Comprehensive Scores of Legal Risk by Using Clustering Analysis

Ying Zhang
Southwest Petroleum University,
Chengdu, China
1186752815@qq.com

Dongsheng Xu
Southwest Petroleum University,
Chengdu, China

Juanjuan Geng
Southwest Petroleum University
Chengdu, 610500, China

Abstract—To evaluate the legal risk of big enterprise in China for the sake of avoiding unnecessary loss, multivariate statistical analysis method was adopted. In the evaluation process, a criterion which contains 967 observation points was raised. All the 967 observation points were carried out by priority setting and risk assessment. To complete the evaluation process hierarchical clustering analysis was applied. It can be concluded that there is no possibility of significant legal risk happening in the enterprise, but its potential legal risk should not be ignored. The result of this paper is a reminder for the enterprise to pay attention to its legal risk. The case showed in the paper indicates that the method used was feasible to evaluate the legal risk. Thus, the enterprise is able to reduce its legal risk in accordance with the provided result.

Keywords—Legal risk; Clustering analysis; Minimum variance unbiased estimation; Distance

I. INTRODUCTION

When the legal risk of an enterprise was mentioned, it refers to the disadvantage impact on its business objectives with certain background, such as the legal provisions, the regulatory requirements or contract, also with the change of the external environment and internal environment, or sometimes even the act or omission of the enterprise and its stakeholders. In this paper, research of quantitative analysis has been carried out on the legal risk severity of the same level, the severity of the legal risk and the possibility of legal risk happening three aspects, by means of clustering analysis, we assessed the level of the legal risk of the enterprise, thus the enterprise can realize the current legal risk which deficiencies and defects.

II. THE DESCRIPTION OF LEGAL RISK SEVERITY

At first, 967 risk points were ascertained. For each observation point, research is carried out on three aspects: criminal legal risk, administrative legal risk and civil legal risk. The criterions to separate different levels of legal risk are listed as follows. [1]-[3]

A. *Loss of property*

The following guidelines are used to measure administrative legal risk:

More than 1,000,000 RMB is significant;
250,000 RMB to 1,000,000 RMB is important;
Less than 250,000 RMB is general.

The following guidelines are used to measure civil legal risk:

More than 2,000,000 RMB is significant;
500,000 RMB to 2,000,000 RMB is important;
Less than 500,000 RMB is general.

B. *Range of influence*

The following principles are adopted:

Regard influence range is bigger than the region as significant;

Regard influence range is the entire region as important;

Regard Influence range is only part of the region as general.

C. *Other circumstances*

The risk of administrative law which involves revoking business license, stopping production or closing down, is significant; The risk of criminal law, which involves the enterprise's main leaders is important;

The other situations are general.

The second step, consideration should be given to different kinds of legal risks and different severity. Scores are quantified as follows(table 1):

TABLE I. SCORES OF LEGAL RISK SEVERITY

	significant	important	general
Risk of criminal law	10	8	0
Risk of administrative law	8	5	2
Risk of civil law	7	4	1

According to above criterion, considering the 967 observation points which selected from the reality of the

enterprise, if there is possibility of risk of criminal law, the following should be assigned:[4]-[5]

$$x_{i1} = \begin{cases} 10, & \text{significant possibility of risk of criminal law,} \\ 0, & \text{no significant possibility of risk of criminal law,} \end{cases} i=1,2,\text{L } 967$$

$$x_{i2} = \begin{cases} 8, & \text{important possibility of risk of criminal law,} \\ 0, & \text{no important possibility of risk of criminal law} \end{cases} i=1,2,\text{L } 967$$

If there is possibility of risk of administrative law, the following should be assigned:

$$y_{i1} = \begin{cases} 8, & \text{significant possibility of risk of administrative law,} \\ 0, & \text{no significant possibility of risk of administrative law,} \end{cases} i=1,2,\text{L },967$$

$$y_{i2} = \begin{cases} 5, & \text{important possibility of risk of administrative law,} \\ 0, & \text{no important possibility of risk of administrative law,} \end{cases} i=1,2,\text{L },967$$

$$y_{i2} = \begin{cases} 2, & \text{general possibility of risk of administrative law,} \\ 0, & \text{no general possibility of risk of administrative law,} \end{cases} i=1,2,\text{L },967$$

If there is possibility of risk of civil law, the following should be assigned:

$$z_{i1} = \begin{cases} 7, & \text{significant possibility of risk of civil law,} \\ 0, & \text{no significant possibility of risk of civil law,} \end{cases} i=1,2,\text{L },967$$

$$z_{i2} = \begin{cases} 4, & \text{important possibility of risk of civil law,} \\ 0, & \text{no important possibility of risk of civil law,} \end{cases} i=1,2,\text{L },967$$

$$z_{i3} = \begin{cases} 1, & \text{general possibility of risk of civil law,} \\ 0, & \text{no general possibility of risk of civil law,} \end{cases} i=1,2,\text{L },967$$

III. THE DESCRIPTION OF LEGAL RISK HAPPENING POSSIBILITY

In practice, although the enterprise is operating well in management at present, together with the support of country and society, the profit of the enterprise is considerable, there still exist a lot of observation points which have potential legal risk but exposed no legal risk at present, which brings difficulties to ascertain the legal risk of the enterprise. Therefore, the following aspects has been creatively put forward to examine legal risk: ① the supervision enforcement of the external; ② the maturity degree of the internal system; ③ the legal degree of its personnel; ④ the comprehensive condition of its stakeholders; ⑤ the activity frequency of its business.

Then 7 experts independently ratings for each sub-project, according to the criterion below:

TABLE II.   SCORE OF THE POSSIBILITY OF LEGAL RISK

The probability	Scores	The basis of score
Significant	5	1.the supervision enforcement of the external; 2.the consumption degree of the internal system; 3.the legal quality of personnel; 4.the comprehensive condition of the stakeholders 5,the activity frequency of its business
Important	4	
General	3	
Minor	2	
Minimal	1	

$$V_{2ij}, i=1,2,\text{L },967; j=1,2,\text{L } 7$$

Where V_{2ij} was the number j expert score for the i observation point.

The description of loss ratio of legal risk

The loss ratio of legal risk could be expressed as:

$$V_{3i} = \frac{the \text{ actual loss}}{\text{the theoretical loss}} \times 100\% \quad i=1,2,\text{L } 967$$

Its value is similar to the average value of the possibility of legal affairs happening, of which numerical interval is from 1% to 100%.

IV. SEVERITY CALCULATION OF LEGAL RISK OF THE SAME LEVEL

In theory, at a certain observation point, there is the possibility of risk of criminal law, risk of administrative law and risk of civil law happening at the same time. But in practice, people always particularly value the highest sentencing. Furthermore, risk evaluation reflect a certain degree of differentiation, if using the sum of different kinds of legal risk of the same level to calculate the legal risk is likely to lead to the next level of risk in value over the former level, which is not in conformity with the actual situation. Therefore, for the same observation point, when different types of legal risk occur simultaneously, the highest score should be chosen.[6]

To conclude: when a certain observation point has a "significant" level of risk,

$$r_{i1} = \max\{x_{i1}, y_{i1}, z_{i1}\}, i=1,2,\text{L },967$$

When a certain observation point has an "important" level of risk,

$$r_{i2} = \max\{x_{i2}, y_{i2}, z_{i2}\}, i=1,2,\text{L },967$$

When a certain observation point has a "general" level of risk,

$$r_{i3} = \max\{x_{i3}, y_{i3}, z_{i3}\}, i=1,2,\text{L },967$$

Comprehensive score of legal risk severity

In practice, when an observation point has the possibility of legal risk, different severity of the consequences is independent, so the comprehensive legal risk severity score should have been the three level severity scores combined ,i.e.[7]

$$V_{1i} = r_{i1} + r_{i2} + r_{i3} \text{ , i=1,2,L ,967}$$

Comprehensive score of happening possibility

There is certain randomness of subject scoring, but it can be proved that the arithmetic mean is the minimum variance unbiased estimation of the comprehensive score, hence

$$V_{2i} = \frac{1}{7}\sum_{j=1}^{7} V_{2ij} \text{ ,i=1,2,L ,967}$$

Where V_{2ij} was the number j expert score for the i observation point.

V. THE COMPREHENSIVE SCORE OF LEGAL RISK

Considering the three dimensions assuming the severity dimension is V_{1i} , the possibility dimension is V_{2i}, the loss rate dimension is V_{3i} ,then the comprehensive score of legal risk could be written as:

$$H_i = V_{1i} \times V_{2i} \times V_{3i}$$

A. Clustering theory

The clustering analysis mainly solves the problem of classification of multi-index sample. It is natural to consider the dimension of severity of the above as V_{1i}, to consider the dimension of possibilities as V_{2i}, the dimension of loss rate as V_{3i} . The problem being researched can be considered as hierarchical clustering of 3 different indexes of a sample of 967, according to the group-average method, considering the pseudo F-statistic and pseudo T^2-optimal, the best result is to divide the 967 observation points into 4 groups. Then the following two problem appears: each group containing roughly the same amount of sample; in each group, the comprehensive score of different observation points are very scattered. [8]-[10]

It can be concluded that such practice is not reasonable, because in the process of hierarchical clustering, the three indicators itself is independent from each other. So the relative change determines the size of the "distance" between two groups. However, the dimension of the severity changes from 1 to 20, forms the miraculous interval of 20 times. While the possibility ranges only from 1 to 4, 4 times of the largest gap, so its utility is severely reduced.

From the above study, it can be drawn that the comprehensive score of legal risk is a result combined three indications which have obvious practical significance. Especially on the risk of civil law problems, this could be directly transformed into the amount of money. [11]

B. Determination of the distance

The following provides useful formulas most frequently used in single-index clustering problem:

Euclidean distance:
$$d_{ij}(\text{P}) = \left[\sum_{k=1}^{p} (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}}$$

Minkowski distance:

$$d_{ij}(q) = \left[\sum_{k=1}^{p} (x_{ik} - x_{jk})^q \right]^{\frac{1}{q}}$$

Chebyshev distance: $d_{ij}(\infty) = \max\left| x_{ik} - x_{jk} \right|$

Mahalanobis distance:
$$d_{ij}(M) = (x_i - x_j)' \sum^{-1} (x_i - x_j)$$

where Σ was covariance matrix of p-dimention random vector.
Here we employ the distance of

$$d_{ij}(p) = \sum_{k=1}^{p} \left| x_{ik} - x_{jk} \right| \text{ , for } p = 1 \text{, so the distance we used}$$
could be expressed as

$$d_{ij} = \left| H_i - H_j \right| \text{ , } \forall 1 \le i < j \le 967$$

VI. CLUSTERING METHOD

Firstly comprehensive score of legal risk of 967 observation points is recorded in the SAS system considering these scores as 967 "classes" （In order to reduce the amount of calculation we can properly order the comprehensive scores of the legal risk）

Secondly calculating $\left| H_i - H_j \right|$, $\forall 1 \le i < j \le 967$ （If you have already ordered, subtracting adjacent two number）, put the value of the smallest group as a class; if the minimum value appears more than one time, these "class" should be merged.

The third step the already combined group and no combined class are considered as some new classes. If the "original class" is not merged the center of class will not change. If the "original class" has already been merged, then take the average value of all the scores as its new center.

Repeat the above process, until all the 967 scores merged in a "class".

According to the maximum distance variation principle, we can determine the number of "class" and its corresponding classification.

VII. RESULTS

We apply this idea to program and calculate in the SAS statistical software. The result can be seen below. (Fig. 1)At the same time we can obtain the 5 groups as the optimal classification.
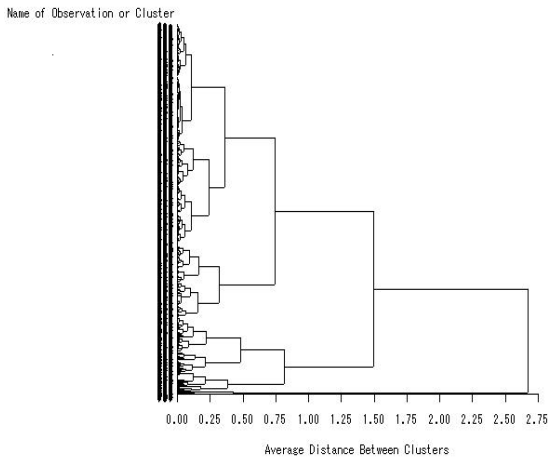
Fig. 1.   pedigree chart of cluster analysis

At the same time we can obtain the clustering process (Only the last few clustering process are chosen) as follows:

```
Norm  T
RMS  i
NCL ----Clusters Joined---FREQ PSF PST2 Dist. e
9 CL15 CL16  196  4299   613  0.3174
8 CL24 CL10  570  2158  1745  0.3602
7 CL12 CL14   51  2406   117  0.3807
6 CL18 CL79    8  2857  94.2  0.4259
5 CL11 CL13  142  2859   543  0.4839
4 CL8  CL9   766  1115  2035  0.7407
3 CL5  CL7   193  1317   384  0.8176
2 CL4  CL3   959   119  2230  1.4979
1 CL2  CL6   967    .    119  2.675
```

After hierarchical clustering, it can be concluded:
The first group: 0.1—9.3, 485 observation points;
The second group: 9.3—24.0, 298 observation points;
The third group: 24.0—41.0, 133 observation points;
The fourth group: 41.0—57.5, 43 observation points;
The fifth group: more than 57.5, 8 observation points.

VIII. CONCLUSION

In this paper, we apply the scores of 967 observation points to hierarchical clustering method. We obtain that there is no possibility of significant legal risk happening to the enterprise at present, but from the past history we can learn that many enterprises experienced crisis for overlooking petty events. It is caused by that little probability event may lead to big probability event with the time going by. So the result is a reminder for the enterprise to notice little probability event. Consider all the 3 dimensions that include the dimensions of severity, happening possibility and loss rate. The dimension of happening possibility is controllable. Accordingly, the enterprise should check the supervision enforcement of the external, the consumption degree of the internal system, the legal quality of personnel, the comprehensive condition of the stakeholders, and the activity frequency of business regularly.

REFERENCES

[1] Ronald P. Cody, Jeffrey K. Smith.(2011). Applied Statistics and the SAS Programming Language, 5e .People's post press.

[2] M. Halkidi , Y. Batistakis, and M. Vazirgiannis. (2002).Clustering algorithms and validity measures.IEEE Transactions on pattern analysis and machine intelligence, 24.

[3] Liming Wang. (2001).Property law legislation of a number of issues. Politics and Law Forum, 4.

[4] Xiaoqun He.Multivariate Statistical Analysis. (2004).China Renmin University Press,

[5] Fei Xiang,Youchun Chen. (2006).Enterprise legal risk assessment,Law Press.

[6] Weiguang Wang. (2010).Construction of enterprise legal risk and prevention systerm. Modern Enterprise Culture,19.

[7] A. Ben-Hur, D. Horn, H. Siegelmann, and V. Vapnik. (2001).Support vector clustering. Journal of Machine Learning Research , 2,125-137.

[8] K. Fish and P. Ruby. (2009).An artificial intelligence foreign market screening method for small businesses. International Journal of Entrepreneurship, 13, 65–91.

[9] S. Wang. (2001).Cluster analysis using a validated self-organizing method: cases of problem identification.International Journal of Intelligent Systems in Accounting, Finance and Management, 10, 127–138.

[10] J. T. Wei, S. Y. Lin, and H. H. Wu. (2010).A review of the application of RFM model, African Journal of Business Management, 4, 4199–4206.

[11] Hsin-Hung Wu, Shih-Yen Lin, Chih-Wei Liu.(2014).Analyzing Patients' Values by Applying Cluster Analysis and LRFM Model in a Pediatric Dental Clinic in Taiwan, The Scientific World Journal, 14.