Research on the Method of Forecasting Annual Electricity Sales Based on Co-Integration Theory

Junhui Liu1, *, Jiangbo Wang1, Hongkun Bai1, Shuo Yin1, Peng Jia1, Fang Guo1 and Xinhui He2

1State Grid Henan Economic Research Institute, Zhengzhou, Henan 450000, China
2Henan Airport Group, Zhengzhou, Henan 450000, China

*Corresponding author e-mail: luh5620418@126.com

Abstract. Annual sales forecast have the great significance for the power company to determine the growth point of the sales market in advance and maintain the market share. Based on the analysis of many external factors, such as economy, energy and industry, this paper establishes a multivariate modeling prediction method based on co-integration theory, and carries out an empirical study on the forecast of electricity sales in Henan Province in 2020. The example shows that the prediction method is reasonable and feasible to avoid false regression and has a high fitting effect, which can effectively improve the forecast accuracy of medium and long term sales.

1. Introduction

Currently widely used in the medium and long term sales forecasting methods are elastic coefficient method, time series method, exponential smoothing method and so on. These traditional modeling techniques often require the analysis of time series to be stable. In order to improve the prediction accuracy of the electricity sales in the medium and long term, it is necessary to consider the influence of various external factors such as economy, energy and industry, and some of these time series are non-stationary. Although the model results have very high fitting degree and significant statistics, the inference based on these statistics is not correct, resulting in a meaningless false return. It is necessary to make use of the co-integration theory to establish a balanced relationship between two or more non-stationary time series so as to achieve better forecasting results.
2. Method Description

Figure 1. A flow chart of the annual sales forecast method for various external factors is considered.

Including the following steps: Step 1, select external variables for data preprocessing. Select the dimensions of economic development, industrial structure, population growth, urbanization, household consumption and energy consumption, and logarithm the selected variables; Step 2, The stability of the time series of variables is tested, and the order of different orders is given through ADF test; Step 3, carry out "co-integration test" on variables and propose a co-integration model. The method is the Engle-Granger two-step test (EG test) and the maximum likelihood method (Johansen-Juselius, J-J test). Step 4, error correction and causality test. Through the Granger causality test to verify the causal relationship between the variables in the "co-integration relationship", and the error correction will be considered and the "co-integration relationship" model will be proposed after determining the "co-integration relation" of the "same order smooth sequence". Step 5, the forecast results are analyzed, and the forecast results of the corresponding medium and long term sales are calculated according to the logarithmic prediction results.

3. The Case Study
This paper analyzes the overall economy and the electricity market operation status in Henan province, introduces external data such as economy, energy and industry, identifies the dominant factors affecting power demand, studies the co-integration relationship between external environment factors and sales volume, and conducts modeling and forecasting.

3.1. Variables and Data Selection
The sample interval was taken from 1997 to 2016. The data comes from the central database and the middle macro database. Through the analysis of the level of economic development, investment in fixed assets, social consumption, foreign trade and the development of key industries, the power demand function in Henan Province can be expressed as follows:

\[ Q = f (GDP, M1, M2, M3, EC, CP, CSP, U, CC) \]  

(1)

In the formula, \( Q \)—Electricity demand; \( GDP \)—Gross domestic product; \( M1 \)—The primary industry added value; \( M2 \)—The secondary industry added value; \( M3 \)—The tertiary industry added value; \( EC \)—Economic development; \( CP \)—Capital expenditure; \( CSP \)—Social consumption; \( U \)—Urbanization; \( CC \)—Energy consumption.
value; EC—Energy consumption; CP—Coke production; CSP—Crude steel production; PP—Population; U—Urbanization rate; CC—Coal consumption.

3.2. Establish co-integration Equation
In this paper, we use ADF test method to test the time series of variables. First, perform a horizontal test, the power demand LNQ that as explained variable is not stationary series in table 1, the explanatory variables that are non-stationary series will be first-order difference test with the explanatory variables.

| Table 1. ADF unit root test - level test results. |
|-----------------|-------|-------|-------|-------|
| sequence        | ADF   | 1%    | 5%    | 10%   | state |
| LNQ             | -1.659050 | -4.004425 | -3.098896 | -2.690439 | unsteady |
| LNCP            | -0.812922 | -3.959148 | -3.081002 | -2.681330 | unsteady |
| LNCS            | -1.564978 | -3.959148 | -3.081002 | -2.681330 | unsteady |
| LNEC            | -2.174368 | -3.959148 | -3.081002 | -2.681330 | unsteady |
| LNU             | -1.575469 | -3.959148 | -3.081002 | -2.681330 | unsteady |
| LNCC            | -1.401732 | -3.959148 | -3.081002 | -2.681330 | unsteady |
| LNPP            | -2.575946 | -4.057910 | -3.119910 | -2.701103 | unsteady |
| LNGDP           | -4.140500 | -4.121990 | -3.144920 | -2.713751 | steady |
| LNM1            | -0.209356 | -3.959148 | -3.081002 | -2.681330 | unsteady |
| LNM2            | -1.621296 | -3.959148 | -3.081002 | -2.681330 | unsteady |
| LNM3            | 1.181508  | -3.959148 | -3.081002 | -2.681330 | unsteady |

Table 2 shows that the first-order differential sequence of the power demand LNQ is a stationary sequence. The coke yield LNCP, energy consumption LNEC, urbanization rate LNU, secondary industry added value LNM2, and population LNPP that are stationary series can be tested in a "co-integration relationship" with LNQ.

| Table 2. ADF unit root test - first-order difference test results. |
|-----------------|-------|-------|-------|-------|
| sequence        | ADF   | 1%    | 5%    | 10%   | state |
| LNQ             | -3.982883 | -4.800080 | -3.791172 | -3.342253 | steady |
| LNCP            | -4.262520 | -5.124875 | -3.933084 | -3.420030 | steady |
| LNCS            | -2.118150 | -4.800080 | -3.791172 | -3.342253 | unsteady |
| LNEC            | -3.625300 | -4.800080 | -3.791172 | -3.342253 | steady |
| LNU             | -3.965838 | -4.800080 | -3.791172 | -3.342253 | steady |
| LNCC            | -2.291850 | -4.886426 | -3.828975 | -3.362984 | unsteady |
| LNM1            | -2.502413 | -4.886426 | -3.828975 | -3.362984 | unsteady |
| LNM2            | -6.660966 | -5.124875 | -3.933084 | -3.420030 | steady |
| LNM3            | -4.167507 | -5.124875 | -3.933084 | -3.420030 | unsteady |

The J-J test is used to determine the number of "co-integration" vectors in the model. The model is similar to full rank matrix when adding LNU variables to establish a "co-integration relationship" test with LNQ, and can not be obtained with the same "co-integration relationship" with other explanatory variables. Therefore, the following "co-integration test" will eliminate LNU variable.
Table 3. J-J "Co-integration test" trace results.

| Assumed number of co-integration | Eigenvalue | Statistics | 5% Threshold | 1% Threshold |
|---------------------------------|------------|------------|--------------|--------------|
| None **                         | 0.998459   | 94.52083   | 69.81889     | 77.81884     |
| At most 1 **                    | 0.958816   | 54.48036   | 47.85613     | 54.68150     |
| At most 2                       | 0.795245   | 32.96940   | 29.79707     | 35.45817     |
| At most 3                       | 0.653355   | 17.97314   | 15.49471     | 19.93711     |
| At most 4                       | 0.362363   | 5.651362   | 3.841466     | 6.634897     |

Note: * (**) indicates that the significance level of 5% (1%) rejects the null hypothesis; the trace statistics, the maximal statistics result show the presence of a co-integration equation with a significance level of 1% at the same time.

Table 4. "Co-integration test" maximum value statistics results.

| The assumed number of CEs | Eigenvalue | Maximum statistics | 5% Threshold | 1% Threshold |
|---------------------------|------------|--------------------|--------------|--------------|
| None **                   | 0.998459   | 40.04047           | 33.87687     | 39.37013     |
| At most 1 **              | 0.958816   | 21.51096           | 27.58434     | 32.71527     |
| At most 2                 | 0.795245   | 14.99626           | 21.13162     | 25.86121     |
| At most 3                 | 0.653355   | 12.32177           | 14.26460     | 18.52001     |
| At most 4                 | 0.362363   | 5.651362           | 3.841466     | 6.634897     |

Note: * (**) indicates that the significance level of 5% (1%) rejects the null hypothesis; the trace statistics, the maximal statistics result show the presence of a co-integration equation with a significance level of 1% at the same time.

After the above test, the coefficient of co-integration vector is estimated as follows:

\[
\beta' = \{1, -1.106017, -0.132907, 0.021235, -2.273010\} \tag{2}
\]

Therefore, the equation can be expressed as:

\[
\ln Q = 1.106017 \ln \text{NEC} + 0.132907 \ln \text{CPC} - 0.021235 \ln \text{NM2} + 2.273010 \ln \text{PP} - 24.969273 \tag{3}
\]

The above results show that there is a stable long-term equilibrium relationship among the variables in the model within the sample interval.

3.3. Prediction and Correction
According to the "co-integration" equation, the logarithmic data of electricity sales in Henan Province can be predicted. Due to coke output four years ahead of sales, energy consumption and the second industrial added value two years ahead of sales, the population one year ahead of schedule, through the information research such as algorithm fitting and news announcements, into the formula to obtain the corresponding sales forecast data.
Table 5. The results of the logarithmic data forecast of the sales volume.

| Year  | 2015   | 2016   | 2017   | 2018   | 2019   | 2020   |
|-------|--------|--------|--------|--------|--------|--------|
| Actual value | 7.7592 | 7.7796 | —      | —      | —      | —      |
| Predicted value | 7.8765 | 7.8113 | 7.8529 | 7.8824 | 7.9189 | 7.9564 |
| error (absolute value) | 0.0151 | 0.0041 | —      | —      | —      | —      |

An ECM based synthesis equation is used to define the customization:

\[ \Delta LNQ=0.554287\Delta LNPP-0.23972\Delta LEC+0.221246\Delta LNCP-0.07345\Delta LNM2-0.00459 \]  \hspace{1cm} (4)

Table 6. The results of the logarithmic data forecast of the sales after the error correction.

| Year  | 2015   | 2016   | 2017   | 2018   | 2019   | 2020   |
|-------|--------|--------|--------|--------|--------|--------|
| Actual value | 7.7592 | 7.7796 | —      | —      | —      | —      |
| The initial measured value | 7.8765 | 7.8113 | 7.8529 | 7.8824 | 7.9189 | 7.9564 |
| Error correction | 0.0281 | 0.0223 | 0.0051 | 0.0007 | 0.0023 | -0.0012 |
| Predicted value | 7.7784 | 7.7890 | 7.8478 | 7.8817 | 7.9166 | 7.9576 |

After restoring to sell electricity we can see the table:

Table 7. The forecast result of the sales quantity after reduction.

| Year  | 2015   | 2016   | 2017   | 2018   | 2019   | 2020   |
|-------|--------|--------|--------|--------|--------|--------|
| Actual value | 2343   | 2391   | —      | —      | —      | —      |
| Predicted value | 2388   | 2414   | 2560   | 2648   | 2742   | 2857   |
| accelerate | 1.9%   | 0.9%   | 7.1%   | 3.4%   | 3.6%   | 4.2%   |

4. Summary

This paper presents an annual sales forecasting method based on co-integration theory, which can effectively solve the problem of false regression caused by the instability of time series. By quantitatively analyzing the influence of various external factors such as economy, energy and industry on the electricity sales, can effectively improve the medium and long term sales forecast accuracy. The case study of Henan electricity sales forecast shows that energy consumption and population are the main factors that affect Henan electricity demand. With the optimization of industrial structure and the decrease of the share of second industry, the demand of electricity will change in the opposite direction.

References

[1] Jiajia Yang, Junhua Zhao. Core Structure of Power Retailing and Decision-making on Electricity Purchase and Sale [J]. Automation of Electric Power Systems, 2017,41 (14): 10-23.
[2] Di Gao. Comparative Analysis of Power Company Sales Volume Forecast [J]. Channel Technology and Industry, 2017.06: 93-95.
[3] Lei Ma, Weichang Zhang, Xin Liu. Predictive Analysis of Electricity Sales in Power Grid [J]. Modern Manufacturing Economy, 2015.03: 373-374.
[4] Dongping Mo, Jing Li. Power Company Sales Forecast [J]. Electronic Manufacturing, 2016.04: 47-48.
[5] Chen Hui, Angang Zheng, Heiping Zou, Mi Zhang. Optimized Management Simulation of Regional Sales Volume Forecast [J]. Computer Simulation, 2016,33 (10): 310-314.
[6] Chenshu Shen, Huiping Tao. Prediction of Middle and Long Term Electricity Sales in a Province Based on Gray Forecasting Technology [J]. Journal of Anhui Water Conservancy
and Hydropower Engineering Vocational College, 2017, 17 (2): 26-30.

[7] Li Xueling Li, Wenxia Liu, Xin Luo, Dehua Guo. Application of Gray Hierarchical Model Based on Neural Network to Optimize Residuals in Power Forecast [J]. Electrical Times, 2015 (1): 120-122.

[8] Yiru Jiang. Research on Sales Forecasting Based on Dual-model Coordination [J]. Electric Abstracts, 2017.05: 8-10 + 22.

[9] Jiao Fan, Hao Feng, Dongxiao Niu, Xiaoyu Wang, Fuyan Liu. Monthly Sales Forecasting Based on Wavelet Analysis and GM-ARIMA Model [J]. Journal of North China Electric Power University, 2015,42 (4): 101-105.

[10] Chao Xu, Yinghui Li, Baijian Yu, Jian Zhang, Xu Wan. Temperature and Meter Reading Time on the Sale of Electricity [J]. Shandong Electric Power Technology, 2017, 44 (230): 41-45.

[11] Yongyi Shi, Weijia Ling, Yang Guo, Hongtao Xia. Sales Analysis and Forecast [J]. Manager, 2016 (6): 189. Journal of Hunan University (Natural Science), 2016, 43 (10): 62-69.

[12] Xinran Li, Honglin Chen, Hua Leng. Fourier-Markov Correction Model for Middle and Long Term Electricity Forecast [J]. Journal of Hunan University (Natural Science), 2016, 43 (10): 62-69.

[13] Guangsen Yang. Based on the Seasonally Adjusted Industry Expansion Analysis Forecast Analysis [D]. Shandong: Shandong University, 2016.