A Long-Term Forecast Method for the Investment Demand of Power Grid Based on Linear Regression and Error Correction of Variable

Yu-qing HE, Xing ZHANG*, Qin JIANG, Shu-juan CHEN, Ying-jie LI, Han-yu YANG and Jia-yong LI

1Economic and Technological Research Institute of Hunan Power Grid, State Grid Hunan Power Co., Ltd, Changsha, China
2Key Laboratory of Intelligent Information Analysis and Comprehensive Optimization of Energy Internet in Hunan Province Hunan University, Changsha, China
3Key Laboratory of Intelligent Information Analysis and Comprehensive Optimization of Energy Internet in Hunan Province Hunan University, Changsha, China
*Corresponding author

Keywords: Power grid, Investment demand, Linear regression, Forecast.

Abstract. In the long-term forecast for the investment demand of power grid, the error of independent variable will result in a large error between forecast result and actual result. In this paper, the correction parameters of factors which have an influence on investment demand of power grid is introduced and the correction model of long-term forecast is built. Firstly, through the correlation analysis and the principle comment analysis (PCA), the variables which the model needs are selected. Secondly, the long-term forecast model based on linear regression is built and the variable coefficient is calculated through least square method. Thirdly, after analyzing historical data of forecast value and actual value, correction model of long-term forecast is built through introducing correction parameters of each variables. Finally, compared with the calculation results which get from correction model and the long-term forecast model, it is proved that the error of correction model is smaller.

Introduction

The construction of power grid has an important influence on the stability of regional economic, sustainable development, the life and production of inhabitants [1]. With the rapid development of economic and the deep reform of electric system, the investment climate has changed a lot, the accuracy of forecasting investment demand is higher. The more accurate forecasting investment demand of power grid is, the more obvious the grid investment planning is. Therefore, how to appropriately predict the demand of power grid investment becomes an urgent matter.

Many scholars have provided different measures to calculate the investment demand of power grid. It establishes variables which have an influence on power grid investment through Granger and builds a long-term equilibrium model between investment demand and influence factors in [2, 3]. However the selection of variables isn’t comprehensive which only includes the aspect of power grid investment and economic. It takes the method of grey correlation degree to forecast the investment demand of power grid and builds short-term forecast for the investment demand of power grid in [4]. It uses Monte Carlo Simulation to build model for investment demand of power grid, and the model considers many kinds of influential factors [5]. Predicting the trend of indicators based on the model of grey prediction, we get the influence of indicators on the investment of power grid construction through the method of Analytic Hierarchy Progress (AHP). While AHP has the disadvantage of subjectivity [6]. It considers different characters of investment demand in different stage of power grid development and includes bell-shaped curve to describe the economic and political influence on the investment demand of power grid [7].

On the basis of [8-9], the process, select variable build long-term forecast model for the investment demand of power grid introduce correction parameters of each variables and build
correction model of long-term forecast, is proposed. Firstly, through the correlation analysis and PCA, the variables which the model needs are selected. Secondly, the long-term forecast model based on linear regression is built. Finally, after analyzing historical data of forecast value and actual value, the correction model of long-term forecast is built through introducing correction parameters of each variables. Reducing dimension of variable system makes calculation simple and introducing correction parameters of each variables reduces the error of results, the accuracy of forecasting investment demand of power grid is improved. It’s more significant for power grid planning and investment decision.

**Theoretical Basis**

**Multiple Linear Regression (MLR)**

(1) The expression of MLR’s equation

The model of MLR can show the correlation between independent various and dependent various. The investment demand of power grid is related to many kinds of factors, in this case, it’s more appropriate to use the model of MLR. And the expression is:

$$ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon $$

where $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are variable coefficients; $x_0, x_1, x_2, \ldots, x_k$ are different independent variables; $y$ is the dependent variable.

(2) The expression of MLR’s coefficient

If putting specific value into Eq. 1, the expression is :

$$ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i $$

And the least square method is used to solve MLR’s coefficient, the expression is:

$$ \beta = (X^T X)^{-1} X^T Y $$

where

$$ Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_{i1} & x_{i2} & \ldots & x_{in} \\ \ldots & \ldots & \ldots & \ldots \\ x_{11} & x_{12} & \ldots & x_{1n} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \ldots \\ \beta_k \end{bmatrix} $$

**Variable Correction**

In the MLR model, if one independent variable changes, the dependent variable also will change. When the model is right and the independent variables have predictive value, accuracy of the model will be improved through the improving the variables’ accuracy. When the independent variables are predictive value in the MLR model, the correction parameters can be considered to improve the accuracy of independent variables. After introducing correction parameters of each variables, the equation is:

$$ y = \beta_0 + \beta_1 x_1 \times (1 + \lambda_1) + \beta_2 x_2 \times (1 + \lambda_2) + \ldots + \beta_k x_k \times (1 + \lambda_k) + \varepsilon $$

Where $\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_k$ are the correction parameter of the corresponding variables.

**The Construction of the Correction Model**

**The Procedure of the Modeling**

It’s a complex process to forecast the investment demand of power grid. Firstly, the variables which
the model needs should be select. Secondly, because the difference of magnitudes are big between variables, the dimensional difference should be eliminate through logarithm. Only finish these, the long-term forecast model and correction model can be build. From variable selection to building correction model for forecasting investment demand of power grid, the flow chart of model is as follows.

![Flow Chart of Model Construction](image)

**Figure 1.** The flow chart of model construction.

**The Selection of Variables**

There are many variables affecting the result of foresting investment demand of power grid, and some of them have a large correlation. The MLR model will lose predictive function if all variables are introduced. Therefore it is necessary to select variables and the processes are as follows.

1. Inputting historical data. The correlation analysis can be conducted in the IBM SPSS Statistics. The person correlation coefficients is used to calculate the correlation between variables and investment demand of power grid, and then the variable system is established through the calculation of correlation degree between variables. And the classification of variables are finished.

2. Applying PCA in IBM SPSS Statistics 24, the score table of common factor is acquired and the contribution of each variable to the power grid investment is determined. Variables that don’t have correlation or have less correlation are selected through variable system and the score table of common factor.

This method can overcome the weakness that one variable can’t show comprehensive information of model. Introducing some variables in different aspect and dividing complex factors into a few important principles, significant variables which are scientific can be selected and the comprehensive information of model can be shown.

**The Building of Long-Term Forecast Model**

The difference of magnitudes are big between variables and investment demand. Doing logarithm for each numerical value don’t affect the stability while this method can eliminate dimensional differences.

Long-term forecast model for the investment demand of power grid based on linear regression can be build. The expression is same as Eq. 1.

**The Introduction of Parameter Correction**

The model whose variables are predictive value is built for long-term forecast for the investment demand of power grid. Because of economic fluctuation and policy fluctuation, there is an error between predictive value and actual value, which is the risky value expressed by $\Delta r_{kn}$. The rate of risky value and predictive value is defined as risky rate, expressed by $k$. The value of each $k$ is:
\[ \begin{bmatrix} k_{11} & k_{21} & \ldots & k_{k1} \\ k_{12} & k_{22} & \ldots & k_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ k_{1n} & k_{2n} & \ldots & k_{kn} \end{bmatrix} = \begin{bmatrix} \Delta x_{11} & \Delta x_{21} & \ldots & \Delta x_{k1} \\ x_{11} & x_{21} & \ldots & x_{k1} \\ \Delta x_{12} & \Delta x_{22} & \ldots & \Delta x_{k2} \\ x_{12} & x_{22} & \ldots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta x_{1n} & \Delta x_{2n} & \ldots & \Delta x_{kn} \\ x_{1n} & x_{2n} & \ldots & x_{kn} \end{bmatrix} \] (5)

where \( x_{kn} \) is the value of Kth predictive variable and Nth value, \( \Delta x_{kn} \) is the corresponding error between predictive value and actual value.

The correction parameter is represented by \( \lambda \). The value of \( k \) is small, the average of variable over the years is defined as the value of \( \lambda \). Each \( \lambda \) can be shown as follows.

\[
\lambda = \begin{bmatrix} \lambda_1 & \lambda_2 & \ldots & \lambda_k \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^{n} k_{i1} & \frac{1}{n} \sum_{i=1}^{n} k_{i2} & \ldots & \frac{1}{n} \sum_{i=1}^{n} k_{ik} \end{bmatrix}
\] (6)

Eq. 7 can be gotten from simultaneous Eq. 5 and Eq. 6.

\[
\lambda = \begin{bmatrix} \lambda_1 & \lambda_2 & \ldots & \lambda_k \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^{n} \Delta x_{i1} & \frac{1}{n} \sum_{i=1}^{n} \Delta x_{i2} & \ldots & \frac{1}{n} \sum_{i=1}^{n} \Delta x_{ik} \end{bmatrix}
\] (7)

Some conclusion can be gotten from Eq. 7. If \( \lambda < 0 \), the predictive value of the corresponding variable is higher than the actual value, we can use \((1+\lambda)\) to reduce the predictive value. If \( \lambda > 0 \), the predictive value of the corresponding variable is lower, we can use \((1+\lambda)\) to increase the predictive value.

**Case Study**

The data from 2003 to 2018 are selected to be analyzed, which include investment demand of power grid (ID), per capita electricity consumption (PEC), GDP, per capita conversion capacity (PCC), per capita disposable income (PCD), permanent resident population (PRP), Reliability of power supply (RPS), comprehensive voltage pass rate (CVPR), Urbanization rate (UR), proportion of new energy (PNE) and so on. The process is:

**Data Analyzing**

(1) The software IBM SPSS Statistics 24 is used to calculate the degree of correlation between various and investment demand of power grid through person correlation coefficients. The results is shown in table 1.

| Various   | Correlation coefficient | Various   | Correlation coefficient |
|-----------|-------------------------|-----------|-------------------------|
| GDP       | 0.882                   | PCCC      | 0.877                   |
| PEC       | 0.895                   | RPS       | 0.839                   |
| PRP       | 0.872                   | CVPR      | 0.867                   |
| UR        | 0.869                   | PNE       | 0.821                   |
| PCD       | 0.856                   |           |                         |

(2) The software IBM SPSS Statistics 24 is used to select various through person correlation coefficients and PCA. The PCA is used to get check score table of each various, and then, the contribution of each variables can be known from the table.
Table 2. Check score table of each various.

|    | Initial value | Exact value | Variable score |
|----|---------------|-------------|----------------|
| GDP | 1.000         | 0.984       | 0.122          |
| PCD | 1.000         | 0.987       | 0.122          |
| PEC | 1.000         | 0.993       | 0.123          |
| PRP | 1.000         | 0.989       | 0.123          |
| UR  | 1.000         | 0.992       | 0.123          |
| PCCC| 1.000         | 0.990       | 0.122          |
| RPS | 1.000         | 0.969       | 0.120          |
| CVPR| 1.000         | 0.345       | 0.072          |
| PNE | 1.000         | 0.870       | 0.115          |

Analyzing the degree of correlation between various and referring to table 2, the various which we want can be selected. Economic and social variable, power grid variable, which are the first order variable is select. Economic and social variable includes PEC and UR. Power grid variable includes RPS.

**Long-term Forecast Model for the Investment Demand of Power Grid Building**

After taking logarithm, the variables that we select and investment demand of power grid can be showed as follow.

$$\ln ID = 0.533366 \ln PEC + 480.7226 \ln RPS - 0.563488 \ln UR - 2209.876$$  \hspace{1cm} (8)

**The Correction Model of Long-Term Forecast Building**

The software IBM SPSS Statistics 24 is used to analyze data. According to Eq. 7, $\lambda$ of corresponding variables can be gotten. The results are showed as follow.

Table 3. The results of corresponding variables’ $\lambda$.

| variables | LnPEC       | LnRPS       | LnUR         |
|-----------|-------------|-------------|--------------|
| $\lambda$| 0.0015142   | 7.25311X10^{-3} | -5.30352X10^{-4} |

Introducing $\lambda$, the correction equation can be showed:

$$\ln ID = -2209.876 + 0.533366 \ln PEC \times (1 + 0.0015142) + 480.7226 \ln RPS \times (1 + 7.25311 \times 10^{-5}) - 0.563488 \ln UR \times (1 - 5.30352 \times 10^{-4})$$  \hspace{1cm} (9)

Simplifying Eq. 9, correction equation of long-term forecast can be shown.
\[ \ln ID = -2209.876 + 0.534173 \ln PEC + 480.757467 \ln RPS - 0.563189 \ln UR \] (10)

**Compared with the Forecast Results**

Put the relevant data into Eq. 8 and 10, different results can be gotten according to different models. Compared with the forecast value of models and actual value, it is showed in figure 3. The result which belongs to long-term forecast equation for the investment demand of power grid is represent by forecast \( \ln ID 1 \). The result which belongs to correction equation of long-term forecast is represent by forecast \( \ln ID 2 \).

![Figure 3. Comparing different results of \( \ln ID \).](image)

Compared with different results of \( \ln ID \), the following conclusions can be gotten. Compared with long-term forecast model, the correction model can get better forecast result which is closed to the actual result. In the long-term forecast for the investment demand of power grid, the correction model has higher accuracy.

**Conclusion**

Correlation analysis and PCA are used to get the key factors, and the key factors make the established model more persuasive. Compared with the long-term forecast results for investment demand of power grid, one is from long-term forecast model and the other is from correction model, we can find the two models can make long-term forecast for investment demand of power grid, while the correction model has higher accuracy on the long-term forecast for investment demand of power grid. And the correction model is more significant for power grid planning.

We put correction parameter of valuables into long-term forecast model for the investment demand of power grid in this paper. This method gets good result. While we only use a simple way to calculate correction parameter of each valuable and don’t consider the influence of policy and economics. In the future, we’ll take into account the influence of policy and economics on correction parameter of valuable. Try our best to make the correction parameter more accuracy and make the correction model more accuracy on the forecast value for the investment demand of power grid.

**Acknowledgement**

This research was financially supported by the state grid science and technology projects (Grant: SGSHJY00GPJS1800060)

**References**

[1] Athanasios I C, Petros A P. Sensitivity analysis of technological economic and sustainability evaluation of power plants using the analytic hierarchy process [J]. Energy Policy, 2009, 37(3):788-798.
[2] Ling Wu, Hui Cai, Yan Lu. Application of co-integration theory in electricity-demanding forecast modelling in Jiangsu Province[J]. International Conference on Renewable Power Generation (RPG 2015). Beijing, China. 17-18 Oct. 2015.

[3] He Xiaoying, Zhou Hui. Investment Analysis of Electricity Industry Based on Cointegration Model[J]. Water Resources and Power, 2011, 29(6):180-182.

[4] A.H. Gholami, S.S. Mohtavipour. Long term investment strategy of disco in a Smart Grid platform[J]. 2017 Iranian Conference on Electrical Engineering (ICEE). Tehran, Iran. 2-4 May 2017.

[5] Zeng Ming, Zhang Ping, Wang Liang, et al. Generation Investment Evaluation Model Under Uncertainty Based on Monte Carlo Simulation[J]. Automation of Electric Power System, 2015, 39(5):54-60.

[6] Hu Bai-Chu, Hu Gang, Hu Chao-Hua, Qing Song, Li Ming-Wei, Peng Chao. Grid Infrastructure Investment Calculation Model Based on Gray Forecast[J]. Journal of University of Electronic Science and Technology of China, 2015, 27(8):8-13.

[7] Zhou Ying, Zou Bo, Wen Fushuan, et al. Investigation on Power Network Investment Demands at Different Development Stages[J]. Proceedings of the Chinese Society of Universities for Electric Power System and Automation, 2015, 27(8):8-13.

[8] Lu Yu, Application of Co-integration Theory in Electricity-demanding Forecast Modeling in Jiangsu Province[J]. Jiangsu Electrical Engineering, 2015, 34(03):49-51.

[9] A. Divia Paul, S.M. Ashraf, J. Ezhilan, S. Vijayakumar, Anuj Kapadiya. A milestone in forecast of the coronary artery dimensions from the multiple linear regression equation[J]. Indian Heart Journal, 2019.

[10] Sandeep S. Sonawane, Santosh S. Chhajed, Sana S. Attar, Sanjay J. Kshirsagar. An approach to select linear regression model in bioanalytical method validation[J]. Journal of Analytical Science and Technology, 2019, 10(1).

[11] Fan Yue, Ma Wenli, Jiang Ping, Huang Jinlong, Chen Kewei, Pan Nian. Improving Angular Accuracy of a Scanning Mirror Based on Error Modeling and Correction.[J]. Sensors (Basel, Switzerland). 2019, 19(2).

[12] Lei Gao, Ling Guan. A Discriminant Two-Dimensional Canonical Correlation Analysis[J]. 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). Edmonton, AB, Canada, Canada. 5-8 May 2019.