# Electricity sales forecasting of Z province based on SARIMA model and temperature correction

Feng Hao[1,a], Fan Jiao[2,b], Zhu Guo-rong[1], Xu Hui[1], Niu Dong-xiao[2],Guo Xiao-dan[2]

[1.] Zhejiang Electric Power Company Economic Research Institute, Hangzhou 310008, China

[2.] School of Business &Management, North China Electric Power University, Beijing 102206, China

[a]superfj@163.com, [b]fanjiao19@163.com

**Keywords:**SARIMA Model; Electricity Sales Forecasting; Temperature Correction

**Abstract.** The sales of electricity have great effect on operation decisions and budget arrangement of power grid companies directly, so the accuracy of electricity sale forecasting is quite crucial. Since the sales of electricity data has obvious increase trend with time and seasonal variation characteristics, single ARIMA model cannot get ideal forecasting results. According to the monthly electricity sales data of Z Province from 2011 January to 2014 October, this paper builds SARIMA model and get high forecast precision. In addition, since the sales of electricity is affected by temperature obviously and there are many temperature anomaly days in Z Province in 2014 compared with previous years, this paper processed the temperature correction and further improved the prediction accuracy.

## Introduction

With the development of economic and electric market, the sales of electricity became the crucial factor that affects the operation status of electricity power enterprises. Thus, reasonable and accurate forecasting of electricity sales can contribute to budget arrangement and orderly grid construction investment with great significance.

Nowadays, there are various forecasting methods, such as multiple linear regression, Holt-Winters and grey model. Multiple linear regression model cannot simulate nonlinear case. And since the sale of electricity is affected by many influence factors, so it is not suitable to forecast short-term electricity[1]; Grey Model is using mathematical method to solve problems of imperfect information system[2-3]; Time series forecasting method is extrapolation method based on series features. SARIMA model belongs to time series forecasting method, which is entirely based on historical data and is more suitable for monthly or daily electricity sale[4].

## SARIMA Forecast Model

SARIMA model derived from ARIMA model (also known as autoregressive moving average model), proposed by Box and Jenkins in the early 70s of last century. ARIMA (p, d, q) model contains three cases, AR (p) autoregressive model, MA (q) moving average model and ARMA (p, q) autoregressive moving average model[5-7]. The ARIMA model is transforming non-stationary time series into stationary time series, so we need to conduct ADF test before the establishment of model. The ARIMA model is the D order difference of ARMA mode, and the corresponding algebraic of ARMA model is as follows.

$$y_t = c + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (1)$$

We express ARIMA model after D order difference with the lag operator as follows.

$$\Phi_p(L)(1-L)^d y_t = c + \Theta_q(L)\varepsilon_t \quad (2)$$

However, when the time trend has obvious seasonal variation, it is not ideal to use simple ARIMA model to realize high precision. So we combine ARIMA model with stochastic seasonal model, namely SARIMA. In SARIMA model, the original series is differenced by d order to eliminate increase trend, and is differenced by D order with the length of s to eliminate seasonal trend. The model expression is as follows.

$$\phi_P(L)\Phi_P(L^s)(1-L)^d(1-L)^D y_t = c + \theta_q\Theta_Q(L^s)\varepsilon_t \quad (3)$$

Among them, P is the seasonal autoregressive order number, Q is the seasonal moving average order number, $(1-L)^d$ represents D order phased difference, $(1-L)^D$ means D order seasonal difference, and $\Phi_P(L^s)$ and $\Theta_Q(L^s)$ are seasonal P order autoregressive operator and Q order moving average operator respectively.

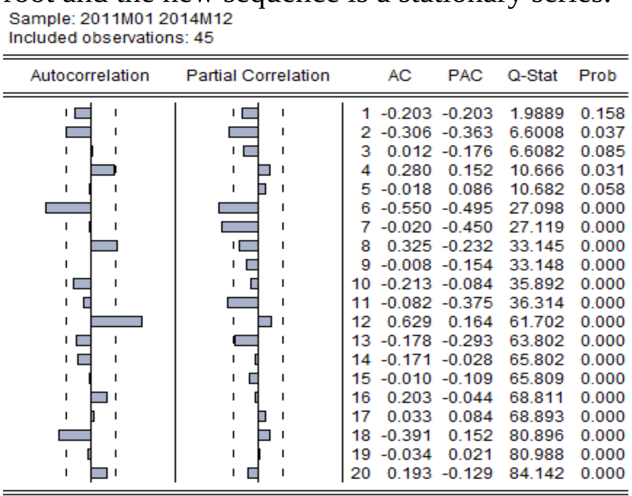
## DeterminING model parameterS

The electricity sales of Z province between 2011 January to 2014 October are shown in Figure 1. We can know that the sale of electricity has increasing trend and seasonal fluctuation. Therefore, we can use SARIMA model to make the series stationary.

First, the original sequence by one order difference is to eliminate the time trend and the coefficients of autocorrelation and partial correlation are shown in Figure 2. The new sequence autocorrelation exceeded the 95% confidence interval at lag 12 period, so we processed season difference with the step length of 12. The autocorrelation coefficients of new series are shown in Figure 3. We can see that the autocorrelation coefficient fell into a random interval quickly without seasonal fluctuations, which means that the new sequence is smooth.


Figure 1. The electricity sale of Z province between 201101-201410

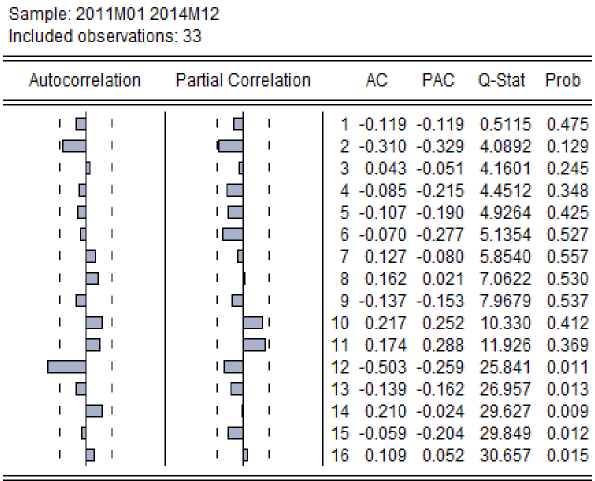
We carried out ADF test on the new series after the difference of trend and season and the results are shown in Figure 4. It reflected that we should reject the null hypothesis, namely there is not a unit root and the new sequence is a stationary series.



Sample: 2011M01 2014M12
Included observations: 45

| Lag | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|
| 1 | -0.203 | -0.203 | 1.9889 | 0.158 |
| 2 | -0.306 | -0.363 | 6.6008 | 0.037 |
| 3 | 0.012 | -0.176 | 6.6082 | 0.085 |
| 4 | 0.280 | 0.152 | 10.666 | 0.031 |
| 5 | -0.018 | 0.086 | 10.682 | 0.058 |
| 6 | -0.550 | -0.495 | 27.098 | 0.000 |
| 7 | -0.020 | -0.450 | 27.119 | 0.000 |
| 8 | 0.325 | -0.232 | 33.145 | 0.000 |
| 9 | -0.008 | -0.154 | 33.148 | 0.000 |
| 10 | -0.213 | -0.084 | 35.892 | 0.000 |
| 11 | -0.082 | -0.375 | 36.314 | 0.000 |
| 12 | 0.629 | 0.164 | 61.702 | 0.000 |
| 13 | -0.178 | -0.293 | 63.802 | 0.000 |
| 14 | -0.171 | -0.028 | 65.802 | 0.000 |
| 15 | -0.010 | -0.109 | 65.809 | 0.000 |
| 16 | 0.203 | -0.044 | 68.811 | 0.000 |
| 17 | 0.033 | 0.084 | 68.893 | 0.000 |
| 18 | -0.391 | 0.152 | 80.896 | 0.000 |
| 19 | -0.034 | 0.021 | 80.988 | 0.000 |
| 20 | 0.193 | -0.129 | 84.142 | 0.000 |

Figure 2. Autocorrelation and partial correlation results of series after the difference of trend



Sample: 2011M01 2014M12
Included observations: 33

| Lag | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|
| 1 | -0.119 | -0.119 | 0.5115 | 0.475 |
| 2 | -0.310 | -0.329 | 4.0892 | 0.129 |
| 3 | 0.043 | -0.051 | 4.1601 | 0.245 |
| 4 | -0.085 | -0.215 | 4.4512 | 0.348 |
| 5 | -0.107 | -0.190 | 4.9264 | 0.425 |
| 6 | -0.070 | -0.277 | 5.1354 | 0.527 |
| 7 | 0.127 | -0.080 | 5.8540 | 0.557 |
| 8 | 0.162 | 0.021 | 7.0622 | 0.530 |
| 9 | -0.137 | -0.153 | 7.9679 | 0.537 |
| 10 | 0.217 | 0.252 | 10.330 | 0.412 |
| 11 | 0.174 | 0.288 | 11.926 | 0.369 |
| 12 | -0.503 | -0.259 | 25.841 | 0.011 |
| 13 | -0.139 | -0.162 | 26.957 | 0.013 |
| 14 | 0.210 | -0.024 | 29.627 | 0.009 |
| 15 | -0.059 | -0.204 | 29.849 | 0.012 |
| 16 | 0.109 | 0.052 | 30.657 | 0.015 |

Figure 3. Autocorrelation and partial correlation results of series after the difference of season

Null Hypothesis: ESED3 has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=11)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -6.355182 | 0.0000 |
| Test critical values: | 1% level | -3.661661 | |
| | 5% level | -2.960411 | |
| | 10% level | -2.619160 | |

*MacKinnon (1996) one-sided p-values.

Figure 4. The ADF results of new series after difference of trend and season

In the model of SARIMA (p, d, q) (P, D, Q), since the original sequence by a difference, d=1, after a seasonal difference, D=1. We can see from Figure 4 that the partial correlation coefficients of new series ESED3 are trailing and the autocorrelation coefficients are censoring. Since partial correlation coefficient is not zero obviously in the second order, so we decided p=2, q=2 or p=2, q=1. Due to the lag of 12 order, partial correlation coefficient is not zero obviously, so P=1, Q=1. According to Table 1, after comparing key indicators AIC Value, SC Value and Adjusted R2, we agreed that the model of ARIMA (2,1,2) (1,1,1) 12 is better.

Table 1. Comparison results of models

| Models | Adjusted R2 | AIC Value | SC Value |
|---|---|---|---|
| ARIMA(2,1,2)(1,1,1)12 | 0.2897 | 26.9598 | 27.1448 |
| ARIMA(1,1,2)(1,1,1)12 | 0.2755 | 27.0222 | 27.2054 |

**Forecasting electricity sales**

We use ARIMA(2,1,2)(1,1,1)12 model to forecast the electricity sale of 201401-201410， and the forecasting results are shown in Table 2

We can see from Table 2 that the biggest error rate is 9.33%, the smallest one is 0.37%, and average error rate is 3.97%. Through compared with actual data, the results of prediction have good agreement with the observed one

**Analysis of temperature amendment**

Except for the influence of economic factors, the sale of electricity is also greatly affected by temperature factors. The average value of monthly temperature in 2013 and 2014 are quite different. The temperature of province Z in the summer of 2013 keeps high. But the temperature in the summer of 2014 is quite low. So we tried to analyze the influence of temperature on the sale of electricity. Firstly, we fitted the quadratic function according to the scatter diagram of monthly average temperature. The scatter diagram and quadratic function are shown in Figure 5.

Then we calculated the average monthly temperature of January to October for 2011 to 2013 and the corresponding electricity sales using above fitting function. Also, we forecast the electricity sale of 2014 (January to October) using actual temperature data. Then we get the temperature adjustment and put it on the previous forecasting results by SARIMA. The final forecasting value and error rate are shown in Table 3.

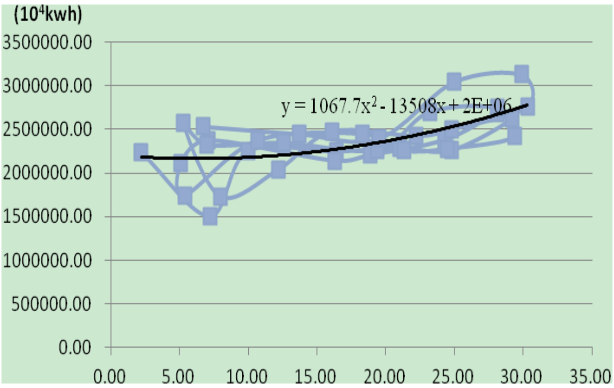



Figure 5. Fitting figure of monthly temperature and electricity sale in 2011-2013

After temperature adjustment, the accuracy of electricity sale forecasting is further increased. The forecasting error after adjustment is 3.67%.

Table 2. The forecasting results of electricity sale(104 kwh)

| Month | Actual electricity sale | Forecasting electricity sale | Absolute error | Error rate |
|---|---|---|---|---|
| 201401 | 2574410.33 | 2685835.34 | 111425.01 | 4.33% |
| 201402 | 1697129.88 | 1839830.70 | 142700.82 | 8.41% |
| 201403 | 2371086.16 | 2149746.82 | -221339.34 | -9.33% |
| 201404 | 2632836.11 | 2587321.92 | -45514.19 | -1.73% |
| 201405 | 2505566.95 | 2535246.51 | 29679.57 | 1.18% |
| 201406 | 2624499.95 | 2610884.35 | -13615.60 | -0.52% |
| 201407 | 2653837.02 | 2625160.60 | -28676.42 | -1.08% |
| 201408 | 3052356.87 | 2951035.48 | -101321.39 | -3.32% |
| 201409 | 2861040.84 | 2910600.58 | 49559.74 | 1.73% |
| 201410 | 2497602.66 | 2477243.22 | -20359.44 | -0.82% |
| 201411 | - | 2561242.97 | - | - |
| 201412 | - | 2534283.70 | - | - |

Table 3. Adjustment results of electricity sale using temperature (104 kwh)

| Month | Actual value | Forecasting value | Temperature adjustment | Forecasting value after adjustment | Error rate |
|---|---|---|---|---|---|
| 201401 | 2574410.33 | 2685835.34 | -8008.66 | 2677826.68 | 4.02% |
| 201402 | 1697129.88 | 1839830.70 | 327.65 | 1840158.35 | 8.43% |
| 201403 | 2371086.16 | 2149746.82 | 85111.76 | 2234858.58 | -5.75% |
| 201404 | 2632836.11 | 2587321.92 | -753.86 | 2586568.06 | -1.76% |
| 201405 | 2505566.95 | 2535246.51 | -6437.91 | 2528808.60 | 0.93% |
| 201406 | 2624499.95 | 2610884.35 | -11674.82 | 2599209.53 | -0.96% |
| 201407 | 2653837.02 | 2625160.60 | -49144.93 | 2576015.67 | -2.93% |
| 201408 | 3052356.87 | 2951035.48 | -86923.77 | 2864111.71 | -6.17% |
| 201409 | 2861040.84 | 2910600.58 | 12489.38 | 2923089.95 | 2.17% |
| 201410 | 2497602.66 | 2477243.22 | 26721.86 | 2503965.07 | 0.25% |

## Conclusion

As the core of power grid enterprises' operating decision, the accurate forecasting of electricity sale has great significance. Proper forecasting model and amendment of external factors can help increase forecasting accuracy. This paper forecasted the monthly electricity sale using SARIMA mode. Further, it took the influence of temperature factor into consideration. The influence of temperature on monthly electricity sale is obvious in winter and summer, especially the days with unusual temperature. This paper corrected the electricity data by fitting the average value of monthly temperature and electricity sale. Because of the low accuracy of temperature forecasting, the amendment of temperature is mainly used in postmortem analysis. It can't be used in the forecasting of monthly electricity sale before the accurate forecasting of itself. In order to increase the accuracy of forecasting, we should do further research of the influence of other external factors on monthly electricity sale, and correct the original data.

**Acknowledgment**

In this paper, the research was sponsored by the National Science Foundation of China (71071052, 71471059)

## References

[1] Ding Hao;Rong Rong. Electricity Consumption Forecasting of Shandong Based on Multivariate Linear Regression Model and Gray Theory[J] Henan Science, 2013, 31(9):1535-1539Henan Sciences Henan Sciences

[2] Deng Julong. Grey system theory[M].Wuhan: Press of Huazhong science and Technology University.1990.

[3] Wang Tao, Zhang Wei, Lv Mengming. Study on Electricity Sales Forecasting Method of Regional Power Grid Based on Time Series[J]. Electrical technology, 2010(11) : 9-12.

[4] Zhang Jian. The application of SARIMA Model in forecastiong CPI of China[J]. Statistics and decision, 2011(5): 28-30.

[5] Li Yanyi, Liu Rong, Ding Weidai. EVIEWS Statistics and Application[M]. Beijing: Electronic Industry Press,2013

[6] Zhou M,Yan Z,Ni Y X, et al.Electricity Price Forecasting with Confidence-interval Estimation through an Extended ARIMA Approach [J].IEE Proc-Gener,Transm ,and Distrib, 2006,153(2):187-195.

[7] Huang S J, Shih K R. Short-term Load Forecasting via ARMA Model Identification Including Non-Gaussian Process Considerations[J].IEEE Trans on Power Systems, 2003, 18(2):673-679