

Article

Constructing a Measurement Method of Differences in Group Preferences Based on Relative Entropy

Shiyu Zhang ¹, Wenzhi Liu ^{1,3,*}, Qin He ¹ and Xuguang Hao ²

¹ Management College, Beijing Union University, Beijing 100101, China; gltshiyu@buu.edu.cn (S.Z.); ldtheqin@buu.edu.cn (Q.H.)

² International business school, University of International Business and Economics, Beijing 100029, China; haoxg@uibe.edu.cn

³ Institute of Transportation Engineering, Tsinghua University, Beijing 100084, China

* Correspondence: liuwenzhi@buu.edu.cn; Tel.: +86-139-1046-4976

Academic Editor: Kevin H. Knuth

Received: 13 November 2016; Accepted: 3 January 2017; Published: 6 January 2017

Abstract: In the research and data analysis of the differences involved in group preferences, conventional statistical methods cannot reflect the integrity and preferences of human minds; in particular, it is difficult to exclude humans' irrational factors. This paper introduces a preference amount model based on relative entropy theory. A related expansion is made based on the characteristics of the questionnaire data, and we also construct the parameters to measure differences in the data distribution of different groups on the whole. In this paper, this parameter is called the center distance, and it effectively reflects the preferences of human minds. Using the survey data of securities market participants as an example, this paper analyzes differences in market participants' attitudes toward the effectiveness of securities regulation. Based on this method, differences between groups that were overlooked by analysis of variance are found, and certain aspects obscured by general data characteristics are also found.

Keywords: relative entropy; preferences; group; difference; securities regulation

1. Introduction

In social science research, certain attitudes of social groups are often studied using a questionnaire [1], such as voters' attitudes toward presidential candidates, individuals' attitudes toward a certain policy, customer satisfaction, and social public satisfaction with the government. Due to the comprehensive effects of a person's psychology, personality, emotions, and other characteristics, the evaluation of various indicators often reveals some irrational information. For example, there are three evaluation indicators that constitute a phenomenon: A, B, and C. The respondents are required to give a score evaluation based on these three indicators, and the score points range from one to five, with one representing the worst and five representing the best. When the respondents make their score evaluations, there will be two cases. In the first, the gap between the evaluations and indicators is very small, i.e., the scores are all high scores, low scores, or medium scores; in the second, there is a large gap in the evaluations given to the various indicators.

For example, in research on people's attitudes toward and views of the management of a community, we often divide the community management into a number of indicators, including sanitation, afforestation, public security, maintenance services, the shopping environment, and so on. Each item can be rated from one to five, with one representing the worst and five the best. There may be some in the community who are always dissatisfied with the community's sanitation, thus bringing their opinions to the management. However, these people may not give the lowest score on the questionnaire regarding the community's sanitation level. For example, these people's average score

may be 2 or 3 points, and those who give an average score of 1–2 points may be silent on the matter in everyday life, not voicing their opinions to community management. Why does this phenomenon occur? It cannot be explained by the mere evaluation of the “degree of sanitation”. However, the overall opinions of the two types of evaluations can be easily explained. Regarding the group that voices its opinion regarding the community’s sanitation problems, compared to others, these people may not give the lowest score for the “degree of sanitation” compared to other community members but their scores for sanitation are lower than the scores they give other indicators. This group’s evaluations of the security situation may be five points and, for the degree of afforestation, four points, but the group’s evaluation of the degree of sanitation is the lowest, perhaps two points. In other words, this group gives scores with a larger gap in various indicators, which can be interpreted such that this group of people takes a serious attitude and has clear ideas concerning what they hate or love. Although the groups with the lowest degree of sanitation evaluation scores do not take any action or voice their opinions, their evaluations on all indicators are very low, and the gap among the evaluation scores is not large. In fact, if the community only has a sanitation problem and the other aspects of community life are good, a group’s evaluations that consist solely of low scores on all indicators will exhibit a small score gap, but their evaluations may not constitute entirely rational evaluations. This group may be driven by certain irrational factors, such as a pessimistic temperament, general social grievances, and so on. Those who give evaluations with large score gaps actually show their true opinions. Similar situations are common, such as polls on election candidates, polls on government performance, and customer satisfaction with companies’ services.

Almost all research is designed to genuinely understand respondents’ attitudes and preferences, but conventional statistical methods can hardly reflect such attitudes and preferences and, therefore, contradict the study’s purposes. In this paper, we believe that to achieve the purposes of a study—to effectively highlight the attitudes and opinion preferences of respondents—in the process of data analysis, we should pay more attention to those individuals who mark different scores with large gaps in the evaluation of different indicators and take them as more important factors among the results during the calculation process. Similarly, we posit that we should pay less attention to those respondents who evaluate various indicators with a small gap (some of whom even give the same scores for all indicators) and take these respondents as less important factors in the calculation process.

The source of this paper is a study on the evaluation of the effectiveness of securities regulations in China. To evaluate the regulations by researching their effect on market participants, this paper classifies market participants into five groups [2,3]: regulators, managers of listed companies, individual investors, managers of fund companies, and managers of securities companies. The respondents are asked to evaluate all aspects using a score. The essence is not highlighted during the first-round analysis of the study, which is conducted by a traditional method. In the second-round analysis, we propose a calculation method based on relative entropy theory, which solves the problem and also has a certain universality.

2. Literature Review

Preference is a widely used concept. Arrow [4], a Nobel laureate in economics, describes preference in his book as follows. When faced with an alternative consisting of a number of selection items, these options can be represented by x, y, z, \dots , thereby forming the set S . The preference of the selector is represented not by treating each of the alternatives in S equally but instead by sorting them after comparison or by preferentially selecting an item. Arrow first defines the reference or indifference relationship; this relationship is an axiom, which is expressed as follows:

For any two optional objects x and y , there must be xRy or yRx .

xRy means “ x is not inferior to y ”, and yRx means “ y is not inferior to x ”; this relationship is called the “weak ordering” relationship. The weak ordering relationship contains a preference or an indifference, and if the indifference is excluded, then only the preference relationship is defined. Arrow’s definition is as follows:

If yRx is false, then it is called xPy .

“ xPy ” means “ x is better than y ”, and Arrow calls it a strict preference relation.

Thus, comparison is a prerequisite for a preference, and a preference is produced in comparison and through people’s selection of different things. The concept of preference is widely used in the field of marketing in customer research. Since each person has preferences when a facing a choice, Arrow notes that to obtain the public attitude toward a problem, this problem becomes a gathering of individual preference types, producing the only society preference type based on these individuals.

There are a number of methods to reflect human preferences by gathering large amounts of data. The most traditional method involves traditional statistical methods, which calculate the average of each person’s evaluation of each variable or the ratio of selections for each option and examines the significance of the differences in the mean value or ratio of different groups of data. This is the most basic method, but its limitations are obvious. When gathering data with the mean value and ratio, each row of data in the data table (i.e., the choice of each respondent for a series of variables) is given equal status, regardless of the rows’ differences in size, which makes their influence on the results consistent, weakening the impact of preferences on the results to a certain extent. The limitations of mean values and standard deviations in data analysis—in addition to entropy theory and its approach, which can compensate for these limitations—have already been noted by scholars [5].

Association rules [6] in the data mining method can find a series of related project groups, as opposed to gathering a large amount of data into a parameter, which cannot solve this problem.

The variance in statistics reflects only discrepancies in the data, and it is difficult for it to reflect the preferences of different groups of data. Gathering a large number of the respondents’ data using entropy theory and the relative entropy method can effectively reflect respondents’ preferences.

The work published by Shannon in 1948 [7] is an important symbolic moment of the birth of information theory. The other two important milestones in the paradigm-shifting development of entropy theory are as follows: (1) the principle of maximum entropy proposed by Jaynes in 1957 [8], i.e., that states that the distribution that is closest to the uniform probability will be chosen in a probabilistic distribution that satisfies all given constraints, namely, the maximum entropy probability distribution; and (2) relative entropy theory, which was developed based on the early directed divergence proposed by Kullback [9]. Relative entropy can be used to measure the proximity of two probability distributions. In this paper [10], eight entropy-based methods, including Shannon’s entropy and relative entropy, are compared using the entropy-based image threshold technique, and the relative entropy method is further developed. Another form of information entropy is mutual information. In their discussion of the purpose of mutual information, Baratpour et al. note that mutual information can be used to measure the degree of dependence of the recorded values of two variables [11].

The development of information entropy, the main theory, and its application have been discussed in detail by Gray, Qiu, and Zhong [9,12,13]. Since the relationship between entropy and information was formed, information entropy has been widely used in economics, managerial studies, and the social sciences [14–18].

The latest research is more in-depth and broader. Two scholars, Baez and Pollard [19] review various information-theoretic characterizations of the approach to equilibrium in biological systems. The replicator equation, evolutionary game theory, Markov processes and chemical reaction networks all describe the dynamics of a population and of a probability distribution. Under suitable assumptions, the distribution will approach equilibrium with the passage of time. Relative entropy provides a quantitative measure of how far from equilibrium the system is [19]. The social preferences of different groups studied in this paper can also be regarded as an ecological system in equilibrium; therefore, entropy theory can also be used to measure how far the actual observation state and the system equilibrium. In terms of specific algorithms and models, Dziurosz-Serafinowicz [20] further affirmed the traditional principle of maximum relative entropy (MRE) and extended its application to the expression of new degrees of beliefs as a result of learning. The study closest

to this study is “Preference Inconsistence-Based Entropy” by Pan et al. [21]. These authors note as follows: “As available information is usually obtained from different evaluation criteria or experts, the derived preference decisions may be inconsistent and uncertain. Shannon entropy is a suitable measurement of uncertainty.” Although their study has the same principle as the present study, it is different in application. In their study, the theory of relative entropy was used to establish models that distinguished the preferences of decision-makers in decision making, which involved individual preferences. In this study, a large volume of social survey data was used to distinguish the preferences of different social groups. Makowski et al. [22] studied the issue of transitivity of preferences in an argument between two people. A recent study, “Information-Theoretic-Entropy Based Weight Aggregation Method in Multiple-Attribute Group Decision-Making” from He et al. [23] is also a study of decision makers’ preferences. Thus, studying individual preferences in the decision-making process using relative entropy theory is more common than studying group preferences.

Nonetheless, the study of social group preferences is a research hotspot over the years, although most of this scholarship involves consumer preferences [24] and/or establishing models [25]. Although the theory of entropy has many applications in terms of handling social survey data [26–34], we have not found that previous applications of the theory of entropy are basically the same as that used in this paper, but it is absolutely possible to extend these applications based on existing results.

According to information entropy theory [9,12,13], if something has m types of states, then the probability of each state is P_i ($i = 1, 2, \dots, m$), and the information entropy of the system can be defined as follows:

$$H = - \sum_{i=1}^m p_i \ln p_i, \text{ where } : 0 \leq p_i \leq 1, \sum p_i = 1 \quad (1)$$

When the entropy value of the system is high, the implication is that there is a high degree of chaos in the system; when the data distribution is even and the degree of variation is small, the implication is that the amount of information is small. When system entropy is low, the degree of variation of the data is large, and the amount of information contained is large. According to the principle of maximum entropy, in the state of nature or in the absence of outside intervention, the entropy of a system tends to increase; if entropy is to be reduced, external forces must be applied to the system. For a data system that reflects the state of something, the data should be evenly distributed in the random condition. If the data distribution is not even and the degree of variation is large, then it is often influenced by certain system factors, and there are reasons to investigate. Regarding the particular problem in this paper, when the respondents must choose among a number of indicators, if there are no preferences, then the data distribution is even, and there will be no differences among the options; if the data distribution is uneven and there are differences among the various options, then the ratio of the different options is different and inevitably affected by preferences. Therefore, the basic definition of the entropy theory formula reflects individuals’ preferences, and the concept of relative entropy should still be adopted to measure the preferences of different people or different groups.

The definition of relative entropy is based on the ratio of two probability distributions. If there are two probability distributions P and Q , then the relative entropy of P to Q can be expressed as $D(P:Q)$ and can be defined as follows:

$$D(P : Q) = \sum_i P_i \ln \frac{P_i}{Q_i} \quad (2)$$

Relative entropy $D(P:Q)$ defines the degree of closeness between probability distributions P and probability distributions Q . The smaller the absolute value of the relative entropy, the closer the two are. In the extreme case, if $P_i = Q_i$, then $D(P:Q) = 0$. According to the principle of relative entropy, to obtain optimal results, the result of gathering should be the closest to the probability (a priori probability) of a choice distribution among all the probability distributions satisfying a given constraint.

When the maximum entropy principle and the conditions of the relative entropy principle are met simultaneously, the two issues in the data analysis of this study can apparently be solved; in other

words, the final parameters will reflect the integrity requirements and preference characteristics of people. In line with the maximum entropy principle, the final parameter can better reflect the integrity requirements, and conforming to the relative entropy principle, it can better reflect the preference characteristics of people.

3. Materials and Methods

3.1. Questions

The above-mentioned requirements for social group preference research can be converted into three specific objectives:

First: to measure the respective attitude tendency degrees of different groups;

Second: to sort the variable based on the attitudes of different groups; and

Third: to measure the differences in attitude among the different groups.

3.2. Problem Model and Solution Target

A realistic problem can be measured with indicators of number j , $j = 1, 2, \dots, k$, and we assume that the numbers of all indicators have common values. Five levels or seven levels are most commonly used in the values, and typical values are five levels [1]: (1) very bad; (2) not good; (3) generally good; (4) very good; and (5) excellent, with the numbers 1, 2, 3, 4, and 5 indicating the variable level measurement. Each respondent can be recorded as i , $i = 1, 2, 3, \dots, n$; each participant evaluates each of the indicators that constitute the problem, and all of the respondents' evaluation data form an $n \times m$ matrix x :

$$X = (X_{ij})_{n \times k} = \begin{bmatrix} X_{11} & \cdots & X_{1k} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nk} \end{bmatrix} \quad (3)$$

Simultaneously, the respondents are divided into g groups, $g = 1, 2, 3, \dots, m$. There are two types of commonly used data analysis methods. One method is based on the mean value, which includes calculating the average score of each indicator, the average score of all indicators of each group, and the average score of each group of each indicator and then conducting the significance test and analysis of variance. The second method is non-parametric statistical analysis, i.e., calculating the number and ratio of various groups of various indicators' values and their corresponding chi-square tests. Each of these two methods have their own focus. In this paper, we only discuss the comparison of methods based on the mean value and the relative entropy method to analyze the study object from a unique perspective.

The aim of problem-solving is to gather all the raw data X into a parameter that can represent the data. As noted above, the mean value method and simple ratio statistics are such classical parameters; despite their natural advantages, they have limitations in terms of reflecting human preferences. In the process of data gathering, this method is expected to accord greater weight to obvious preference data to improve their influence on the final result.

3.3. Model Structure

3.3.1. Data Gathering Based on the Relative Entropy Method

The data analysis of scores on a number of indicators by different groups is essentially a data-gathering problem in which the evaluation data of many people and groups, which are expressed as numbers, are aggregated into a parameter that should reflect individual preferences, i.e., the distance between each person's preferences and the parameters should be minimized such that participants' opinions are reflected to the greatest extent. Since each person's preferences are different, the final gathered parameter is minimized with respect to each person's preference, namely, minimizing

the inconsistencies between the gathered parameters (decision results) and individual preferences and maximizing the consistency of group preferences to maximize the consistency of the group selection. This optimization problem is mathematically classified as the optimal solution to the model. Establishing the model is based on the following two principles: the maximum entropy principle and the relative entropy principle.

According to the principle of maximum entropy, if we seek to obtain optimal results, then in all satisfied given constraints of the probability distribution, the probability distribution that is closest to the uniform distribution will be chosen, i.e., the maximum entropy probability distribution.

In solving the group decision problem, Qiu [12] proposes the relative entropy model and offers the solution. Regarding the decision scheme set $A = \{a_j, j = 1, 2, \dots, n\}$; policymakers set group $E = \{e_i, i = 1, 2, \dots, m\}$, and x_{ij} represents the evaluation by policy-makers e_i of project a_j . If we assume that a larger value means the project is more certain, supposing that the group preference g can be measured, and its measure value is x_{gj} , then x_{gj} is the mapping of evaluation value a_j . Thus, if the preference amount of the group preference is $X_g = (x_{g1}, x_{g2}, \dots, x_{gn})^T$, when x_{gj} is obtained, the decision scheme can be sorted and the group preferences compared. The main application in the decision-making field is to compare the decision-making scheme, and as this paper is for group decision-makers—that is, comparing groups who are making evaluations—then the scheme set $W = \{w_i, i = 1, 2, \dots, m\}$ is a set of decision-makers weights and is combined to 1. Since the continuous variables are too complex, it must be assumed here that the variables used to evaluate the scheme are discrete variables and that different groups are making independent evaluations. The programming model (p) is formulated below as Equation (4):

$$(p) \begin{cases} \min Q(X_g) = \sum_{i=1}^m w_i \sum_{j=1}^n \left[\log x_{gj} - \log \frac{x_{ij}}{\sum_{j=1}^n x_{ij}} \right] x_{gj} \\ \text{s.t. } \sum_{j=1}^n x_{gj} = 1, x_{gj} > 0 \end{cases} \quad (4)$$

The intuitive interpretation of this objective function is to minimize the deviation between the preference utility value and the group preference amount for each decision maker such that each decision-maker’s preference efficiency can be compared with the group preference efficiency. Then, the method of calculating the preference amount X_g is obtained by solving this objective function.

By solving this mathematical programming model, the optimal solution $X_g^* = (X_{g1}^*, X_{g2}^*, \dots, X_{gN}^*)^T$ of scheme problem P, which is called the preference amount, can be solved:

$$x_{gj}^* = \frac{\prod_{i=1}^m (b_{ij})^{w_i}}{\sum_{j=1}^n \prod_{i=1}^m (b_{ij})^{w_i}}, b_{ij} = \frac{x_{ij}}{\sum_{j=1}^n x_{ij}} \quad (5)$$

This model is solved by Qiu [12], and the process is rather complex, as it introduces the Lagrange function. We will not repeat this process here. In group decision-making, this model is successfully applied to solve the gathering problem of a certain number of evaluation data on the scheme of a number of experts. The purpose is to sort the different schemes. This paper shifts to assume another angle by sorting the distance of attitude preferences of different groups. Under the same principle, the results can be used to analyze different groups’ evaluations of a certain social problem, namely, calculating the gathered parameter x of the above $n \times m$ matrix.

In social surveys, the number of respondents is at least several hundred—and can be as many as several thousand, tens of thousands or more, which is far greater than the dozens of decision-makers in group decision-making. Thus, in this study, in fact, we do not care about an individual’s preferences and, instead, care about the grouping variables of data, i.e., groups of people’s attributes, such as gender, occupation, etc., that are used to classify groups of people and then to aggregate data. Although the basic theory is based on this model, the application perspective, data aggregation, and interpretations of the conclusions are different.

3.3.2. Distance of Data Measurement Based on the Relative Entropy Method

The preference amount is the final gathering of a certain group of data; although it contains the group preference factors, it remains a central tendency and requires a corresponding parameter to measure its discrete trend.

1. Measurement of the demand degree. Since the preference amount is a relative index, the total preference amount of line i in the table is valued at 1, which is in line with the conditions of the information entropy of Equation (1). Therefore, information entropy can be used to reflect the discrete degree of each row of data from the preference amount of each indicator in each line. Define the preference entropy of any line i as H_g , based on Equation (1):

$$H_g = - \sum_{j=1}^n x_{gj} \ln x_{gj}, \text{ s.t. } \sum_{j=1}^n x_{gj} = 1, x_{gj} > 0 \quad (6)$$

The preference entropy reflects the preference discrete degree of the indicator data evaluated by an individual or group; a smaller value of H_g means that the data are discrete and contain a large amount of information. As reflected in the reality, indicating that there are substantial differences among the survey respondents in the different indicators of the degree of preference, the “love and hate” degree is large, and more information is demanded. Conversely, a larger value of H_g represents an even distribution of data. As reflected in the reality, the differences in the degree of preference indicators between the different respondents is small; they give nearly the same score to the evaluation of each indicator, and less information is demanded. Thus, the reciprocal of preference entropy H_g , i.e., $1/H_g$, can be used to measure the degree of “love and hate”, which is also known as the intensity of demand.

2. Measurement of the distance between the components and the total. In contrast to the mean value, the preference amount is a relative quantity. Thus, its separate data make little sense; in a meaningful amount of preference array, the total array value is 1. A comparison of the different preference amount array proximities is actually a comparison of the distance between the two groups of data distributions. The overall amount of the preference data array is $X_g^* = (X_{g1}^*, X_{g2}^*, \dots, X_{gj}^*)^T$; when setting a certain amount of the preference component array as $X_i^* = (X_{i1}^*, X_{i2}^*, \dots, X_{ij}^*)^T$, the distance can be measured by Equation (2), which belongs to the K-L measure in mathematics. However, using the K-L measure requires accordance with a condition, namely, for any i , there must be $P_i \geq Q_i$ to guarantee a non-negative conclusion. To solve this problem, one can simply take the absolute value of the method, but people often do not use this method in mathematics, instead preferring the method of extraction of a root after squaring. Based on relative entropy theory and mathematical practice, we define two distributions, and the distance D_i of the component X_i^* to the total X_g^* is:

$$D_i = \sqrt{\sum_{j=1}^m (x_{ij}^* \ln \frac{x_{ij}^*}{x_{gj}^*})^2} \quad (7)$$

3. Measurement of the components' centrifugal force or centripetal force. Regarding a component i , w_i is the measurement of the component in the total amount of weight; multiplying the weight by the distance forms a tendency of the component to deviate from the total indicator, called the centrifugal force. Define the centrifugal force of a component, $i = w_i D_i$. Since the distance variable belongs to the fixed distance variables, addition and subtraction can be used instead of multiplication and division; thus, the corresponding centripetal force = $w_i(1 - D_i)$ is defined. The centrifugal force and the centripetal force have two mathematical characteristics after being defined.

First: the sum of the centrifugal force and the centripetal force of a component is the weight of such component, which is $w_i D_i + w_i(1 - D_i) = w_i$.

This feature indicates that the weight is a neutral force that can not only increase the centrifugal tendency, but also increase the centripetal tendency.

Second: although the centripetal force is inversely proportional to the centrifugal force, the order of the centripetal force is not necessarily the reverse order of the centrifugal force. In other words, for a component whose centrifugal force ranks first, its centripetal force may not be last. This feature is clearly derived from the first characteristic and is the result of the weight.

The above two features illustrate the rationality of the definitions of the centrifugal force and the centripetal force, and they provide a method of analyzing the eigenvalue arrays of the general characteristics of a group to deviate or tend toward an array of the eigenvalues of the general characteristics. Such characteristic values can be a preference amount, ratio, or mean value.

4. Discussion and Results

4.1. The Formation of the Measured Variables

Securities market participants in China are divided into the following five groups for the survey: regulators, general investors, listed companies, fund companies, and securities traders. They are required to evaluate the following five indicators.

- ① Do you think the regulatory authorities' past regulatory policy for the stock market has been effective?
- ② What do you think of the effects of a series of policies and measures that were taken when the stock market crashed?
- ③ Do you think the reform of non-tradable shares has proven successful?
- ④ On 20 April 2008, the China Securities Regulatory Commission (CSRC) issued the "Guidance Opinions on Releasing the Transfer of Restricted Stocks of Listed Companies". What do you think of its effects in practice?
- ⑤ Do you think the regulatory policy of the CSRC on market manipulation and insider trading and its implementation have been successful?

To answer, there are four response options, and each is assigned a score, i.e., the number in the bracket.

A. very successful (7); B. failed (1); C. not completely successful (5); and D. not clear (3).

A total of 139 valid questionnaires are obtained, and the data are analyzed.

Due to length restrictions for this paper, not all raw data can be displayed, but selected data are shown to use in subsequent example calculations. Table 1 shows the raw data (use only line 1 as an example). Table 2 lists the raw data obtained from the regulators (19 respondents) to question 1.

Table 1. Raw survey data (use only line 1 as an example).

No.	Individual	Question 1	Question 2	Question 3	Question 4	Question 5	Total
1	Regulator	1	5	3	3	3	15
...
Total	Regulator	63	103	89	83	77	415

Table 2. Raw data of the answers obtained from the regulators (19 respondents) to question 1.

No.	1	2	3	4	5	6	7	8	9	10
Value	1	1	5	5	5	1	5	5	1	1
No.	11	12	13	14	15	16	17	18	19	—
Value	5	5	5	1	1	1	5	5	5	—

4.2. Comparison of the Mean Value and Preference Amount Sorting

The commonly used mean value and the preference amount proposed in this paper are combined in the same table to make a contraction. As shown in Table 1, both the mean value sorting and the preference amount sorting in the study are consistent in total, i.e., question 2 > question 3 > question 5 > question 4 > question 1, which indicates that regardless of the absolute value of the evaluation or the relative preferences, this sorting order is consistent because the mean value stresses the absolute value of each indicator evaluation separately and the preference amount shows the relative degree of preference from the connection between each indicator. The two perspectives agree and further confirm the reliability of this sorting.

Based on the raw data in Tables 1 and 2, the preference amount and the average score of each group’s answers to each question are calculated. Using the regulator as an example, first the ratio of evaluation value of each person *i* to each survey item *j* is calculated, i.e., *b_{ij}*. The *b₁₁* of the first person (*i* = 1) to question 1 (*j* = 1) is:

$$b_{11} = \frac{x_{11}}{\sum_{j=1}^5 x_{1j}} = \frac{1}{15} = 0.06667$$

Then, the individual evaluation value ratio *b_{ij}* is aggregated to the ratio of the group evaluation values. For example, to calculate the ratio of the evaluation value of the regulator group (*g* = 1) to question 1 (*j* = 1), given the total number of regulators is 19 and the weight of each regulator to the whole regulator group is *w₁* = 1/19, the evaluation value ratio of the whole regulator group (*g* = 1) to question 1 (*j* = 1) is:

$$\prod_{i=1}^m (b_{ij})^{w_i} = \prod_{i=1}^{19} (b_{i1})^{w_i} = 0.1294$$

The evaluation value ratios of the five groups (*g* = 1, 2, 3, 4, 5) to the five questions are, thus, calculated as shown in Table 3.

Table 3. The evaluation value ratio of various groups to each question.

Group(<i>g</i>)	Question 1	Question 2	Question 3	Question 4	Question 5	Total
Regulator (<i>g</i> ₁)	0.1294	0.2572	0.2180	0.2022	0.1871	0.9939
General investors (<i>g</i> ₂)	0.1297	0.2165	0.2116	0.1708	0.1663	0.8949
Listed companies (<i>g</i> ₃)	0.1240	0.2431	0.1919	0.1683	0.1565	0.8838
Fund companies (<i>g</i> ₄)	0.1156	0.2594	0.1837	0.1817	0.1861	0.9265
Securities traders (<i>g</i> ₅)	0.1311	0.2380	0.1868	0.1303	0.1791	0.8654

Then, the preference amount of each group is calculated based on Table 3. For example, the preference amount of regulators in question 1 is:

$$x_{11}^* = \frac{\prod_{i=1}^m (b_{ij})^{w_i}}{\sum_{j=i=1}^n \prod_{i=1}^m (b_{ij})^{w_i}} = x_{11}^* = \frac{\prod_{i=1}^{19} (b_{i1})^{w_i}}{\sum_{j=i=1}^5 \prod_{i=1}^{19} (b_{ij})^{w_i}} = \frac{0.1294}{0.9939} = 0.130$$

Based on the data in Table 2, the average regulator evaluation score is $(63/19) = 3.32$, and the remainder of the data is calculated in the same way. The preference amounts and average scores are listed in Table 4 for comparison.

Table 4. Evaluation indicators of the five types of market participants for the securities regulatory problems.

		Q1	Q2	Q3	Q4	Q5
Regulators (19 persons)	Preference amount	0.130	0.259	0.219	0.203	0.188
	Mean value	3.32	5.42	4.68	4.37	4.05
General investors (40 persons)	Preference amount	0.145	0.242	0.236	0.191	0.186
	Mean value	3.25	4.50	4.15	3.55	3.65
Listed Companies (32 persons)	Preference amount	0.140	0.275	0.217	0.190	0.177
	Mean value	3.06	4.88	3.88	3.44	3.44
Fund companies (26 persons)	Preference amount	0.125	0.280	0.198	0.196	0.200
	Mean value	3.23	5.69	4.15	4.15	4.15
Securities traders (22 persons)	Preference amount	0.152	0.275	0.216	0.151	0.207
	Mean value	3.64	5.73	4.36	3.18	4.36
Total (139 persons)	Preference amount	0.139	0.264	0.219	0.187	0.190
	Mean value	3.27	5.13	4.19	3.69	3.86

As can be seen from Table 4:

- Average score conclusion: question 2 > question 4 > question 6 > question 5 > question 1
- Preference conclusion: question 2 > question 4 > question 6 > question 5 > question 1
- In most cases, the order of each indicator sorted by average score and by preference are the same.

4.3. Main Contradiction Found through a Significance Test

The analysis of variance shows that the average evaluation scores of the five questions from different groups have no significant difference. Using the method of the independent sample ratio difference test, a two-difference significance test is conducted on the preferences of the five types of groups on the different respective issues (taking a level of significance of 0.05), and no significant difference is found. Then, a conclusion is reached; from the two aspects of the absolute value of the mean value and the relative evaluation of the preference amount, the five groups of market participants have no significant differences in the evaluation of each question, which shows that the attitudes of the five groups are basically the same.

Then, consider the differences among the five questions from the two aspects of the mean value and the preference amount. Student's *t*-test is used to test the mean value difference paired with the various questions, and the preference amount is a ratio that can also be tested by the significance in the difference of the ratios. For the sake of comparison, the test results are combined in Table 5.

As shown in Table 5, the five questions have a total of 10 combinations. Regarding the mean value, there are only three pairs of questions, i.e., question 1 and question 4, question 3 and question 5, and question 4 and question 5. There is no significant difference between them, and the other seven combinations of questions have significant differences. Regarding the amount of preference, only three combinations of questions have a significant difference, and the mean values between the three combinations also have a significant difference, i.e., only the following three question combinations: question 1 and question 2, question 1 and question 3, and question 2 and question 4. From the two aspects of the absolute value of the score and the relative preferences, the differences are significant. It can be considered that the gap between the three pairs of questions is the main contradiction of the entire question system, and a qualitative explanation and analysis should be prioritized in the case of the three pairs of questions.

Table 5. The difference test between the mean value and the difference test of the preference amount for all of the questions.

Question Combinations	Mean Value			Preference Value		
	Difference b/w the Two	Sig. (Two-Tailed)	Differences	Difference b/w the Two	Sig. (Two-Tailed)	Differences
Q1–Q2	−1.856	0.000	Significant	−0.125	0.000	Significant
Q1–Q3	−0.921	0.000	Significant	−0.08	0.005	Significant
Q1–Q4	−0.417	0.078	Not significant	−0.048	0.095	Not significant
Q1–Q5	−0.590	0.018	Significant	−0.051	0.075	Not significant
Q2–Q3	0.935	0.000	Significant	0.045	0.245	Not significant
Q2–Q4	1.439	0.000	Significant	0.077	0.052	Significant
Q2–Q5	1.266	0.000	Significant	0.074	0.061	Not significant
Q3–Q4	0.504	0.003	Significant	0.032	0.373	Not significant
Q3–Q5	0.331	0.075	Not significant	0.029	0.418	Not significant
Q4–Q5	−0.173	0.312	Not significant	−0.003	0.928	Not significant

4.4. Characteristic Analysis of Different Groups

If the analysis is based solely on the mean value system, then when the analysis of variance shows that the differences in the scores of different groups are not significant, the conclusion is that the difference of each group is not significant, and the analysis ends there. This conclusion is clearly not objective and is not consistent with reality. This problem can be solved by the indicators constructed in this paper, i.e., preference entropy, center distance, centrifugal force, and centripetal force.

Based on the data in Table 4, and calculated according to the above formula, the preference entropy of the regulator group (g = 1) is:

$$\begin{aligned}
 H_1 &= - \sum_{j=1}^n x_{1j} \ln x_{1j} \\
 &= -(0.13 \ln 0.13 + 0.259 \ln 0.259 + 0.219 \ln 0.219 + 0.203 \ln 0.203 + 0.188 \ln 0.188) \\
 &= 1.59
 \end{aligned}$$

The center distance of the regulator group is:

$$\begin{aligned}
 D_1 &= \sqrt{\sum_{j=1}^m (x_{1j}^* \ln \frac{x_{1j}^*}{x_{1j}^*})^2} \\
 &= \sqrt{(0.13 \times \ln \frac{0.13}{0.139})^2 + (0.259 \times \ln \frac{0.259}{0.264})^2 + (0.219 \times \ln \frac{0.219}{0.219})^2 + (0.203 \times \ln \frac{0.203}{0.187})^2 + (0.188 \times \ln \frac{0.188}{0.190})^2} \\
 &= 0.02
 \end{aligned}$$

The parameters of the other groups are calculated in the same way, producing the data in Table 6.

Table 6. Discrete distribution of the five group’s data.

Groups	Ratio (w _i)	Preference Entropy (H _g)	Center Distance
Regulators	0.137	1.59	0.020
General investors	0.288	1.59	0.029
Listed companies	0.23	1.58	0.017
Fund companies	0.187	1.58	0.032
Securities traders	0.158	1.58	0.041
Total	1	1.59	0.000

As shown in Table 6, the group with the largest center distance is the group of securities traders, i.e., the group of securities traders shows significant differences with the other groups. This finding is consistent with the actual qualitative experience, as all of the other groups (outside of the group of securities traders) basically want the stock market to move steadily upward, whereas the securities traders hope that the stock market has relatively frequent shocks. Thus, this method reveals the deep psychological differences between different groups.

5. Conclusions

5.1. Relative Entropy Theory Solves the Quantitative Measure of Group Preference

Since Arrow [4] proposed his definition of preference, how to quantitatively measure this preference has become a real problem. In social science research, traditional statistical methods based on mean values and standard deviations cannot solve this problem. Relative entropy theory proposes a method of solving this problem.

5.2. Preference Entropy and Center Distance are the Specific Methods for Measurements

Based on similarity entropy theory, in this paper, we propose the two indicators of preference entropy and the center distance. Of these, preference entropy can be used as the degree of preference of the different choice items by groups, and the center distance clusters the different groups based on the preference degree, which can measure the distance between different groups from the perspective of social attitudes.

5.3. The Empirical Research Has Been Successful

By using this method to study the attitudes toward regulations of participants in China securities market, we find that the securities trader group and the other groups have relative particularity, which is in accordance with practical experience.

Acknowledgments: The study is supported by the 60th China postdoctoral Science Foundation funded project “Research on Multi-day Activity Scheduling and Travel Simulation System”; National Social Science Fund Project “Research on the effectiveness of China’s capital market regulation from the perspective of game theory” (05BJL028); Beijing Union University “New Starting Point” project “Research on the balance of Beijing-Tianjin-Hebei” “Green Innovation Industrial Park” (Sk10201503).

Author Contributions: Shiyu Zhang and Wenzhi Liu conceived and designed the experiments; Qin He performed the experiments; Shiyu Zhang and Wenzhi Liu analyzed the data; Xuguang Hao contributed analysis tools; Wenzhi Liu wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fowler, F.J. *Improving Survey Questions: Design and Evaluation*; SAGE Publications: Thousand Oaks, CA, USA, 1995.
2. Hao, X. A study on the effectiveness of China’s securities market regulation. *China Ind. Econ.* **2011**, *6*, 16–25.
3. Hao, X.; Zhu, B.; Zhang, S. Research on the effectiveness of China’s securities market regulatory policy: Based on the analysis of a questionnaire survey. *Manag. World* **2012**, *7*, 44–53.
4. Arrow, K.J. *Social Choice and Individual Values*; Yale University Press: New Haven, CT, USA, 1963.
5. He, D. Research on the application of entropy in data analysis. *Stat. Decis. Mak.* **2005**, *8*, 27–29.
6. Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Academic Press: New York, NY, USA, 2001.
7. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [[CrossRef](#)]
8. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [[CrossRef](#)]
9. Gray, R.M. *Entropy and Information Theory*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2011.
10. Chang, C.I.; Du, Y.; Wang, J.; Guo, S.M.; Thouin, P.D. Survey and comparative analysis of entropy and relative entropy thresholding techniques. *IEE Proc. Vis. Image Signal Proc.* **2006**, *153*, 837–850. [[CrossRef](#)]
11. Baratpour, S.; Ahmadi, J.; Arghami, N.R. Entropy properties of record statistics. *Stat. Pap.* **2007**, *48*, 197–213. [[CrossRef](#)]
12. Qiu, W. *Management Decision and Applied Entropy*; Machinery Industry Press: Beijing, China, 2002.
13. Zhong, Y. *Principles of Information Science*, 3rd ed.; Beijing University of Posts and Telecommunications Press: Beijing, China, 2002.
14. Kadane, J.B.; Krishnan, R.; Shmueli, G. A data disclosure policy for count data based on the COM-Poisson distribution. *Manag. Sci.* **2006**, *52*, 1610–1617. [[CrossRef](#)]

15. Jose, V.R.R.; Nau, R.F.; Winkler, R.L. Scoring rules, generalized entropy, and utility maximization. *Oper. Res.* **2008**, *56*, 1146–1157. [[CrossRef](#)]
16. Zhao, K.; Karsai, M.; Bianconi, G. Entropy of dynamical social networks. *PLoS ONE* **2011**, *6*, e28116. [[CrossRef](#)] [[PubMed](#)]
17. Gandica, Y.; Charmell, A.; Villegas-Febres, J.; Bonalde, I. Cluster-size entropy in the axelrod model of social influence: Small-world networks and mass media. *Phys. Rev. E* **2011**, *84*, 046109. [[CrossRef](#)] [[PubMed](#)]
18. Smith, D.B.; Stettler, H.; Beedles, W. An investigation of the information content of foreign sensitive payment disclosures. *J. Account. Econ.* **1984**, *6*, 153–162. [[CrossRef](#)]
19. Baez, J.C.; Pollard, B.S. Relative entropy in biological systems. *Entropy* **2016**, *18*, 46. [[CrossRef](#)]
20. Dziurosz-Serafinowicz, P. Maximum relative entropy updating and the value of learning. *Entropy* **2015**, *17*, 1146–1164. [[CrossRef](#)]
21. Pan, W.; She, K.; Wei, P. Preference inconsistency-based entropy. *Entropy* **2016**, *18*, 96. [[CrossRef](#)]
22. Makowski, M.; Piotrowski, E.W.; Śładkowski, J. Do transitive preferences always result in indifferent divisions? *Entropy* **2015**, *17*, 968–983. [[CrossRef](#)]
23. He, D.; Xu, J.; Chen, X. Information-theoretic-entropy based weight aggregation method in Multiple-Attribute Group decision-making. *Entropy* **2016**, *18*, 171. [[CrossRef](#)]
24. Iazzi, A.; Vrontis, D.; Trio, O.; Melanthiou, Y. Consumer preference, satisfaction, and intentional behavior: Investigating consumer attitudes for branded or unbranded products. *J. Transnatl. Manag.* **2016**, *21*, 84–98. [[CrossRef](#)]
25. Ragul'skii, A.D. Consumer's preference: Psychological attitude and dynamic modeling. *Econ. Anal.* **2014**, *41*, 59–67.
26. Pistor, K.; Xu, C. Governing emerging stock markets: Legal vs. administrative governance. *Corp. Gov.* **2005**, *13*, 5–10. [[CrossRef](#)]
27. Agrawal, R.; Imieliński, T.; Swami, A. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 25–28 May 1993; pp. 207–216.
28. Hwang, C.L.; Lin, M.L. *Group Decision Making Under Multiple Criteria*; Springer: Berlin/Heidelberg, Germany, 1987.
29. Toque, C.; Terraza, V. Time series factorial models with uncertainty measures: Applications to ARMA processes and financial data. *Commun. Stat. Theory Methods* **2011**, *40*, 1533–1544. [[CrossRef](#)]
30. Hu, M.; Liang, H. Adaptive multiscale entropy analysis of multivariate neural data. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 12–15. [[PubMed](#)]
31. Hollisaaz, M.T.; Khedmat, H.; Effatmanesh-Nik, M.; Yousefvand, M.; Mansouri, S.; Saadat, S.H.; Rafati-Shaldehi, H.; Ebrahiminia, M. Data-entropy analysis of renal transplantation data. *Transplant. Proc.* **2007**, *39*, 930–931. [[CrossRef](#)] [[PubMed](#)]
32. Van Wieringen, W.N.; van der Vaart, A.W. Statistical analysis of the cancer cell's molecular entropy using high-throughput data. *Bioinformatics* **2011**, *27*, 556–563. [[CrossRef](#)] [[PubMed](#)]
33. Richman, J.S. Sample entropy statistics and testing for order in complex physiological signals. *Commun. Stat. Theory Methods* **2007**, *36*, 1005–1019. [[CrossRef](#)]
34. Razmkhah, M.; Morabbi, H.; Ahmadi, J. Comparing two sampling schemes based on entropy of record statistics. *Stat. Pap.* **2012**, *53*, 95–106. [[CrossRef](#)]

