Early warning analysis of electricity sales based on multi-factor correlation analysis

CHEN Hai[1,*], CUI Yan - yan[1], MA li[1], LIU Wei[1], WANG Jun[2], YUAN Shuai[2], Guo Ju[2], HUANG Yi[2] and Dong Xiao – jing[3]

[1]China Electric Power Research Institute, Beijing City, China
[2]State Grid Corporation of China, Beijing City, China
[3]Economic and Technical Research Institute of Ningxia Electric Power Company of State Grid, Ningxia Province, China

Abstract. Under the background of the slowdown in macroeconomic growth and the gradual liberalization of the power system reforming market, the competition pressure of power grid companies in the electricity sales market has intensified, and the growth of power sales is not optimistic. It is necessary to conduct research and analysis of electricity sales. This paper conducts the analysis with the following steps: first, determines the leading, coincident, lagging economic indicators based on multi-factor correlation analysis, then synthesizes early warning index, forecasts electricity sales, finally, achieves early warning of external risks to improve the company's management quality of the electricity sales.

1 introduction

In March 2015, "Several Opinions of the Central Committee of the Communist Party of China and the State Council on Further Deepening the Reform of the Power System" (Zhongfa [2015] No. 9 Document) was issued, which opened the prelude to the reform of the power system. The "Implementation Opinions on Promoting the Reform of the Power-Sold Side" in the six supporting documents pointed out that the construction of the power market will be accelerated and market competition will be introduced. The National Development and Reform Commission and the National Energy Bureau will issue a notice on the pilot reform of the incremental power distribution business. The document of the National Development and Reform Commission [2016] No. 2480) pointed out that the placing of electricity business was released to social capital, the competition in the electricity sales market of power grid companies intensified, and the risk of loss of customers was large. In addition, the growth rate of macroeconomic slowed down in recent years. The rapid growth of electricity consumption is unsustainable, and the growth rate of electricity sales by power grid companies is limited. The growth of electricity sales of the company is not optimistic.

In order to actively respond to the impact of power system reform and external macroeconomic changes caused to the company's electricity sales market, it is necessary to establish a complete external risk early warning analysis system based on external risk indicators, on the one hand to achieve early warning of power sales risks [11] On the other hand, it provides managers with decision support to improve management quality. Based on the multi-factor correlation analysis between electricity sales and external macroeconomic indicators, this paper conducts early warning analysis of electricity sales through the synthesis of early warning indicators.

2 Association analysis theory

According to different attributes of different factors, the correlation analysis is mainly divided into classification, sequencing and distance correlation analysis.

(1) Correlation analysis of two categorical variables

A categorical variable is a nominal variable, which means that the value of the variable is the name or symbol of the research object. Each value represents a category, and there is no difference in size or order between these values, which is equal. For the gender variable, the values are male and female.

$$\lambda = \frac{\sum f_{im} - F_{ym}}{n - F_{ym}} \tag{1}$$

In equation 1: f_{im} is the number of modes of y distributions for each class of x, and F_{ym} is the number of modes for each classification of the variable y; n is the total number of times. In general, the λ coefficient is between 0 and 1, and a larger value indicates a higher degree of correlation.

(2) Correlation coefficient between two ordered variables

The ordering variable is the level variable, and the variable value has the meaning of order. In other words, the value has a grade or order. Spearman and Kendall

correlation coefficients overcome the shortcomings of Pearson's correlation coefficient that cannot analyze ordered variables and discontinuous variables. The Spearman rank correlation coefficient is generally considered to be the Pearson linear correlation coefficient between the arranged variables. In the actual calculation, there is a simpler method for calculating the Spearman correlation coefficient (p_s). Suppose the original data x_i , y_i has been arranged in descending order, denoted as $x_i^{'}$, $y_i^{'}$ is the original x_i , y_i 's position where the data is arranged after the arrangement, then $x_i^{'}$, $y_i^{'}$ is called the rank of x_i , y_i . thus, $d_i = x_i^{'} - y_i^{'}$ can be obtained as the difference between the ranks of x_i , y_i [1].

If there is no identical rank, then p_s can be calculated by equation 2:

$$p_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{2}$$

An identical value must have the same rank in a column of data, so the rank used in the calculation is the average of the locations where the values are ranked from largest to smallest.

(3) Correlation coefficient between two distance variables

1) Pearson correlation coefficient:

A statistical analysis indicator indicating the closeness of the correlation between the two phenomena.

Its calculation equation is equation 3:

$$\rho_{XY} = \frac{COV(X,Y)}{\sqrt{DX}\sqrt{DY}} \tag{3}$$

The correlation coefficient is represented by the Greek letter ρ, and the value of ρ is between -1 and +1. ρ>0 is a positive correlation and ρ<0 is a negative correlation. ρ=0 means no correlation; the larger the absolute value of ρ, the higher the degree of correlation.

2) Grey correlation coefficient:

Also known as the gray correlation degree, it is based on the analysis of the proximity of the curve shape of each factor, and then provides some suggestions for decision makers [2]. It proposes the concept of gray correlation analysis for each subsystem, and intends to seek numerical relationships among subsystems (or factors) in the system through certain methods. In short, the meaning of gray correlation analysis means that if the situation of two factors changes in the process of system development, that is, the degree of synchronization changes is high, then the relationship between the two can be considered to be large; otherwise, the degree of association between the two is small. Therefore, the gray correlation analysis provides a quantitative measure of a system development and change situation, which is very suitable for dynamic process analysis.

The gray correlation analysis calculation steps are as follows:

a. Identify reference series that reflect system behavior characteristics and comparison series that affect system behavior.

b. Dimensionless processing of the reference series and the comparison series.

c. Find the gray correlation coefficient between the reference series and the comparison series ξ(Xi)

If the reference series is: $X_0=[X_{0(1)}, X_{0(2)}, \ldots, X_{0(n)}]$, the evaluation series is $X_i=[X_{i(1)}, X_{i(2)}, \ldots, X_{i(n)}]$, i=1, 2, …, N, the correlation coefficient between the evaluated series X_i and the reference series X_0 can be defined as: the average value of the correlation coefficient between the evaluation type and the reference type is called the correlation between the comparison series X_i and the reference series X_0, as show in equation 4.

$$\xi_1(k) = \frac{\min\limits_i \min\limits_k |x_0(k)-x_1(k)| + \xi \max\limits_i \max\limits_k |x_0(k)-x_1(k)|}{|x_0(k)-x_1(k)| + \xi \max\limits_i \max\limits_k |x_0(k)-x_1(k)|} \tag{4}$$

Δi(k)= Xi(k)-X0(k) difference sequence, ξ is the resolution coefficient, maximum (small) absolute difference (ie Δmax, Δmin) resolution coefficient P, the resolution coefficient is usually in the range of 0~1 .

d. Get the correlation degree.

In order to clarify the different importance of each indicator, according to the the degree of effect of each indicator, different weights W(k) are given respectively, and the degree of association is calculated according to the following equation 5. The weighted correlation degree between the evaluated type and the reference type is obtained, and by comparison, A comprehensive evaluation of the quantity can be performed.

$$\mathrm{GL(i)} = \sum_{k=1}^{n} w(k)\xi_1(k) \tag{5}$$

In this paper, while analyzing the correlation between electricity sales and external macroeconomic indicators, the data characteristics is taken into account. The sequence and distance method are used for correlation analysis.

3 Establish an early warning model

Based on the theory of multi factor correlation analysis and macroeconomic environment prosperity analysis, this paper uses time difference correlation analysis, Grainger causality test, impulse response function and other metering method to analyse the relationship between the selected external economic indicators [3] and electricity demand first, then establishes an external environment warning model. The electricity industry early warning index is used to reflect the fluctuation of demand in the future electricity market, and to provide early warning of external risks in the power grid industry, as shown in Figure 1.

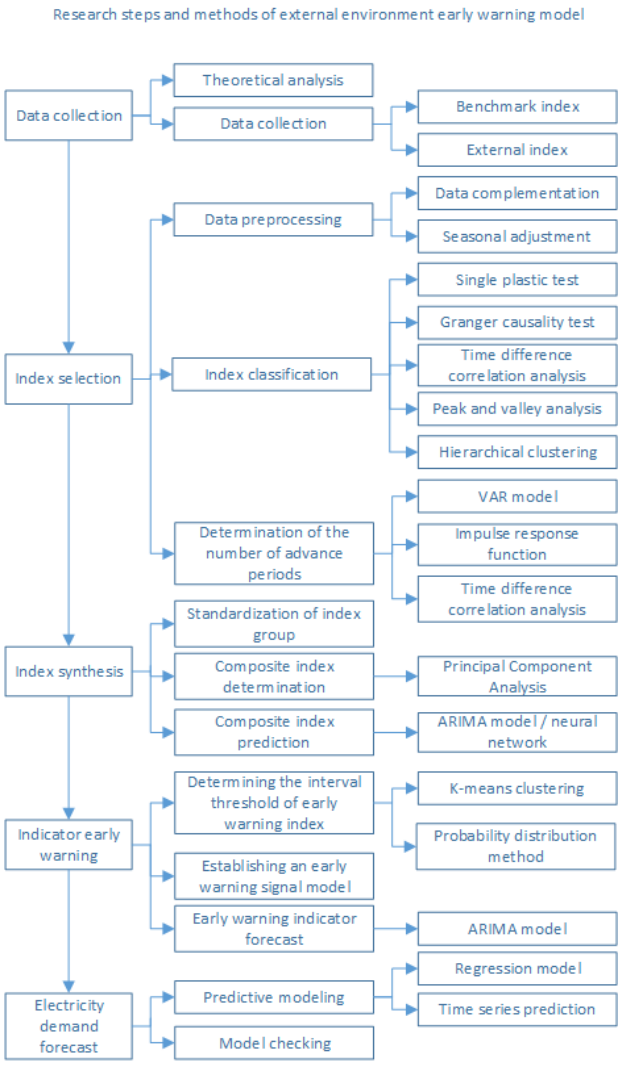

Fig. 1. analysis idea of power sales early warning

4 Case Analysis

4.1 Select external macroeconomic indicators

From the perspective of the close relationship with electricity sales, this paper first selects 45 indicators such as GDP, industrial added value, and industrial product output for subsequent correlation analysis, as shown in Table 1.

4.2 Correlation analysis

The correlation analysis between the selected 45 external macroeconomic indicators and the electricity sales was carried out one by one [4-7], and the analysis results are shown in Tables 2 to 4.

Table 1. Primary selection external index

External index	External index	External index
GDP	Asphalt production	Ten non-ferrous metal production
Industrial added value	Primary selection external index	Yarn production
Steel production	Plastic product production	Kerosene production
Aluminum production	Copper production	AC motor production
Household freezer production	Chemical fiber production	Gas production
Agricultural NPK fertilizer	Vehicle production	Integrated circuit production
Metal cutting machine tool production	Power generation equipment production	Cement special equipment production
Newsprint production	Iron ore raw ore output	Soft drink production
Commodity retail price index	Diesel production	Fuel oil production
Electric hand-held tool production	Household refrigerator production	Primary form of plastic production
Cement production	Coke production	Cloth production
Room air conditioner production	Machine paper and board production	Synthetic detergent production
Mobile communication handset production	Packaging special equipment production	Soda ash. sodium carbonate production
Consumer Price Index	Color TV production	Sulfuric acid production
New construction area of commercial housing	Metal smelting equipment production	Program-controlled switch production

Table 2. Coincident indicator results

Coincident indicator	Correlation coefficient	Delay period
Steel production	0.8367	-2
Aluminum production	0.8107	-2
Synthetic detergent production	0.7518	1
Electric hand-held tool production	0.7302	0
New construction area of commercial housing	0.7238	0
Cloth production	0.7189	-1
Room air conditioner output	0.6713	1

Table 3. Leading indicator results

Leading indicator	Correlation coefficient	Delay period
Household freezer production	0.7942	-3

Metal cutting machine tool production	0.7672	-3
Cement production	0.7017	-3
Asphalt production	0.6850	-6
Chemical fiber production	0.6650	-4
Vehicle production	0.6640	-6

Table 4. Lagging indicator results

Lagging indicator	Correlation coefficient	Delay period
Commodity retail price index	0.7465	5
Consumer price index	0.7182	5

The results show that the coincident indicators of electricity sales include 7 items such as steel production, aluminum production, synthetic detergent production, and electric hand-held tool production, etc. The leading indicators include 6 items such as household freezer production, metal cutting machine tool production, cement production, and vehicle production, etc. The lagging indicators include 2 items such as the commodity retail price index and the consumer price index.

4.3 Synthesize warning index

Step 1: According to the results of the above-mentioned electricity sales first, coincident, and lagging indicators, 15 indicators including industrial added value, steel production, aluminum production, and synthetic detergent production are selected as early warning indicators, and the monthly growth rate of each indicator is calculated.

Step 2: Assume that the year-on-year growth rate of each indicator is in a normal distribution. According to the average and variance of the growth rate of each indicator, the data is calculated to fall in the two-tailed and one-tailed range of 90%, 85%, and 50%. Determine the range of overheating (red light), partial heat (yellow light), stable (green light), partial cold (light blue light), and overcooling (blue light) of each indicator, as shown in Table 5.

Step 3: Take the ratio of the correlation coefficient between each indicator and the electricity sales to the sum of the total correlation coefficients as the weight. The value of the early warning indicator is obtained by multiplying the annual growth rate of each indicator in each month with the corresponding weight, as shown in Table 6 and Table 7.

Step 4: Correspond to the obtained warning value of each indicator and the determined indicator early warning index threshold value (as shown in 8), and obtain the warning value Ai of each indicator, as shown in Table 9.

Step 5: Synthesize the monthly warning index $= \sum_i^{15} A_i W_i / 3 * 100$, as shown in Table 10 and Figure 2, and finally calculate the average value of the monthly warning index to get the total early warning index[8-10], as shown in Table 11.

It can be seen from Table 10 and Figure 2 that the monthly warning index of electricity sales fluctuated greatly. From September 2008 to December 2010, the early warning indicators changed from cold to overheating. After January 2012, the fluctuations tend to be flat, and the early warning index In a stable state, but there is still a tendency for fluctuations to decline, and entered a colder state at the end of 15 years. The total synthetic early warning index is 100.5, which is in the stable green light interval. It shows that since January 2007, the average growth rate of electricity sales has been stable, and many places are in the 4-5% growth rate range.

4.4 Electricity sales warning result

According to the results of the above-mentioned early warning index and macroeconomic forecast, the forecast results of the early warning index on electricity sales will show a volatility trend in the sales of electricity in late 2017. The electricity consumption situation is relatively stable, that is, the risk of electricity sales is relatively small. As show in figure 3.

5 Conclusion suggestion

In this paper, multi-factor correlation analysis is used to determine 15 macroeconomic indicators with high correlation coefficient with electricity sales, and then the early warning index is used to realize the early warning of electricity sales. From the early warning structure, the external macroeconomic indicators have been better since 2017, the electricity consumption situation is more optimistic, and the risk of selling electricity is small.

In view of the fact that the growth rate of electricity sales is the most important indicator for grid companies to input and output, the changes in the future economic environment will inevitably affect the results of the electricity sales index. In addition, the power system reform has been comprehensively promoted in recent years, and there are certain shortcomings in the sale of electricity warnings only from macroeconomic indicators. Therefore, it is recommended that grid companies use the analysis results of the early warning index of electricity sales, and need to establish a set of electricity sales adjustment analysis system based on power system reform, comprehensively assess the risk of electricity sales, and improve the company's risk prevention capabilities.

Table 5. Early-warning lamp setting

State	Overheating	Partial heat	Stable	Partial cold	Overcooling
Early-warning lamp	Red light	Yellow light	Green light	Light blue light	Blue light
Early warning value (A)	5	4	3	2	1
Probability	10%	15%	50%	15%	10%

Table 6. Calculation of weights of each indicator

Early warning indicator	Indicator type	Mean (X)	Standard deviation (ε)	Correlation coefficient	Weights
Industrial added value	Coincident indicator	109.25	4.24	0.85	0.08
Steel production	Coincident indicator	115.02	17.12	0.84	0.08
Aluminum production	Coincident indicator	118.37	17.01	0.81	0.07
Household freezer production	Leading indicator	110.12	22.79	0.79	0.07
Metal cutting machine tool output	Leading indicator	109.49	32.17	0.77	0.07
Synthetic detergent production	Coincident indicator	107.18	10.36	0.75	0.07
Commodity retail price index	Lagging indicator	102.41	2.72	0.75	0.07
Electric hand-held tool output	Coincident indicator	99.51	16.95	0.73	0.07
New construction area of commercial housing	Coincident indicator	108.60	28.57	0.72	0.07
Cloth production	Coincident indicator	106.06	11.70	0.72	0.06
Cement production	Leading indicator	102.05	6.14	0.70	0.06
Asphalt production	Leading indicator	125.56	55.18	0.68	0.06
Room air conditioner production	Coincident indicator	114.55	29.33	0.67	0.06
Chemical fiber production	Leading indicator	111.54	5.59	0.66	0.06
production	Leading indicator	113.16	26.65	0.66	0.06

Table 7. Values of early warning indicators for each indicator (increased rate)

Early warning indicator value	2007-1-1	2007-2-1	2007-3-1	.....	2016-8-1	2016-9-1	2016-10-1
Industrial added value	112.89	113.83	114.86	.....	107.56	107.43	107.32
Steel production	141.36	140.61	140.07	.....	94.48	93.91	93.21
Aluminum production	127.39	132.00	136.28	.....	118.22	120.01	121.80
Household freezer production	137.94	139.31	141.83	.....	99.69	97.66	96.26
Metal cutting machine tool production	192.85	186.68	177.62	.....	100.20	101.25	101.38

Synthetic detergent production	97.72	98.14	99.92	.....	113.42	114.52	115.06
Commodity retail price index	101.64	101.73	101.88	.....	100.00	99.89	99.83
Electric hand-held tool production	99.93	108.75	117.30	.....	104.57	106.73	107.84
New construction area of commercial housing	92.08	94.12	95.52	.....	134.10	134.83	134.53
Cloth production	98.82	101.72	105.63	.....	102.43	102.30	101.84
Cement production	113.42	111.87	110.25	.....	112.02	114.15	115.49
Asphalt production	158.38	147.60	135.15	.....	101.12	100.67	100.57
Room air conditioner production	58.55	73.76	92.31	.....	114.54	113.72	109.82
Chemical fiber production	130.61	130.61	130.61	.....	108.65	108.46	108.18
Vehicle production	125.32	123.19	121.03	.....	109.63	105.62	103.26

Table 8. Threshold values of early warning indicators for each indicator

Early warning indicator	Mean (X)	Standard deviation (ε)	Overheating (red light)	Partial heat (yellow light)		Stable (green light)			Partial cold (light blue light)		Overcooling (blue light)
			Spot probability	Spot probability	Upper limit (X+1.282ε)	Spot probability	Upper limit (X+0.674ε)	Lower limit (X-0.674ε)	Spot probability	Lower limit (X-1.282ε)	Spot probability
Industrial added value	109.25	4.24	10%	15%	114.68	50%	112.11	106.40	15%	103.82	10%
Steel production	115.02	17.12	10%	15%	136.96	50%	126.56	103.49	15%	93.08	10%
Aluminum production	118.37	17.01	10%	15%	140.17	50%	129.83	106.91	15%	96.57	10%
Household freezer production	110.12	22.79	10%	15%	139.34	50%	125.48	94.76	15%	80.90	10%
Metal cutting machine tool production	109.49	32.17	10%	15%	150.73	50%	131.17	87.80	15%	68.24	10%
Synthetic detergent production	107.18	10.36	10%	15%	120.46	50%	114.16	100.19	15%	93.90	10%
Commodity retail price index	102.41	2.72	10%	15%	105.89	50%	104.24	100.58	15%	98.93	10%
Electric hand-held tool production	99.51	16.95	10%	15%	121.24	50%	110.93	88.09	15%	77.79	10%
New construction area of commercial housing	108.60	28.57	10%	15%	145.23	50%	127.86	89.34	15%	71.97	10%
Cloth production	106.06	11.70	10%	15%	121.05	50%	113.94	98.17	15%	91.06	10%
Cement production	102.05	6.14	10%	15%	109.93	50%	106.20	97.91	15%	94.18	10%

Asphalt production	125.56	55.18	10%	15%	196.30	50%	162.75	88.36	15%	54.81	10%
Room air conditioner production	114.55	29.33	10%	15%	152.15	50%	134.32	94.78	15%	76.94	10%
Chemical fiber production	111.54	5.59	10%	15%	118.71	50%	115.31	107.77	15%	104.37	10%
Vehicle production	113.16	26.65	10%	15%	147.32	50%	131.12	95.20	15%	79.00	10%

Table 9. Monthly warning indicators for each indicator

Early warning indicator value	2007-2-1	2007-3-1	2007-4-1	.....	2016-8-1	2016-9-1	2016-10-1
Industrial added value	4	5	5	.....	3	3	3
Steel production	5	5	5	.....	2	2	2
Aluminum production	4	4	4	.....	3	3	3
Household freezer production	4	5	5	.....	3	3	3
Metal cutting machine tool production	5	5	5	.....	3	3	3
Synthetic detergent production	2	2	3	.....	3	4	4
Commodity retail price index	3	3	3	.....	2	2	2
Electric hand-held tool production	3	4	5	.....	3	3	3
New construction area of commercial housing	3	3	3	.....	4	4	4
Cloth production	3	3	3	.....	3	3	3
Cement production	5	5	4	.....	5	5	5
Asphalt production	3	3	3	.....	3	3	3
Room air conditioner production	1	2	3	.....	3	3	3
Chemical fiber production	5	5	5	.....	3	3	3
Vehicle production	3	3	3	.....	3	3	3

Table 10. Monthly warning index

Early warning indicator value	2007-1-1	2007-2-1	2007-3-1	.....	2016-8-1	2016-9-1	2016-10-1
Industrial added value	4	4	5	.....	3	3	3
Steel production	5	5	5	.....	2	2	2
Aluminum production	3	4	4	.....	3	3	3
Household freezer production	4	4	5	.....	3	3	3
Metal cutting machine tool production	5	5	5	.....	3	3	3

Synthetic detergent production	2	2	2	.....	3	4	4
Commodity retail price index	3	3	3	.....	2	2	2
Electric hand-held tool production	3	3	4	.....	3	3	3
New construction area of commercial housing	3	3	3	.....	4	4	4
Cloth production	3	3	3	.....	3	3	3
Cement production	5	5	5	.....	5	5	5
Asphalt production	3	3	3	.....	3	3	3
Room air conditioner production	1	1	2	.....	3	3	3
Chemical fiber production	5	5	5	.....	3	3	3
Vehicle production	3	3	3	.....	3	3	3
Early warning index	116.5	118.9	128.0	.....	101.6	103.9	103.9
Warning index signal light	Yellow light	Yellow light	Yellow light	.....	Green light	Green light	Green light

Table 11. Total Synthetic Warning Index

Early warning indica-tor	Indi-cator type	Mean (X)	Stan-dard devia-tion (ε)	Over-heating (red light)	Partial heat (yellow light)		Stable (green light)			Partial cold (light blue light)		Over-cooling (blue light)
				Spot proba-bility	Spot probabi-lity	Up-per limit (X+1.282ε)	Spot probab ility	Upper limit (X+0.674ε)	Lower limit (X-0.674ε)	Spot probab ili-ty	Lower limit (X-1.282ε)	Spot proba-bility
Early warning index		100.50	22.67	10%	15%	129.56	50%	115.78	85.21	15%	71.43	10%

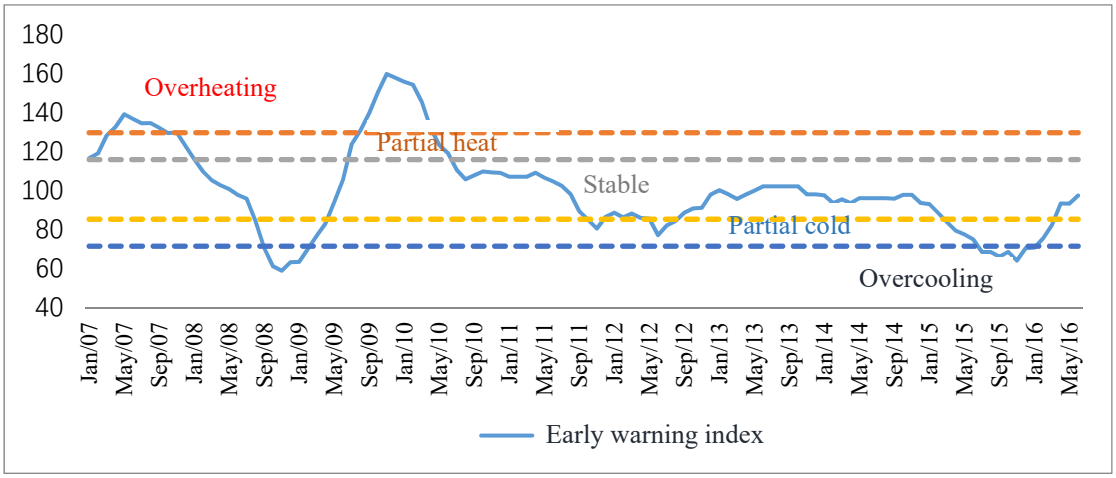

Fig. 2. Monthly warning index

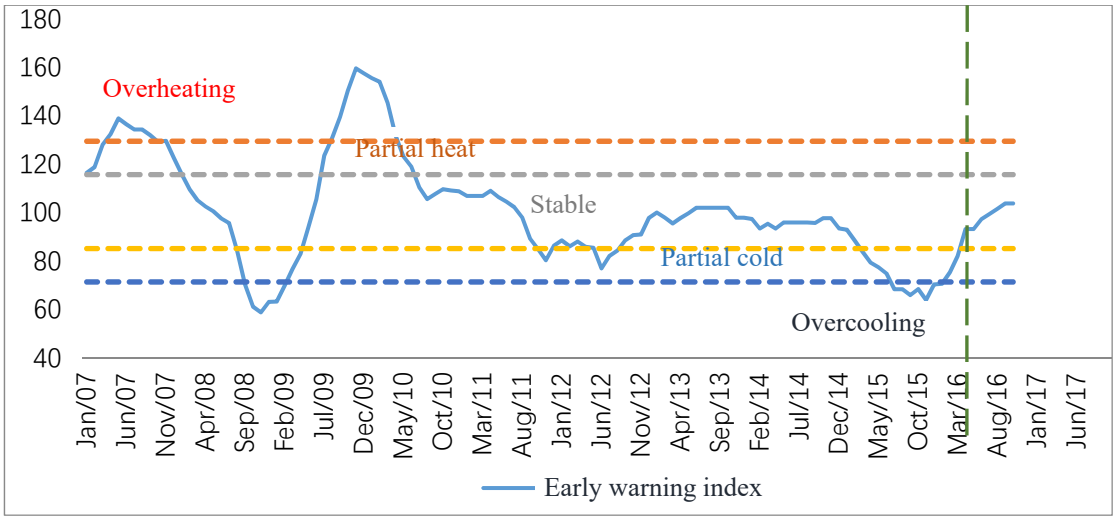

Fig. 3. Electricity sales warning results

Reference

1. Z.L. Dong, Association Analysis of Industrial Convergence and Industrial Structure Optimization and Upgrading. Cooperative Economy and Technology, **16** (2017)
2. Y. Zhang et al., Gray correlation analysis combined with support vector machine for near-infrared spectroscopy. Spectroscopy and Spectral Analysis, **2 (**2013)
3. Y.C. Liu, et al., China's Macroeconomic Analysis and Forecast 2014-2015 – China's Macroeconomics in the New Normal Period. Economic Theory and Economic Management, **3** (2015)
4. J.G. Hua, Advance Indicators and Macroeconomic Fluctuation Prediction. Systems Engineering Theory and Practice, **10** (2014)
5. F.H. Lu, First, Synchronous Synthesis Index. Data, **4** (2011)
6. [6] H. Zhang, Analysis of the Dynamic Relationship between Economic Advance Index and China's Macroeconomic Growth. Economic Issues, **11** (2012)
7. Z. Zhang, A problem that should be paid attention to when using time difference correlation analysis to determine leading, coincident and lagging indicators. Forecast, **3 (**1991)
8. Y.N. Shi, China's Prosperity Forecast Based on Diffusion Index. Finance and Economics, **4** (2014)
9. L.Y. Chen, Li Renbo, Li Chunfeng, Analysis of the Synthetic Index of Current Economic Trends in China. Contemporary Economic Research, **2** (2014)
10. X.L. Kong, Zhang Tongbin, Gao Tiemei, Study on the Periodic Fluctuation Characteristics and the Characteristics of Current Volatility of China's Industrial Economy Based on Prosperity Index. Mathematics Practice and Cognition, **7** (2012)
11. J.M Wang, Research on the Early Warning Function of China's Leading Index on Economic Fluctuation. Jiangsu Social Science, **3** (2014)