Research Article

Operation Performance Evaluation of Urban Rail Transit PPP Projects: Based on Best Worst Method and Large-Scale Group Evaluation Technology

Yuning Wang ¹, Yingzi Liang ², Cong Li ², and Xinyuan Zhang ³

¹School of Geographic and Environmental Sciences, Tianjin Normal University, Tianjin, China

²College of Management and Economics, Tianjin University, Tianjin, China

³Department of Statistics, University of California, Los Angeles, CA, USA

Correspondence should be addressed to Xinyuan Zhang; zxy47@ucla.edu

Received 30 April 2021; Accepted 19 September 2021; Published 29 September 2021

Academic Editor: Castorina S. Vieira

Copyright © 2021 Yuning Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Urban rail transit project is a kind of typical operating public project. Adopting the PPP model can alleviate local financial pressure and improve capital use efficiency. However, many existing urban rail transit PPP projects have fallen into the dilemma of cost overrun, schedule delay, and poor product quality. The lack of reasonable project performance evaluation is considered as an important cause. This research first clarifies the meaning and characteristics of project performance evaluation by comparing and analyzing several basic concepts and relationships. Secondly, an operation performance evaluation system based on urban rail transit PPP project of a three party is constructed from a multistakeholder perspective. Finally, the best worst method and large-scale group evaluation technology are used based on the comparison of multistakeholder evaluation application scenarios and evaluation methods. A quantitative model is constructed to evaluate the operating performance of urban rail transit PPP projects and is tested and explained by a specific case study. Most current studies generally focus on the earlier stage of project performance, and this article mainly discusses and researches the operation performance of PPP projects. There, suggestions are provided for the operation performance evaluation theory and urban rail transit PPP project practices. This article focuses on the operation performance evaluation of PPP projects.

1. Introduction

With the China's economic development and the acceleration of urbanization, the number of urban populations and cities' scale continues to expand. Citizens' demand for urban transportation increases continuously, and the number of private cars in urban area rises significantly. However, the travel mode represented by private cars has caused many urban problems: traffic congestion, energy consumption, greenhouse gas emissions, and air pollution caused by automobile exhaust. These problems seriously threaten citizens' living standards and the development sustainability of the city [1]. To solve these problems, the government hopes to reduce private car ownerships and replace private car travel by increasing the public transport supply. Urban rail

transit has many advantages, such as green, efficient, convenient, large volume, vital accessibility, and optimization of urban spatial structure. Public transportation is widely favored by the government and society, and it is regarded as an essential measure to solve urban traffic problems [2].

As a typical type of public transportation infrastructure, urban rail transit projects have an enormous initial investment demand, which is a massive debt and is difficult for local governments to bear within a short time. From the perspective of economic attributes, urban rail transit projects are quasipublic products. The application of the Public-Private Partnership (PPP) in the construction and operation of urban rail transit projects can effectively alleviate the situation of lacking local government infrastructure funds. It can also enhance the improvement of project management

philosophy, technical capability, and operational efficiency [3]. However, many ongoing infrastructure PPP projects worldwide have encountered several difficulties, such as construction delays, cost overruns, corruption, poor operating performance, and low public satisfaction. The performance of these PPP projects is far behind expectations [4, 5]. Studies have shown that among the many factors affecting the success of infrastructure PPP projects [6], insufficient project performance evaluation is one of the key factors with great importance [7]. The existing performance evaluation tools of PPP projects generally have various shortcomings, such as limited evaluation dimension of performance indicators and ignoring operation and maintenance period performance. These problems lead to low public product supply and public resource utilization efficiency in PPP infrastructure projects [8].

Urban rail transit PPP projects have a considerable investment scale and complex stakeholder relationships, whose success or failure has a profound impact on both the society and economy. Hence, strengthening the performance management of urban rail transit PPP projects is of great significance. It guarantees to achieve project goals, ensures the rational allocation and full utilization of resources, stimulates the vitality of social capital, and promotes social and economic development. The traditional project performance evaluation method based on the “period cost quality” has shown many shortcomings when measuring PPP projects’ performance: (1) In terms of content evaluation, the evaluation dimension is single and unable to reflect the numerous key success factors of infrastructure PPP projects. For instance, the VfM (Value for Money) method is mainly based on financial evaluation, which ignores operating performance from the perspectives of stakeholders. It is challenging to quantify PPP projects’ external influence within VfM. (2) In terms of time evaluation, existing researches cannot fully reflect the dynamic changes of the project’s external environment and the project performance requirements in different periods (especially for the operation period). (3) In terms of the function evaluation, the evaluation results only reflect the insufficiency of the project output. It cannot correctly and comprehensively reflect the project’s specific problems, which allows further targeted and operable improvement measures. Hence, this research first summarizes relevant theories of PPP project performance evaluation. Building on stakeholders’ perspective, this research adopts the best worst method and large-scale group evaluation technology to construct an urban rail transit PPP project’s operation performance evaluation system. A quantitative evaluation model of the large-scale group is proposed and tested through a practical case. From a performance perspective, this article attributes the theoretical basis for achieving a win-win situation within multistakeholders and sustainable development of rail transit PPP projects and provides a quantitative scientific basis for the government to supervise urban rail transit PPP projects.

The structure of the rest parts in this article is as follows. Section 2 provides a literature review for the relevant research. Operation performance evaluation system of urban

rail transit PPP projects is constructed in Section 3. Section 4 introduces the theory of the best worst method and large-scale group evaluation technology. An empirical study based on Beijing Metro Line 4 is conducted in Section 5. Section 6 gives the conclusions.

2. Literature Review

2.1. Performance Evaluation of PPP Projects. Performance management is one of the important topics in the PPP project management field. With the widespread application of the PPP model, PPP project performance management has attracted significant attention in both academic and industry fields. The research on PPP projects’ performance was initially developed based on the performance management of the government, enterprises, and organizations, from which an independent theoretical system is gradually formed. Several methods have been widely implemented in PPP project performance, such as Balanced Score Card (BSC), the performance prism, benchmarking, Critical Success Factors (Critical Success Factors, CSFs), project success degree models, quality function deployment, and Key Performance Indicators (KPIs) [9–11]. Besides, mathematical methods, such as tomographic analysis, structural equation modeling, fuzzy comprehensive evaluation, and data envelopment analysis, have also been applied in performance evaluation researches [12, 13]. Initially, performance evaluation was only a simple postevaluation of the PPP project regarding time, cost, and quality. With the proposition and development of the Value for Money (VfM) concept, the evaluation of project performance extended into the time, input, output, production efficiency, economic benefits, and social and ecological environmental impacts during the project’s entire life cycle. The evaluation of PPP infrastructure projects began to focus on the project’s continuity and influence [14]. Liu et al. proposed a stage-oriented evaluation process that evaluates each stage’s KPIs [15]. Liu et al. summarized three evaluation stages: (1) initiation and planning, (2) procurement, and (3) partnership [16, 17]. Large-scale group decision making (LGDM) involved various participants from different subgroups, which conduct the evaluation comprehensive. As PPP projects involved multistakeholders in the construction and operation periods, LGDM is suitable for the operation performance evaluation of PPP projects [18].

Unlike the “temporary contract organization” perspective adopted by traditional project governance, PPP projects have a long construction and operation cycle. The long-term and dynamic partnership between public sectors and social capital is vital for PPP projects, and thus, evaluating interest requirements of multistakeholder and operation performance is necessary. However, existing performance evaluations for PPP projects in operation and maintenance period is insufficient.

2.2. Performance Evaluation Indicators of PPP Projects. A scientific, reasonable, and quantifiable performance evaluation indicator system is the essential foundation of

performance evaluation. PPP project performance evaluation is achieved by comparing the actual performance of key performance indicators with appropriate performance standards expected by multistakeholders during the project execution period. The PPP project performance evaluation indicator system should reflect the project's implementation and achievement objectives comprehensively and objectively to support the value control of the PPP project life cycle.

Currently, the principles of PPP project performance evaluation indicators are designed based on the effectiveness, efficiency, economy, fairness, and other indicators, combining with the project's types and characteristics. Kagioglou et al. first introduced the tool balanced scorecard for organizational performance management into the construction industry [19]. Yuan et al. analyzed the critical success factors (CSFs) affecting PPP projects' performance and designed a PPP infrastructure project performance system [20]. From multiple perspectives of different stakeholders, Goran et al. developed a two-level performance evaluation indicator system for PPP transportation projects based on CSF [21]. Yuan et al. comprehensively used CSF and KPI methods to extract key performance indicators and established a system for PPP projects based on project process [22]. Villalba-Romero and Liyanage compared the difference between Performance Measurement System (PMS) and KPI through case analyses, which measures PPP highway projects' success, then selected strong indicators based on the analyses [13]. To reflect PPP project performance indicators comprehensively, existing studies have summarized and classified these indicators from different perspectives, including the four-quadrant classification method based on the balanced scorecard, three-stage classification based on the initial period, design and construction period, operation and maintenance period of PPP projects and indicator classification based on the difference of technical, economic, operation, and maintenance [16, 17]. Huang et al. evaluated the operation performance of urban rail transit projects based on entropy-TOPSIS model [23]. Liu et al. developed a performance evaluation system for urban rail transit PPP projects based on the balanced scorecard and PPP projects characteristics [24].

These classification methods reflect the characteristics of project performance indicators in different aspects. However, comprehensive applications of methods are limited. An integrated key performance indicator scheme and performance management system has become the trend of PPP project research, which incorporates the views of different stakeholders, extends to the project operation and maintenance period, and combines internal learning, improvement, and innovation of the project organization.

3. Operation Performance Evaluation System of Urban Rail Transit PPP Projects

3.1. Identification for the Performance Evaluation Subject of Urban Rail Transit PPP Projects. The urban rail transit PPP project is a kind of large-scale public infrastructure project. From the perspective of government functions, the public sector, as a provider of public products and services, should

represent the public's interests and be supervised from the public. Hence, introducing public participation into project performance evaluation reflects that the public sector respects public opinion and adheres to user demand orientation. Consequently, this improves the objectivity, scientificity, and legitimacy of the evaluation results [25]. Public participation is a planned and organized institutional arrangement, which refers to the communication among the government and development departments, and the public that can be achieved through specific forms and channels. Thus, the public can participate in public affairs management and decision-making process involving public interest to resolve the conflicts and maximize social welfare [26]. However, research shows that the content and priority of the project's interest requirements from the public sector and the public are not entirely consistent [22]. It leads to the different definitions of project success, which is the fundamental source of PPP project performance management's complexity, as shown in Table 1.

In summary, many stakeholders of urban rail transit PPP projects have apparently different expectations in the project performance, which have led to the ambiguity of project success criteria and project value evaluation criteria. Therefore, the clarification of each evaluation subject's project performance and evaluation standards is the basis for establishing a project performance evaluation system. According to Mitchell's stakeholder identification theory, this article divides the key stakeholders in the PPP projects into three categories: public sector, social capital, and the public from the perspectives of power, urgency, and legitimacy [27]. The public sector expects that the project can guarantee the value for money (VFM, value for money), which means investment efficiency and positive externality achievement. For social capital, the most critical project goals are profitability, sustainability, and owners and end-users satisfaction. The public expects that the process of the project should be open and transparent. They anticipate that the project should provide high-quality public products and services and achieve other positive externalities. Although the performance goals of three parties seem different, due to the nature of the PPP project partnership, the three parties share a relatively broad consensus on the project's success goals, such as establishing good public-private relationships and ensuring end-user satisfaction. However, the priorities of these indicators vary for different stakeholders [28]. Effective project performance management should coordinate the interests of all parties, reduce differences, and achieve a win-win situation for the multistakeholders [29].

3.2. Operation Performance Evaluation Indicator System of Urban Rail Transit PPP Projects. The mainstream PPP project operation performance evaluation method currently is based on the balanced scorecard and key success factors. However, the balanced scorecard only distinguishes between financial and nonfinancial perspectives and internal and external performance evaluation views. It cannot explain the performance indicators and the relationship between indicators from the perspective of project operation logistic.

TABLE 1: Interest requirements on projects from multistakeholders.

Rank	Social capital	Government sector	Public
1	Acceptable construction quality	Acceptable construction quality	Acceptable construction quality
2	Reduce life cycle costs	Within construction and operation budget	Qualified public service
3	Complete on time	Address insufficient public sector budgets	Provide timely and convenient services
4	Within construction and operation budget	Transfer risk	Satisfy the needs of public facilities
5	Qualified public service	Qualified public service	Completed on time

Moreover, there is a lack of explanation of the relationship between the key success factors, which include various uncontrollable environmental indicators of projects, such as the political and economic environment. Compared with the balanced scorecard and key success factors, the performance prism method emphasizes stakeholders' status in performance evaluation, which reflects the logic of project operation and the interactive relationship between the project and stakeholders [30, 31]. The performance prism is a three-dimensional framework model, and its structure can be imaged as a triangular prism, as shown in Figure 1. The performance prism includes five interrelated aspects: (1) Satisfaction of stakeholders: to maintain the relationship with stakeholders, it is necessary to clearly understand the different interest needs of various stakeholders, such as the project company's technology, management experience, profitability, and end-user satisfaction. (2) Contributions of stakeholders; there is a mutually beneficial relationship among various stakeholders and the project company. Stakeholders obtain certain benefits from the project company and are required to contribute to the project company. For example, the project company hopes to receive supervision and support from the public sector, and the trust and cooperation from stakeholders. (3) Developing a strategy: for the project company, no strategy, or unreasonable strategy will lead to many problems such as blind expansion, weak competition, poor execution, and unsustainable growth. Therefore, considering the internal and external factors of the project company, it is particularly important to develop a value-for-money analysis of the project's whole life cycle. (4) Implementation process: for project company, the process of strategy implementation is the most important driving factor to improve the performance of the project company. It determines the quality, efficiency, cycle, and cost of products and services. (5) Capacities: it mainly reflects the capacities to operate these processes. Among them, strategy, process, and capability constitute the three sides of the performance prism. The satisfaction of stakeholders and the contributions of stakeholders constitute the two bottom sides of the performance prism. The five indicators permeate and interact with each other to form a complete system. Strategy formulation, implementation process, and ability are interdependent, mutually conditional, and dynamic cyclic. Strategy formulation is the direction, and the implementation process is the means, giving full play to the four most basic functions of performance evaluation: judgment, prediction, selection, and orientation. The performance prism starts on stakeholders' needs; determines the strategic planning, process setting, and capacity building of the project; and finally returns to the stakeholders' contributions to form a complete closed loop.

This method fits the PPP project performance evaluation concept proposed in this article, which is based on process management and stakeholder orientation [16].

Based on the perspective of the performance prism, from the definition of project success and project operation performance goals, this research constructs an operation performance evaluation indicator system of urban rail transit PPP projects that combines the viewpoint of continuous improvement, as shown in Table 2. Compared with traditional performance evaluation methods, the proposed system (1) includes process and efficiency indicators, which are more comprehensive, (2) introduces public participation to fill in the lack of stakeholders, and (3) emphasize project governance and continuous improvement to ensure the timeliness and availability of evaluation results.

4. The Method

4.1. Determination of Evaluation Indicator Weight. Multicriteria Decision Making (MCDM) refers to evaluating multiple candidate options based on a series of different criteria to find the optimal solution [32]. Weighting the evaluation criteria is the most important segment. After decades of research, several relatively thorough and systematic methods are developed to conduct the weight vector of evaluation criteria. At present, there are mainly two categories of conventional weighting methods: (1) subjective weighting methods based on subjective opinions from experts, such as the Delphi method, analytic hierarchy process (AHP), fuzzy comprehensive evaluation, and gray clustering, and (2) objective weighting methods based on the characteristics of the evaluation data, such as entropy weight method, OWA operator, and principal component analysis/factor analysis [33]. However, there are numerous indicators and large-scale evaluation data in the performance evaluation of urban rail transit PPP projects. Also, stakeholders are classified in prior, rather than identified using the evaluation information. It is challenging for methods mentioned above to process large-scale evaluation data, and most of the objective weighting methods based on data feature classification can hardly perform satisfyingly. In response to these two problems, Rezaei proposed a new multicriteria decision-making method—the best worst method (BWM) [34]. The decision maker finds the best and most important criteria and the worst and least important criteria in the criteria group according to the ultimate target. Then, pairwise comparisons are conducted between the best worst criteria and other criteria. The maximum nonlinear programming problem is constructed according to the judgment vector. The weight of each criteria corresponding to the final goal can be obtained

FIGURE 1: The performance prism of urban rail transit PPP project.

TABLE 2: Performance evaluation indicator system during operation and maintenance period.

Dimension	Performance evaluation indicators during operation and maintenance period
Stakeholder needs	KPI1: introduce advanced equipment, technology, and management experience (public sector) KPI2: the user's appropriate rate, price adjustment, and compensation mechanism to ensure the sustainability of the project and the reasonable profit of the social capital (public sector and social capital)
Strategy	KPI3: end-user satisfaction (the public) KPI4: value for money during the project life cycle KPI5: project's profitability, solvency, and financial sustainability
Process	KPI6: compliance, contract performance, and management and dispute resolution KPI7: social, environmental, and economic externalities, including health, safety and environment (HSE) management, relations with surrounding communities, reduction of carbon and exhaust emissions, noise and vibration pollution, promotion of employment and local economic development, etc. KPI8: operational management and service quality, including transportation volume, comfort level, punctuality rate, complaint rate, failure rate, accident rate, average waiting time, etc. KPI9: application of advanced technology equipment, management, and maintenance of facilities KPI10: effective interface management
Capacity	KPI11: the qualifications and capabilities of the project team, including experienced employees, training and learning systems, information management systems, etc. KPI12: innovation and sustainable development, including operation innovation, facility maintenance innovation, etc.
Stakeholder contribution	KPI13: supervision and support of the public sector (operation, maintenance) KPI14: effective public participation (evaluation) KPI15: trust and cooperation of stakeholders

by solving the problem. Statistical results show that BWM's performance is significantly better than AHP in terms of evaluation difficulty, consistency ratio, minimum violation, maximum deviation, etc. [35]. Comparing with the existing MCDM method, BWM has several advantages: (1) less comparative data are required. Other matrix-based MCDM methods (such as AHP) require $n(n-1)/2$ comparisons. BWM is a vector-based method requiring only $2n-3$ comparisons and thus reduces calculations; (2) the comparison vector is more consistent, which means the result is more accurate [36]. Therefore, this research chooses BWM as the weighting method for evaluating urban rail PPP projects' performance evaluation indicators.

Based on the idea of BWM, the steps to solve for weight vectors of performance evaluation indicators of the urban rail transit PPP projects are as follows:

Step 1: invite evaluation experts in multiple fields of urban rail transit, including authoritative scholars, consulting experts, senior practitioners, deputies of community people's congress, media, and Non-governmental Organization (NGO) representatives, with sufficient professional knowledge and scientific literacy required for evaluation.

Step 2: determine the project performance evaluation indicator system $(C_1, C_2, \dots, C_n)$ based on the stage of the project.

Step 3: ask the experts to select the most and least important indicators. Specific comparison values are not required in this step.

Step 4: determine the priority of the most important indicators relative to other indicators, obtain the

relative importance from language scale transformation and according to which obtain the comparison vector $A_B = (a_{B1}, a_{B2}, \dots, a_{Bn})$. Besides, a_{Bj} represents the priority of the most important indicator C_B to another indicator C_j , while $a_{BB} = 1$.

Step 5: determine the inferiority degree of the least important indicators relative to other indicators and get the comparison vector $A_W = (a_{1W}, a_{2W}, \dots, a_{nW})^T$. a_{jW} represents the priority of other indicator C_j to the least important indicator C_W , while $a_{WW} = 1$.

Step 6: calculate the weight vector. To determine the optimal weights of the indicators, the maximum absolute error $\{|w_B - a_{Bj}w_j|, |w_j - a_{jW}w_W|\}$ for all indicators should be the minimized, and thus, a nonlinear programming function is constructed as follows:

$$\left\{ \begin{array}{ll} \min & \max_j \left\{ \left| \frac{w_B}{w_j} - a_{Bj} \right|, \left| \frac{w_j}{w_W} - a_{jW} \right| \right\}, \\ \text{s.t} & \sum w_j = 1, \\ & w_j \geq 0 \text{ for all } j. \end{array} \right. \quad (1)$$

For the convenience of solving, the nonlinear programming problem can be transformed into the following form:

$$\left\{ \begin{array}{ll} \min & \xi, \\ & \left| \frac{w_B}{w_j - a_{Bj}} \right| \leq \xi, \text{ for all } j \\ \text{s.t} & \left| \frac{w_j}{w_W - a_{jW}} \right| \leq \xi, \text{ for all } j \\ & \sum w_j = 1, \\ & w_j \geq 0 \text{ for all } j. \end{array} \right. \quad (2)$$

The weight vector $(w_1, w_2, \dots, w_n)$ and the minimum consistency error ξ^* can be obtained by solving this nonlinear programming problem. It can be proved that the problem has a unique solution [36]. A smaller ξ^* indicates a smaller consistency error of the comparison information provided by the evaluator and results in stronger consistency. Therefore, ξ^* can be used to construct a consistency coefficient CR of quality evaluation, that is, the ratio of the maximum error obtained when ξ^* and a_{BW} take different values (Table 3).

$$\text{consistency ratio} = \frac{\xi^*}{\text{consistency index}_{\max}}. \quad (3)$$

4.2. Construction of Evaluation Model

4.2.1. Problem Description and Variable Definition. To establish the large-scale evaluation model for operation performance of urban trail transit PPP projects, this research first provides relevant variables and their definitions:

$G = \{G_1, G_2, \dots, G_m\}$: G represents the total set of evaluators, G_i represents the i th subset that participates the evaluation, $i = 1, 2, \dots, m$.

$Q = (Q_1, Q_2, \dots, Q_m)$: Q represents the number of evaluators in each subset of total set G , Q_i represents the number of evaluator in the subset G_i , $i = 1, 2, \dots, m$.

$C = (C_1, C_2, \dots, C_n)$: C represents evaluation indicator system, C_j represents the j th indicator, $j = 1, 2, \dots, n$.

$W = (W_1, W_2, \dots, W_n)$: W represents the weight of evaluation indicator, W_j represents the j th indicator weight, $j = 1, 2, \dots, n$.

$S = \{S^1, S^2, \dots, S^l\}$: S represents the set of evaluation scale, S^l represents the l th evaluation level, $l = 1, 2, \dots, t$. Generally, set S is positively ordered, which means, when $l > l'$, $l, l' \in \{1, 2, \dots, t\}$, $S^l > S^{l'}$. Set S can include both language scale, such as {very bad, bad, fair, good, very good}, and numerical evaluation value, such as {1, 2, 3, 4, 5}.

$D_{ij} = (d_{ij}^1, d_{ij}^2, \dots, d_{ij}^{h_i})$: D_{ij} represents the evaluation vector of subset G_i within indicator C_j , d_{ij}^h represents the evaluation information of h th evaluator in subset G_i , $d_{ij}^h \in S$, $i = 1, 2, \dots, m$, $h = 1, 2, \dots, Q_i$. For research convenience, if S is language scale, then convert it into the corresponding numerical evaluation value. For example, {very bad, bad, fair, good, very good} can be transferred as {1, 2, 3, 4, 5}.

4.2.2. Steps of Evaluation

(1) *Determine the Frequency Distribution of Project Performance Evaluation Information for Each Subset under a Certain Indicator.* According to the evaluation information on evaluation indicator C_j provided by evaluator in subset G_i , which is vector $D_{ij} = (d_{ij}^1, d_{ij}^2, \dots, d_{ij}^{h_i})$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, the count matrix $I_{ij} = [I_{ij}^{hl}]_{Q_i \times t}$ can be obtained. Besides, I_{ij}^{hl} is a 0-1 variable in every row vector $I_{ij}^{hl} = (I_{ij}^{h1}, I_{ij}^{h2}, \dots, I_{ij}^{ht})$. If $d_{ij}^h = S^t$, $I_{ij}^{ht} = 1$. Otherwise, $I_{ij}^{hl} = 0$, $h = 1, 2, \dots, Q_i$, $l = 1, 2, \dots, t$.

Define q_{ij}^l as the number of evaluators using language scale S^t when evaluating the project in subset G_i under indicator C_j . Then

$$q_{ij}^l = \sum_{h=1}^{Q_i} I_{ij}^{hl}. \quad (4)$$

Define p_{ij}^l as the frequency of evaluators using language scale S^t when evaluating the project in subset G_i under indicator C_j . Then

TABLE 3: Consistency coefficient.

a_{BW}	1	2	3	4	5	6	7	8	9
$\xi_{\max}$	0.00	0.44	1.00	1.63	2.30	3.00	3.73	4.47	5.23

$$p_{ij}^l = \frac{q_{ij}^l}{Q_i} \quad (5)$$

Obviously, $\sum_{l=1}^t p_{ij}^l = 1$, $0 \leq p_{ij}^l \leq 1$.

Therefore, the frequency distribution of evaluation in subset G_i under indicator C_j is

$$P_{ij}(x) = \begin{cases} p_{ij}^1, & x = S^1, \\ p_{ij}^2, & x = S^2, \\ \dots & \dots \\ p_{ij}^t, & x = S^t, \end{cases} \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n. \quad (6)$$

To facilitate calculation, define $B = (B^1, B^2, \dots, B^t)$ as the set of evaluation value transferred from evaluation scale $S = (S^1, S^2, \dots, S^t)$. If S is a numeric set, then $B^l = S^l$, $l = 1, 2, \dots, t$. If S is language scale set, then B^l represents the transferred evaluation value corresponding to scale S^l , $l = 1, 2, \dots, t$. Based on this transformation, the frequency distribution of evaluation information based on the language scale can be transformed into a frequency distribution based on the evaluation value.

$$P_{ij}(x) = \begin{cases} p_{ij}^1, & x = B^1, \\ p_{ij}^2, & x = B^2, \\ \dots & \dots \\ p_{ij}^t, & x = B^t, \end{cases} \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n. \quad (7)$$

For the convenience of research, it is expressed in vector form as $p_{ij}^l = (P_{ij}^1, P_{ij}^2, \dots, P_{ij}^t)$.

In the context where a large number of evaluators or extreme opinions appear in the subset, a few extreme opinions can be identified as noise and eliminated. In this way, the consistency of the evaluation results can be increased and the authenticity of evaluation information can be guaranteed. Define u_{ij} and σ_{ij} as the mean value and variance of the evaluation information's frequency distribution $p_{ij}^l = (P_{ij}^1, P_{ij}^2, \dots, P_{ij}^t)$, and

$$u_{ij} = \sum_{l=1}^t P_{ij}^l B^l, \quad (8)$$

$$\sigma_{ij} = \sum_{l=1}^t P_{ij}^l (B^l - u_{ij})^2. \quad (9)$$

According to the Rajda criterion (3σ rule) for judging data abnormalities, when the number of measurements is sufficiently large (the number of samples > 10), if the sample data distribution is normal or approximately normal, data

outside the interval $(\mu - 3\sigma, \mu + 3\sigma)$ should be regarded as gross errors and eliminated, in which standard error is calculated through data processing. For the stakeholder's subset with similar background and interest requirements, the provided evaluation information can be regarded as following the normal distribution. Thus, denoise the evaluation information I_{ij} and gain I'_{ij}

$$I'_{ij} = \begin{cases} 0, & |B_{ij}^l - u_{ij}| > 3\sigma_{ij}, \\ I_{ij}^l, & |B_{ij}^l - u_{ij}| \leq 3\sigma_{ij}. \end{cases} \quad (10)$$

After eliminating abnormal data, frequency needs to be amended

$$q_{ij}^l = \sum_{h=1}^{Q_i} I'_{ij}^h, \quad (11)$$

$$p_{ij}^l = \frac{q_{ij}^l}{\sum_{l=1}^t q_{ij}^l}.$$

Then, the demanded frequency distribution $P_{ij}(x)$ can be obtained. It should be mentioned that the denoise process is not compulsory. If the number of evaluators is relatively small or the evaluation organizer considered that the valuable information would lose during the denoise process, the denoise procedure should not be performed.

(2) *Determine Each Subset's Evaluation Weight for the Project Evaluation under a Particular Evaluation Indicator.* When aggregating evaluation information, it is essential to assign appropriate weights to each subset. Two factors should be considered when determining each subset's weight for project evaluation: Firstly, the evaluators are stakeholders on different behalf with different knowledge, experience, background, and interest requirements. The importance and discourse power of each subset in the process should be considered according to the actual situation to obtain subjective weights of the subsets; Secondly, with a large number of participants in the evaluation process, the consistency of the evaluation information given by the evaluators in each subset can reflect its quality [37]. Higher consistency of the evaluation information means that the evaluators from the same subset have highly consistent knowledge and interest needs. Consequently, the evaluation information they provided is more important, and higher weight should be assigned to obtain the objective weight of the subset. Therefore, the comprehensive consideration of the subset evaluation weights within subjective and objective factors can ensure the scientificity and reliability of the weighting process.

Define $w^{\text{sub}} = (w_1^{\text{sub}}, w_2^{\text{sub}}, \dots, w_m^{\text{sub}})$ as the subjective weight of each subset under indicator C_j , w_i^{sub} represents the subjective weight value of subset G_i , and then $\sum_{i=1}^m w_i^{\text{sub}} = 1$ and $0 \leq w_i^{\text{sub}} \leq 1$, $i = 1, 2, \dots, m$.

The objective weight of subset is obtained based on the evaluation information's consistency given by subset. Define $\sigma(P')$ as the theoretical maximum variance in all frequency distribution. Then

$$p'(x) = \begin{cases} 0.5, & x = B^1, \\ 0, & x = B^2, \\ \dots, & \dots, \\ 0.5, & x = B^t, \end{cases} \quad (12)$$

$$\sigma(P') = 0.5[B^1 - (0.5B^1 + 0.5B^t)]^2 + 0.5[B^t - (0.5B^1 + 0.5B^t)]^2 = \frac{(B^t - B^1)^2}{4}. \quad (13)$$

Define the consistency coefficient CI_{ij} as the consistence of the evaluation information given by subset G_i under indicator C_j [37]. Then

$$CI_{ij} = 1 - \sqrt{\frac{\sigma(P_{ij})}{\sigma(P'_{ij})}} = 1 - \frac{2\sqrt{\sigma(P_{ij})}}{B^t - B^1}. \quad (14)$$

According to (5) and (9), (a) $0 \leq CI_{ij} \leq 1$; (b) $CI_{ij} = 1$ if and only if $\sum_{l=1}^t p_{ij}^l = 1$ and $p_{ij}^g = 1, g = 1, 2, \dots, t$; (c) $CI_{ij} = 0$ if and only if $P_{ij} = P'$. Generally, the larger consistence coefficient CI_{ij} , the better information quality, the higher objective weight. Define w_{ij}^{ob} as the objective weight of each subset under indicator C_j . Then

$$w_{ij}^{ob} = \begin{cases} \frac{CI_{ij}}{\sum_{k=1}^m CI_{kj}}, & \text{if } \sum_{k=1}^m CI_{kj} > 0, \\ W_{ij}^{sub}, & \text{if } \sum_{k=1}^m CI_{kj} = 0. \end{cases} \quad (15)$$

Obviously, $\sum_{i=1}^m w_{ij}^{ob} = 1$ and $0 \leq w_{ij}^{ob} \leq 1, i = 1, 2, \dots, m$.

Define w_{ij} as the evaluation weight of subset G_i under indicator C_j . Then

$$w_{ij} = \alpha w_{ij}^{ob} + \beta w_{ij}^{sub}. \quad (16)$$

In this equation, α and β are weight combination coefficients, $0 \leq \alpha, \beta \leq 1, \alpha + \beta = 1$. Evidently, $\sum_{i=1}^m w_{ij} = 1$ and $0 \leq w_{ij} \leq 1, i = 1, 2, \dots, m, j = 1, 2, \dots, n$. Values of α and β are obtained with the comprehensive consideration of the effects of subjective and objective factors on the importance and discourse power of subset.

(3) *According to the Frequency Distribution of the Evaluation Information and the Subset's Evaluation Weight, Evaluation Information and Fairness Are Considered Comprehensively, and the Evaluation Results Are Aggregated under a Particular Evaluation Indicator.* When aggregating evaluation information, besides the evaluation information, differences in opinions between different subsets should also be noted to ensure the fairness of the evaluation. When different subsets under a particular evaluation indicator have different perceptions of project performance, it is indicated that the project performance under this indicator does not reach consensus among all stakeholders. Hence, it is difficult to

evaluate this project as successful. For urban rail transit PPP projects, it is particularly important for project performance to ensure fairness and good partnership in projects. High-performance urban rail transit PPP projects not only should perform outstandingly in performance indicators but also be highly regarded by different stakeholders. By defining the fairness indicator of the two subsets' opinions, the difference of their opinions can be measured. After integrating the fairness evaluation into the project performance evaluation results, the evaluation results' authenticity and completeness can be guaranteed.

According to evaluation information, evaluation score η_j under the evolution indicator C_j is

$$\eta_j = \frac{\sum_{i=1}^m w_{ij} u_{ij}}{B^t}. \quad (17)$$

Then, $0 \leq \eta_j \leq 1, i = 1, 2, \dots, m, j = 1, 2, \dots, n$.

Cumulative frequency distribution $H_{ij}(x)$ is constructed to balance the difference between frequency distribution

$$H_{ij}(x) = \begin{cases} 0, & x < B^1, \\ p_{ij}^1, & B^1 \leq x < B^2, \\ p_{ij}^1 + p_{ij}^2, & B^2 \leq x < B^3, \\ \dots, & \dots, \\ p_{ij}^1 + p_{ij}^2 + \dots + p_{ij}^{t-1}, & B^{t-1} \leq x < B^t, \\ 1, & B^t \leq x. \end{cases} \quad (18)$$

Define D_j^{ir} as the diversity factor between subset G_i and subset G_r , $G_i, G_r \in G$, $H_{ij}(x)$ and $H_{rj}(x)$ presents the evaluation information's cumulative distribution of two groups of evaluators under indicator C_j . Then

$$D_j^{ir} = \sum_{l=1}^t |[H_{ij}(B^l) - H_{rj}(B^l)][B^{l+1} - B^l]|, \quad (19)$$

$i, r = 1, 2, \dots, m, i \neq r, j = 1, 2, \dots, n, B^{t+1} = B^t + 1.$

Define the difference D_l between subset G'_i and subset G'_r as the theoretical maximum difference of pairwise comparison in cumulative distribution. $H'_{ij}(x)$ and $H'_{rj}(x)$ represent the evaluation information's cumulative distribution from two group of evaluators under indicator C_j . Then

$$H'_{ij}(x) = \begin{cases} 0, & x < B^1, \\ 0, & B^1 \leq x < B^2, \\ 0, & B^2 \leq x < B^3, \\ \dots, & \dots, \\ 0, & B^{t-1} \leq x < B^t, \\ 1, & B^t \leq x, \end{cases}$$

$$H'_{rj}(x) = \begin{cases} 1, & x < B^1, \\ 1, & B^1 \leq x < B^2, \\ 1, & B^2 \leq x < B^3, \\ \dots, & \dots, \\ 1, & B^{t-1} \leq x < B^t, \\ 1, & B^t \leq x, \end{cases} \quad (20)$$

$$D' = \sum_{l=1}^t \left| [H_{ij}(B^l) - H_{rj}(B^l)] [B^{l+1} - B^l] \right| = (B^t - B^{t-1}) + (B^{t-1} - B^{t-2}) + \dots + (B^2 - B^1) = B^t - B^1. \quad (21)$$

According to the evaluation information, the fairness coefficient f_j^{ir} between subset G_i and subset G_r is defined under evaluation indicator C_j . Then

$$f_j^{ir} = 1 - \frac{D_j^{ir}}{D'} = 1 - \frac{D_j^{ir}}{B^t - B^1}, \quad i, r = 1, 2, \dots, m, i \neq r, j = 1, 2, \dots, n. \quad (22)$$

According to (13) and (14), (a) $0 \leq f_j^{ir} \leq 1$; (b) $f_j^{ir} = 1$ if and only if $H_{ij}(x) = H_{rj}(x)$; (c) $f_j^{ir} = 0$ if and only if $D_j^{ir} = D'$; (d) $f_j^{ir} = f_j^{ri}$. Generally, with larger fairness coefficient f_j^{ir} , the disagreement between the two subsets will be smaller and the evaluation information will be fairer.

By calculating the fairness coefficient between any two subsets in G under evaluation indicator C_j , the fairness coefficient matrix F_j between each subset under the evaluation indicator C_j can be obtained as

$$F_j = [f_j^{ir}]_{m \times m} = \begin{bmatrix} 1 & f_j^{12} & \dots & f_j^{1m} \\ f_j^{21} & 1 & \dots & f_j^{2m} \\ \vdots & \vdots & \ddots & \vdots \\ f_j^{m1} & f_j^{m2} & \dots & 1 \end{bmatrix}. \quad (23)$$

Based on this, define the fairness score f_j of each subset under evaluation indicator C_j , the influence of self-contrast is eliminated, and then

$$f_j = \frac{\sum_{i=1, i \neq r}^m \sum_{r=1}^m f_j^{ir}}{m \times (m-1)}. \quad (24)$$

And $0 \leq f_j \leq 1$, $i, r = 1, 2, \dots, m, i \neq r, j = 1, 2, \dots, n$. The larger f_j , the more consistency of opinions between subsets.

Considering the evaluation score η_j and fairness score f_j , the evaluation result V_j under evaluation indicator is aggregated.

$$V_j = a\eta_j + bf_j. \quad (25)$$

In this equation, a and b are weight combination coefficient, $0 \leq a, b \leq 1$, $a + b = 1$. Obviously, $0 \leq V_j \leq 1$. The values of a and b are defined based on the influence of the evaluation score and fairness score of evaluators on the evaluation results. It should be noticed that there is a situation in which the evaluation result may be good if the subset has a consistent negative evaluation of the project. Hence, the value of b should not be overly high to avoid this situation.

(4) *Based on the Indicator Weight and the Evaluation Results under Each Indicator, the Final Comprehensive Evaluation Conclusion Is Obtained.* According to the indicator weight vector W gained by the best worst method, the final evaluation result U is aggregated based on the evaluation results V_j under each evaluation indicator C_j

$$U = \sum_{j=1}^n W_j V_j. \quad (26)$$

Then, $0 \leq U \leq 1$. The obtained evaluation scores can be classified according to actual needs. Then, the performance under the indicator and the total performance can be evaluated.

5. Case Study

5.1. Overview of Beijing Metro Line 4 Project. Beijing Metro Line 4 is one of the main lines operating north-south in Beijing's urban rail transit network, with a total length of 28.2 kilometers. It starts from Anheqiao Bei(N) Station in the north, passing through major urban areas such as Haidian District, Xicheng District, Fengtai District, to the terminal station of Gongyi Xiqiao Station in the south. With 24 stations in total, it connects essential facilities and

transportation hubs such as well-known universities, Zhongguancun Science and Technology Park, Xidan Commercial District, places of interest, and Beijing South Railway Station. The overall planned investment for the Beijing Metro Line 4 project is about 15 billion yuan. The project officially started in August 2004 with a construction period of 5 years. The project completed and opened for trial operation in September 2009, and until the end of 2019, it has been in operation for more than ten years.

Beijing Metro Line 4 is the first PPP project in the field of urban rail transit in China. Beijing Infrastructure Investment Co., Ltd lead the project and introduced foreign-funded MTR Corporation Limited and state-owned Beijing Capital Group as joint shareholders. The company achieved multiple project management agreements with Beijing government and Beijing Rail Transit Construction Management Co. Ltd, franchise, in aspects of project investment, financing, construction, operation, maintenance, and other related matters. Regarding specific arrangements for the PPP model, the three companies set up Beijing MTR Corporation Limited (BJMTR) as a joint venture at a ratio of 2:49:49.

The project construction investment is divided into two parts: main body civil engineering and vehicle and ancillary equipment. The former is handled by a subsidiary wholly owned by Beijing Urban Investment Platform. The latter is in charge of BJMTR. After completing and accepting the project, BJMTR obtained the right to use Part A through the lease agreement. BJMTR is responsible for the operation, maintenance, and renewal of all the project assets through the franchise agreement. The franchise period is 30 years. Afterward, the AB parts will be transferred freely to the municipal government and the urban investment company.

5.2. Performance Evaluation of Beijing Metro Line 4 Project in Operation and Maintenance Period

5.2.1. The Determination of Evaluation Indicator Weight. This research uses the best worst method to determine the weight of the evaluation indicator. To ensure the scientificity and fairness of the evaluation indicator value assignment, five experienced experts specialized in related fields, such as urban rail transit investment, construction, and operation, were invited. They have no direct interest in the project. These five experts formed an evaluation team for value assignment (Table 4).

Taking the relative importance vector into formula (2), a nonlinear programming function is constructed, and LINDO software is used to get the solution. Weight vector $(w_1, w_2, \dots, w_n)$ and minimum consistency error ξ^* are obtained. As $a_{BW} = 9$, $\xi_{\max} = 5.23$, take them into (3) to calculate consistency coefficient CR (Table 5).

Then, $W = (0.0476, 0.1385, 0.0733, 0.1813, 0.1125, 0.0537, 0.0756, 0.0949, 0.0362, 0.0286, 0.0302, 0.0271, 0.0339, 0.0275, 0.0391)$. Through consistency coefficient, the average error is 0.81%, the maximum error does not exceed 1%, and the results have very good consistency.

5.2.2. The Collection and Processing of Evaluation Information. The total set of evaluators G for urban rail PPP projects comprises evaluators from three subsets: public sector G_1 , social capital G_2 , and social public G_3 . The evaluators are experts and representatives invited from the three sectors. To closely integrate project performance evaluation with project management and better apply the function of project performance evaluation, the evaluators include direct participants in the project from the public sector and social capital. This study invited 20 experts from three parties to participate in the operation and maintenance performance evaluation of the Beijing Metro Line 4 project. The evaluation was mainly based on the performance of the project in 2018-2019. According to the nature of the original information, the qualitative indicators should be divided into different performance evaluation levels under the consultation of scholars, experts, and senior practitioners. For research convenience, this study set the project performance at five levels, namely, {very poor, poor, qualified, good, very good}. For each level, a specific performance description was given to facilitate the judgment of the evaluator. The original qualitative data were obtained using several methods, such as case study, field investigation, symposium, and questionnaire survey. For quantitative indicators, levels are also determined according to corresponding ratios and levels with professional knowledge. The original quantitative data were obtained through the project company's database, daily reports, statements, field measurements, questionnaire surveys, and the like. After transforming the original qualitative information and quantitative data into an evaluation language scale based on measurement and comparison, it was further converted into corresponding evaluation values, such as $\{1, 2, 3, 4, 5\}$.

(1) *Determine the Frequency Distribution of Project Performance Evaluation Information from Each Subset under a Particular Indicator.* For indicator KPI1, "Introduce advanced equipment, technology, and management experience," according to formula (4) and (5), the evaluation information and frequency distribution of the three parties are shown in Table 6.

From (8) and (9),

$$\begin{aligned} u_{11} &= 4.45, \\ \sigma_{11} &= 0.4475, \\ u_{21} &= 4.35, \\ \sigma_{21} &= 0.5275, \\ u_{31} &= 4.45, \\ \sigma_{31} &= 0.66. \end{aligned} \quad (27)$$

(2) *Determine Each Subset's Evaluation Weight for the Project Evaluation under a Particular Evaluation Indicator.* Combining the importance and discourse power of each stakeholder in operation and maintenance stages of urban rail

TABLE 4: The best and worst comparison vectors provided by experts.

No	Best	Worst	Comparison vector	KPI1	KPI2	KPI3	KPI4	KPI5	KPI6	KPI7	KPI8	KPI9	KPI10	KPI11	KPI12	KPI13	KPI14	KPI15
1	4	12	a_{Bj} a_{jW}	7	2	3	1	3	6	5	3	7	8	8	9	7	8	7
2	2, 4	11	a_{Bj} a_{jW}	3	7	7	9	7	4	5	7	3	2	2	1	3	2	3
3	4	12	a_{Bj} a_{jW}	4	1	4	1	2	3	3	3	6	7	9	8	5	8	6
4	4	10, 14	a_{Bj} a_{jW}	6	9	6	9	8	7	7	7	4	3	1	2	4	2	4
5	2, 5	14	a_{Bj} a_{jW}	6	2	2	1	4	5	2	3	7	7	6	9	8	7	4
			a_{Bj} a_{jW}	4	8	8	9	6	5	8	7	3	3	4	1	2	3	5
			a_{Bj} a_{jW}	4	3	4	1	2	5	4	2	6	9	7	7	8	9	8
			a_{Bj} a_{jW}	6	7	6	9	7	5	6	8	3	1	3	3	2	1	2
			a_{Bj} a_{jW}	5	1	5	2	1	4	3	2	7	8	8	6	8	9	7
			a_{Bj} a_{jW}	5	9	5	8	9	6	7	8	3	2	2	4	2	1	3

TABLE 5: Indicator weight and consistency coefficient.

Experts number	1	2	3	4	5	Average
ξ^*	0.05	0.04	0.04	0.04	0.04	0.04
W1	0.0385	0.0539	0.0386	0.0627	0.0442	0.0476
W2	0.1347	0.1770	0.1159	0.0835	0.1815	0.1385
W3	0.0898	0.0539	0.1159	0.0627	0.0442	0.0733
W4	0.2226	0.1770	0.1905	0.2059	0.1105	0.1813
W5	0.0898	0.1077	0.0580	0.1253	0.1815	0.1125
W6	0.0449	0.0718	0.0464	0.0501	0.0552	0.0537
W7	0.0539	0.0718	0.1159	0.0627	0.0737	0.0756
W8	0.0898	0.0718	0.0773	0.1253	0.1105	0.0949
W9	0.0385	0.0359	0.0331	0.0418	0.0316	0.0362
W10	0.0337	0.0308	0.0331	0.0179	0.0276	0.0286
W11	0.0337	0.0154	0.0386	0.0358	0.0276	0.0302
W12	0.0195	0.0269	0.0166	0.0358	0.0368	0.0271
W13	0.0385	0.0431	0.0290	0.0313	0.0276	0.0339
W14	0.0337	0.0269	0.0331	0.0279	0.0158	0.0275
W15	0.0385	0.0359	0.0580	0.0313	0.0316	0.0391
CR	0.0090	0.0074	0.0079	0.0086	0.0075	0.0081

TABLE 6: Evaluation information.

Evaluator	1	2	3	4	5
Public sector	0	0	2	7	11
Social capital	0	0	3	7	10
Public	0	1	2	9	8

PPP projects, experts gave the subjective weight of each subset $w^{sub} = (0.3, 0.35, 0.35)$. According to (9),

$$\begin{aligned} CI_{11} &= 0.6655, \\ CI_{21} &= 0.6368, \\ CI_{31} &= 0.59380. \end{aligned} \quad (28)$$

According to (10), objective weight of each subset is $w_1^{ob} = (0.3510, 0.3559, 0.3131)$. Considering the subjective and objective factors, experts gave the weight combination coefficient $\alpha = 0.5$, $\beta = 0.5$. According to (11), the evaluation weight of subset G_i under indicator C_1 is

$$\begin{aligned} w_{11} &= 0.3225, \\ w_{21} &= 0.3429, \\ w_{31} &= 0.3316. \end{aligned} \quad (29)$$

(3) According to the Frequency Distribution of the Evaluation Information and the Subset's Evaluation Weight, Evaluation Information and Fairness Were Considered Comprehensively, and the Evaluation Results Were Aggregated under a Particular Evaluation Indicator. From (17), according to the evaluation information, the evaluation score η_1 under the evaluation indicator C_1 is

$$\eta_1 = 0.8666. \quad (30)$$

Cumulative frequency distributions $H_{11}(x)$, $H_{21}(x)$, and $H_{31}(x)$ are constructed.

From (13), the diversity factor between the pairs of three subsets. According to (14), gain the fairness coefficient matrix F_1 of each subset under evaluation indicator C_1

$$F_1 = \begin{pmatrix} 1 & 0.975 & 0.9375 \\ 0.975 & 1 & 0.9625 \\ 0.9375 & 0.9625 & 1 \end{pmatrix}. \quad (31)$$

From (24), project performance under evaluation indicator obtained the fairness score $f_1 = 0.9583$, which indicates that the fairness of evaluation is very good.

Based on the effects of evaluation information and fairness on evaluation results, experts assigned value for weight combination coefficient as $a = 0.8$ and $b = 0.2$. From (25), the comprehensive score of evaluation indicator C_1 "Introduce advanced equipment, technology and management experience" for Beijing Metro Line 4 in operation and maintenance periods can be calculated as

$$V_1 = 0.8849. \quad (32)$$

Similarly, the comprehensive scores of other indicators for Beijing Metro Line 4 in operation and maintenance periods are

$$\begin{aligned} V_1 &= 0.8849, \\ V_2 &= 0.9224, \\ V_3 &= 0.8447, \\ V_4 &= 0.9152, \\ V_5 &= 0.9331, \\ V_6 &= 0.9174, \\ V_7 &= 0.9225, \\ V_8 &= 0.8263, \\ V_9 &= 0.8842, \\ V_{10} &= 0.8877, \\ V_{11} &= 0.9048, \\ V_{12} &= 0.8917, \\ V_{13} &= 0.9132, \\ V_{14} &= 0.8039, \\ V_{15} &= 0.8732. \end{aligned} \quad (33)$$

The indicator weight vector W can be gained using the best worst method from (26), the final evaluation results U was obtained based on the aggregation of evaluation V_j under each evaluation indicator C_j

$$U = 0.8962. \quad (34)$$

The project performance were divided into 5 categories based on the scores, which are {very bad, bad, qualified, good, very good}, corresponding to the scores ranges $\{[0-0.6), [0.6-0.7), [0.7-0.8), [0.8-0.9), [0.9-1]\}$. The project's total evaluation results were obtained, including eight "good" indicators and seven "very good" indicators.

5.3. Results Analysis and Suggestions Based on Operational Performance Evaluation of Beijing Metro Line 4 Project. Through overall model calculation, the Beijing Metro Line 4 project's entire performance in operation and maintenance periods was evaluated as good. Most existing PPP projects rely on financial subsidies as important project income. According to "Administrative Measures for Infrastructure and Public Utilities Franchising" and "Government and Social Capital Cooperation Project Performance Management Operational Guidelines," the project's performance evaluation is closely linked to financial subsidies. The findings of this study provide a scientific, reasonable, and accurate quantitative foundation for PPP projects financing with government payments and feasibility gap subsidies to better implement "pay for performance" incentive mechanism.

Through the evaluation results of each indicator, the project gains high evaluation results in indicators, including "Project's profitability, solvency, and financial sustainability." "The user's appropriate rate, price adjustment, and compensation mechanism ensured the sustainability of the project and the reasonable profit of the social capital" and "Social, environmental, and economic externalities." These results mean that the project has a good overall financial situation and strong sustainability. The project's pricing and compensation mechanism is relatively complete and balances all parties' interests. The project is expected to have a positive impact on the social economy. However, the project performs normally in indicators, including "End-user satisfaction," "Operational management and service quality," and "Effective public participation." Line 4 has always had one of the top three passenger volumes among urban rail transit lines in Beijing and one of the top ten in China. With further increases in passenger volume in recent years, excessive congestion and full load have become critical. As a consequence, current transportation service is not adequate for the requirements of passengers and unexpected breakdowns and delays also lead to a decline in passenger satisfaction. It shows that the project needs further improvement in meeting users' requirements, improving operation efficiency and service quality, and helping the public participate in the performance evaluation. The evaluation results of indicators can be better integrated with the actual project management to enhance the performance improvement and provide benchmark data and experience. The evaluation results can also facilitate maintaining transparency to the public and stakeholders and improving service quality, finding and solving problems, and support project governance continuously from the process management perspective.

For existing problems, two suggestions are proposed: (1) Attach importance to public participation in project performance. It is necessary to collect users' opinions and conduct satisfaction surveys during the project operation and maintenance periods, to reveal the problems affecting the quality of transportation services and the satisfaction of end-users. After that, the problems should be solved timely, and practical feedback should be provided. It is essential to encourage public participation in the project performance

evaluation and to enhance the acceptance toward the evaluation results, from which a virtuous circle of project performance between the project and the user can be formed. (2) Improve operational capabilities and strengthen equipment maintenance. In view of the problem of insufficient transportation volume and overcrowding during the morning and evening peak hours, the carrying capacity during the peak period should be increased through institutional arrangements such as increasing the number of departures, shortening the departure interval, introducing more efficient vehicle equipment, and management systems and other technical means. In usual, it is necessary to strengthen the overhaul of project equipment and signal systems and assign special personnel responsible for particular tasks to avoid delays caused by vehicle and equipment failures.

6. Conclusion

In the context of the standardized and refined development of the PPP model, this research starts focus on the operation performance evaluation of PPP projects and improve the operation performance evaluation-related theory. The following studies have been conducted: (1) Research on the evaluation system of urban rail transit PPP projects. According to the economic and technical characteristics of urban rail transit PPP projects and PPP-related research, a logical model and an indicator system of project performance evaluation based on the performance prism are constructed. (2) Construct a project performance evaluation model based on the best worst method and large-scale group evaluation technology and illustrate with case studies. Comparing the advantages and disadvantages of existing project performance evaluation methods and analyzing the application situation, a project performance evaluation model is constructed using BWM and large-scale group evaluation technology and is verified using a practical example.

Through the research and case study, the main conclusions of this research are as follows. (1) The multi-stakeholder-oriented urban rail transit PPP project performance logic model and the three-party evaluation indicator system can scientifically and comprehensively cover the project performance during the operation and maintenance period to ensure project success. (2) The urban rail transit PPP project performance evaluation model based on the best worst method and large-scale group evaluation technology can meet the performance evaluation circumstance's requirements and ensure scientificity and authenticity of evaluation results.

Although several works have been conducted on the performance evaluation of urban rail PPP projects in this study, some limitations still exist and need to be further improved. (1) The interrelationship between multidimensional indicators was not considered when constructing urban rail PPP projects' operation performance evaluation indicators. There may be duplication, correlation, or causation between indicators. Further empirical research is needed to analyze the mutual influence mechanism between

indicators. (2) Multiple indicators constructed in this study are comprehensive and general indicators, such as satisfaction and external impact. These indicators need to be further refined into measurable constructs for the evaluators to measure. (3) Some of the data in the quantitative evaluation model proposed by this research were obtained through questionnaire surveys. The obtainment and calculation of these data will inevitably lead to subjective errors. In future research, the relevant evaluation index system needs to be further improved. Besides, there are two directions for further researches. (1) Empirical research on the relationship between indicators at various stages of the project based on process management. The research on the key success factors of PPP projects has been relatively mature. However, the relationship between the indicators in each stage of the project still needs further investigation. How to successfully achieve project success through project management is a question that project managers need to answer. (2) The encouragement and improvement of public participation in PPP projects evaluation. For current infrastructure PPP projects, the public participation system for the performance evaluation is still blank in China, which weakens the credibility of project performance evaluation to some extent. How to design a particular system to overcome the dilemma mentioned above and realize effective public participation still needs further research. (3) The construction of BIM-based project performance information management system. An entire project performance evaluation system relies on comprehensive and detailed measurement and accumulation of daily data throughout the project cycle. As a comprehensive platform and visualization tool for project information management, BIM can promote the organic combination of project management and project performance evaluation. It can help project managers and performance appraisers avoid subjective errors and improve the performance evaluation.

Data Availability

The numerical application data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Youth Program of Humanities and Social Sciences Foundation of Ministry of Education of China (21YJCZH169).

References

- [1] C. Cao and J. Chen, "An empirical analysis of the relationship among the service quality, customer satisfaction and loyalty of high speed railway based on structural equation model," *Canadian Social Science*, vol. 7, no. 4, pp. 67–73, 2011.
- [2] N. Aydin, E. Celik, and A. T. Gumus, "A hierarchical customer satisfaction framework for evaluating rail transit systems of Istanbul," *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 61–81, 2015.
- [3] S.-Y. Phang, "Urban rail transit PPPs: survey and risk assessment of recent strategies," *Transport Policy*, vol. 14, no. 3, pp. 214–231, 2007.
- [4] M. Regan, J. Smith, and P. Love, "Infrastructure procurement: learning from private-public partnership experiences 'down under'," *Environment and Planning C: Government and Policy*, vol. 29, no. 2, pp. 363–378, 2011.
- [5] S. P. Ho and C. W. Tsui, "When are public-private partnerships not an appropriate governance structure? Case study evidence," *American Society of Civil Engineers*, no. 373, pp. 817–826, 2010.
- [6] P. E. D. Love, J. Liu, J. Matthews, C.-P. Sing, and J. Smith, "Future proofing PPPs: life-cycle performance measurement and building information modelling," *Automation in Construction*, vol. 56, pp. 26–35, 2015.
- [7] T. M. Koontz and C. W. Thomas, "Measuring the performance of public-private partnerships," *Public Performance and Management Review*, vol. 35, no. 4, pp. 769–786, 2012.
- [8] M. Regan, P. E. D. Love, and J. Jim, "Public infrastructure procurement: a review of adversarial and non-adversarial contracting methods," *Journal of Public Procurement*, vol. 15, no. 4, pp. 405–438, 2015.
- [9] R. Greatbanks and D. Tapp, "The impact of balanced scorecards in a public sector environment," *International Journal of Operations & Production Management*, vol. 27, no. 8, pp. 846–873, 2007.
- [10] O. O. Ugwu and T. C. Haupt, "Key performance indicators and assessment methods for infrastructure sustainability—a South African construction industry perspective," *Building and Environment*, vol. 42, no. 2, pp. 665–680, 2007.
- [11] D. Lawrence, J. Houghton, and A. George, "International comparisons of Australia's infrastructure performance," *Journal of Productivity Analysis*, vol. 8, no. 4, pp. 361–378, 1997.
- [12] J. Liu, P. E. D. Love, J. Smith, M. Regan, and E. Palaneeswaran, "Review of performance measurement: implications for public-private partnerships," *Built Environment Project and Asset Management*, vol. 5, no. 1, pp. 35–51, 2015.
- [13] F. Villalba-Romero and C. Liyanage, "Evaluating success in PPP road projects in Europe: a comparison of performance measurement approaches," *Transportation Research Procedia*, vol. 14, pp. 372–381, 2016.
- [14] C. Henjewe, M. Sun, and P. Fewings, "Critical parameters influencing value for money variations in PFI projects in the healthcare and transport sectors," *Construction Management & Economics*, vol. 29, no. 8, pp. 825–839, 2011.
- [15] J. Liu, P. E. D. Love, J. Smith, M. Regan, and M. Sutrisna, "Public-private partnerships: a review of theory and practice of performance measurement," *International Journal of Productivity and Performance Management*, vol. 63, no. 4, pp. 499–512, 2014.
- [16] H. J. Liu, P. E. D. Love, J. Smith, J. Matthews, and M. C. P. Sing, "Evaluation of public-private partnerships: a life-cycle performance prism for ensuring value for money," *Environment and Planning C: Politics and Space*, vol. 36, no. 6, pp. 1–21, 2018.
- [17] H. J. Liu, P. E. D. Love, J. Smith, Z. Irani, N. Hajli, and M. C. P. Sing, "From design to operations: a process management life-cycle performance measurement system for

- public-private partnerships," *Production Planning & Control*, vol. 29, no. 1, pp. 1–16, 2017.
- [18] Y. Liu, Z.-P. Fan, T.-H. You, and W.-Y. Zhang, "Large group decision-making (LGDM) with the participators from multiple subgroups of stakeholders: a method considering both the collective evaluation and the fairness of the alternative," *Computers & Industrial Engineering*, vol. 122, pp. 262–272, 2018.
 - [19] M. Kagioglou, R. Cooper, and G. Aouad, "Performance management in construction: a conceptual framework," *Construction Management & Economics*, vol. 19, no. 1, pp. 85–95, 2001.
 - [20] J. Yuan, C. Wang, M. J. Skibniewski, and Q. Li, "Developing key performance indicators for public-private partnership projects: questionnaire survey and analysis," *Journal of Management in Engineering*, vol. 28, no. 3, pp. 252–264, 2012.
 - [21] G. Mladenovic, N. Vajdic, B. Wüdsch, and A. Temeljotov-Salaj, "Use of key performance indicators for PPP transport projects to meet stakeholders' performance objectives," *Built Environment Project and Asset Management*, vol. 3, no. 2, pp. 228–249, 2013.
 - [22] J. Yuan, A. Y. Zeng, M. J. Skibniewski, and Q. Li, "Selection of performance objectives and key performance indicators in public-private partnership projects to achieve value for money," *Construction Management & Economics*, vol. 27, no. 3, pp. 253–270, 2009.
 - [23] W. Huang, B. Shuai, Y. Sun, Y. Wang, and E. Antwi, "Using entropy-TOPSIS method to evaluate urban rail transit system operation performance: the China case," *Transportation Research Part A: Policy and Practice*, vol. 111, pp. 292–303, 2018.
 - [24] C. Liu, X. Guo, and R. Men, "Performance evaluation of urban rail transit PPP mode project," in *IEIS 2020*, M. Li, G. Bohács, G. Hua, D. Gong, and X. Shang, Eds., Springer, New York, NY, USA, 2020.
 - [25] Y. Zhou, L. Hou, Y. Yang, H.-Y. Chong, and S. Moon, "A comparative review and framework development on public participation for decision-making in Chinese public projects," *Environmental Impact Assessment Review*, vol. 75, pp. 79–87, 2019.
 - [26] D. M. Connor, "A new ladder of citizen participation," *National Civic Review*, vol. 77, no. 3, pp. 249–257, 1988.
 - [27] R. K. Mitchell, B. R. Agle, and D. J. Wood, "Toward a theory of stakeholder identification and salience: defining the principle of who and what really counts," *Academy of Management Review*, vol. 22, no. 4, pp. 853–886, 1997.
 - [28] W. Xiong, J.-F. Yuan, Q. Li, and M. J. Skibniewski, "Performance objective-based dynamic adjustment model to balance the stakeholders' satisfaction in PPP projects," *Journal of Civil Engineering and Management*, vol. 21, no. 5, pp. 539–547, 2015.
 - [29] J. Liu, P. E. D. Love, P. R. Davis, J. Smith, and M. Regan, "Conceptual framework for the performance measurement of public-private partnerships," *Journal of Infrastructure Systems*, vol. 21, no. 1, pp. 1–15, 2015.
 - [30] A. Neely, C. Adams, and P. Crowe, "The performance prism in practice," *Measuring Business Excellence*, vol. 5, no. 2, pp. 6–13, 2001.
 - [31] A. Neely, J. Mills, K. Platts et al., "Performance measurement system design: developing and testing a process-based approach," *International Journal of Operations & Production Management*, vol. 20, no. 10, pp. 1119–1145, 2000.
 - [32] J. Rezaei, "Best-worst multi-criteria decision-making method," *Omega*, vol. 53, pp. 49–57, 2015.
 - [33] Y. Xiang and S. Mou, "Research review on performance evaluation of PPP projects," *Fiscal Science*, vol. 44, no. 8, pp. 97–106, 2019.
 - [34] J. Rezaei, "Best-worst multi-criteria decision-making method: some properties and a linear model," *Omega*, vol. 64, pp. 126–130, 2016.
 - [35] J. Rezaei, T. Nispeling, J. Sarkis, and L. Tavasszy, "A supplier selection life cycle approach integrating traditional and environmental criteria using the best worst method," *Journal of Cleaner Production*, vol. 135, pp. 577–588, 2016.
 - [36] J. Rezaei, A. Hemmes, and L. Tavasszy, "Multi-criteria decision-making for complex bundling configurations in surface transportation of air freight," *Journal of Air Transport Management*, vol. 61, pp. 95–105, 2017.
 - [37] Y. Liu, Z.-P. Fan, and X. Zhang, "A method for large group decision-making based on evaluation information provided by participators from multiple groups," *Information Fusion*, vol. 29, pp. 132–141, 2016.