Research on Reasonable Interval Analysis Method of Power Grid Technical Transformation Project Based on Kernel Density Estimation Method

Jianguo Tang [a], Huixing Liu [b], Yizhen Zhou [c], Wenjing Xiao [d*]

State Grid Hubei Extra High Voltage Company, Hubei, Wuhan 430050, China

[a]e-mail: 13808635306@139.com, [b]e-mail: 32138352@qq.com, [c]e-mail: 232635831@qq.com, * Corresponding author's [d]e-mail: juzen123@163.com

Abstract. With the rapid development of the electric power industry, grid technology transformation projects have become an important way to enhance grid operational efficiency and improve the reliability of power supply. However, due to the wide range of equipment involved and the technical complexity of these projects, how to reasonably assess and control project investment costs has become an urgent issue. As the cost components of technological transformation projects are numerous and the influencing factors are complex, precise determination becomes challenging. Therefore, this paper proposes a novel approach based on Kernel Density Estimation. By utilizing KDE, it is possible to estimate the data distribution more accurately without relying on assumed models, thus providing a more scientific basis for determining the cost range of power grid technology transformation projects.

Keywords: Power Engineering, Technological Transformation Technology, Nuclear Density Estimation Method, Cost Interval, Analysis Technology

1 Introduction

With the new stage of China's economic development, the rapid development of the power industry, the speed of power grid construction shows a rapid expansion trend. In particular, in the field of grid technology renovation, the emergence of new technologies, advanced equipment, and innovative materials has led to certain deviations between the actual costs of renovation projects and the related standard benchmarks, which hinders the scientific management and control of project budgets. Furthermore, the growing demand for technological upgrades in the grid has made the need for precise cost control more urgent. Effectively managing costs and scientifically defining the reasonable boundaries of project expenses not only affects the input-output efficiency of individual projects but also has a significant impact on the operational development of power grid companies [1].

Combined with the current development situation, relying on advanced big data sta-

S. M. Ghadiri et al. (eds.), *Proceedings of the 2025 4th International Conference on Engineering Management and Information Science (EMIS 2025)*, Advances in Computer Science Research 121,
https://doi.org/10.2991/978-94-6463-736-6_3

tistical analysis technology, the precise determination of the cost input range and scope of power grid technical transformation projects is the basis for ensuring the orderly development of project construction and the precise control of cost input. Literature [2] proposes a method utilizing neural networks for optimizing the rapid estimation of power grid engineering costs and quantities is introduced. This approach enables accurate estimation of both costs and quantities within a short timeframe, offering valuable insights for the relevant stakeholders. Literature [3] presents a stacking-based ensemble model prediction technique. This approach combines the strengths of individual classifiers, significantly enhancing the model's generalization ability and prediction accuracy. In Literature [4], the entire process of power grid engineering cost management is examined, with a focus on selecting an evaluation index system and constructing a comprehensive cost management model, which serves as a reference for improving precise investment control. Literature [5] investigates the factors contributing to cost deviations, explores strategies and methods for optimizing power grid engineering cost management, and enhances the overall effectiveness of cost management. Literature [6] presents a method for selecting technology upgrades in energy systems, focusing on cost estimation and ranking of alternatives. Literature [7] discusses an integrated approach for evaluating design upgrades, focusing on lifecycle impact assessment and optimization. Literature [8] proposes a systematic economic evaluation model for power grid technological upgrades, enhancing investment accuracy and sustainability.

These papers emphasize methods for optimizing cost estimation and enhancing cost control, while this paper stands out by proposing a scientific statistical analysis-based economic evaluation model for power grid technological transformation projects, focusing on precise cost control and investment return assessment.

2 Data-driven approach to probability distribution modeling

Parametric distribution modeling makes statistical inference through the structural properties of the preset probability model family, while the free distribution inference framework builds the probability structure directly based on the features of the observed data, typically represented as $f(x|\theta)$, where θ, are the parameters defining the distribution. This approach uses a small number of parameters to describe the distribution of the sample. However, the main limitation of this method is that it requires prior knowledge of the data distribution. If the assumed distribution does not match the actual data, significant estimation errors may occur. In contrast, the free distribution inference framework implements probabilistic modeling through the analysis of data intrinsic features, completely escaping the structural constraints of the preset probability model. Instead, it directly estimates the probability density function based on the sample data itself. By analyzing the actual distribution of the data, this method can more accurately approximate the true probability density function, avoiding the biases introduced by assumed distributions. Kernel density estimation is one of the commonly used non-parametric methods. It smooths the data to estimate the density curve, providing a more accurate reflection of the data's distribution. The specific formula is shown as follows:

$$f(x)=\frac{1}{nl}\sum_{i=1}^{n}Q\left(\frac{x-X_i}{l}\right) \tag{1}$$

In the above formula, n represents the number of calculated samples, h represents the smoothing parameter, X_i represents a certain sample determined, $Q(u)$ representing the numerical value of the function. For special functions, the following principles must be followed:

$$\begin{cases} Q(u)\geq 0 \\ \int_{-\infty}^{+\infty}Q(u)du=1 \\ \sup Q(u)<+\infty \\ \int_{-\infty}^{+\infty}Q2(u)du<-\infty \\ \lim\limits_{x\text{-}\infty} Q(u)u=0 \end{cases} \tag{2}$$

l represents the bandwidth, when the special function is determined, the variance of the calculation results is mainly affected by the bandwidth Once the kernel function is determined, its variance is controlled by the bandwidth l. The smooth characteristics of the estimated curve of kernel density is mainly affected by the parameter h. The parameter hhh governs the smoothness of the kernel density estimation curve. A smaller value of l results in a steeper curve, as the influence of each sample point decreases. On the other hand, a larger l leads to a smoother curve, where the contribution of each sample point is more evenly spread. The choice of l is crucial as it directly affects the trade-off between overfitting and underfitting in the density estimation.

3 Construction of a Model for Reasonable Interval Determination in Technology Transformation Projects

This study using the spatial interpolation principle of bell-shaped feature functions, based on the unconstrained probability field modeling, the statistical density distribution is determined by an adaptive smoothing process of the observed data.

$$f(x)=\frac{1}{\sqrt{2\pi}}\frac{1}{nl}\sum_{i=1}^{n}\exp\left[-\frac{1}{2}\left(\frac{x-X_i}{l}\right)2\right] \tag{3}$$

If the bandwidth l is set too small, the density estimate becomes overly localized around the observed data points, causes a lot of unpredictable mistakes. Conversely, if l is set too large, important features are ignored, resulting in biased estimates.

In this case, the value of the base function is strongly correlated with the value of the parameter l. In this paper, the bandwidth is chosen using the optimal bandwidth determination technique offered by MATLAB.

The project cost control domain should match the preset reliability threshold. After establishing the engineering cost statistical distribution surface based on the freely distributed probability field reconstruction technology, the economic evaluation interval corresponding to the confidence parameter can be derived through the cumulative field construction algorithm, which can be implemented specifically:

$$F(R) = \int_{-\infty}^{R} f(t)dt \tag{4}$$

Finally, the confidence interval is defined by $1-\alpha$, depending on the chosen confidence level.

4 Example analysis

The equipment upgrade project data of a provincial power grid in the past three years (2020-2022) was selected as the research object, and its reliability in the engineering economic evaluation was verified through the non-preset distribution modeling technology.

The study sample covers 180 cases of transformer iteration engineering, and the dynamic data shows that: In 2020,50 cases will be implemented, involving 450 groups of equipment updates, with a total investment of 102 million yuan; In 2021,75 cases of technology upgrading were completed, covering 720 sets of equipment, with an investment scale of 165 million yuan; In 2022,55 cases of renovation will be promoted, and 480 sets of key equipment will be updated, with 123 million yuan invested.

Figure 1 shows the thermal map of the cost distribution of a single equipment generated based on the probability field reconstruction algorithm. whose statistical characteristics are highly consistent with the economic law of engineering, which confirms the practical value of this method in the quantification of cost interval.

Figure 1 shows the statistical distribution of unit costs for disconnector replacement projects. The blue bars represent individual unit costs, while the red line indicates the average cost. The cost distribution is highly varied, with some projects having significantly higher costs, while others are closer to the average.

Based on the probability field modeling method, the single cost statistical characteristics of the equipment iterative engineering are extracted, and the cumulative probability integral curve is constructed by using the three-times Hermite interpolation algorithm. Figure 2 shows a single cost distribution surface generated by feature-preserving density analysis technology, which accurately reflects the statistical law of engineering economic parameters.

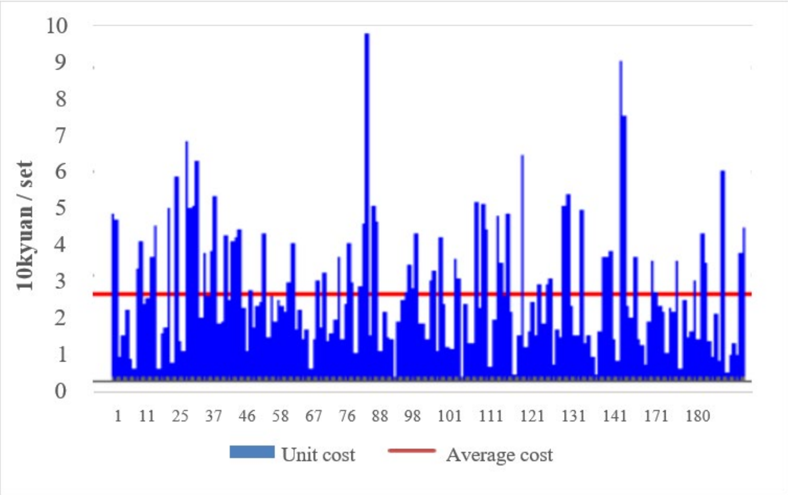

Fig. 1. Statistics of unit cost of disconnector replacement projects.

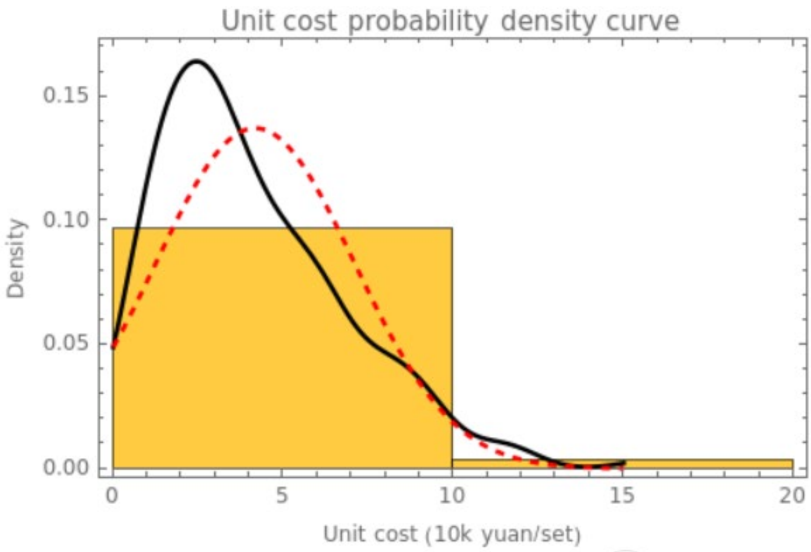

Fig. 2. Results of density curves

The val coverage efficiency (ICP) is used as the confidence verification index, and its Bayesian verification formula is defined as:

$$OW = \frac{1}{N}\sum_{i=1}^{N}\beta_i \tag{5}$$

Where, N representing the sum of the number of samples; β_i is a Boolean value, taking the value of 1 if the sample is within the interval, and 0 otherwise. If ICP=1, it implies that all samples lie within the cost interval, providing the highest reliability. However, this makes the cost interval lose its practical relevance. Typically, the interval coverage efficacy index should be statistical consistent with the preset reliability threshold. The empirical data show that when the confidence domain coverage efficiency and the target confidence degree show convergence characteristics, the constructed engineering economic decision interval has the optimal statistical efficiency. Figure 3 verifies the theoretical completeness of the confidence domain construction method by analyzing the coverage characteristics under multi-level reliability parameters.

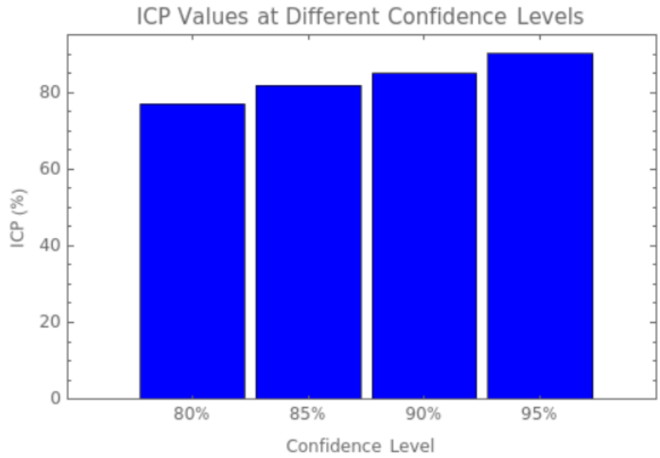

Fig. 3. Confidence Level vs. ICP Distribution

5 Conclusion

This study applies the kernel density estimation KDE) method to determine a reasonable cost interval for power grid technological transformation projects. By analyzing transformer replacement cost data, KDE proves to be a flexible and accurate alternative to traditional parametric models. The introduction of interval coverage probability ICP further ensures cost interval reliability at various confidence levels. Results highlight the importance of selecting an appropriate bandwidth to balance estimation accuracy. Overall, this research enhances cost control and investment decisions. Future work may integrate KDE with machine learning for improved prediction accuracy.

References

1. Li Yafei. Analysis on cost deviation and evaluation of management effectiveness [J]. China Equipment Engineering, 2022, (22): 58-60.
2. YUAN Ya. Discussion on the application of neural network in power grid engineering cost and rapid estimation of engineering quantity[J].Low Carbon World, 2023, 13(06): 175-177. DOI:10.16844/j.cnki.cn10-1007/tk.2023.06.033.)
3. YUAN Zhaoxiang, ZHANG Xueqin, GUO Yujun, et al. Research on cost prediction of power grid transmission project based on stacking ensemble model[J].Shandong Electric Power Technology,2022,49(12):14-19.)
4. Peng Luwei, Bao Yijun, Wu Yao, et al. Research on cost management evaluation model of the whole process of power grid engineering[J]. Project Management Technology, 2023, 21(08):127-133.)
5. Li Yafei. Research on cost deviation analysis and management effectiveness evaluation of power grid engineering [J]. China Equipment Engineering,2022(22):58-60.)
6. Kulak Y V, Petrova V T ,Novichikhin V A .Development of decision- making mechanism in engineering design of phased coal mines technical upgrade[J].IOP Conference Series: Earth and Environmental Science,2017,84(1):012038-012038.
7. Vasilis A , Ioannis M ,Andreas S , et al.An integrated life cycle assessment and life cycle costing approach towards sustainable building renovation via a dynamic online tool [J]. Applied Energy,2023,334.
8. Yang Yunhua, Yin Bingbing, Spring Tour, et al. Research on investment economy measurement of power grid production technical transformation project [J]. Automation application, 2021, (11): 169-172.