

Research on Correlation analysis of industry electricity quantity

Wang Yongjun¹, Yao Gang¹, He Yanying²

(1 Guizhou Power Grid Corporation, Guiyang, Guizhou, China; 2 School of Economics and Management, North China Electric Power University, Beijing 102206, China)

Keywords: Correlation analysis, Pearson correlation coefficient, partial correlation coefficient, the multiple correlation coefficient

Abstract: Power sale quantity of Power grid enterprise is affected by many factors. First, this paper introduces the correlation analysis method, including the Pearson correlation coefficient, partial correlation coefficient and the multiple correlation coefficient; Then analyzing the external influence factor, such as population, climate and internal factors, such as industry power sale, and designing the industry power correlation analysis method; finally, this paper made an empirical analysis in Guizhou province as an example, The results cleared about the influence factor of quantity change and showed this method could help identifying the focal point of electric power marketing, and be used as auxiliary power prediction work reference.

Introduction

Power forecasting is a basic work for the power supply enterprise to develop electric power marketing. Existing power prediction model often only inspect its own characteristics of history data^[1-3], but lack in-depth analysis on the influence factors, so the prediction results reliability is questionable.

Power changes are influenced by many factors. If we can calculate the specific incidence of influence factors on power changes, then we can modify the power forecasting results, dig out the key factors of influencing the power change, and identifying the focal point of electric power marketing.

There are many factors that can result in the change of power, which can be divided into external factors and internal factors^[4]. Some external factors, such as population, climate, economy have certain data records, so can make correlation analysis. However, other factors, such as energy policy, environmental standards and scientific have no quantitative data, and can be only expressed by language, so it is difficult to analyze its impact on power directly^[5]. But, the impact of these factors can often reflected by the change of some economic indicators, so you can do correlation analysis of economic index and quantity, indirectly reflect the compositive impact of these factors on electricity.

Every industry power can be seen as internal factors that result in the change of power^[6]. A region's electricity is made up of different industries of electricity. The development of industries is different, then the energy change rule is not the same^[7]. We can know which industries are the important factors that affect the total quantity change by an analysis of the correlation industry power and total power.

This paper introduced correlation analysis method, using the Pearson correlation coefficient, partial correlation coefficient and multiple correlation coefficient to analyze the interrelation between the single factor and multiple factors, so as to find out the key factors which affect the power change, and provide important basis for electricity marketing work well.

Correlation analysis method

If the power prediction influence degree of population, climate and other uncertain factors is studied, correlation analysis on them is necessary. The common correlation analysis method is to calculate the correlation coefficient between the variables. The correlation coefficient can be divided into linear correlation coefficient and nonlinear correlation coefficient^[8-9]. This paper only studies the linear correlation coefficient including Pearson correlation coefficient, partial correlation coefficient and multiple correlation coefficient

Pearson correlation coefficient

Pearson correlation coefficient is also called coefficient of the product, which can explain correlation degree and correlation direction between two variables of linear relationship. The correlation coefficient of n dimensional row vector X and Y are as follows:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{\sum_{i=1}^n X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\sum_{i=1}^n X_i^2 - \frac{(\sum X_i)^2}{n}} \sqrt{\sum_{i=1}^n Y_i^2 - \frac{(\sum Y_i)^2}{n}}} \quad (1)$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (2)$$

Where, r is sample correlation coefficient, P is total correlation coefficient; \bar{X}, \bar{Y} is the average value of the two index respectively.

The range of correlation coefficient r can be judged the correlation degree between the two indicators: if r=0, it is completely irrelevant; If 0<r<0.3, it is incomplete relevant; if 0.3<r<0.5, it is low relevant; if 0.5<r<0.8, it is significant correlation; if 0.8<r<1, it is highly relevant; if r=1, it is completely relevant.

Partial correlation coefficient

Partial correlation coefficient, is also known as "net" or "pure" When too many variables need to be considered, any two variables may present correlation. The two variables is highly relevant due to they are all related with the third variable, so Pearson correlation coefficient often cannot truly reflect the true relationship between the two variables. Partial correlation coefficient can truly reflect the linear correlation degree of any two variables in the variable X₁, X₂,..., X_p when the rest of the p-2 variables is fixed.

Given an assumption of variable unit, {X₁, X₁, ..., X_p} ,the correlation coefficient matrix is as follows:

$$R = \begin{bmatrix} r_{11} & r_{12} & \mathbf{L} & r_{1p} \\ r_{21} & r_{22} & \mathbf{L} & r_{2p} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ r_{p1} & r_{p2} & \mathbf{L} & r_{pp} \end{bmatrix} \quad (3)$$

Where, r_{ij} is Pearson correlation coefficient of $X_i, X_j, \forall i, j = 1, 2, \dots, p$.

The partial correlation coefficient of variable X_i and X_j under the condition of other variables fixed based on the theory of multiple fall:

$$r(X_i, X_j | X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_p) = \frac{-\Delta^{ij}}{\sqrt{\Delta^{ii}} \sqrt{\Delta^{jj}}} \quad (4)$$

Where, $\Delta^{ij}, \Delta^{ii}, \Delta^{jj}$ is algebraic complement of r_{ij}, r_{ii}, r_{jj} in $|R|$

$$R^{-1} = \frac{1}{|R|} R^* = \frac{1}{|R|} (\Delta^{i,j})_{p \times p}, i, j = 1, 2, \dots, p \quad (5)$$

If $R^{-1} = (r^{i,j})_{p \times p}, i, j = 1, 2, \dots, p$, formulation (4) is equivalent to:

$$r(X_i, X_j | X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_p) = \frac{-r^{ij}}{\sqrt{r^{ii}} \sqrt{r^{jj}}} \quad (6)$$

Multiple correlation coefficient

Multiple correlation coefficient is a statistical indicators which reflect the degree of linear correlation between one variable and other multiple variables. Multiple correlation coefficient can't calculate directly, and need to find out the multiple linear regression equation firstly.

n dimensional vector Y and other n dimensional vector X_1, X_2, \dots, X_m , the calculation process of the multiple correlation coefficient between Y and X_m is as follows:

Firstly, make the multiple linear regression equation about Y and X_1, X_2, \dots, X_m .

$$\hat{Y} = a_0 + a_1 X_1 + \dots + a_m X_m \quad (7)$$

Then reach the multiple correlation coefficient about Y and X_1, X_2, \dots, X_m .

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}} \quad (8)$$

We can find out Pearson correlation coefficient is a special case when multiple correlation coefficient m is one, so Pearson correlation coefficient can be reached after calculating unary linear regression equation of X and Y . Due to the change trend of multiple variables is difficult to maintain consistent, plus or minus of multiple correlation coefficient loses the significance, so commonly consider the absolute value of the multiple correlation coefficient. the greater $|\rho|$, the more closely is degree of linear correlation between the variable Y and X_1, X_2, \dots, X_m .

Empirical analysis

The paper made a correlation analysis on all data assets power sales in Guizhou from the year 2011 to 2014. According to the electricity price category and the national standard <the national economy industry classification>, electricity industry power statistics of Guizhou power grid are classified into resident non-resident, business, the common industrial, industrial, agricultural irrigated, wholesale power sale quantity.

Economy, population, meteorological data of External factors can be analyzed. Economic indicators include GDP, total industrial production, and total export-import volume; the population indicators include census register population, the population of permanent residents, and floating population; meteorological indicators include annual mean temperature, monthly mean temperature, monthly high (low) temperature. The history of the power value of various industries of Internal factors can be analyzed. So influence factors analyzed valuably have monthly average temperature, GDP, industrial production, export-import volume and power value of different industry and region.

Pearson correlation analysis

According to all data assets power sales in Guizhou from 2011 to 2014, reach the trend. diagram as shown in Fig.1, as can be seen, the power sales rise year by year, electricity power fell in February, electricity generally to rise in summer.

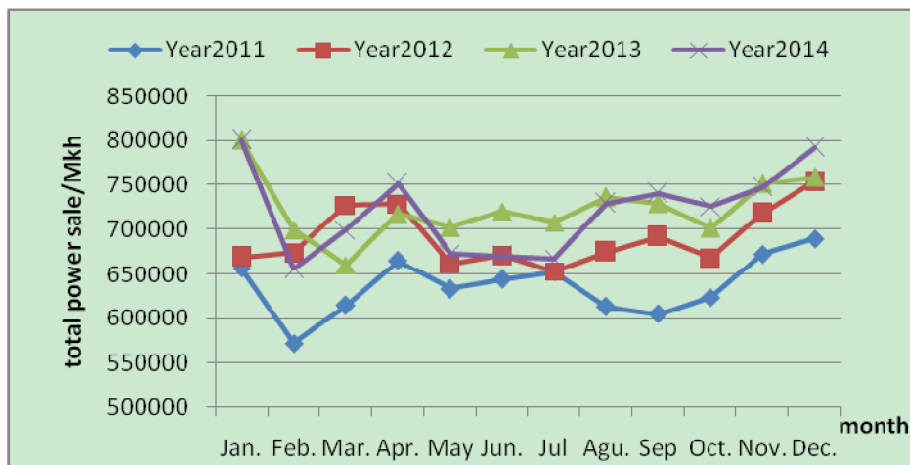


Fig.1 tendency chart of total power sales in Guizhou province from 2011 to 2014

The population and meteorological data is lack, combined with the conditions of total power sales in Guizhou, so GDP, industrial production, consumer price index, and temperature in external factors is only selected. And hydroelectric power generation amount and the total thermal power generation in the internal factors is selected.

Calculate Pearson correlation coefficient between total power sales in Guizhou with local GDP, industrial added value and export-import volume, the results are shown in table 1. Thus, the total quantity sold in Guizhou and GDP of Guizhou province were positively correlated, and Pearson correlation coefficient is higher than 0.8; The Pearson correlation coefficient between total power sales and the temperature is irrelative, and the correlation coefficient with the industrial added value, hydroelectric power generation and total thermal power generation is not high; total power sales and the consumer price index or producer price index showed a negative correlation.

Table.1 Analysis result of Pearson correlation coefficient

total power sale	influence factor	Pearson
power sales	Local GDP/million	0.912
power sales	industrial production / million	0.659
power sales	Local CPI	-0.657
power sales	mean temperature	-0.209
power sales	Hydroelectric power generation	0.639
power sales	thermal power generation/Mkh	0.645

partial correlation analysis

According to table 1, the total power sales in Guizhou have a higher correlation with local GDP. Controlling the GDP variable, make partial correlation analysis on total power sales in Guizhou province and industrial added value using SPSS software, the results are shown in table 2.

Table.2 partial correlation analysis about total power sales in Guizhou province and industrial production

control variable		total power sales	industrial production	GDP	
-no- ^a	total power sales	correlation	1	0.9	0.912
		Significance level(both sides)	.	0.1	0.088
		df	0	2	2
industrial production		correlation	0.659	1	0.983
		Significance level(both sides)	0.1	.	0.017
		df	2	0	2
GDP		correlation	0.912	0.783	1
		Significance level(both sides)	0.088	0.017	.
		df	2	2	0
GDP	total power sales	correlation	1	0.08	
		Significance level(both sides)	.	0.968	
		df	0	1	
industrial production		correlation	0.08	1	
		Significance level(both sides)	0.968	.	
		df	1	0	

a. Pearson correlation included in each cell

As can be seen from Tab 3, partial correlation of coefficient between the total power sales and industrial added value is 0.8, the probability p value is 0.968 under the condition of Guizhou province GDP as control variables. The total power sales and industrial added value have a positive relationship under the condition of 0.01 significant level. Compared with Pearson correlation coefficient, partial correlation analysis is more rational.

Multiple correlation analysis

Make multiple linear regression equation about the power sales in Guizhou province and local GDP, industrial added value, hydroelectric power generation, and thermal power generation using SPSS software, reach the significance level of hydroelectric power generation, and thermal power generation is greater than 0.05, so this paper consider multiple linear regression equation and

multiple correlation coefficient about the power sales in Guizhou province and local GDP, industrial added value. Besides, make multiple linear regression equation and multiple correlation coefficient about the power sales in Guizhou province and the five industry including electrolytic aluminium, electrolytic manganese, iron alloy, chemical industry, as well as building materials, the calculation results as shown in table 3.

As seen in Tab.3, multiple correlation coefficients reached by multiple factor correlation analysis are above 0.8, it belongs to the highly correlated. multiple correlation coefficients between total power sales and power of five industry reach 0.91, which shows that as pillar industry, the five industries have an obvious leading role in Guizhou, and should be a point of focus.

Table.3 Correlation analysis results of various factors

Dependent variable	influence factor	multiple correlation coefficient	multiple linear regression equation
total power sales	GDP, industrial production	0.83	$Y=62+253x_1+91.2x_2$
total power sales	The five industry	0.91	$Y=3.963x_1+1.249x_2+1.816x_3+1.575x_4+1.052x_5-21.8$

In conclusion, the selected influence factors are highly related with total power sales in Guizhou province in the empirical analysis, so they should be as the focus of power marketing. Referring to the trends of these factors, the power quantity change trend can be estimated. Even combining the multiple linear regression equation in table 2, based on the forecast of GDP, industrial added value and other economic indicators from the government's economic related departments, total power sales in Guizhou can be directly predicted.

Conclusion

The paper first analyze the factors that result in the change of power sale, and the factor is divided into external factors and internal factors, quantifiable factors and quantitative factors; then introduce the correlation analysis method, compute the Pearson correlation coefficient, partial correlation coefficient and multiple correlation coefficient between power sales and the various affecting factors, and determine the degree of linear correlation of power and various influencing factors. According to the classification of the electric power industry, make the correlation analysis method in the power industry. Examples show that the method is simple, practical, and can provide theory support for the power marketing focus. Finally, obtain the linear regression equation obtained from the correlation analysis; combined with the influence factor, power prediction can be undertook, providing reference for electric power marketing personnel.

References

- [1] Eisa Almeshaie, Hassan Soltan. A methodology for Electric Power Load Forecasting[J]. Alexandria Engineering Journal, 2011, 50(2):137-144.
- [2] Y.X. He, J.X. Zhang, Y. Xu, Y. etc. GaoForecasting the urban power load in China based on the risk analysis of land-use change and load density[J]. International Journal of Electrical Power & Energy Systems, 2015(7):71-79.
- [3] Maria Grazia De Giorgi, Paolo Maria Congedo, Maria Malvoni, etc. Error analysis of hybrid

- photovoltaic power forecasting models: A case study of mediterranean climate[J]. *Energy Conversion and Management*,2015,100:117-130.
- [4] Han Dan, Zhang hongbo, Jia Yong. Analysis and Forecasting for Effect of Economic Factor on Electricity Consumption[J]. *Jilin electric Power*,2009,37(3):16-20.
- [5] Liga Poznaka, Ilze Laicane, Dagnija Blumberga etc.Analysis of Electricity User Behavior: Case Study Based on Results from Extended Household Survey[J]. *Energy Procedia*,2015,72:79-86.
- [6] Diego Pavanello, Willem Zaaiman, Alessandra Colli.Statistical functions and relevant correlation coefficients of clearness index[J]. *Journal of Atmospheric and Solar-Terrestrial Physics*,2015,130:126-133.
- [7] I Xiu-min,JIANG Wei-hua. Research on Linear Correlation and Dependence Measure[J]. *mathematics in practice and theory*, 2006,36(12):188-192.
- [8] ZHANG Shi-qiang, LV Jie-neng, JIANG Zheng, etc.Study of the Correlation Coefficients in Mathematical Statistics[J], *mathematics in practice and theory*,2009,39(19):102-107.
- [9] Miao-Yu Tsai. Comparison of concordance correlation coefficient via variance components, generalized estimating equations and weighted approaches with model selection[J]. *Computational Statistics & Data Analysis*,2015,82:47-58.