Research on Key Index System and Investment Scale Calculation Model of Power Grid Investment

Ying Du1, Tiannan Ma1, Mingyong Wan1, Jinyong Chen1, Guangxiu Yu1, Yaling Sai1 and Guangchuan Wu2, *

1State Grid Sichuan Economic Research Institute, Chengdu Sichuan, China
2School of North China Electric Power University, Baoding Hebei, China

*Corresponding author: 2182218075@ncepu.edu.cn

Abstract. In order to avoid invalid investment and improve the investment efficiency of power grid companies, this paper constructed the power grid investment index system. Then extracted the key indicators by using the grey correlation analysis method. Finally, the regression analysis based on the multi scenario theory is used to predict the investment scale of power grid.

Keywords: Index system, Key indicators, Forecast, Multi scenario, Investment scale.

1. Introduction
In the face of a new round of power system reform, the provincial grid companies as the object of local government to strengthen supervision, its investment and planning management will face serious challenges. The power reform requires provincial companies to determine the investment scale reasonably, ensure that the deviation between the actual operating income and the approved permitted income is small, and strengthen the management and control ability of various inputs to achieve accurate investment.

This paper constructed an indicator system that affects power grid investment, and then extracts key indicators by using gray correlation analysis. Finally, the scale of power grid investment is calculated based on the key indicators. The rest of this article is structured as follows: Section2 constructed the index system affecting the power grid investment, Section3 extracted the key indicators, Section4 calculated of the power grid investment scale based on the key indicators, then the conclusions are drawn in Section 5.

2. Constructed the Index System.
The calculation of power grid investment involves many internal and external factors, and the relationship between the factors is complex, and many time series data are needed to support the calculation of the total investment scale of the company. Fang Jianliang et al. [1] in the research on the influencing factors of investment scale of power grid main grid, nine influencing factors were selected from three aspects of economic development, company operation and investment performance, such as regional GDP, total electricity consumption of the whole society, maximum load of power consumption of the whole society, electricity sales, operating income, total assets, total profit, unit grid investment and sales increment, unit grid investment load increment.
Luo Bin et al. [2] systematically analyzed the influencing factors and transmission path of power grid investment from the aspects of economy, society, politics, technology and power industry. In this study, the grey correlation method is used to calculate the correlation between each index and power grid investment, and 21 key indicators are obtained, such as GDP, total social population, fixed asset investment of the whole society, transmission line length, power generation, electricity sales, average sales price, unit line sales, line loss rate, etc.

In this paper, the above influencing factors are sorted into four categories: policy factors, economic and social factors, industry factors and enterprise operation. Considering the factors such as data sources and repeated indicators, the index system affecting the company's power grid investment is summarized as follows:

### Table 1. Index system of power grid investment

| Primary indicators          | Secondary indicators                                      |
|-----------------------------|-----------------------------------------------------------|
| Policy                      | Revenue                                                  |
|                             | Expenditure                                               |
| Economic and Social         | GDP                                                       |
|                             | CPI                                                       |
|                             | The total population                                      |
|                             | The total electricity consumption                         |
|                             | Average wage level                                        |
| Industry Factors            | Installed capacity of power generation                   |
|                             | Circuit length of 35kV and above transmission line        |
|                             | Capacity of 35kV and above substation equipment           |
|                             | New installed generating capacity                         |
|                             | Length of newly added AC 110kV and above transmission lines |
|                             | Newly increased capacity of AC 110kV and above substation equipment |
|                             | Electricity sales                                         |
|                             | Line loss rate                                            |
|                             | Urban power supply reliability                            |
|                             | Reliability of rural power supply                         |
| Enterprise Operation        | Return on net assets                                      |
|                             | Asset liability ratio                                     |
|                             | Operating income                                          |
|                             | Total profit                                              |

3. Extracted the Key Indicators

The grey correlation analysis method is used to calculate the correlation between each index and power grid investment to determine whether it is the key index affecting the power grid investment decision. The grey correlation analysis method is used because the method does not require high data sample size. The calculation steps are as follows:

3.1. Determine reference sequence and comparison sequence.

Set reference sequence as $x_0 = \{x_0(k)|k = 1, \ldots, n\} = \{x_0(1), \ldots, x_0(n)\}$, comparison sequence $x_i = \{x_i(k)|k = 1, \ldots, n\} = \{x_i(1), \ldots, x_i(n)\}, i = 1, \ldots, m$, $x_0(k)$ represents the observed value of the reference sequence at time $k$, $x_i(k)$ denotes the observed value of the comparison sequence $x_i$ at time $K$.

3.2. Transformed the reference sequence and comparison sequence to the dimensionless form.

Due to the different physical meanings of the factors in the system, the dimensions of the data are not necessarily the same, which is not easy to compare, or it is difficult to get a correct conclusion when
comparing. Therefore, it is necessary to carry out the grey relational analysis. At present, there are three common dimensionless processing methods: initial value processing, average processing and standardization processing. The dimensionless reference sequence and comparison sequence are obtained by initial value processing as (3-1):

\[ x_i(k) = \frac{x_i(k)}{x_i(1)}, \quad k = 1, ..., n; i = 0, 1, ..., m \]  

\[ (1) \]

3.3. Calculated the correlation coefficients of the corresponding elements of each comparison sequence and the reference sequence respectively.

\[ e_i(k) = \frac{\min_k \min_i |x_0(k) - x_i(k)| + \rho \max_k \max_i |x_0(k) - x_i(k)|}{\max_k |x_0(k) - x_i(k)| + \rho \max_k \max_i |x_0(k) - x_i(k)|} \]

\[ (2) \]

The resolution coefficient \( \rho \) ranges from 0 to 1.

3.4. Calculate the correlation degree

Because the correlation coefficient is the value of the correlation degree between the comparison sequence and the reference sequence at each time, it has more than one number, and the information is too scattered for the overall comparison. Therefore, it is necessary to concentrate the correlation coefficients of each time into one value, that is to calculate the average value, as the quantitative expression of the correlation degree between the comparison sequence and the reference sequence. The formula is as follows:

\[ y_i(k) = \frac{1}{n} \sum_{k=1}^{n} e_i(k) \]

\[ (3) \]

The larger the correlation degree, the closer the correlation between indicators. The criteria for defining the degree of association are shown in table 2.

| Relevance Degree | Weak  | Moderate | Strong | Very Strong |
|------------------|-------|----------|--------|-------------|
| Degree           | 0 ≤ r ≤ 0.35 | 0.35 ≤ r ≤ 0.65 | 0.65 ≤ r ≤ 0.85 | 0.85 ≤ r ≤ 1 |

3.5. Data Validation

In this paper, the data of 21 indicators listed in Table 1-1 that may have an impact on grid investment are extracted from China Statistical Yearbook, China Electric Power Yearbook and social responsibility report of State Grid Corporation. The time span is from 2010 to 2017. The grey correlation analysis method is used to calculate the correlation between each index series and the power grid investment series. The calculation results are shown in table 3, The resolution coefficient \( \rho \) in grey correlation analysis method is 0.5.

Figure 1 further compares the dimensionless power grid investment sequence and five index series with strong correlation (total social power consumption, 35kV and above transmission line circuit length, electricity sales, new installed generation capacity and operating revenue). The results show that all index series are linearly correlated with power grid investment series. Therefore, multiple linear regression model can be further constructed to predict the growth of power grid investment.
Table 3. Correlation degree results of related indicators

| Relevance Degree | Indicators                                                                                                                                 |
|------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Very Strong      | Circuit length of 35kV and above transmission line, The total electricity consumption, New installed generating capacity, Operating income, Electricity sales |
| Strong           | Line loss rate, Installed capacity of power generation, Capacity of 35kV and above substation equipment, The total population, Reliability of rural power supply, Urban power supply reliability, CPI, GDP, Asset liability ratio, Return on net assets, Newly increased capacity of AC 110kV and above substation equipment, Length of newly added AC 110kV and above transmission lines, Revenue, Expenditure |
| Moderate         | Total profit                                                                                                                                   |

Figure 1. Sequence diagram

Figure 1 further compares the dimensionless power grid investment sequence and five index series with strong correlation (total social power consumption, 35kV and above transmission line circuit length, electricity sales, new installed generation capacity and operating revenue). The results show that all index series are linearly correlated with power grid investment series. Therefore, multiple linear regression model can be further constructed to predict the growth of power grid investment.

4. Research on Calculation of Power Grid Investment Scale Based on Key Indicators.

4.1. Regression Analysis

Based on the analysis in the previous section, the key indicators affecting the power grid investment are determined as the index series with strong correlation (the correlation degree is greater than 0.85), including circuit length of 35kV and above transmission line, the total electricity consumption, new installed generating capacity, operating income and electricity sales.

By using regression analysis method, the calculation model of the above indicators to power grid investment is established, and the effectiveness and popularization of the method and model are guaranteed by cross test.

(1) The Calculation Model

Let $y$ be the dependent variable, $x_1, x_2, ..., x_m$ is an independent variable, When the relationship between independent variables and dependent variables is linear, the multiple linear regression model is as follows:
\[ y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_mx_m + e \]  

\( b_0 \) is a constant, \( b_1, b_2, \ldots, b_m \) is the regression coefficient, \( e \) is the error term. Parameter \( b_0, b_1, b_2, \ldots, b_m \) can be solved by the least square method on the premise that the sum of squares of errors \( (\sum e^2) \) is the minimum.

When establishing multivariate regression model, in order to ensure that the regression model has good explanatory power and prediction effect, we should first pay attention to the selection of independent variables. The criteria are: ① the independent variables must have a significant impact on the dependent variables and have a close linear correlation; ② the linear correlation between the independent variables and the dependent variables must be true, not formal; ③ the independent variables should have a significant impact on the dependent variables There is a certain degree of mutual exclusion, that is, the correlation between independent variables should not be higher than that between independent variables and the cause of dependent variables; ④ independent variables should have complete statistical data, and their predicted values can be easily determined

(2) Cross Check

In the given power grid investment data, most of the samples are used to construct the regression model, and a small part of the samples are used to test the established regression model.

Take \( n \) years of historical actual power grid investment \( y_j (j = 1, \ldots, n) \), The grid investment in the remaining years is predicted by using the grid investment in \( n-1 \) years and the corresponding index data \( y'_j (j = 1, \ldots, n) \) And calculate the deviation between the actual input and the actual input_ \( j \). Where \( \sigma_j = (y_j - y'_j)/y_j \). The details are shown in table 4.

| Year |
|------|
| 1    |
| 2    |
| \ldots |
| \( n \) |

**Table 4. Cross test results and deviations**

| Years | Predicted value(Billion) | Actual value(Billion) | Error(%) |
|-------|--------------------------|-----------------------|----------|
| 1     | \( y'_1 \)               | \( y_1 \)             | \( \sigma_1 \) |
| 2     | \( y'_2 \)               | \( y_2 \)             | \( \sigma_2 \) |
| \ldots | \ldots                   | \ldots               | \ldots    |
| \( n \) | \( y'_n \)               | \( y_n \)             | \( \sigma_n \) |

4.2. Analysis of power grid investment scale based on scenario theory

Considering the influence factors of uncertainty in the future, in order to better predict the investment scale of the company's power grid, this project plans to use the scenario analysis method to calculate the company's power grid investment scale under multiple scenarios.

Based on the company's historical power grid investment and relevant theoretical research, this project mainly considers the impact of GDP and operating income on the company's grid investment, and sets 9 scenarios with different growth rates of two indicators, as shown in table 5.

Taking 2016 as the base year, the GDP and operating income in 2017 are calculated by the growth rates set under different scenarios. Under the benchmark scenario (scenario A), the growth rates of GDP and operating income are 7% and 11% respectively.
### Table 5. Scenario Design

| Scene settings | Specific situation                                                                 | GDP growth rate | Growth rate of operating revenue |
|----------------|------------------------------------------------------------------------------------|-----------------|----------------------------------|
| scene A        | The economy maintained the current growth and the company's revenue grew steadily | 7%              | 11%                              |
| scene B        | The economy maintained the current growth, and the growth rate of the company's revenue decreased | 7%              | 9%                               |
| scene C        | The economy maintained the current growth, and the company's revenue increased rapidly | 7%              | 13%                              |
| scene D        | Economic growth slowed down, the company's revenue grew steadily                     | 6%              | 11%                              |
| scene E        | The company's growth rate slowed down, and the company's revenue declined           | 6%              | 9%                               |
| scene F        | Economic growth slowed down, the company's revenue increased rapidly                | 6%              | 13%                              |
| scene G        | Economic growth accelerated, the company's revenue grew steadily                     | 8%              | 11%                              |
| scene H        | Economic growth accelerated, the company's revenue growth decreased                 | 8%              | 9%                               |
| scene I        | Economic growth accelerated, the company's revenue increased rapidly                | 8%              | 13%                              |

### 4.3. Extracted Model

According to the historical data of GDP, Operating income and Power grid investment, the regression analysis shows that the relationship is as follows:

\[
Y = d_1E + d_2M + e
\]  

(5)

\(d_1, d_2\) is a coefficient and \(e\) constant term.

According to the nine scenarios, the GDP value of the base year (2016) is set as \(E_{\text{base}}\), the value of operating revenue in the base year is \(M_{\text{base}}\). According to the relationship between power grid investment and GDP and the company's operating income, the total investment \(y\) under nine scenarios is predicted. Total investment growth \(Y'_y\), The growth rate of total investment is shown in table 6.
Table 6. Power grid investment forecast under multi scenario growth

| Scene | Forecast GDP | Forecast operating revenue | Forecast power grid investment | Forecast the growth of power grid investment | Forecast the growth rate of power grid investment | Actual grid investment value | Error |
|-------|--------------|---------------------------|--------------------------------|---------------------------------------------|--------------------------------------------------|-----------------------------|-------|
| Scene A | $1.07E_{\text{base}}$ | $1.11M_{\text{base}}$ | $Y_1$ | $Y'_1$ | $\alpha_1$ | $Z_1$ | $\sigma_1$ |
| Scene B | $1.07E_{\text{base}}$ | $1.09M_{\text{base}}$ | $Y_2$ | $Y'_2$ | $\alpha_2$ | $Z_2$ | $\sigma_2$ |
| Scene C | $1.07E_{\text{base}}$ | $1.13M_{\text{base}}$ | $Y_3$ | $Y'_3$ | $\alpha_3$ | $Z_3$ | $\sigma_3$ |
| Scene D | $1.06E_{\text{base}}$ | $1.11M_{\text{base}}$ | $Y_4$ | $Y'_4$ | $\alpha_4$ | $Z_4$ | $\sigma_4$ |
| Scene E | $1.06E_{\text{base}}$ | $1.09M_{\text{base}}$ | $Y_5$ | $Y'_5$ | $\alpha_5$ | $Z_5$ | $\sigma_5$ |
| Scene F | $1.06E_{\text{base}}$ | $1.13M_{\text{base}}$ | $Y_6$ | $Y'_6$ | $\alpha_6$ | $Z_6$ | $\sigma_6$ |
| Scene G | $1.08E_{\text{base}}$ | $1.11M_{\text{base}}$ | $Y_7$ | $Y'_7$ | $\alpha_7$ | $Z_7$ | $\sigma_7$ |
| Scene H | $1.08E_{\text{base}}$ | $1.09M_{\text{base}}$ | $Y_8$ | $Y'_8$ | $\alpha_8$ | $Z_8$ | $\sigma_8$ |
| Scene I | $1.08E_{\text{base}}$ | $1.13M_{\text{base}}$ | $Y_9$ | $Y'_9$ | $\alpha_9$ | $Z_9$ | $\sigma_9$ |

5. Case Analysis.

5.1. The regression analysis

According to the index data from 2008 to 2017, the quantitative relationship between the power consumption of the whole society, the circuit length of 35kV and above transmission lines, the electricity sales, the operating income, the newly added installed generating capacity and the grid investment are established respectively. The calculation results are shown in table 7.

Table 7. Linear regression analysis between the key indicators and the grid investment

| Key indicators | Univariate linear regression equation | Maximum deviation (%) | Average deviation (%) |
|----------------|--------------------------------------|-----------------------|-----------------------|
| The power consumption of the whole society | $\text{The grid investment} = 0.09839 \times \text{The power consumption of the whole society} - 1000.5434$ | 10.4069 | 6.2887 |
| The circuit length of 35kV and above transmission lines | $\text{The grid investment} = 0.004248 \times \text{The circuit length of 35kV and above transmission lines} - 2459.4295$ | 8.1081 | 5.2658 |
| The electricity sales | $\text{The grid investment} = 0.1445 \times \text{The electricity sales} - 2052.3496$ | 11.5349 | 7.0221 |
| The operating income | $\text{The grid investment} = 0.07634 \times \text{The operating income} + 1514.776$ | 16.0847 | 8.7011 |
| The newly added installed generating capacity | $\text{The grid investment} = 0.3713 \times \text{The newly added installed generating capacity} + 885.6117$ | 11.7078 | 6.4529 |
The regression equation based on a single key index has large average error and maximum error, so it is difficult to accurately predict the future power grid investment scale. Therefore, this project further considers the establishment of a multiple linear regression equation with the whole society's electricity consumption, the circuit length of 35kV and above transmission lines, electricity sales, business income and newly added installed generating capacity as independent variables, and the grid investment as the dependent variable. The calculation results are as follows:

The grid investment = -0.2454* The power consumption of the whole society +0.008006* The circuit length of 35kV and above transmission lines +0.272613* The electricity sales -0.1499* The operating income -0.07683* The newly added installed generating capacity -5140.77

For the data from 2010 to 2017, the results are shown in table 8, and the deviation is within 4%, which indicates that the above multiple linear regression model can be used to predict the power grid investment to a certain extent.

| Years | predicted value | Actual value | Error(%) |
|-------|----------------|--------------|----------|
| 2010  | 335.5304       | 344.8        | 2.69     |
| 2011  | 372.2793       | 368.7        | 0.97     |
| 2012  | 372.674        | 366.1        | 1.80     |
| 2013  | 383.7005       | 385.6        | 0.49     |
| 2014  | 411.3488       | 411.9        | 0.13     |
| 2015  | 474.3125       | 464.0        | 2.22     |
| 2016  | 498.6025       | 497.7        | 0.18     |
| 2017  | 524.252        | 533.9        | 1.81     |

5.2. Analysis on investment scale of multi scenario power grid

Using the data from 2010 to 2017, this paper makes regression analysis on GDP, company's operating income and power grid investment. The $R^2$ is 0.9832, and the goodness of fit is good. The quantitative formula is obtained according to the coefficient table:

The grid investment = 0.011286*GDP -0.32873* The operating income +3790.649

According to the nine scenarios set up and taking 2016 data as the base year, the grid investment in 2017 is predicted as shown in table 9. Considering only the impact of GDP and operating income on grid investment, according to the 2016 GDP and the company's operating income, we can find that under scenario a to scenario F, when the domestic economy maintains the current growth rate or the growth rate slows down, the predicted value of grid investment is smaller than the actual value, while in scenario g to scenario I, that is, under the situation of rapid domestic economic growth, the predicted value of grid investment is smaller than the actual value, The predicted value of power grid investment is larger than the actual value. In 2017, the actual power grid investment was 533.9 billion yuan, and the deviation between the predicted value and the actual value was small under each scenario, which showed that the model had good robustness and certain application value.
Table 9. Forecast of total investment under multi scenario growth

| Scene | Forecast power grid investment (Billion) | Forecast the growth of power grid investment (Billion) | Forecast the growth rate of power grid investment | Actual total investment (Billion) | Error |
|-------|-----------------------------------------|-----------------------------------------------|-------------------------------------------------|---------------------------------|-------|
| Scene A | 5410 | -20.00 | -0.37% | 5339 | 1.35% |
| Scene B | 5484 | 52.89 | 0.97% | 5339 | 2.71% |
| Scene C | 5338 | -92.90 | -1.71% | 5339 | -0.02% |
| Scene D | 5361 | -70.30 | -1.29% | 5339 | 0.41% |
| Scene E | 5434 | 2.59 | 0.05% | 5339 | 1.77% |
| Scene F | 5288 | -143.20 | -2.64% | 5339 | -0.96% |
| Scene G | 5461 | 30.29 | 0.56% | 5339 | 2.29% |
| Scene H | 5534 | 103.19 | 1.90% | 5339 | 3.66% |
| Scene I | 5388 | -42.60 | -0.78% | 5339 | 0.93% |

Therefore, Therefore, this method can be used to calculate the investment scale of power grid.

6. Conclusion
In this paper, the gray correlation analysis method is used to select the most relevant indicators of the power grid investment as the key indicator that affected the power grid investment. And then, based on the key indicators, regression analysis combined with multi scenario theory is used to calculate the investment, which greatly reduces the redundancy of data and the difficulty of analysis and calculation. At the same time, it can help power grid enterprises avoid invalid investment and improve investment efficiency.

Acknowledgments
This paper is supported by the science and technology project of State Grid Corporation of China (SGSCJY00JJJS2000019, 5102-2019563111A-0-0-00)

References
[1] J.L Fang, Y.J Xie, Y Jin, et al, Research on the prediction model of investment scale of power company's main network, J. China high tech enterprises, 2016, (8) 20-21.
[2] B Luo, W.J Su, X.X Zhang, et al. Research on key influencing factors and transmission path of power grid investment, J. Price theory and practice, 2019 (5)