Estimation of the Price Range of Power Grid Technological Transformation Based on Walsh Average

Fan Yang*, Xiaohu Zhu and Jianqing Li
State Grid Anhui Electric Power Co. LTD Economic Research Institute, Hefei, China

*Corresponding author e-mail: yangf1828@ah.sgcc.com.cn

Abstract. With the continuous improvement of the construction of power grid projects, the pressure to transform power grid equipment is increasing. It is particularly important to perform reasonable data processing for the transformation costs of various types and voltage levels of power grid equipment. The data processing method based on Walsh average has an efficiency advantage in dealing with the non-normal parameter distribution. Based on the traditional statistical processing of power grid technical renovation price data, this paper combines the Walsh average comprehensive data processing method to obtain the final technical renovation price data for the overall range of various equipment. First, conduct a preliminary analysis of the sample, use the box plot to eliminate outliers, and then use the Walsh average model to obtain an estimate of the power grid technological transformation price range. In this paper, the real data is processed to verify the feasibility of the comprehensive data processing method and use this method to obtain a more appropriate confidence interval for the unit cost of each type of power grid equipment at each voltage level. The condition of the equipment is of great significance to guide the subsequent technological transformation projects and provide references for them.

Keywords: Walsh Average, Unit Cost, Confidence Interval, Power Grid Technological Transformation, Data Processing

1. Introduction
As a major power country, China has a huge demand for electricity at all times. The geographical environment factors caused the uneven distribution of electricity geographically. The power grid has become a guarantee for the safe transmission of electricity. The management of fixed assets by grid companies is also an important part of corporate management. At the same time, in the asset management of enterprises, the two most important expenditure items are expenditures for technological transformation and expenditures for equipment maintenance [1]. Many power supply companies in my country have not integrated and optimized the management of technological transformation projects, so they have insufficient awareness of cost budgeting and control in the management of technological transformation projects [2]. If the project cost is not managed properly, the assets of the power supply company will be more or less affected and the grid technical renovation project will be difficult to carry out normally, which will ultimately affect the grid efficiency and
operational safety and risks. Therefore, reasonable data processing changes will be made to the technical renovation project cost. It is particularly important to guide the subsequent technological transformation projects and provide evaluation references for them.

Among various methods for estimating the unit cost of power grid equipment, many researchers have proposed different methods for similar problems. Li Yuanyuan and others used basic mathematical statistics to classify power grid equipment and obtained the unit cost of various types of equipment [3]. Although the analysis of the project cost according to the construction parameters of different projects is more accurate and detailed, the classification conditions are too harsh to provide insufficient help for future projects, and the credibility of the processing results is reduced due to insufficient data processing. For further analysis of project cost data and interval estimation, Yao Gang et al. regarded the sample data of power grid project cost as approximately normal distribution and obtained the estimation interval according to the probability density function [4], but it was obviously in a non-normal distribution or sample size. There are errors in the results obtained by this method in the lower samples, which leads to the unsatisfactory effect of this method and the estimation result is not robust enough. In the coefficient of the cost of power grid technical renovation equipment, most of the data distribution is unknown. It is inappropriate to use traditional statistical processing methods, and the Walsh function and its extension can better deal with unknown distribution data [5]. Tong Rujun and others introduced the Monte Carlo simulation analysis theory and constructed a reasonable interval analysis model for overhead line engineering [6]. However, the defect of this model is that it is difficult to operate. For statisticians who may not understand the principle, there may be some difficulty in actual operation. Therefore, this paper selects the combination of Walsh average and traditional statistical methods to calculate the unit cost of technical renovation equipment, which can not only solve the problem of the non-normal distribution of samples but also facilitate operation and can obtain more accurate results.

Walsh function is a function used in digital signal processing introduced by American mathematician Walsh. It can provide a representation method similar to Fourier transform. Many results are easy to implement on a computer. In some respects, it is superior to Fourier analysis. Walsh functions are widely used in signal processing, pattern recognition, and genetic algorithms [7]. Inspired by these applications, this article considers applying the Walsh function and its extension theorem to the field of data processing. The Walsh function has unique advantages in processing mode characteristics. For example, some characteristic values of the model can be obtained through Walsh transformation, which is expressed as the average fitness value of the model in the genetic algorithm. Similarly, it is envisaged that in the data processing of the unit cost of the technical transformation equipment of the power grid, it should be possible to find a reasonable unit cost interval through Walsh function transformation.

Therefore, this paper establishes a comprehensive data processing method that combines basic statistical processing (3δ principle, box plot method) and Walsh average, using computer data processing software and program editing to obtain a reasonable interval between the screening results of the data and the final unit cost. This model can provide a very convenient and effective reference for asset management of future power grid technological transformation projects.

2. Steps of Data Processing Method Based on Walsh Average
Collecting statistics on the unit cost of the power grid technical renovation project in the past, and obtaining the interval estimate of the unit cost of the power grid technical renovation project through abnormal value elimination and Walsh average, provides a very convenient and effective reference for the asset management of future power grid technical renovation projects [8].

Based on the comprehensive data processing method of Walsh average [9], the data analysis of all technical transformation price data is mainly divided into three parts: descriptive statistics of the sample, outlier removal, abnormal data analysis and overall interval estimation [10].

2.1. Descriptive Statistics of the Sample
Before processing and analyzing the data of the unit cost of the power grid technical renovation project, it is necessary to understand the data sample to get the general situation of the sample, mainly covering
the basic statistical information of the sample. It should be understood when the data of the unit cost of the power grid technical renovation project is compiled the overall distribution of the data in the sample provides theoretical and logical support for the elimination of outliers and the overall interval estimation below.

In the descriptive statistics of the sample, this study mainly counts the mean, median, variance, standard deviation, minimum, maximum, range, interquartile range, skewness, and kurtosis of the sample, and plots scatter points the figure combines the above indicators to perform basic descriptive statistics on the sample.

2.2. Outlier Elimination

In general, there are a certain number of abnormal values in the initial data obtained. These samples are different from other samples due to some external conditions, resulting in deviations between the data of this sample and the data of other normal samples. In the process of interval estimation of engineering unit cost, this deviation may lead to unreasonable interval estimation of the unit cost of power grid technological transformation project. Therefore, the accuracy of sample interval estimation is improved by eliminating outliers.

In this study, the box plot method was used to filter out abnormal values. The box plot consists of five parts, namely the minimum, median, maximum, and two quartiles $Q_1$ and $Q_2$. And to filter out abnormal values is to remove the value of the sample that is less than $Q_1 - 1.5IQR$ or greater than $Q_2 + 1.5IQR$ ($IQR$ is the interquartile range of the sample).

The first step is to calculate the first quartile $Q_1$ of the sample, which is equal to the 25th percentile of all the values in the sample in descending order.

The second step is to calculate the median $F$ of the sample, which is equal to the 50th percentile of all the values in the sample in descending order.

The third step is to calculate the quartile $Q_2$ of the sample, which is equal to the 75% number after all the values in the sample are arranged from small to large.

The fourth step is to calculate the interquartile range $IQR$ of the sample:

$$IQR = Q_2 - Q_1$$

Finally, the values of $Q_1 - 1.5IQR$ and $Q_2 + 1.5IQR$ are calculated, and outliers outside the two intervals are eliminated.

When drawing and calculating, sometimes the ends of the two whiskers in the box plot are not exactly 1.5 times the box length, but the farthest value that does not exceed the length. At the same time, the cost data is positive, so the outlier is calculated using programming For the range, if the lower limit is a negative value, change to the minimum value of the interval, and the final determination of the outlier is based on the output box plot result.

After the abnormal value is selected, it is necessary to analyze the abnormal value according to the specific project and analyze the cause of the abnormality, to reduce the abnormal rate of the subsequent technical renovation price data.

2.3. Population Interval Estimate

To obtain a reasonable interval for the unit cost of the power grid technical renovation project, it is necessary to estimate the overall interval of the results after data processing and obtain the confidence interval of the unit cost to guide the future cost control management of the power grid technical renovation project.

Interval estimation starts from the point estimate and sampling standard error and establishes an interval containing the parameter to be estimated according to a given probability value. The given probability value is called the confidence level, and this established interval containing the function to be estimated is called the confidence interval, which refers to the probability that the overall parameter value falls within a certain area of the sample statistical value.

Because the sample size of many data is not large and the overall distribution is unknown, the
confidence interval of the center of symmetry is constructed using Walsh average order statistics. This method does not depend on the overall distribution and can estimate the confidence interval more accurately.

Let \( x_1, x_2, \ldots, x_n \) be the original data, which is a continuous symmetric independent and identically distributed random sample, and \( \theta \) is the center of symmetry of the distribution. First, calculate the Walsh average:

\[
W(\cdot) = \frac{x_i + x_j}{2}, \quad 1 \leq i \leq j \leq n
\]  

Arrange the ascending powers of the Walsh average, denoted as \( W(1), W(2), \ldots, W(N) \)

\[
N = \frac{1}{2} n(n + 1)
\]  

Given a confidence level of \( 1 - \alpha \), estimate the interval of \( \theta \) \( (\widehat{\theta}_l, \widehat{\theta}_u) \), which is determined by the following formula.

\[
\widehat{\theta}_l = W(K), \quad \widehat{\theta}_u = W(N - K + 1)
\]

\[
K \approx \frac{N}{2} - \frac{Z_{\alpha/2}(n + 1)\sqrt{n}}{2\sqrt{3}}
\]

\( Z_{\alpha/2} \) is the \( \alpha/2 \) critical value of a normal random variable. If \( \Phi(\cdot) \) is the standard normal distribution function, then

\[
1 - \Phi\left(\frac{Z_{\alpha/2}}{2}\right) = \frac{\alpha}{2}
\]

When performing calculations, proceed as follows:

1. Check the normal distribution table by \( \alpha \) to get \( Z_{\alpha/2} \);
2. Calculate \( K \) and \( N-K+1 \) by formula (4);
3. Arrange (1) from small to large, and find the K-th value \( W(K) \) of (2) from small to large, and then find the K-th value of (2) from back to forward from large to small, That is, the N-K+1th value \( W(N-K+1) \);
4. From (3), we get \( \widehat{\theta}_l, \widehat{\theta}_u \).
5. Software implementation: Use R software to calculate interval estimation results. For some small voltage samples, R cannot calculate the 95% or 99% confidence interval level and the program does not give the result of interval estimation.

3. Case Analysis

This section takes the actual price data of a transformer replacement project in a province as an example for analysis.

3.1. Replacement Transformer Engineering Sample Situation

First, the basic situation of the samples of various transformer replacement projects in a certain province is calculated, as shown in Table 1.

| Voltage level | Data year interval | Number of samples | Average value | Sample data interval |
|--------------|-------------------|-------------------|--------------|---------------------|
| 220kV        | 2015-2019         | 1                 | 660.67       | [660.67]            |
| 110kV        | 2015-2019         | 3                 | 297.99       | [260.49,323.58]     |
| 35kV         | 2015-2019         | 19                | 70.41        | [43.1,121.95]       |
Due to the lack of data for 220kV and 110kV projects, this time only the 35kV transformer replacement project data for 2015-2019 are compiled, totaling 19 items. Before sorting out, the sample interval of the unit cost of each voltage level is between 431,000 yuan/unit/group and 1219.5 thousand yuan/unit/group. The sample description statistics for the voltage level of 35kV are shown in the following table.

**Table 2.** Descriptive statistics of 35kV samples (before collation) (ten thousand yuan/unit)

| Replace the transformer 35kV | Statistics | Standard error |
|-----------------------------|------------|----------------|
| Median                      | 64.98      |                |
| Variance                    | 544.75     |                |
| Standard deviation          | 23.34      |                |
| Minimum                     | 43.10      |                |
| Maximum                     | 121.95     |                |
| Range                       | 78.85      |                |
| Interquartile range         | 22.93      |                |
| Skewness                    | 1.17       | 0.52           |
| Kurtosis                    | 0.59       | 1.01           |

The data scatter plot is shown in Figure 1 below. At the same time, the sample histogram obtained according to the sample data is shown below. It can be seen that the distribution of the 35kV sample data is not an accurate normal distribution.

![Figure 1. 35kV transformer replacement project unit cost scattered point box type histogram](image)

3.2. 35kV Transformer Engineering Data Collation Process

Outlier elimination: Draw a 35kV transformer engineering unit cost box line diagram, as shown in Figure 1, and calculate the normal value range of 164,200 yuan/unit/group to 1.0812 million yuan/unit/group, and the figure shows that there are 2 abnormal values.

The interval estimation of the population: The software calculation result range of 508,100 yuan/unit to 737,300 yuan/unit (99%).

Abnormal data analysis: The abnormal data is 2 sets of 35kV transformers, the equipment costs are both 760,500 yuan, and the installation costs are both close to 600,000 yuan, resulting in unit cost reaching 1.2195 and 1.2017 million yuan/unit/group respectively, and the average value of 704,100 yuan/unit/group Many deviations should be caused by the improper splitting of project costs. Removed in this data collation.

3.3. Results of 35kV Transformer Engineering Samples

After removing the outliers, the remaining samples are sorted and counted again, and the statistical
results are shown in the following table.

**Table 3. Sample collation results (ten thousand yuan/unit)**

| Data year interval | Number of samples | Average value | Sample data interval |
|--------------------|-------------------|---------------|----------------------|
| 2015-2019          | 17                | 64.45         | [43.1-96.68]         |

A total of 19 items of 35kV replacement transformer unit cost data from 2015 to 2019 were compiled this time. After the collation, 2 items of abnormal data were deleted, and 17 sample data were retained. After collation, the 35kV replacement transformer engineering unit cost sample interval is range of 431,000 yuan/unit/group to 966,800 yuan/unit/group. The descriptive statistical results of the 35kV sample after sorting are shown in the following table.

**Table 4. Descriptive statistics of 35kV samples (after collation) (ten thousand yuan/unit)**

| Replace the transformer 35kV | Statistics       | Standard error |
|-----------------------------|------------------|----------------|
| Median                      | 64.97            |                |
| Variance                    | 254.309          |                |
| Standard deviation          | 15.94707         |                |
| Minimum                     | 43.1             |                |
| Maximum                     | 98.68            |                |
| Range                       | 55.58            |                |
| Interquartile range         | 22.1075          |                |
| Skewness                    | 0.982            | 0.55           |
| Kurtosis                    | 0.838            | 1.063          |

The 35kV overall interval estimation result is shown in the figure below.

![Figure 2. 35kV transformer replacement engineering unit cost interval estimation](image)

After excluding the outliers, the probability that the true value of the sample population falls between 491,500 yuan/unit to 723,700 yuan/unit is 99%. That is, the unit cost of a 35kV transformer is reasonable in the range of 491,500 yuan/unit to 723,700 yuan/unit.

4. Conclusion
Calculating a reasonable project cost and controlling the project cost within a reasonable range is crucial for power companies. Based on the traditional statistical processing of power grid technical renovation price data, this paper combines the box-and-plot method and Walsh average comprehensive data processing method to obtain the final technical renovation price data of various equipment overall interval estimates.

Based on the method described, this paper uses the actual data of a 35kV power grid replacement transformer equipment project in a certain province, the basic information of statistical data, and the reasonable range of unit cost is calculated. Analyzing the cost of power grid equipment in the current
sample is of great significance for establishing a reasonable cost range for engineering projects with power grid companies and guiding subsequent technological transformation projects.

However, the methods involved in this study need the support of a large number of samples. To obtain a more accurate estimation interval, a large amount of data is necessary.

References

[1] Post C, Wentingmann N, Bramsiepe C, et al. Using design spaces for more accurate cost estimation during early engineering phases [J]. Chemical Engineering Research and Design. 2020, 153: 592-602.
[2] Odeck J, Kjerke A. The accuracy of benefit-cost analyses (BCAs) in transportation: An ex-post evaluation of road projects [J]. Transportation Research Part A: Policy and Practice. 2019, 120: 277-294.
[3] Vig R, Chauhan S S. Speech Compression using Multi-Resolution Hybrid Wavelet using DCT and Walsh Transforms [J]. Procedia Computer Science. 2018, 132: 1404-1411.
[4] Ahn J, Ji S, Ahn S J, et al. Performance evaluation of normalization-based CBR models for improving construction cost estimation [J]. Automation in Construction. 2020, 119: 103329.
[5] Ning F, Shi Y, Cai M, et al. Manufacturing cost estimation based on the machining process and deep-learning method [J]. Journal of Manufacturing Systems. 2020, 56: 11-22.
[6] Moins B, France C, Van den Bergh W, et al. Implementing life cycle cost analysis in road engineering: A critical review on methodological framework choices [J]. Renewable and Sustainable Energy Reviews. 2020, 133: 110284.
[7] Jong G D, Vignetti S, Pancotti C. Ex-post evaluation of major infrastructure projects [J]. Transportation Research Procedia. 2019, 42: 75-84.
[8] Mousazadeh A, Shekofteh Y. Cost function based on the self-organizing map for parameter estimation of chaotic discrete-time systems [J]. Engineering Applications of Artificial Intelligence. 2020, 94: 103817.
[9] Scherer J N, Schuch J B, Rabelo-Da-Ponte F D, et al. Analytical reliability of four oral fluid point-of-collection testing devices for drug detection in drivers [J]. Forensic Science International. 2020, 315: 110434.
[10] Gao Q, Tang H, Xiang J, et al. A Walsh transform-based Teager energy operator demodulation method to detect faults in axial piston pumps [J]. Measurement. 2019, 134: 293-306.