

Research on optimization of index system design and its inspection method

Indicator design and expert assessment quality inspection

Yin Kedong, Shiwei Zhou and Tongtong Xu
Ocean University of China, Qingdao, China

Optimization
of index
system design

1

Received 16 October 2019
Revised 19 October 2019
Accepted 19 October 2019

Abstract

Purpose – To construct a scientific and reasonable indicator system, it is necessary to design a set of standardized indicator primary selection and optimization inspection process. The purpose of this paper is to provide theoretical guidance and reference standards for the indicator system design process, laying a solid foundation for the application of the indicator system, by systematically exploring the expert evaluation method to optimize the index system to enhance its credibility and reliability, to improve its resolution and accuracy and reduce its objectivity and randomness.

Design/methodology/approach – The paper is based on system theory and statistics, and it designs the main line of “relevant theoretical analysis – identification of indicators – expert assignment and quality inspection” to achieve the design and optimization of the indicator system. First, the theoretical basis analysis, relevant factor analysis and physical process description are used to clarify the comprehensive evaluation problem and the correlation mechanism. Second, the system structure analysis, hierarchical decomposition and indicator set identification are used to complete the initial establishment of the indicator system. Third, based on expert assignment method, such as Delphi assignments, statistical analysis, *t*-test and non-parametric test are used to complete the expert assignment quality diagnosis of a single index, the reliability and validity test is used to perform single-index assignment correction and consistency test is used for KENDALL coordination coefficient and *F*-test multi-indicator expert assignment quality diagnosis.

Findings – Compared with the traditional index system construction method, the optimization process used in the study standardizes the process of index establishment, reduces subjectivity and randomness, and enhances objectivity and scientificity.

Originality/value – The innovation point and value of the paper are embodied in three aspects. First, the system design process of the combined indicator system, the multi-dimensional index screening and system optimization are carried out to ensure that the index system is scientific, reasonable and comprehensive. Second, the experts' background is comprehensively evaluated. The objectivity and reliability of experts' assignment are analyzed and improved on the basis of traditional methods. Third, aim at the quality of expert assignment, conduct *t*-test, non-parametric test of single index, and multi-optimal test of coordination and importance of multiple indicators, enhance experts the practicality of assignment and ensures the quality of expert assignment.

Keywords Expert assessment quality inspection

Paper type Research paper

1. Introduction

The indicator system is a set of organic wholes with relevant, scientific, dynamic and purpose characteristics. It consists of a series of interrelated, complementary, clear and well-structured statistical indicators. The index system mainly includes the theoretical

© Yin Kedong, Shiwei Zhou and Tongtong Xu. Published in *Marine Economics and Management*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial & non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This research is supported by the National Social Science Fund Major Projects (14ZDB151), Ministry of Education Philosophy and Social Sciences Development Report Cultivation Project (13JBG005).



Marine Economics and
Management
Vol. 2 No. 1, 2019
pp. 1-28
Emerald Publishing Limited
2516-158X

DOI 10.1108/MAEM-10-2019-0010

index system and the evaluation index system. The theoretical index system is a set of complete indicators; it analyzes and evaluates the research objects from the macro level systematically and comprehensively according to the relevant theoretical system. The evaluation index system is based on the actual phenomenon of the researched problem. It analyzes the theoretical index system or related literature; the design is according to the availability of the indicator data, the accuracy of the definition of the index and the feasibility of the operation of the index. Fully consideration of design principles (Table I), using a series of consistency, validity, credibility test methods, set up a set of simple, practical and scientific indicators to optimize through multi-layer screening.

Saisana and Tarantola introduced polymerization systems, multiple linear regression models, principal component analysis and factor analysis (Saisana and Tarantola). Murias *et al.* (2008) used data enveloped analysis method to construct index aggregation and weighting in the index system, which could determine the weights of some indexes internally. Shiau and Liu (2013) used fuzzy cognitive map and analytic hierarchy process (AHP) to construct the causal relationship between key indicators and evaluate sustainable transportation strategies (Shiau and Liu, 2013). Zheng (2017) proposed a new method using fuzzy theory and established an evaluation index system based on the development model of water BOT (Zheng, 2017). Domestic scholars have also made great achievements. Bao Yidan proposed an improved AHP, combined with the orthogonal design ideas, to further improve the objectivity and impartiality of the evaluation process (Yidan *et al.*, 2005). Yu Shunkun proposed a multi-index system for evaluating the performance of power supply enterprises and applied the neural network and entropy method to the correlation analysis of the index system. The results showed that the model and method had high accuracy (Lisha *et al.*, 2011). Liu Weijun combined the coefficient of variation method and GIS spatial analysis to establish a floor water inrush risk assessment model (Weitao *et al.*, 2016). Xi Peiyu *et al.* established an evaluation index system for power purchase schemes, which combined AHP, coefficient of variation method and gray triangle whitening weight function to comprehensively evaluate different power purchase schemes (Peiyu *et al.*).

In summary, domestic and foreign scholars have studied a lot of methods for index system construction. In general, they have used the multi-statistical analysis methods such as DEA method, principal component analysis and AHP to construct the index system and conduct comprehensive evaluation. However, at present, there is no scholar system to propose the optimization and test of the index system design, which is the core of index system design and construction; among them, the expert assignment quality test is one of the most important links of the index system optimization and testing. In the optimization and testing of the indicator system, if there is no systematic expert assessment quality inspection method, no matter how accurate and objective the indicator data are, how innovative and scientific the index screening method is and how forward-looking and comprehensive the index system test is, the final indicator system application will inevitably be worse. In view of this, based on the analysis of the index system construction and testing

Logical-related design principles

Scientific and holistic principles
Principle of compatibility and comprehensiveness
Comprehensive and balanced principle
Potential and forward-looking principles

Data-related design principles

Comparability principle
Principle of operability
Evaluation principle

Trend-related design principles

Trend reflection principle
True objective principle

Table I.
Design principles of
evaluation index
system

methods, this paper systematically studies and proposes the optimization and testing process of the index system design based on the expert evaluation quality test, in order to provide a comprehensive evaluation index system for each field.

2. Relevant theoretical analysis

Theoretical analysis is an important reference system of index system design. It is mainly aimed at revealing the prior theoretical features of the development and change process of the studied object, analyzing relevant influencing factors and describing the internal structure of the studied object and the basic principles and physical processes followed by the dynamic evolution process.

The role of theoretical analysis is mainly reflected in the following aspects (Haibo, 2004):

- (1) Analyzing the statistical results according to the theory. The statistical analysis of the survey data needs to be further explained through theoretical analysis in order to further reveal the essential phenomena of the problem.
- (2) Verifying and demonstrating the basic assumptions. When the statistical analysis results are inconsistent with the research hypothesis, it needs to be explained by theoretical analysis. Even if the statistical analysis results are consistent with the research hypothesis, it is also necessary to combine the specific theoretical analysis to demonstrate the correctness of the hypothesis from different angles.
- (3) Theoretical abstraction and sublimation of practical experience. Some typical case analysis data are limited, and only by means of relevant theoretical analysis, problems can be found from specific phenomena, and the essential characteristics of the research problems can be grasped through representation.

2.1 Theoretical basis analysis

Theoretical basis analysis is a scientific analysis method put forward by Li QingZhen in 1999. Compared with the empirical analysis method, it is a method to explore and study the nature of problems and their development rules by using rational thinking on the basis of perceptual knowledge (Qingzhen, 1999). In 2011, HeYun further defined theoretical analysis as a series of systematic analysis on the basic principles, related concepts, connotation extension, attribute characteristics, classification characteristics, basic theories and representation of practical problems involved in the research object (Yun, 2011).

Through theoretical analysis, the research object can be further explored in a systematic, scientific, logical, relevant, general and frontier way, and the process, rules and mechanism of the occurrence, development and evolution of the research object can be distinguished and analyzed, and its future development trend can be explored and predicted. However, due to the timeliness, particularity, limitation and complexity of practical conditions, there are still some differences between relevant theoretical analysis and research objects, objective facts and constraints. Therefore, theoretical analysis must be closely related to reality, must deeply analyze the attributes of the research object, must master sufficient data, must ensure the objectivity, accuracy and scientificity of the analysis conclusion and must be able to pass the test of relevant theory and practice.

Theoretical analysis is *a priori* or empirical indirect research method, which is the basis of analysis of many practical problems. After failing to fully know, understand and grasp the internal mechanism and the essential characteristics of the actual research object, the theoretical analysis provides researchers with an empirical research train of thought, reveals the common property of this kind of research object and also provides researchers with a relatively efficient analysis method, on the basis of in-depth and creative research. When the theoretical analysis system becomes more and more abundant and perfect,

researchers can make full use of the previous research literature and existing materials to get relevant research conclusions quickly through theoretical analysis, which is more efficient than empirical analysis.

2.2 Correlation factor analysis

Factor analysis is a multivariate analysis method first proposed by C. Spearman in 1904 in the field of psychology. In addition to being influenced by potential factors, a certain research object is more influenced by external factors. If there is no external condition constraint, the research can only be conducted according to the observed data (Spearman, 1923). Spearman's two-factor theory of intelligence holds that intelligence is a combination of general factors and special factors. Later, Thurston *et al.* proposed "thurston's intelligence group factor theory" in 1938 (Xiting *et al.*, 2003). In 1961, Vernon divided intelligence into different levels of factor branches through factor analysis and established the factor hierarchy structure (Mingyuan, 1998). Carroll (1993) proposed a three-level model of cognitive ability by integrating some viewpoints of two-factor theory, multi-factor theory and information processing theory. In 2012, Yao Ligen *et al.* defined factor analysis method as a qualitative analysis method, which mainly analyzed various factors affecting the research object based on the empirical knowledge of researchers, and then carried out the analysis (Ligen and Xuewen, 2012).

Factor analysis is an empirical analysis method. In fact, it analyzes the correlation factors of the research objects and judges the relationship among various factors. The correlation between variables is mathematically described by the correlation coefficient. The correlation coefficient is taken as $[-1, 1]$, -1 is the complete negative correlation between the two variables and $+1$ is the complete positive correlation between the two variables. The larger the absolute value of the correlation coefficient, the stronger is the correlation.

The factor analysis method can carry out the internal and external features of the research object, such as roughing and refining, de-authentication, and thus processing from the table and the inside, and then excavating the inherent essential relationship of objective things. However, the factor analysis method can simplify the research problem, ensure a certain amount of information and greatly improve the efficiency of research.

2.3 Description of physical process

Physical description was proposed by Meigang Zhang and others in 2008, which is mainly a systematic analysis and description of the state of things, including the attributes, characteristics and corresponding terminology of things (Meigang *et al.*, 2008). It mainly includes general description, description of object and substance and description of concept. It mainly explains and describes the meaning of relevant terms and concepts in detail. It can be compared and described with other research objects, and at the same time, keywords can be appropriately used to simplify the description process or brief explanation can be made through cases.

Process expression was put forward by Lan Wang and others in 2004. It means to express the theoretical knowledge and research method of the studied problem into a solution process. A series of in-depth studies can be carried out smoothly only when the analysis process of the problem is fully mastered (Wei and Yujun, 2004). Process expression includes some objective laws, development changes, correlation relations and research progress of the research object. All research information is included in the process expression. Process presentation clearly represents the problem background, theoretical knowledge, solution method, etc., of the research object as a series of research processes, which is a process and focuses on the description of dynamic process. However, some other description methods mainly focus on the static description of the research object.

Description of relationship mechanism (analysis of mechanism) is based on the analysis of the system's internal reasons (mechanism). It is a scientific research method that studies the intrinsic working mode of each element in a certain system structure and the operating rules and principles of the interconnection and interaction of various elements under certain environmental conditions, so as to find out the law of its development and change.

3. Identification of complete set of indicators

3.1 *Analysis of system structure*

The concept of system exists in all things in human society and nature, and system thought runs through the whole practice process of human society. A system is an organic whole composed of two or more organic elements with a certain structure and specific functions. System structure analysis (Yun, 2011) is mainly used to identify the components of the system elements of the research object and the relationship between the elements and the system, so as to provide analysis basis for revealing the hierarchical structure among the elements, the hierarchical division of the elements and the causal relationship among the elements.

First, system method starts from wholeness; it comprehensively inspects the relationship between system and elements, elements and elements, system and environment as well as its rule of movement, so as to achieve the purpose of optimizing the overall function of the system. It was first proposed by Bertalanffy in 1937.

The main principles of the general system theory analysis are as follows: integrity, connectedness, order, structure, hierarchy, dynamic, environmental adaptability, optimization and so on. Integrity is the first basic principle, which requires that the understanding of the system must start from the integrity, thus achieving the integrity of the understanding; connectedness emphasizes the organic connection between each element and the system, and emphasizes the structural and integral integrity of the system. The structure of the system is hierarchical, and the hierarchy of the structure shows the high order of the system. Any objective system is an open and evolving one. In order to develop from low order to high order, the system in dynamic change must adapt to the environment and constantly optimize its structure and function.

The general system theory analysis method must not only raise the problem comprehensively and systematically, but it must also accurately explain the problem and establish the system goal. At the same time, it must formulate the plan and establish and verify the relevant mathematical model, evaluate and select the optimal solution to solve the problem. Making decisions and implementing plans require constant review, feedback and revision from practice.

Second, method of cybernetics is a scientific method for applying control theory to research, identify and solve system control problems. It mainly analyzes the interaction between system and environment, elements and elements, and elements and systems (positive and negative feedback relationship), and it reveals and describes the behavioral characteristics of the system. It was first proposed by Norbert Wiener in 1948. At present, a method group including a feedback method, a control method, a function simulation method, a black box method, a white box method, a gray box method and a system identification method has been formed.

Third, the method of information is a modern scientific research method that abstracts the motion change process of the object system into the information transformation process and then recognizes the law of the system motion according to the information characteristics. It was first proposed by Shannon in 1948.

Fourth, methods of synergetics are scientific methods for studying how systems spontaneously produce ordered structures and study the common laws of how elements (or subsystems) in various types of systems can produce overall effects through synergy. It was founded by the famous German theoretical physicist H. Haken in the 1960s.

Haken discovered that the system evolved from disorder to order. Non-equilibrium phase transitions and equilibrium phase transitions have their commonalities. They are the coordinated results of each element (or subsystem) through nonlinear interaction and coherence effects, which can be processed by the same mathematical model.

Fifth, the method of catastrophe theory is a scientific method for studying the non-continuous qualitative transformation process and its laws by using mathematical tools such as topology, singularity theory and structural stability. It was first proposed by the French Mathematician R. Thom in 1972. Qualitative change is a common phenomenon in the objective world. The combination of catastrophe theory and synergistic system methods is conducive to the formation, structure and development of the system of in-depth study and reveals the system's transformation mechanism from disorder to order.

3.2 Hierarchical decomposition of the system

The social and economic complex system generally has the characteristics of deep structure, higher system order, strong element correlation and more object conflicts. "Decomposition-coordination" method is an effective approach to the study of complex large system. "Decomposition" is to decompose a complex system into simple subsystems (or elements) and solve them separately. Due to the complex association between the elements, the solution of the subsystem is not a system solution and even has a conflict. "Coordination" coordinates the consistency and compatibility of subsystem solutions, and it is achieved through the overall goals of the system and associated constraints. The "decomposition-coordination" of complex systems can make the system present a hierarchical multi-level structure, which is a hierarchical structure. The results of systematic hierarchical decomposition provide an important reference for hierarchical structure and classification of index system.

3.2.1 *System hierarchical structure.* Messervik proposed three types of hierarchical structures (Yonghua, 1997):

- (1) Multiple hierarchical structure, mainly focusing on multi-level structure division of functions of natural attributes of complex systems. The function of the system needs to start from the specific principle corresponding to various hierarchies and multiple descriptions of the corresponding features of the system, so as to fully analyze and describe the overall function of the system. The multiple hierarchies of system functions have their own different rules and principles, inputs and outputs, and finally the multiple hierarchies are coupled.
- (2) Multi-level hierarchical structure, mainly focusing on the hierarchical division of the control, management, decision-making and other objectives of complex systems. The stratification of goals such as control, management and decision-making focuses on the stratification of the concept of "what to do" rather than the stratification of "task details" and "how to." The higher the target level, the greater is the strategic significance and the more complex is the decision; the lower the level, the more frequent is the decision.
- (3) Multi-level hierarchical structure, mainly focusing on the hierarchical decomposition of the organizational composition of complex systems, including horizontal decomposition and vertical decomposition. The multi-level hierarchical structure is a pyramid-shaped mesh structure or a tree-shaped mesh structure. The lower level structural units are coordinated by local controllers, and the higher level structural units are coordinated by associations and constraints to achieve global optimization. The descriptions and problems of higher level structural units are less structural, have greater uncertainty and are generally more complex than lower level units.

3.2.2 Decomposition and coordination of hierarchical structure. Decomposition of hierarchical structure: in general, the total objective function of complex system optimization is the objective function that can be decomposed into multiple subsystems, and the subsystem is the local optimization problem when the coordination parameters are relatively fixed. It is necessary to determine the inter-relationship between subsystems and their local objective function and constraints.

Coordination of hierarchical structure: the optimal result of the subsystem is closely related to its coordination parameters. Therefore, it is necessary to verify or improve the coordination parameters according to the iterative coordination rules to make the global optimization.

3.2.3 Conceptual model design. The conceptual model is a description of real-world problems or things (Environmental Engineering Evaluation Center of the Ministry of Environmental Protection, 2016), which is the first layer of abstraction from the real world to the information world. The conceptual model must have four characteristics: first, rich in semantic expression, able to express various needs of users, fully reflect the real world, including the connection between things and the user's processing requirements for data; second, easy to communicate and understand, expressed naturally, intuitive and easy to understand, in order to exchange ideas with users who are not familiar with the computer; third, easy to modify and expand, can be flexibly changed to reflect changes in user needs and real-world environment; fourth, easy to convert to various data models and can be easily converted to various data models such as relational models and hierarchical models.

3.3 Identification of the complete set of indicators

An indicator is a concept that describes the quantity or quality characteristics of the population. It is the unit or method of measuring the target. It represents the expected index, specification, target or standard, and it is generally expressed by data (quantitative data and qualitative data). The index is generally composed of two parts: the name of the indicator and the value of the indicator (data). It embodies the characteristics of the prescribing nature of the substance and the prescriptiveness of the quantity.

Indicators and signs are different and cannot be mixed. First, the indicators indicate the overall characteristics, reflecting the overall quantity or quality characteristics, and must use the quality or quantity data (quantitative, qualitative, etc., quantitative, qualitative description) to answer questions, cannot answer questions with questions, the indicator data is after certain The summary obtained, a complete indicator should have the time, place, scope, data and other conditions. Second, the logo indicates the characteristics of the overall unit, which reflects both the characteristics of the overall unit quantity and the quality characteristics of the overall unit. Only the quantity mark uses the number to answer the question, and the quality mark uses the word to answer the question. The quantity mark in the mark may not be aggregated or obtained directly. The logo generally does not have conditions such as time and place.

The identification of the complete set of indicators is mainly for the multi-dimensional analysis and analysis of factors in the process of "theoretical analysis, factor analysis, process analysis, structural analysis, hierarchical decomposition" and so on, to clarify the meaning of indicators, the role of indicators, the attributes of indicators, indicator data, calculation methods, measurement units, as well as indicator type, indicator relationship and other characteristics, build indicator collection warehouse or index set dictionary and provide a basis for screening, testing and optimization of indicator system.

Indicator attributes include subjectivity and objectivity indicators, randomness, certainty indicators, discreteness, continuity indicators, quantitative and qualitative

indicators, and positive, negative and neutral indicators. Indicator types include the following: semantic indicators, statistical indicators, calculation indicators, comprehensive indicators, end indicators, time point indicators, period indicators, physical indicators, value indicators, descriptive indicators, evaluation indicators, early warning indicators, technical indicators, quantitative indicators, qualitative indicators.

Quantitative indicators are mainly a quantitative result, statistical data, such as total indicators, relative indicators, average indicators, incremental indicators and growth indicators. Qualitative indicators, usually unstructured, empirical, revealing and difficult to categorize, are generally subjective, descriptive and semantic:

- (1) Total indicator: it reflects the size, quantity and total difference of a research object, which is the basic statistical indicator. The size of this indicator is directly affected by the overall size, reflecting the problem with a certain one-sidedness.
- (2) Relative indicators: also known as “relative numbers,” the ratio of two related indicator values reflects the degree of development, structure, intensity, prevalence or proportionality of the research object. Common relative indicators are structural relative number (specific gravity), proportional relative number, comparative relative number, intensity relative number, dynamic relative number and elastic coefficient.
- (3) Average indicators: analytical indicators reflecting the general level of the research object, divided into static average and dynamic average. The static average is the general horizontal state at a certain time, and the dynamic average is the horizontal state at different points in a period of time.
- (4) Variability indicators: indicators that reflect the extent of the overall difference in the study subjects. Common ones are extreme difference, mean difference, standard deviation, coefficient of variation, etc.

3.4 Indicator system primary selection

The construction of the indicator system is generally divided into three steps: initial construction of indicator system, screening of indicator system and optimization of indicator system structure. The initial construction methods of the indicator system include qualitative methods and quantitative methods. At present, most of the practical applications use qualitative methods to select indicators.

For the initial construction of the indicator system, first, it is necessary to clarify the evaluation object and the purpose of the evaluation. The evaluation object and the evaluation purpose directly determine the selection of the indicator system and the evaluation method. Second, it is necessary to determine the primary selection method of the indicator system (such as the system analysis method, Delphi method), determine the “top-down” or “bottom-up” index system construction sequence, obtain the evaluation index set, determine the mutual index and structural relationships, such as using the target hierarchy to evaluate the level and using factor decomposition structure (DuPont indicator system) to analyze the factors. Third, it is necessary to determine the connotation of indicators, the role of indicators, data sources, calculation methods and measurement units.

The initial construction methods of the index system in the existing literature are shown in Table II. The traditional methods have their own characteristics and different emphasis. They are subjective and random, and lack the basis of the normative method. It is difficult to adapt to the research requirements of the rapidly developing indicator system.

The primary selection method of the index system in this paper follows a series of standardization and scientific processes from “the theoretical basis analysis, relevant factor analysis, process mechanism description” to “system structure analysis, system hierarchical

Method	Principle	Application field	Advantages	Disadvantages
The comprehensive method	Refer to the existing indicators, comprehensively analyze and summarize to form new indicators, and classify the more mature socioeconomic evaluation index system.	More mature socioeconomic evaluation index system	Drawing on research experience and results, avoiding subjective randomness and integrating multiple perspectives	Indicators are old or even misleading and difficult to reflect new changes
Analysis	Subdivide research objects and measurement targets, partition subsystems and their functional modules	Sustainable development evaluation index system, economic benefit evaluation index system	Systematic analysis of different feature attributes of research objects	Influenced by knowledge structure, research level, familiarity, etc., it has certain subjectivity
Target level	According to the research objectives, build the target layer and the criteria layer to form an indicator system.	Comprehensive evaluation system for planning scheme	Easy to understand, simple and practical, the target structure can reduce the crossover of indicators	Subjective and subjective categorization
Cross method	Through multi-dimensional cross-analysis, a series of indicators are derived to form an indicator system	Economic benefit evaluation, evaluation index system of coordinated development of social economy and science	Can reflect the contrast or coordination between multiple elements	Cross-analysis is difficult to control and has a limited range of applications
Index attribute grouping	Determine the indicator composition based on the indicator properties (dynamic, static, absolute, relative)	Unemployment status evaluation index system, etc.	The indicator attributes are clear and considered more comprehensive	Indicators are easy to repeat, single indicator type

Table II.
Initial construction
method of the
indicator system

decomposition, indicator complete set identification.” On the basis of comprehensive analysis and reference to traditional methods, the system engineering and hierarchical structure decomposition theory are used to optimize the standardization process and analysis method of the primary selection of index system, which reduces subjectivity and arbitrariness, enhances objectivity and scientificity, and agrees with the normative and transparent nature of the primary selection method.

4. Indicator expert assignment

In the design of the indicator system, there are three tasks that require consulting experts or an expert seminar to confirm the relevant content. First, through the primaries, the indicators are fully concentrated, the reliability and feasibility, importance and necessity of each indicator, validity and rationality, particularity and forward-looking characteristics require expert evaluation scores in relevant fields to classify each indicator. Second, there are no quantitative indicators and qualitative indicators of indicator attribute values. Experts in related fields need to be assigned scoring. Third, the weighting of each indicator in the indicator system requires expert assignment scoring in the relevant field.

4.1 Expert background analysis

Expert assignment or scoring is an important task in the process of index design. Whether it is the attribute value score of qualitative indicators or the importance and rationality of indicators, the determination of index weights, etc., it is necessary to consult experts in relevant fields and pass expert seminars. In other forms, the relevant experts of the organization assign or score the different attribute requirements of the indicator. However, due to the different academic background, influence and judgment basis of the experts, it is necessary to uniformly measure the representativeness and authority of the expert assignment or scoring value. Therefore, it is necessary to fully understand the academic background and research fields of relevant experts.

The representativeness or authority of expert opinions is generally determined by three main factors: first, the educational background, age and professional title of the experts; second, the academic level, influence and credibility of the experts; third, the practical experience and problems of the experts, and the level of familiarity and the basis for judging. The degree of representation or authority of expert opinions is mainly self-evaluation.

The educational background, age and title of the experts themselves: the educational background of the experts themselves mainly includes the basic situation of domestic and foreign educational experience, age, administrative duties, professional technical titles and academic tutor level.

The academic level, influence and credibility of the experts themselves: the academic level of the experts themselves mainly include the number of academic articles, the number and level of monographs and their citations, the number and level of relevant scientific research topics, the number and level of awards for major scientific research achievements, the number of academicians and other talents in related fields, the level of academic group organization in related fields, recognition and popularity of peer experts.

The expert's practical experience and familiarity with the problem mainly include the time of research in the professional field, the nature and level of the work unit of the expert, the number of patents applied in related fields, the number of scientific research projects, the social and economic benefits, and review expert level, etc.

The attribute value range or scoring interval of the three main factors of the expert is [0,100] points. The attribute assignment or scoring of each major factor is formed by the project organizer based on the relevant information of each expert and soliciting the opinions of other relevant experts.

The relative importance of the three main factors of the degree of authority of expert opinions can be determined by a pairwise comparison method. Among the three main factors, "the academic level and influence of the experts themselves and the credibility" are the most important and can be assigned 100 points. Then, "the expert's practical experience and familiarity with the problem" can be assigned 85 points. The importance of background, age and title for the experts themselves can be assigned 70 points. According to a large number of practical comparisons, consultations and empirical judgments, the weight coefficients of the three factors are 0.40–0.45, 0.30–0.35 and 0.25–0.30.

The degree of authority of i expert opinion is expressed by E_i , which can be calculated by weighted average. The larger the weighted average value, the higher is the authority of the expert opinion (Table III).

4.2 Indicator attribute value domain

The indicator attribute range is the value range of the expert's evaluation of the importance, rationality and forward-looking attributes of the indicator. It is often determined according to the meaning and role of the attribute of the indicator, and it is also an important reference for the expert assignment score. Indicators have different attribute types, such as

quantifiable indicators, non-quantitative indicators, positive indicators, negative indicators, neutral indicators, etc. The attribute values or data types of its indicators can be appropriately adjusted and transformed accordingly.

Non-quantitative indicator attribute value ranges can take many forms, and other indicators that lack data can determine the value range based on their indicator attributes:

- (1) Quantity formula: 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 0 or 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.
- (2) Interval: 90–100, 80–90, 70–80, 60–70, 50–60, 40–50, 30–40, 20–30, 10–20, 0–10.
- (3) Quantifiers: excellent, good, medium, qualified, poor, very poor or good, good, average, poor, very poor.
- (4) Level: special, first, second, third, fourth, and fifth (level scale can also be represented by different numbers).
- (5) Definition formula: A, B, C, D, E.

4.3 Expert assignment of indexes

The expert assignment or score of the index is mainly based on the importance of the index, the attribute value of the qualitative index and the weight of the index. In essence, it is a form of consultation, correspondence, discussion, research, questionnaire, discussion, etc., to organize experts and scholars in related fields, respectively, to conduct diagnostic evaluation and score on relevant indicators, mainly including direct evaluation, pairwise comparison, fuzzy evaluation and interval evaluation.

The assignment score of experts can take the independent assignment score of each index in the same level index under a certain kind of research (evaluation) object, and it can also take the pairwise comparison assignment score of the importance of the same level index under a certain kind of research (evaluation) object. Common expert assignment scoring methods are detailed next.

4.3.1 Delphi assignment method. Delphi assignment method is essentially a feedback anonymous consultation method, which is a collective anonymous thought exchange process in the form of correspondence consultation. It was first pioneered by Olaf Helmer and Norman Dalkey in the 1940s. In 1946, RAND Corporation first used this method to conduct qualitative rating and prediction in order to avoid the defects of subjection to authority or blind obedience to the majority in group discussion. Later, this method was quickly and widely adopted, and enjoyed high credibility in evaluation, decision-making and planning (Environmental Engineering Evaluation Center of the Ministry of Environmental Protection, 2016). By improving Delphi's process, Vakil *et al.* (2006) adopted the method of repeated iterative voting to reach consensus and improve the transparency and reliability of decision-making (Yuxiang and Donghua, 1990). Navdeep Kaur *et al.* proposed a multi-round progressive improvement of Delphi method to enhance the universality of analysis (Vakil *et al.*, 2006).

Factor	Expert's own educational background, age, title (X_1)	The academic level and influence of the experts themselves and their credibility (X_2)	Expert experience and familiarity with the problem (X_3)	The degree of authority of expert opinions (E_i)
Weight (value range)	0.25–0.30 (0–100)	0.40–0.45 (0–100)	0.30–0.35 (0–100)	1.00 (0–100)

Table III.
Authoritative opinion degree and its factor weight distribution table

Wang Chengbin (1991) used Bayesian estimation to improve the Delphi method to enhance the prediction accuracy (Kaur and Pluye, 2019). Lian Jinyi (1992) quantified expert opinions to improve Delphi method (Chengbin, 1991). Cai Hui *et al.* (1995), based on the credibility of expert opinions and the overall evaluation of experts, improved the coordination degree of expert opinions (Jinyi, 1992). Zong Jinfeng *et al.* introduced the idea of set value statistics to convert the evaluation results that are difficult or impossible to be quantitative into interval valuation, thus improving the Delphi method (Hui *et al.*, 1995). Daowu *et al.* (2004) adjusted the improved decision-making procedure of the Delphi method based on the evolutionary game, so as to reach a consensus group game example (Jinfeng and Shuqin, 1997). Guan Chun *et al.* modified the Delphi method based on the weighted average calculation model of BP neural network, so as to reduce the influence of subjective factors on weight distribution (Daowu *et al.*, 2004). Yuan Jixue *et al.* improved Delphi method based on consistency analysis and listed the personnel composition and working procedure diagram of group decision-making (Chun and Jun, 2006). Zhang Qun *et al.* (2010) introduced Vague set and its similarity measurement formula and established the Delphi method process based on Vague set theory, so as to make more use of expert opinions (Jixue and Weijin, 2010). Yuan Qinjian *et al.* used Citespace to analyze the development and application of Delphi method in China and found that “index system” and “comprehensive evaluation” were the first application fields (Qun *et al.*, 2010).

Delphi assignment method generally requires 8–20 experts and has obvious decision-making advantages:

- (1) The authority of experts: it makes full use of the experience, background and expertise of famous experts.
- (2) Anonymity of experts: experts are anonymous and back-to-back, and each expert is free to make his or her own judgment.
- (3) Convergence of assignments: the opinions of experts are summarized and fed back, and the opinions of experts are adjusted and modified.
- (4) Statistical quantification: one median and two quartiles of all expert assignments are counted, 50 percent of which are in the two quartiles and 50 percent are outside the two quartiles.

4.3.2 Likert score. In 1932, Likert scale was proposed on the basis of the improvement of the original total addition scale. Likert scale is composed of a set of problem statements and their expert ratings, indicating the strength or state of the expert’s attitude (Qinwei *et al.*, 2011). Schuman and Presser (1981) used words such as “uncertain” or “unclear” to set up no response options and improved Likert scale to make the separation test of information more accurate (Chonghua and Wenfu, 2013). Bai Kai (2011) added “no response” option in Likert scale to improve the integrity of data information (Schuman and Presser, 1981). Han Guanghua *et al.* studied the influence of different semantic expressions on scientific measurement in Likert scale and used numerical analysis and mathematical reasoning methods to determine the range of semantic differences with reliability (Kai, 2011).

Common Likert score include four-stage grading method, five-stage grading method, six-stage grading method, seven-stage grading method and sometimes ten-stage grading method. The determination of grade is often more arbitrary, and the recognition of the advantages and disadvantages of various grades is based more on perceptual experience (Table IV).

4.3.3 Fuzzy semantic scale scoring method. Fuzzy semantic scale score was produced after American mathematician L.A. Zadeh proposed “fuzzy mathematics” in 1965. In real life, in answer to a question or judgment, there are often “is also the essential” in the middle

Grade	Content of grading options
Four-stage grading method	1 – not important; 2 – generally important; 3 – important; 4 – very important
Five-stage grading method	1 – very unimportant; 2 – not important; 3 – generally important; 4 – relatively important; 5 – very important
Six-stage grading method	1 – very unimportant; 2 – not important; 3 – not too important; 4 – important; 5 – relatively important; 6 – very important
Seven-stage grading method	1 – very unimportant; 2 – not important; 3 – not too important; 4 – generally important; 5 – important; 6 – relatively important; 7 – very important
Ten-stage grading method	Use any number from 1 to 10 to indicate importance. A score of 10 out of 10 indicates very important

Table IV.
Likert score method

of the state, people use natural language description of “very clear, clear, not too clear, not clear, very not clear” (very good, good, relatively good, average, poor, very poor), as well as likert rating, etc., with strong fuzziness, expert assignment or score results are often must rely on subjective judgment in fuzzy several options to determine an option.

It is more objective to assign fuzzy language with fuzzy semantic scale (Guanghua and Bo, 2017). The fuzzy semantic scale not only considers the general fuzziness of language, but also synthesizes the subjective difference of expert cognition, so the score is more accurate.

It is assumed that there are k states for the answer or evaluation of a certain question X , and the corresponding score of each state is X_1, X_2, \dots, X_k . According to his/her own psychological feelings or cognitive level, expert i needs to fill in his/her perception level for each state j , which is expressed as percentage A_{ij} . The score quantification value of expert’s fuzzy semantic scale is expressed as follows:

$$F_i = \sum_{j=1}^k (k+1-j) \times A_{ij} = \sum_{j=1}^k X_j \times A_{ij},$$

where the corresponding score of state j is $X_j = k + 1 - j$.

For example, there are five states, “very satisfied, somewhat satisfied, satisfied, dissatisfied, very dissatisfied” on a scale of 0–5 (or 0–10), concerning the question “are you satisfied with your ability to do your job?”. According to the degree of perception, respondents (or experts) filled in the corresponding degree of perception (percentage) for each of the five states. The fuzzy semantic scale of this respondent is shown in Table V.

The score quantification value is $F = 5 \times 5\% + 4 \times 80\% + 3 \times 5\% + 2 \times 10\% = 3.62$.

This score indicates that the expert’s true perception is between “satisfied” and “slightly satisfied,” rather than “satisfied” with the traditional assignment rating options. Therefore, the fuzzy semantic scale scoring method is more realistic and accurate.

4.3.4 Pairwise comparison assignment. Based on the analysis of the background information of the problem, each expert analyzed and sorted the target indicators, gave 100 points for the best indicator and 0 points for the worst indicator, and then compared other indicators with these two indicators and gave corresponding assignment or score.

Fuzzy semantic scale score	Very satisfied	Satisfied	Slightly satisfied	Dissatisfied	Very dissatisfied
State score	5	4	3	2	1
Percentile of perception (%)	5	80	5	10	0

Table V.
Expert fuzzy
semantic scale

It is also possible to assign values according to the pairwise comparison method of 1–9 scales in the AHP method. The scale values of pairwise comparison and quantification are shown in Table VI.

5. Expert assignment quality inspection of individual indicators

Through expert assignment scoring, each indicator has an expert assignment set, and all sets can form an expert assignment matrix. The quality evaluation of the expert evaluation of a single indicator is mainly for the validity of the evaluation of the expert, the consistency of the assignment of multiple experts, etc., and the rationality of the design of the indicator can also be tested.

The quality of assignment refers to the degree to which the consistency, correctness and integrity of the expert assignment are satisfied in the information system. The quality test of assignment is mainly to test the reliability of the expert assignment and to check the abnormal points in the expert assignment in order to exclude or correct the error assignment. At the same time, it is also possible to judge whether there is a difference in the understanding of the indicator by the quality test result of the expert assignment and to test the rationality of the indicator design.

Suppose n represents the number of experts, k for the number of indicators, and x_{ij} for the value of the i th expert assignment to the j th indicator. $x_j(\max)$ represents the maximum value in the j th indicator value(data) range.

5.1 Statistical analysis of expert assignments with single indicators

The statistical analysis of expert assignment of individual indicators mainly includes three aspects: expert participation, concentration of expert opinions and coordination of expert opinions. Among them, the expert participation degree is measured by the positive coefficient of the expert assignment; the expert opinion concentration degree is measured by the average number, the perfect rate and the dispersion degree of the expert assignment; and the coordination degree of the expert opinion is measured by the coefficient of variation and the coordination coefficient.

5.1.1 Expert participation analysis. The enthusiasm or participation of experts is reflected by the positive coefficient P_j . P_j is used to reflect the level of concern, participation, and cooperation of experts on research issues, which can be measured by the ratio of the number of experts participating in the evaluation to the total number of experts generally. This ratio can be measured by the recovery rate and efficiency of the questionnaire. In general, $P_j \geq 50\%$ is the lowest ratio that can be used for analysis, and $P_j \geq 80\%$ indicates that the expert’s participation is relatively high.

5.1.2 Analysis of the concentration of expert opinions. The importance of the indicator can be expressed by the average value \bar{x}_j and the full rate K_j of the j th indicator by the m_j th experts. The degree of concentration of expert opinions indicates the degree of consistency of experts on the assignment of indicators, which can be expressed by the dispersion of the assignment of the j indicator by m_j experts.

Table VI.
Pairwise comparison of 1–9 quantitative scales of importance of indicators

Scale	1	3	5	7	9	2, 4, 6, 8
Index i relative to index j	Equally important	Slightly important	Obviously important	Strong important	Extreme important	Somewhere in the middle of the previous number

The average of expert assignment is given as follows:

$$\bar{x}_j = \frac{1}{m_j} \sum_{i=1}^{m_j} x_{ij}, \quad (1)$$

where m_j represents the total number of experts assigned to the j th indicator. The closer the ratio of $\bar{x}_j/x_{j(\max)}$ is to 1, the more important the experts think the j th indicator is.

The full rate of expert assignment is given as follows:

$$K_j = \frac{\tilde{m}_j}{m_j}, \quad (2)$$

where m_j represents the total number of experts assigned to the j th indicator and \tilde{m}_j represents the number of experts who have a full score for the j th indicator. The value of \tilde{K}_j ranges from 0 to 1, and the full rate K_j can be used as a supplementary indicator of average \bar{x}_j . The larger the K_j , the larger is the proportion of experts who give the full value to the indicator and the more important is the indicator.

The dispersion of expert assignment is given as follows:

$$\delta_j^2 = \frac{1}{m_j} \sum_{i=1}^{m_j} (x_{ij} - \bar{x}_j)^2. \quad (3)$$

The dispersion δ_j^2 of the expert assignment is the variance of the expert assignment, indicating the degree of dispersion of the expert assignment to the j th indicator and the concentration of the expert assignment from another aspect. The greater the dispersion δ_j^2 , more inconsistent is the expert's awareness of the importance of the j th indicator. The smaller the dispersion δ_j^2 , more consistent is the experts' cognition of the importance of the j th indicator and higher is the concentration of expert opinions.

5.1.3 Analysis of coordination of expert opinions. The degree of coordination of expert opinions is based on the coefficient of variation (V_j). The coefficient of variation describes the extent to which the m_j th experts coordinate the assignment of the j th indicator. By calculating the coefficient of variation, it can be judged whether there is a big disagreement between the experts on the assignment (score) of a single indicator. The specific calculation method is as follows:

$$V_j = \frac{\delta_j}{\bar{x}_j}, \quad (4)$$

where V_j represents the coefficient of variation of the m_j th experts assigned to the j th indicator; δ_j represents the expert assignment's standard deviation of the j th indicator; and \bar{x}_j represents the average assignment of the expert to the j th indicator. The coefficient of variation indicates the degree of fluctuation of the expert assignment to the j th indicator. The smaller the V_j , the higher is the consistency of the expert assignment to the j th indicator and the higher is the degree of coordination of the expert assignment.

5.2 Single-index expert assignment t-test

In general, there are two types of ideas for quality inspection methods. One is the overall impact analysis. The basic idea is to delete one or more assignment points in the assignment set of an indicator and then examine the impact on the statistical inference

after deleting the assignment point. The other type is local impact analysis. The basic idea is to make a small perturbation of some assignment points in the assignment set. Then, the influence of the assignment point perturbation on statistical inference is examined.

The quality test of the assignment of a single indicator expert is based on the overall impact analysis, testing the consistency, validity and reliability of the assignment of multiple experts, judging whether the experts have a large disagreement on the score of each indicator and checking and analyzing experts with abnormal assignments. When the expert assignment distribution obeys the normal distribution, the commonly used statistics are used to test.

5.2.1 Consistency test of expert assignment. Whether all expert assignments are consistent can be tested with t -statistic, that is t -test. The inspection steps are as follows:

- (1) Randomly dividing all expert assignments of an indicator into two groups (or dividing into two groups according to the size of the assignment).
- (2) Establishing assumptions:
 - The null hypothesis H_0 . All expert assignments are consistent.
 - Alternative hypothesis H_1 . All expert assignments are not consistent.
- (3) Constructing t -statistic:

$$t = \frac{\bar{x}_{1j} - \bar{x}_{2j}}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n-2), \quad (5)$$

where \bar{x}_{1j} is the average of all expert assignments of the first group of the j th indicator, \bar{x}_{2j} is the average of all expert assignments of the second group of the j th indicator, S_1^2 is the variance of all experts assigned to the first group of the j th indicator, S_2^2 is the variance of all experts assigned to the second group of the j th indicator, n_1 is the number of the first grouping experts and n_2 is the number of the second group of experts.

- (4) Checking the t distribution table to obtain the critical value $t_{(\alpha/2)(n-2)}$, given the significance level α .
- (5) Comparing and judging:
 - If $|t| \leq t_{(\alpha/2)(n-2)}$, then H_0 is accepted and H_1 is rejected, thus considering the expert assignment to be consistent.
 - If $|t| > t_{(\alpha/2)(n-2)}$, then H_0 is rejected and H_1 is accepted, thus considering the expert assignment to be not consistent.

5.2.2 Validity test of expert assignment. The validity and reliability of an expert assignment can also be tested using the t -statistic, that is verifying that the expert assignment is consistent with the mean of all expert assignments for that indicator. The t -test steps are as follows:

- (1) Arbitrarily selecting an expert assignment x_{ij} to establish the following hypotheses:
 - The null hypothesis H_0 . The expert assignment has consistency with the mean, and the assignment is valid.
 - Alternative hypothesis H_1 . The expert assignment has no consistency with the mean, and the assignment is invalid.

(2) Constructing t -statistic:

$$t = \frac{x_{ij} - \bar{x}_j}{\delta_j} \sim t(n-2), \quad (6)$$

where x_{ij} represents the assignment of the j th indicator by the i th expert, \bar{x}_j is the average value of all the experts of the j th indicator and δ_j is the standard deviation of the assignment of all the experts of the j th indicator.

(3) Checking the t distribution table to obtain the critical value $t_{(\alpha/2)}(n-2)$, given the significance level α .

(4) Comparing and judging:

- If $|t| \leq t_{(\alpha/2)}(n-2)$, then H_0 is accepted and H_1 is rejected; thus, the expert assignment is consistent with the mean, and the assignment is valid.
- If $|t| > t_{(\alpha/2)}(n-2)$, then H_0 is rejected and H_1 is accepted; thus, the expert assignment and the mean have no consistency, and the assignment is invalid.

5.2.3 Outlier test of expert assignment. The t -test is only to test the consistency and validity of the expert assignment of a single indicator, but it is impossible to test whether an expert's abnormal assignment has a significant impact on the overall evaluation of the indicator. Therefore, on the basis of the t -test, it is also necessary to perform an F -test on a single indicator. The F -test steps are as follows:

(1) Sorting the expert assignments of individual indicators from small to large. The assignment sequence of the original N experts setting the j th indicator is represented as $x_j = \{x_{1j}, x_{2j}, \dots, x_{nj}\}$, $j = 1, 2, \dots, m$, and the sequence after sorting is $x'_j = \{x'_{1j}, x'_{2j}, \dots, x'_{nj}\}$, $j = 1, 2, \dots, m$.

(2) Establishing assumptions:

- The null hypothesis H_0 . There is no significant difference between the expert assignment after removing the maximum and minimum values and all expert assignments.
- Alternative hypothesis H_1 . There is a significant difference between the expert assignment after removing the maximum and minimum values and all expert assignments.

(3) Constructing F -statistic:

$$F = \frac{(SS_1 - SS_2)/2}{SS_1/(m-3)} \sim F_\alpha(2, m-3), \quad (7)$$

where $SS_1 = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ represents the sum of squared deviations of all expert assignments and $SS_2 = \sum_{i=2}^{n-1} (x'_{ij} - \bar{x}'_j)^2$ represents the sum of squared deviations after removing the maximum and minimum values of expert assignments.

(4) Checking the F distribution table to obtain the critical value $F_\alpha(2, m-3)$, given the significance level α .

(5) Comparing and judging:

- If $F \leq F_\alpha(2, m-3)$, then H_0 is accepted and H_1 is rejected, thus considering that there is no significant difference between the expert assignment after removing the maximum and minimum values and all expert assignments.

- If $F > F_{\alpha}(2, m-3)$, then H_0 is rejected and H_1 is accepted, thus considering that the expert assignment after removing the maximum and minimum values and the total expert assignment is significantly different.

5.3 Experts assignment normal distribution transformation

If the expert assignment does not obey the normal distribution, a normal distribution transformation should be considered. Common transformation methods include logarithmic transformation, square root transformation, reciprocal transformation, square root inverse sine transformation, etc., and the appropriate transformation method should be selected according to the nature of expert assignment data (Wei, 2012).

5.3.1 Data normal distribution transformation method

- (1) Calculating the distribution of the data and two parameters: skewness and kurtosis.
- (2) Determining whether transformation is needed according to the distribution shape and parameters of the data (Wei, 2012):
 - Symmetric judgment: the value of skewness should be observed. If the skewness is 0, it is completely symmetrical (but rare); if the skewness is positive, the distribution of the data is positively skewed; and if the skewness is negative, it is negatively skewed. However, the skewness value cannot fully judge whether the distribution of the skewness is significantly different from the normal distribution, and it is also necessary to make a significance test. If the test results are significant, the transformation can be used to achieve or approach symmetry.
 - Kurtosis test: the kurtosis is an indicator that judges the curve to be steep and gentle. If the kurtosis is 3, the shape of the distribution peak is the same as the normal distribution; if the kurtosis is greater than 3, the distribution is steep. On the contrary, it shows that the distribution of data is gentle. The kurtosis also needs to be judged by significance tests to see if there is a significant difference from the normal distribution. If the test results are significant, the conversion can be used to achieve or approximate a normal distribution.
- (3) Determining the corresponding transformation formula according to the distribution shape of the data, if a normal transformation is required. There are three common normal transformation methods:
 - Moderately skewed: if the skewness is two to three times its standard error, the root value for transformation should be considered.
 - Highly skewed: if the skewness is more than three times its standard error, it can be transformed in logarithmic form.
 - Bimodal or multimodal data: the Blom function is used to calculate the normal score, and then the data are transformed into a normal distribution.
- (4) Retesting the distribution after data conversion. If the problem is not solved or even worsened, it is necessary to restart from (2) or (3) and then carry out relevant tests until a satisfactory result is achieved.

5.3.2 Several problems of normal distribution transformation

- (1) The normal transformation method of data is not universal. It is necessary to select a suitable or improved transformation method according to the data distribution. After the transformation, the transformation effect must be verified, and finally the transformation purpose is achieved.

- (2) Not all non-normally distributed data can be normally distributed. Non-normally distributed data can also be analyzed using non-parametric methods.
- (3) The standard error of skewness and kurtosis is directly related to the sample size; generally, the larger the sample size, the smaller is the standard deviation of skewness and kurtosis.
- (4) When the expert assignment set does not obey the normal distribution, the consistency and validity of the experts assignment can use the Mann–Whitney U -test and the Wilcoxon Signed rank test.

5.4 Non-parametric test of experts assignment

When the expert assignment data do not satisfy the normal distribution and cannot be normally transformed, the consistency of the assignment data cannot be used for parameter verification. The following non-parametric test methods can be used: Mann–Whitney U -test and Wilcoxon Signed rank test.

5.4.1 *Mann–Whitney U -test.* The Mann–Whitney U -test, also known as the “Mann–Whitney rank sum test,” was proposed by H.B. Mann and D.R. Whitney in 1947. This test is mainly to test whether two independent samples are from the population of the same mean, that is to test the assignment scores of different experts. The specific steps (Guoxiang, 2014) are as follows:

- (1) Randomly searching for two groups of experts to assign indicators or randomly assigning expert values to two groups of samples A and B. The expert assignments of the two groups are graded in ascending order according to the numerical value. The minimum data level is 1, the second smallest data level is 2, and so on. If there is a case where the assignments are equal in the sorted data, the rank values of the same data should be the same and take the average of the unranked ranks. If the data are sorted {3, 5, 5, 9}, then their rating value should be {1, 2.5, 2.5, 4}.
- (2) Obtaining, respectively, the rank and sum R_A, R_B of the two sets of assignment samples, according to the rank value of (1).
- (3) Establishing hypothesis:
 - H_0 . There is no significant difference between the mean values of the two groups.
 - H_1 . There is a significant difference between the mean values of the two groups.
- (4) Mann–Whitney U -test. Calculating U : its sum is always equal to $n_A n_B$, that is $U_A + U_B = n_A n_B$. If $n_A \leq 20, n_B \leq 20$, its test statistic is given as follows:

$$U_A = n_A n_B + \frac{n_A(n_A + 1)}{2} - R_A, \quad (8)$$

$$U_B = n_A n_B + \frac{n_B(n_B + 1)}{2} - R_B, \quad (9)$$

where n_A and n_B represent the sample sizes of samples A and B, respectively.

The threshold table for the Mann–Whitney U -test only gives a small threshold, so the smaller U value in U_A, U_B , b is used as the test statistic.

- (5) Choosing the smaller of these values to compare with the threshold, given the level of significance α :
 - If U is less than U_α , H_0 is accepted and H_1 is rejected.
 - if U is greater than U_α , H_0 is rejected.

The U -test are also small and large samples for inspection. In small samples, the critical value of U is shown in the table. In large samples, the distribution of U approaches a normal distribution, so normal approximation can be used.

5.4.2 *Wilcoxon Signed rank test.* Wilcoxon Signed rank test, also known as Wilcoxon's sign rank test, was proposed by F. Wilcoxon in 1945. This method is developed on the basis of the symbolic test of paired observation data and is suitable for testing the data of two sets of associated samples. The specific steps are as follows (Xuemin, 2010):

- (1) Randomly finding a group of experts to assign two independent scores to the same index and getting two groups of assigned samples.
- (2) Establishing assumptions:
 - H_0 . There is no significant difference in the assignment results between the two groups.
 - H_1 . There are significant differences in the assignment results between the two groups.
- (3) Calculating the difference d_i of expert pair assignment and arranging the rank R_i of absolute value of d_i in order from large to small. If d_i is equal, the rank of the same data is the same and the average of the unranked array is taken. If d_i takes $\{9, 5, 5, 3\}$, then its grade value should be $\{1, 2.5, 2.5, 4\}$.
- (4) Restoring the positive and negative signs and calculating the sum of positive grades T^+ and negative grades T^- , respectively, after the completion of grade numbering. The smaller of T^+ and T^- was selected as the Wilkerson test statistic:

$$T = \min(T^+, T^-). \quad (10)$$

Thereinto, the sum of positive grades $T^+ = \sum_{d_i > 0} R_i$ and negative grades $T^- = \sum_{d_i < 0} R_i$.

- (5) Making a judgment: the critical value table should be checked according to the significance level α and the critical value T_α should be obtained; if $T < T_\alpha$, we will reject null hypothesis H_0 .

6. Reliability test and validity test

6.1 Reliability test of expert assignment

Reliability refers to the consistency and reliability of the evaluation index, and the consistency of the results obtained when the same method is used for repeated evaluation of the same index, that is the degree to reflect the actual situation. The higher the reliability coefficient, the more consistent and reliable are the evaluation results. The most commonly used reliability measurement method is Cronbach's α reliability, which was put forward by American educator Lee Cronbach in 1951. The calculation formula of α reliability coefficient is based on the early internal consistency calculation formula developed by G. Frederick kuder (Johnson *et al.*, 2014) and M.W. Richardson in 1937. Its calculation formula is as follows:

$$\alpha = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \sigma_{Y_j}^2}{\sigma_X^2} \right), \quad (11)$$

where m represents the number of indicators, $\sigma_{Y_j}^2$ represents the variance of the expert assignment of a single indicator and σ_X^2 represents the variance of the sum of the expert assignments of all indicators.

Cronbach's α coefficient is usually between 0 and 1. If the coefficient α does not exceed 0.6, it is generally considered that the internal reliability is insufficient. When it reaches 0.7~0.8, the scale has a good reliability, and when it reaches 0.8~0.9, the scale has a very good reliability.

Steps of reliability analysis test are given as follows:

- (1) According to the designed index system, the reliability of the index should be assessed by experts twice and the evaluation results should be counted.
- (2) Cronbach's α coefficient should be calculated.
- (3) Indicators should be selected by reliability.

6.2 Validity test of expert assignment

Validity is a complex concept that continues to evolve. In 1952, the committee on testing standards of the American psychological association published preliminary technical recommendations on psychological testing and diagnostic techniques, which indicated for the first time that the types of validity included predictive validity, simultaneity validity, content validity and consistency validity (Committee on Test Standards of American Psychological Association, 1952). In 1974, the standards of education and psychological testing defined validity clearly for the first time, believing that validity is the appropriateness of speculating on test scores and other evaluation results (American Psychological Association, American Educational Research Association, National Council on Measurement in Education, 1974). American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1985) supplemented and improved the definition of validity, believing that validity refers to the appropriateness, richness and usefulness of specific predictions made by indicators on test scores (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1985). American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014) pointed out that the validity is degree of evidence and theory support for the interpretation of test scores, and includes content based on the test of evidence, evidence based on the reaction process, based on the internal structure of the evidence, evidence based on the relationship with other variables and based on the test results of five kinds of validity of the evidence sources of evidence (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014).

Reliability is the premise of validity, which means the degree to which the index screening results are consistent with the target index system. It cannot be said that it is absolutely "invalid" or "effective." The value of validity ranges from -1 to 1, and the larger the value, the higher is the validity of the index.

The validity test of screening indicators usually adopts "content validity ratio," whose formula is given as follows:

$$CVR = \frac{\sum_{i=1}^N e_{ij} - 10 \times N/2}{10 \times N/2}, \quad (12)$$

where e_{ij} is the score of necessity and importance of indicator j given by expert i and N is the number of experts.

It is generally believed that when CVR is greater than 0.8, content validity is better, and this index is adopted, otherwise it is removed or modified.

The steps of validity analysis and test are as follows:

- (1) According to the designed index system, the necessity (importance) of the index should be evaluated by experts and the evaluation results should be counted.

- (2) Content validity ratio, CVR, should be calculated.
- (3) Indicators should be filtered through CVR.

6.3 Single-index expert assignment correction

When the validity of the expert assignment with a single indicator cannot pass the *t*-test, if the failure rate is greater than 20 percent, the theoretical indicator system is unreasonably constructed, and the indicators and classification stratification need to be re-screened; if the failure rate is less than or equal to 20 percent, it means that there is a large deviation in the assignment of a few experts, and it is necessary to re-evaluate the assignment.

When the expert assignment of a single indicator cannot pass the *F*-test or the reliability and validity test, in order not to lose the expert assignment, the expert assignment needs to be corrected. The correction methods are as follows:

- (1) The average value of the maximum value, the minimum value and the adjacent value of the single indicator expert assignment should be taken as the new value instead of the original assignment.
- (2) After the introduction of the new value, the test *t*, *F*-test and the reliability and validity test should be re-expanded. If the new *t* and *F* tests are still not passed, it means that the expert's cognition of the indicator is controversial and needs to be re-assigned.
- (3) If the second expert assignment is still controversial, it means that the design of the indicator is wrong, and the indicator needs to be redefined.

7. Multi-index expert assignment quality test

The multi-indicator expert assignment quality test is mainly based on the validity of the expert assignment of multiple indicators and the consistency of multiple expert assignments.

n Represents the number of experts, the number of the representative index *k*, *x_{ij}* represents the *i*th expert scores assigned to the *j*th index. *x_j(max)* represents the maximum value in the *j*th indicator attribute value range (data).

7.1 Multi-index expert assignment consistency test

Because of the correlation between multiple attributes between indicators, it is often difficult for an expert to accurately judge the complex relationship between multiple indicators. For example, it has been judged that *C₁-C₂* is important, and *C₂-C₃* is important. Naturally, *C₁-C₃* is more important, but if the expert considers *C₃-C₁* important or an equally important conclusion, a logical error will occur. Therefore, it is necessary to judge the reliability of the expert assignment or the accuracy of the judgment by judging the consistency of the matrix.

The consistency test (Shufeng, 2006) is mainly to test whether each expert assignment of multiple indicators has a logical consistency test (or a consistency test of multiple experts' judgment results) to avoid the contradiction between the two evaluation results. Consistency indicators and random consistency ratios are used in the AHP for consistency testing.

7.1.1 Consistency index. For the evaluation of multiple indicators, one is to test whether each expert's evaluation scores are consistent, and the other is to test whether all experts judge the consistency of these indicators. For the *m* indicators that need to be assigned scores, based on the quality test of the individual indicator experts and the expert evaluation scores of the two pairs, the judgment matrix is constructed, and the maximum eigenvalue λ_{max} of the judgment matrix is calculated. If the number is *m*, then the consistency index of the judgment matrix is CI:

$$CI = \frac{\lambda_{max} - m}{m - 1} \tag{13}$$

If multiple experts are required to perform consistency check on the assignment of multiple indicators, it is necessary to perform the mean processing on the expert assignment of each indicator, such as arithmetic mean, geometric mean and then refer to the assignment of multiple indicators by a single expert. The consistency check process can be tested.

7.1.2 *Random consistency ratio.* As k increases, the error of the judgment matrix consistency test index increases, so the random consistency ratio CR is introduced to correct CI:

$$CR = \frac{CI}{RI} \tag{14}$$

where RI is the average random consistency index, which is the average of the consistency indicators calculated by a large number of experimental experts of the National Oak Ridge National Laboratory. The value RI of the 1–15th order is shown in Table VII, where the order represents the number of indicators in the judgment matrix, and when the order exceeds 3, the consistency test is often required.

When $CR < 0.1$, the judgment matrix is considered to have reasonable consistency; otherwise, the expert assignment of the pairwise comparison needs to be adjusted until sufficient consistency is achieved.

7.2 Coordination test of multi-index expert assignment

For multiple indicators or even for each index, there will always be differences in the evaluation scores of different experts. Some experts' evaluation is generally high, some experts' evaluation is generally low and some experts' evaluation is generally different. How to judge whether the overall evaluation scores of multiple indicators are stable and effective, and whether experts have the same understanding of the importance of each index? Coordination test of expert assignment is needed.

Coordination coefficient W reflects the consistency of all experts' evaluation of all indicators, and it is also an indicator of the credibility and stability of expert evaluation results. In this paper, Kendall synergy coefficient is used to test the overall consistency of all experts' evaluation scores. The consistency test of Kendall's synergistic coefficient (Dong *et al.*, 1997) requires ranking transformation after each expert ranks the evaluation results of each index. The specific method refers to the non-parametric test of 4.4 expert assignment.

Kendall's synergistic coefficient formula is divided into two forms on the basis of whether there is the same rank or not:

- (1) If each expert has different rating for each index, that is no same rank, the Kendall synergy coefficient W is given as follows:

$$W = \frac{12S}{n^2(m^3 - m)} \tag{15}$$

- (2) If an expert has the same rank in the evaluation of m indexes, the Kendall synergy coefficient W is given as follows:

$$W = \frac{12S}{n^2(m^3 - m) - n \sum_g (\tau_k^3 - \tau_k)} \tag{16}$$

where n is the number of experts, m is the number of indicators, R_j is the rank (or rank) sum of each expert evaluation grade of the index j , τ_k is the length of k equal-order columns and

Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.52	1.54	1.56	1.58	1.59

Table VII.
RI value table

g is the number of all equal-order columns and S is the sum of squares of deviations between the total rank and the average rank of the index:

$$S = \sum_{j=1}^m \left(R_j - \frac{n(m+1)}{2} \right)^2.$$

The range of Kendall's synergy coefficient is $[0, 1]$. When all the experts' opinions on index j are completely consistent, the value of R_j is $jn (j=1, 2, \dots, m)$, the sum of squares of deviations S is the largest and the Kendall's synergy coefficient W is close to 1.

The specific steps (Linquan, 2017) of Kendall synergistic coefficient significance test are as follows:

- (1) Establishing original hypothesis and alternative hypothesis:
 - Original hypothesis H_0 . Expert evaluation score is random.
 - Alternative hypothesis H_1 . The expert evaluation score is stable.
- (2) Constructing χ^2 -statistic K :

$$K = n(m-1)W \rightarrow \chi^2_\alpha(m-1). \tag{17}$$

Under the original assumption that H_0 holds, the statistic K approximately obeys the chi-square distribution with degree of freedom $m-1$:

- (3) Conducting significance test.

At a given saliency level α , the critical value table $\chi^2_\alpha(m-1)$ of χ^2 is consulted and the original hypothesis is compared and judged.

When $K > \chi^2_\alpha(m-1)$, the original hypothesis H_0 is rejected and the alternative hypothesis is accepted, thus considering that the expert score has overall stability under the significant level α .

For example, in order to study the operation status of enterprises, three indicators are preliminarily screened: corporate social responsibility, customer satisfaction and customer loyalty. Through in-depth interviews with enterprises, experts and entrepreneurs are invited to evaluate the importance of these three indicators (Tables VIII–X).

Because of the same rank, the Kendall synergistic coefficient is calculated by formula (16). The results show the following:

$$W = \frac{12S}{n^2(m^3-m) - n \sum_g (\tau_k^3 - \tau_k)} = \sum_{j=1}^3 \frac{12(R_j - (6(3+1)/2))^2}{6^2(3^3-3) - 6 \sum_{k=1}^5 (\tau_k^3 - \tau_k)} = 0.2321.$$

Significance test results show that statistic $K = n(m-1)W = 6 \times (3-1) \times 0.2321 = 2.7852$. The score of six entrepreneurs is random because it is less than the critical value $\chi^2_{0.05} = 5.991$.

	Corporate social responsibility	Customer satisfaction	Customer loyalty
Entrepreneur 1	4	4	5
Entrepreneur 2	6	6	6
Entrepreneur 3	3	2	4
Entrepreneur 4	1	1	1
Entrepreneur 5	4	3	3
Entrepreneur 6	4	4	4

Table VIII.
Indicators of expert appointment scores

7.3 Distribution of importance of all indicators

When the quality diagnosis of the single indicator expert assignment is passed, it indicates that the expert assignment is consistent and reliable; and the expert assignment indicates that the expert's recognition of the indicator is greater, indicating that the expert believes that the indicator is more important. However, how to use the value of the indicator to measure the importance of the indicator in the whole indicator system, especially in the same level or less important but need to be retained, does not stipulate the standard of inspection, so it is necessary to test the importance of all indicators to achieve the purpose of preliminary screening indicators.

First, the average value of the experts for each index is averaged. When the mean sequence approximates the normal distribution, the test is used. When the normal distribution is not followed, the distribution can be performed using the method in Section 5.3.

The specific steps of all indicators based on the importance distribution test of the F -test are as follows:

First, calculating the mean- x_j of the expert assignments for each indicator and arranging the mean values from small to large. The ascending mean sequence is obtained as $X'_j = \{x'_1, x'_2, \dots, x'_m\}$.

Second, establishing hypothesis:

- H_0 . An indicator of the minimum assigned mean m_0 is important.
- H_1 . The indicator of the minimum assigned mean m_0 is not important.

Third, constructing F statistics:

$$F = \frac{(SS_1 - SS_2)/m_0}{SS_1/(m - m_0 - 1)} \sim F_{\alpha}(m_0, m - m_0 - 1), \tag{18}$$

where $SS_1 = \sum_{j=1}^m (x_j - \bar{x}_j)^2$ is the sum of squared deviations on behalf of all indicator experts and $SS_2 = \sum_{j=2}^{m-1} (x'_j - \bar{x}'_j)^2$ represents the sum of squared deviations after removing the maximum and minimum values assigned by the expert.

Fourth, checking the F distribution table to get the critical value $F_{\alpha}(m_0, m - m_0 - 1)$, given a level of significance α .

	Corporate social responsibility	Customer satisfaction	Customer loyalty
Entrepreneur 1	1.5	1.5	3
Entrepreneur 2	2	2	2
Entrepreneur 3	2	1	3
Entrepreneur 4	2	2	2
Entrepreneur 5	3	1.5	1.5
Entrepreneur 6	2	2	2
Entrepreneur R	12.5	10	13.5

Table IX.
Rank of grade
conversion of expert
appointment scores

	Sequence 1	Sequence 2	Sequence 3	Sequence 4	Sequence 5
τ_k	2	3	3	2	3
$\tau_k^3 - \tau_k$	6	24	24	6	24

Table X.
Serial modified
calculating table with
the same rank
for index converted
from expert
appointment grade

Fifth, comparing and judging:

- If $F \leq F_{\alpha}(m_0, m-m_0-1)$, H_0 is accepted and H_1 is rejected. It is considered that the indicator corresponding to the minimum m_0 assignment mean values is important and should be retained.
- If $F > F_{\alpha}(m_0, m-m_0-1)$, H_0 is rejected and H_1 is accepted. It is considered that the assignment of the indicator corresponding to the smallest m_0 of the assigned mean values is too small and can be filtered out.

$m_0 = j-1, j-2, \dots, 2, 1$ should be taken separately. The above F -test should be performed until the one F -statistic is less than the critical value.

8. Conclusion

In this paper, through the scientific process that from “analysis of theory basic, analysis of related factors, description of process mechanism” to “analysis of system structure, decomposition of system hierarchical, identification of the complete set of indicators” to primary selection the indicators system, using system engineering and hierarchical structure decomposition theory, optimize the index system construction process and analysis method; aim at the expert assessment quality test, judge the authoritative degree of expert opinion and clarify the index attribute value range, based on Delphi assignment method and other four methods for expert assignment, expert evaluation results for individual indicators assignment quality test, reliability test and validity test, and multi-index expert assignment quality test.

Through the above-mentioned series of tests for the evaluation of experts in the indicator system, the paper completes the optimization of the index system design. In the optimization process, the overall inspection process is detailed and clear. Compared with the traditional index system construction method, the process standardizes the process of index establishment, reduces subjectivity and randomness, and enhances objectivity and scientificity. In the future, the focus of optimization and testing of the design of the indicator system should be on the importance test and classification based on the indicator data, that is how to use the principal component analysis method, factor analysis method and orthogonal design method to reduce the dimension of the index system, how to combine the gray comprehensive evaluation, fuzzy comprehensive evaluation and AHP to comprehensively evaluate and how to innovate the golden section method to effectively classify.

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1985), *Standards for Educational and Psychological Testing*, American Psychological Association, Washington, DC, p. 9.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014), *Standards for Educational and Psychological Testing*, American Psychological Association, Washington, DC, p. 14.
- American Psychological Association, American Educational Research Association, National Council on Measurement in Education (1974), in Davis, F.B. (Ed.), *Standards for Educational and Psychological Tests*, American Psychological Association, Washington, DC, p. 26.
- Carroll, J.B. (1993), *Human Cognitive Abilities: A Survey of Factor Analytic Studies*, Cambridge University Press, Cambridge.
- Chengbin, W. (1991), “Empirical Bayes estimation for Delphi quantitative forecasting”, *Journal of Northern Jiaotong University*, No. 2, pp. 70-75.
- Chonghua, W. and Wenfu, J. (2013), *Chinese Medical Statistics Encyclopedia – Health Measurement Volume*, China Statistics Press, Beijing, pp. 62-63.

- Chun, G. and Jun, H. (2006), "Improvement of the Delphi method based on artificial neural network", *Microcomputer Information*, No. 30, pp. 171-173.
- Committee on Test Standards of American Psychological Association (1952), "Technical recommendations for psychological tests and diagnostic techniques: a preliminary proposal", *American Psychologist*, No. 7, p. 467.
- Daowu, Z., Shukun, T. and Zhaoben, F. (2004), "An evolutionary group's decision-making mechanism", *Prediction*, No. 3, pp. 36-40.
- Dong, X., *et al.* (1997), *Management Evaluation*, Vol. 3, National University of Defense Technology Press, Changsha, p. 134.
- Environmental Engineering Evaluation Center of the Ministry of Environmental Protection (2016), *Technical Guidelines for the Design of Basic Databases for Environmental Impact Assessment*, China Environmental Science Press, Beijing, p. 9.
- Guanghua, H. and Bo, F. (2017), "The effect of semantic difference of Likert-type questionnaires on scientific measurements", *Science & Technology Progress and Policy*, Vol. 34 No. 20, pp. 1-6.
- Guoxiang, X. (2014), *Editor-in-Chief. Statistics*, Vol. 9, 2nd ed., Gezhi Press, Shanghai, pp. 243-244.
- Haibo, Z. (2004), *Survey Data Analysis*, China Financial and Economic Publishing House, Beijing, pp. 281-283.
- Hui, C., Ying, Z. and Zongxi, N. (1995), "Screening of evaluation experts in Delphi method", *China Health Care Management*, No. 1, pp. 49-50+55.
- Jinfeng, Z. and Shuqin, Z. (1997), "Introducing set value statistics to improve data processing process in Delphi method", *Statistics & Decision*, No. 3, pp. 10-11.
- Jinyi, L. (1992), "Comprehensive evaluation with improved Delphi method", *Construction Mechanization*, No. 6, pp. 31-33.
- Jixue, Y. and Weijin, W. (2010), "An advanced Delphi method of resolving a kind of group decision making", *Mathematics in Practice and Theory*, Vol. 40 No. 20, pp. 89-95.
- Johnson, B., Christensen, L., Jiansheng, M., *et al.* (2014), *Quantitative, Qualitative and Mixed Approaches to Educational Research*, Vol. 9, 4th ed., Chongqing University Press, Chongqing.
- Kai, B. (2011), "On the application and testing of non-response option of Likert-type scales in tourism research", *Tourism Journal*, Vol. 26 No. 4, pp. 29-35.
- Kaur, N. and Pluye, P. (2019), "Delineating and operationalizing the definition of patient-oriented research: a modified e-Delphi study", *Journal of Patient-Centered Research and Reviews*, Vol. 6 No. 1, pp. 7-16.
- Ligen, Y. and Xuewen, W. (2012), *Introduction to Engineering*, Publishing House of Electronics Industry, Beijing, pp. 231-232.
- Linquan, Z. (2017), "The principle of Kendall's synergy coefficient test and its application analysis", *Modern Business and Industry*, No. 10, pp. 175-176.
- Lisha, Z., Chen, L. and Shunkun, Y. (2011), "Research on performance evaluation of low carbon management employees in power grid enterprise based on rough sets", *Hydropower Energy Science*, Vol. 29 No. 12, pp. 184-187+215.
- Meigang, Z., Jing, Y. and Yuxing, L. (2008), *Rhetoric of Science and Technology English*, National Defense Industry Press, Beijing, pp. 189-194.
- Mingyuan, G. (1998), *Education Dictionary*, Shanghai Education Press, Shanghai, pp. 1289-1293.
- Murias, P., de Miguel, J.C. and Rodriguez, D.J.S.I.R. (2008), "A composite indicator for university quality assessment: the case of Spanish higher education system", *Social Indicators Research*, Vol. 89 No. 1, pp. 129-146.
- Peiyu, X., Lei, W., Shuhong, S., Qian, Z., Chuncheng, G. and Qian, H., "Research on the construction of comprehensive evaluation system for power grid purchase scheme – based on grey triangle whitening weight function", *Price Theory And Practice*, pp. 1-3.
- Qingzhen, L. (1999), *Dictionary of Science and Technology Methods*, Science Press, Beijing, pp. 350-352.

- Qinwei, Y., Ganjin, Z. and Hongzhou, S. (2011), "Review of development and application of Delphi method in China – one of series papers of Nanjing University knowledge mapping research group", *Modern Intelligence*, Vol. 31 No. 5, pp. 3-7.
- Qun, Z., Xiaolei, H. and Guangxing, S. (2010), "Implementation of Delphi method based on vague set theory", *Statistics & Decision*, No. 1, pp. 158-160.
- Saisana, M. and Tarantola, S., "State-of-the-art report on current methodologies and practices for composite indicator development", CiteSeer.
- Schuman, H. and Presser, S. (1981), *Questions and Answers in Attitude Surveys*, Academic Press, New York, NY, pp. 33-56.
- Shiau, T.-A. and Liu, J.-S.J.E.I. (2013), "Developing an indicator system for local governments to evaluate transport sustainability strategies", *Ecological Indicators*, Vol. 34, November, pp. 361-371.
- Shufeng, J. (2006), "The algorithm of mean random consistency index in AHP and its implementation", *Journal of Taiyuan Normal University (Natural Science Edition)*, No. 4, pp. 45-47.
- Spearman, C. (1923), *The Nature of "Intelligence" and Principles of Cognition*, Macmillan, London.
- Vakil, N., van Zanten, S.V. and Kahrilas, P. (2006), "The Montreal definition and classification of gastro esophageal reflux disease: a global evidence-based consensus", *American Journal of Gastroenterology*, Vol. 101 No. 8, pp. 1900-1920.
- Wei, T. (2012), *Normality Test Based on Skewness and Kurtosis*, Shanghai Jiaotong University.
- Wei, W. and Yujun, L. (2004), *Computer Automatic Reasoning and Intelligent Teaching*, Beijing University of Posts and Telecommunications Press, Beijing, pp. 58-63.
- Weitao, L., Shanghui, L., Wencheng, S. and Xiangxiang, X. (2016), "Coal seam floor water inrush evaluation model based on variation coefficient method and GIS spatial analysis", *Mining Safety and Environmental Protection*, Vol. 43 No. 2, pp. 59-62.
- Xiting, H., Zhiliang, Y. and Chongde, L. (2003), *Dictionary of Psychology*, Shanghai Education Press, Shanghai, pp. 784-786.
- Xuemin, W. (2010), *Statistical Analysis Methods and Applications*, Vol. 6, Shanghai University of Finance and Economics Press, Shanghai, pp. 103-105.
- Yidan, B., Yanping, W., Yong, H. and Xiaofeng, G. (2005), "A new AHP integrated with orthogonal design", *System Engineering-Theory & Practice*, No. 1, pp. 123-127+133.
- Yonghua, W. (1997), *Hydropower Energy System Planning*, Vol. 11, China Electric Power Press, Beijing, pp. 15-20.
- Yun, H. (2011), "A comparative study of empirical analysis methods and normative analysis methods in economics", *Business Culture (2nd Half)*, No. 4, pp. 103-104.
- Yuxiang, C. and Donghua, Z. (1990), "Applied research of Delphi method", *Future and Development*, No. 5, pp. 42-44.
- Zheng, Y. (2017), "Fuzzy-based composite indicator development methodology for evaluating overall water BOT project performance", *International Conference on Service Systems and Service Management, IEEE*.

Corresponding author

Yin Kedong can be contacted at: yinkedong@ouc.edu.cn