the power system, this paper studies the multiple linear regression model and its application in the regional economy of electricity. In this paper a prediction for power load was presented on the basis of multiple liner regression model. The model in this paper applied the variant GDP and population to predict the total electricity consumption of Hebei from 2000-2014. The result showed that the model is effective to the prediction of power load and further can provide foundation for the controlling and prediction of power load.

Keywords—Multiple liner regression; Hebei network; electricity consumption forecast; power regional economy

I. INTRODUCTION

In recent years, with the rapid economic growth in China, both energy consumption and greenhouse gas emissions have attracted the attention of domestic and foreign scholars as well as Chinese government. In 2013, China's energy consumption has reached 3750 million tons of coal equivalents, while the CO₂ emissions from fossil fuel combustion were 9.023 billion tons, which accounted for 28.03% of total global emissions [1].

The electricity consumption is huge energy consumption. The development of national economy cannot be separated from the growth of electricity consumption. Making scientific electric development planning and industrial policy making will promote the sustainable development of regional economy. This paper aims to discuss the relationship between economic development and social power consumption growth in Hebei Province in recent 15 years, and analyzes various factors that affect this relationship [2-3].

In order to forecast electricity consumption more rapidly and accurately, this article focuses on the least-squares regression method to data collection and analysis. It is a new multivariate statistical data analysis method which combines with multiple linear regressions (MLR) [3-4]. MLR has more advantages over traditional regression analysis method on accuracy and stability of the models.

II. ANALYSIS ON CURRENT SITUATION OF ELECTRICITY CONSUMPTION AND GDP

From the Hebei Statistical Yearbook, historical data of Hebei about electricity consumption, GDP and population from 2000 to 2014 are obtained. The electricity consumption, population and GDP data are shown in TABLE I. The data is from statistical yearbook of Hebei Province 2014. Hebei has a population of 73.84 million and its total electricity consumption is 311.411 billion kilowatt-hours in 2014. The GDP value of Hebei in 2014 is 2942.115 billion.

TABLE I. HEBEI'S ELECTRICITY CONSUMPTION, POPULATION AND GDP IN 2000-2014

Year	Electricity consumption (ten thousand kilowatt hours)	Population (ten thousand)	GDP (billion yuan)
2000	809.34	6,674.00	5043.96
2001	867.55	6,699.00	5516.76
2002	965.83	6,735.00	6018.28
2003	1,099.00	6,769.00	6921.29
2004	1,291.40	6,809.00	8477.63
2005	1,501.92	6,851.00	10012.11
2006	1,734.83	6,898.00	11467.6
2007	2,013.67	6,943.00	13607.32
2008	2,095.02	6,989.00	16011.97
2009	2,343.85	7,034.00	17235.48
2010	2,691.52	7,194.00	20394.26
2011	2,984.90	7,241.00	24515.76
2012	3,077.73	7,288.00	26575.01
2013	3,251.19	7,333.00	28442.95
2014	3,314.11	7,384.00	29421.15

III. TO SET UP THE FORECAST METHOD

A. Introduction of Multiple Liner Regression Model

Multivariate linear regression analysis is a method to analyze the linear relationship between one dependent variable and multiple independent variables. The general form of the multiple linear regression model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_j X_{j3} + \dots + \beta_k X_k + \mu \quad (1)$$

- Where, k, μ represents number of variables and random error.

The matrix expression for the random equation is:

$$Y = X\beta + \mu \quad (2)$$

$$\left. \begin{aligned} \text{Where, } y_1 &= \beta_0 + \beta_1 X_{11} + \dots + \beta_k X_{1k} + \mu_1 \\ y_2 &= \beta_0 + \beta_2 X_{21} + \dots + \beta_k X_{2k} + \mu_2 \\ y_3 &= \beta_0 + \beta_1 X_{31} + \dots + \beta_k X_{3k} + \mu_3 \\ &\dots \\ y &= \beta_0 + \beta_1 X_{n1} + \dots + \beta_k X_{nk} + \mu_n \end{aligned} \right\} \quad (3)$$

Formula (1) provides the matrix expression for N random equations. If the rank of X is full, then, the estimated values of least-squares regression method are:

$$\beta^* = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} = (X^T X)^{-1} X^T Y \quad (4)$$

$$\text{Where, } Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_n \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{bmatrix}$$

After the parameter estimation of multiple linear regression model, it is necessary to test the regression functions of the model to determine the reliability of the estimate. To forecast the relationship between electricity consumption and economy, three tests need to be carried out. The three examination models are: the goodness of fit tests, the F-test, and the t-test.

The fitting value of goodness of the model is R², the total sum of squares of deviations is TSS, the sum of squares of regression is ESS, and the residual sum of squares is RSS. The equation of R² is:

$$R^2 = \frac{ESS}{RSS} = 1 - \frac{RSS}{TSS} \quad (5)$$

The F-test follows the freedom of (k, n-k-1). Given the significant level of α , we can get the critical value of F_α by looking up the related table. We can reject or accept the original hypothesis by the relationship of F and $F_\alpha(k, n-k-1)$. The F-test model is:

$$F = \frac{ESS/k}{RSS/(n-k-1)} \quad (6)$$

Based on the t-test, a t statistics is structure to calculate and analyze the original data. We assume that: $H_0: \beta_j = 0, H_1: \beta_j \neq 0$, given the significant level of α . The reject area is: $t_j = t_{\frac{\alpha}{2}}(n-k-1)$. $t_{\frac{\alpha}{2}}(n-k-1)$ is the criteria to determine the acceptance or rejection of the original hypothesis. If $t_j \geq t_{\frac{\alpha}{2}}(n-k-1)$, we can reject H_0 . If $t_j < t_{\frac{\alpha}{2}}(n-k-1)$, we can accept H_0 .

B. Hebei Electricity Consumption Forecast

SPSS 13.0 was used to deal with the multiple linear regression process. The model operation results are shown in TABLE II, TABLE III and TABLE IV.

TABLE II. THE VARIANT ATTRIBUTES

Model	independent variable	Excluded variables	Data
1	population, GDP	null	all

TABLE III. THE MODEL Operation Results

Model	R	R ²	Standard error	Statistics				
				The change Of R ²	F	Freedom 1	Freedom 2	Silence value
1	0.994	0.989	102.866	0.989	534.161	2	12	0.000

TABLE IV. THE CORRELATIVE COEFFICIENT OF MODEL EQUATION

Model	Parameters	Coefficients			T	Silence value	Statistics	
		Standard deviation					Tolerance	VIF
1	Constant	-18161.903	8073.903	-2.249	0.044	-	-	
	Population	2.832	1.230	2.303	0.040	0.008	118.596	
	GDP	0.024	0.034	0.701	0.497	0.008	118.596	

As shown in TABLE II, the population and GDP are the two arguments and the electricity consumption is the dependent variable. As shown in TABLE III, the correlation coefficient of the model is 0.994. The corresponding significant level is 0.000. The R² of the model is 0.989, which shows that the electricity

consumption of Hebei is determined by 98.9% of the population and GDP.

As shown in TABLE IV, the electricity consumption = -18161.903 + 2.832 * population + 0.024 * GDP. The equation can forecast the electricity consumption of the whole society. If we

need to predict the electricity consumption in Hebei in 2015, only the data of population and GDP in 2015 is needed.

There is a certain gap between actual electricity consumption and forecasting electricity consumption. Comparison of predictive value and actual value of multivariate linear regression model was shown in TABLE V.

TABLE V. COMPARISON OF PREDICTED VALUE AND ACTUAL VALUE

Year	Actual Electricity consumption (ten thousand kilowatt hours)	Predicted Electricity consumption (ten thousand kilowatt hours)	Relative Error %
2000	809.34	859.92	0.062495
2001	867.55	942.0672	0.085894
2002	965.83	1056.056	0.093418
2003	1,099.00	1174.016	0.068258
2004	1,291.40	1324.648	0.025746
2005	1,501.92	1480.42	-0.01432
2006	1,734.83	1648.455	-0.04979
2007	2,013.67	1827.249	-0.09258
2008	2,095.02	2015.232	-0.03808
2009	2,343.85	2172.037	-0.0733
2010	2,691.52	2700.967	0.00351
2011	2,984.90	2932.987	-0.01739
2012	3,077.73	3115.513	0.012276
2013	3,251.19	3287.784	0.011256
2014	3,314.11	3455.693	0.042721

The comparison between actual value and predicted value of Hebei electricity consumption from 2000 to 2014 are shown in TABLE V. The maximum relative error of the model was 9.34%, and the minimum relative error was 0.351%. The multiple liner regression model has higher prediction accuracy. So, the model is used to predict the future city electricity power consumption of Hebei.

The error fitting of electricity consumption is shown in Fig. 1:

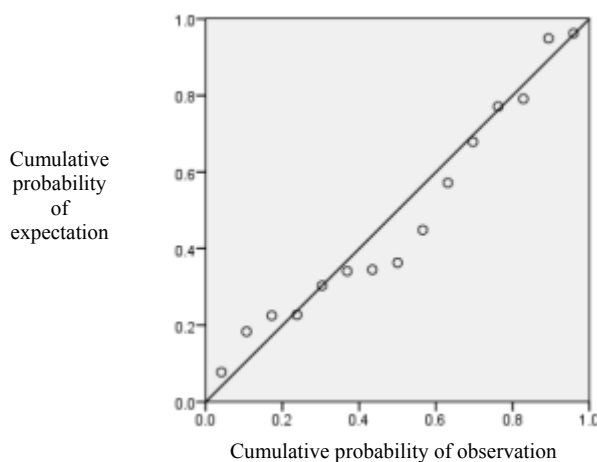


Fig. 1. The error fitting of electricity consumption

As shown in Fig.1, the electricity consumption fluctuates between narrow ranges. There is a certain deviation in the observation cumulative probability and the expected cumulative probability.

The development of economy has led to the development of electrical industry. Equally, the electricity consumption will promote the growth of GDP. The multiple liner regression is an effective method for power load forecasting. The accuracy of the national GDP forecast data will fluctuate due to changes in the economic environment at home and abroad [5-6]. The international financial crisis may lead to a marked slowdown in the growth rate of GDP in China [7-8]. Thus, these factors will affect the accuracy of the prediction and bring the errors to the model. The model needs to control the errors within a certain range to guarantee the validity of the results.

IV. CONCLUSION

The development of economy is closely related to electricity consumption. Electricity consumption is an important factor to promote the growth of GDP. To achieve long-term development, techniques for prediction of electric loads are important for the realization of power market in China. After 12th Five-Year, Baoding undertakes industry transfer of Beijing and Tianjin. As a regional central city in Hebei, energy consumption will increase significantly. Thus, the accuracy of load forecasting has a direct impact on system distribution network system planning, quality of power supply and operation scheduling.

Traditional technology-centric electricity consumption forecast approach can't directly reflect the influence of population and GDP. To improve the accuracy of load forecasting, by use of multiple liner regression model based on real number output and fitting residual error by trigonometric function, three test models are put forward. In order to ensure the healthy development of the city, energy demand in Hebei is predicted. The multiple liner regression method not only considers the population growth, but also takes into account the future of the industrial structure. The method is reasonable and high reliability. With the development of modern technologies, how to effectively improve the power load forecasting accuracy is a critical question for scholars.

REFERENCES

- [1] Bagchi, A., &Paul, J.A. Optimal allocation of resources in airport security: Pro filing vs. screening. *Journal of Operations Research*, vol. 62, no. 2, pp. 219-233, 2014.
- [2] Baoding Municipal Development and Reform Commission. *Energy status and energy saving work in Baoding City*, 2015.
- [3] G. Fan, "A method to estimating the parameters of general Logistic curve," *Journal of Liaoning Normal University*, vol. 32, no. 4, pp. 426-429, Dec. 2009. (In Chinese)
- [4] G. Selvaraju, G.A. Balasubramaniam. *Multiple Linear Regression Model for Prediction of Sheep Pox Outbreaks**, vol. 92, no. 3, pp.78-85, 2015.
- [5] L. Jiao, D. Dong, W. Zheng, X. Zhao, S. Zhang, C. Shen. *Determination of thiophanate-methyl using UV absorption spectra based on multiple linear regression*, vol. 125, no. 1, pp.78-85, 2014.
- [6] National Bureau of Statistics (NBS), 2015. *Chinese statistics year book 2015*. China Stat. Press.

- [7] National Development and Reform Commission. "12th Five-Year" plan of Renewable energy development, 2012
- [8] Yuan Yanhua. "Comparisons of parameter estimation on logistic model". Journal of Heilongjiang institute of science & technology, vol. 18, no. 4, pp: 474-477, 2008. (In Chinese)