Research on modeling of government debt risk comprehensive evaluation based on multidimensional data mining

Li ChaoYing1 · Wu Xiang Da2 · Zhao En Hui3

Accepted: 21 October 2021 / Published online: 1 December 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
In order to solve the problems of low accuracy of data mining, high relative error rate of evaluation, and long time of evaluation in traditional government debt risk evaluation methods, this paper proposes a modeling method of government debt risk comprehensive evaluation based on multidimensional data mining. The MAFIA algorithm is used for multidimensional mining of government debt risk data, and K-means clustering algorithm is used for clustering processing of mined data. The KMV model is built based on the clustering findings, and the uncertainty factor is utilized to alter the model in order to provide a complete assessment of government debt risk using the modified KMV model. The experimental results show that the accuracy rate of government debt risk data mining is always above 91%, the relative error rate of evaluation is always below 3.4%, and the average evaluation time is 0.71 s, the practical application effect is good.

Keywords Multidimensional data mining · Government debt risk · Comprehensive evaluation · Evaluation modeling · KMV model

1 Introduction

On a global scale, the scale of government debt has become a threat to national economic development, the important factors and with the onset of the global financial crisis, the economic development of many countries has suffered serious impact, and some developed countries governments to ease financial and economic crisis impact on the development of national economy and threat (Xu et al. 2020). However, the implementation of these government fiscal policies has also had a negative impact, resulting in a significant increase in the sources of financial risks. Plus with the rapid development of social economy and accelerating urbanization process, increase infrastructure construction project, makes the government demand for funding growth is nonlinear, which have increased the government’s fiscal pressure, so the borrowing will be looking for ways to solve these problems, such as (Zhang et al. 2021), therefore makes the government debt risk increases. Therefore, in order to further measure the risk of government debt, it is of great significance to study a new government debt risk evaluation method (Croce et al. 2021) (Talha 2020).

At present, for the government debt risk assessment studies have made certain progress and have produced many excellent research results, such as reference (ChaoYing et al. 2021) proposed a government debt risk evaluation method based on pressure-state-response model, this method from the present situation of local government debt as a research foundation, analysis of government debt pressure, debt, debt paying ability, and other indicators, and analyzes the logical relationship between various indicators, in order to build the relevant evaluation index system, based on the analytic hierarchy process (AHP) is used to analyze the evaluation index weight calculation, according to the weight calculation results build pressure-state-response model, in order to complete the government debt risk assessment, but the government debt risk existing in the method of data mining to solve the problem of low accuracy the practical effect is not good. Reference (Fang...
proposed a local government debt risk evaluation method based on the central point triangle whitening weight function, the method by investigating and collecting a number of provinces and cities government financial data, and according to these data to construct the evaluation system of local government debt risk, according to the result of evaluation index weight calculation of center of building triangular whitening weight function. The function is used to enhance the conventional risk assessment model, which is then used to estimate the risk of municipal debt (Talha et al. 2021). However, it is found in the practical application that this method has the problem of low relative error rate of risk assessment of local government debt, and there is a big gap with the ideal application effect. Reference (Wenwei 2019) proposed a local government debt risk evaluation method based on factor analysis, this method will debt pressure, solvency, growth potential as the first-level indicators, in order to build the corresponding evaluation index system of local government debt risk, and by using the factor analysis method to calculate the weight of each evaluation index, to achieve the debt risk assessment and early warning interval. However, this method has a certain complexity, resulting in an increase in the time of risk evaluation of local government debt.

In order to solve the problems existing in the above methods, this paper puts forward a new modeling method of government debt risk comprehensive evaluation based on multidimensional data mining and verifies its application performance in government debt risk evaluation through experiments (Talha et al. 2020).

2 Design of modeling method for comprehensive evaluation of government debt risk

2.1 Data collection and processing based on multidimensional data mining

The data mining method refers to a process in which some computer methods are used to select useful data from massive data according to certain rules. Because the chosen data have a possibility for correlation in this procedure, it is also known as data mining (Loey 2020). This article enhances the data mining technique and offers a new multidimensional data mining method to accomplish the relevant research goals in order to improve the complete assessment accuracy of government debt risk.

It is assumed that the standard set of government debt risk data is $R = \{a_1, a_2, ..., a_n\}$, which $a_i$ is the standard value of the ith data source characteristic, the upper floating value is $u_i$, the lower floating value is $l_i$, in which $n$ is the total dimension of government debt risk data (Mengash 2020). On this basis, the government debt risk data set is assumed to be $D = \{r_1, r_2, ..., r_n\}$, where $r_i = \{x_1, x_2, ..., x_n\}$ represents the data record of the ith government debt risk data source and $x_i$ is the actual value of the characteristics of the ith data source.

Combined with the above analysis, in order to have a clearer understanding of the differences between the data in the above two different data sets, it is necessary to calculate the difference between the standard value and the real value and compare the degree of change between the value and the designed floating value. The specific calculation formula is as follows:

$$ t = \begin{cases} 3, & 0.6 \leq \sigma \leq 1.0 \\ 3, & 0.3 \leq \sigma < 0.6 \\ 1, & 0.1 \leq \sigma < 0.3 \\ 0, & 0 \leq \sigma < 0.1 \\ 4, & -0.1 \leq \sigma < 0 \\ 5, & -0.3 \leq \sigma < -0.1 \\ 6, & -0.6 \leq \sigma < -0.3 \\ 7, & -1.0 \leq \sigma < -0.6 \end{cases} $$

Then, the difference between the real value and the standard value $\sigma$ can be calculated as follows:

$$ \sigma = \frac{x - a}{u} $$

In the above formula, $a$ represents the discretization coefficient and $u$ represents the data standard value.

The MAFIA algorithm is used in conjunction with the aforementioned study to perform multidimensional mining of government debt risk data. To complete the process of creating candidate item sets (Wu 2020), depth first search is needed, and this process is represented as a candidate item set tree. A detailed description of the generation process is shown in Fig. 1.

With each step of tree generation, a single item is expanded into a collection of multiple items. Among them, when the number of items in the item set increases

![Fig 1 Candidate item set tree generation process](image-url)
gradually, the confidence level of the item set decreases accordingly. Finally, the support will be lower than the minimum support of the frequent item set (Talha et al. 2020). When the support is lower than the minimum support of the frequent item set, data mining will be stopped and the government debt risk data will be stored as a series of vertical bitmaps, as shown in Fig. 2.

According to Fig. 2, bits in each bitmap indicate whether or not this piece of data has a matching item set. A bitmap only relates to one item set in the beginning. As a consequence, the “and” calculation between bitmaps may be used to get the frequency calculation result of a single item set or multinomy set. (Liu 2020).

Combined with the above analysis, the process of applying MAFIA algorithm to obtain frequent item sets in government debt risk data is as follows:

1. Calculate the discrete value set of the change degree of government debt risk data according to the above formula, as shown in Table 1.

2. Change the set of discrete values in Table 1 according to the order, and there is only one number for each set.

3. Set the minimum value of support degree. When the support degree of an item set is less than this value, a relatively frequent item set is obtained from the set of discrete values with varying degrees by combining the MAFIA algorithm. This set is the data set \( F \) that is most closely related to the government debt risk (Moayedi et al. 2021).

After the multidimensional government debt risk data mining is completed, the data must be clustered. This study utilizes the K-means algorithm to cluster government debt risk data since it is a partition-based clustering technique that has been extensively used in practice and got positive feedback. The following is the procedure for data clustering (Kumalasari et al. 2020):

Combined with the data mining results, the preliminary cluster set \( T \) formed by DBScan was used to calculate the cluster centroid set \( M = \{m_1, m_2, \ldots, m_k\} \) according to \( T \), and \( k \) feature vectors \( x_1, x_2, \ldots, x_k \in M \) were randomly selected from \( M \) as the initial cluster center.

According to the above calculation formula, the relevant formula is used to call the distance from \( n \)-dimensional centroid to each initial cluster center \( m_i \). The specific calculation formula is as follows:

\[
d_{(i,j)} = \sqrt{\sum_{i=1}^{n} \left( \frac{m_{H_i} - x_{H_j}}{s_i} \right)^2}
\]

(3)

The centroid is grouped into the cluster closest to the point (Liu et al. 2021), and the particular computation is given by the formula:

\[
c_i = \arg \min |d_{(i,j)}|
\]

(4)

Combined with the above analysis, the average coordinate of all points in the clustering process is calculated, and a new clustering center is obtained according to the value. The specific formula is described as follows:

\[
x_j = \frac{\sum_{i=1}^{m} \{c_i = j\} m_i}{\sum_{i=1}^{m} \{c_i = j\}}
\]

(5)

The above process is repeated continuously. When the clustering center no longer moves in a large range and meets the convergence criterion function, the clustering is stopped and the clustering results of government debt risk data are output (14–15). This function is described as follows:

\[
J(c, x) = \sum_{i=1}^{m} \|m_i - x_c\|^2
\]

(6)

### 2.2 Realization of government debt risk comprehensive evaluation modeling

The KMV model is built based on the clustering findings, and the uncertainty factor is utilized to alter the model in order to provide a complete assessment of government debt risk using the modified KMV model. In practice, the KMV model is a model developed from the option pricing formula that can be used to determine a company’s or government’s default probability and perform a thorough assessment of the company’s or government’s debt risk based on the results.

Because the KMV model has many advantages, such as less required parameters, high accuracy, and fast speed, this paper applies the model to the comprehensive evaluation of government debt risk. Therefore, it is assumed that the government debt service revenue is a random variable that follows geometric Brownian motion and can circulate

---

Fig. 2 Concrete structure of vertical bitmap
under infinite conditions, then the variable satisfies the following formula:
\[ R_t = f(Z_t) \]  
(7)

In the above formula, \( R_t \) represents the debt service income in time \( t \). At the initial time, the value is a constant, then \( R_0 = R \) exists. \( Z_t \) represents a random variable and \( f(x) \) represents a geometric Brownian motion function. Therefore, in combination with the above analysis, the following formula exists:
\[ dR_t = \mu R_t dt + \sigma R_t dx_t \]  
(8)

In the above formula, \( dR_t \) represents the growth of government fiscal revenue in the period \( t \), \( dx_t \) represents the geometric Brownian motion growth coefficient, and \( \mu \) and \( \sigma \) represent the growth rate and fluctuation probability of \( R_t \), respectively.

According to ITO’s theorem, the following relation exists:
\[ d = R_0 \exp \left\{ \left( \mu - \frac{1}{2} \sigma^2 \right) t + \sigma \sqrt{t} \right\} \]  
(9)

In general, the random variable \( x_t \) and its logarithmic form obey normal distribution. Therefore, according to the normal distribution density function of the variable, the logarithmic form of \( R_t \) is obtained to obtain the expected value and variance. Then, the calculation formula of the expected value is as follows:
\[ E(\ln R_t) = \ln R_0 + \mu \frac{t}{2} - \frac{1}{n-1} \sum_{i=1}^{n} \ln \left( R_0 * \frac{R_{i+1}}{R_i} \right) \]
\[ = \ln R_0 + \frac{1}{n-1} \sum_{i=1}^{n} \ln \left( R_0 * \frac{R_{i+1}}{R_i} \right) \]  
(10)

The logarithmic variance expression of the debt service income variable \( R_t \) is as follows:
\[ \text{Var}(\ln R_t) = \sigma^2 t \]
\[ = \frac{1}{n-2} \sum_{i=1}^{n} \left[ \ln \left( \frac{R_{i+1}}{R_i} \right) - \frac{1}{n-1} \sum_{i=1}^{n} \ln \left( \frac{R_{i+1}}{R_i} \right) \right]^2 \]  
(11)

Based on the above analysis and combined with the characteristic analysis results of the normal fraction function, the calculation results of \( \mu \) and \( \sigma \) are as follows:
\[ \mu = \frac{1}{n-1} \sum_{i=1}^{n-1} \ln \left( \frac{R_{i+1}}{R_i} \right) + \frac{1}{2} \sigma^2 t \]  
(12)
\[ \sigma = \sqrt{\left\{ \frac{1}{n-2} \left[ \ln \left( \frac{R_{i+1}}{R_i} \right) - \frac{1}{n-1} \sum_{i=1}^{n-1} \ln \left( \frac{R_{i+1}}{R_i} \right) \right]^2 \right\}} \]  
(13)

Suppose that the debt repayment date of the government is described as \( T \), and the amount payable is described as \( B_T \). When the debt repayment date of the government is \( R_T \), the debt repayment income of the government is less than the face value of the debt payable by the government, which is \( B_T \), then the government has the problem of bond default. The probability of this problem is \( P \). The default probability expression and default distance of government debt can be calculated. The specific calculation formula is as follows: When the government’s debt payback date is less than the face value of the debt, called the government bond default problem, the likelihood of the problem is, the combination of Merton model, expression, and default on government debt default probability distance calculation. The specific calculation formula is as follows:
\[ P = P[R_T < B_T] = P[f(x_T < B_T)] = P[x_T < f^{-1}(B_T)] \]
\[ = N \left[ \frac{\ln B_T - \ln R_0 - \frac{1}{2} \sigma^2 T}{\sigma \sqrt{T}} \right] \]  
(14)
\[ DD = \frac{\ln \frac{R_T}{B_T} + \left( \frac{\mu_T}{2} + \frac{1}{2} \sigma^2 \right) T}{\sigma \sqrt{T}} \]  
(15)

In general, the KMV model for the comprehensive evaluation of government debt risk can be constructed by knowing the four parameters: the government’s debt repayment revenue \( R_T \), the debt payable value \( B_T \) and \( \mu, \sigma \) at time \( T \).

However, in practice, the volatility of economic development, capital, and market increases the uncertainty of government debt risk. As a result, the model in this article must be improved in order to enhance the overall assessment accuracy of government debt risk.

This article examines the anticipated price of European no-dividend call options with Knight uncertainty, in

| Characteristics of 1 | Characteristics of 2 | Characteristics of 3 | Characteristics of n |
|---------------------|---------------------|---------------------|---------------------|
| #1                  | 0                   | 4                   | 1                   | 5                   |
| #2                  | 7                   | 4                   | 4                   | …                   |
| …                   | …                   | …                   | …                   | …                   |
| #n                  | 2                   | 1                   | 3                   | 5                   |

Table 1 Set of discrete values of varying degree

\[ \text{Var}(\ln R_t) = \sigma^2 t \]
addition to the previous research. And the specific calculation formula is as follows:

$$C(S_T, K) = S_0e^{-\lambda T}N(d_1) - Ke^{-\lambda T}N(d_2)$$ (16)

Since government debt can also be regarded as a European call option, the traditional KMV model can be modified and optimized by using the uncertainty parameters of Knight, so as to further improve the performance of the model. Therefore, in combination with the above analysis, this paper introduces the uncertainty factor and recalculates the default probability and default distance in the case of uncertainty factor $\lambda \geq 0$, which is expressed by the following formula:

$$P_0 = P[R_T < B_T] = P[f(x_T < B_T)] = P[x_T < f^{-1}(B_T)]\]

$$N\left[\frac{(\ln B_T - \ln R_0) - (\mu_T - \frac{1}{2} \sigma_T^2 - \lambda)T}{\sigma\sqrt{T}}\right]$$ (17)

$$DD' = \left[\ln \frac{B_T}{R_0} + \left(\mu_T - \frac{1}{2} \sigma_T^2 - \lambda \sigma\right)T\right] \sigma\sqrt{T}$$ (18)

As a consequence, the KMV model is updated in accordance with the revised default probability and default distance calculation findings, and this model is used to perform a complete assessment of government debt risk. The detailed description of this model is as follows:

$$V_E = \frac{P'(DD' B_T N(d_1) - R_T N(d_2) + \sigma^2)}{\mu}$$ (19)

### 3 Simulation experiment design and result analysis

#### 3.1 Experimental experiment design

In order to verify the practical application effect of the modeling method of government debt risk comprehensive evaluation based on multidimensional data mining, an experimental test is carried out. In order to ensure the scientifi city and reliability of the experimental results, the experiment should be carried out in the same experimental environment, as shown in Table 2.

The government debt data of a local city is taken as the experimental sample data, and the experimental sample data is cleaned and repaired to reduce the experimental error and improve the accuracy of the experimental results. The experimental comparison methods are the evaluation method based on the pressure-state-response model proposed in reference (Holilah et al. 2021), the evaluation method based on the central point triangle whitening weight function proposed in reference (Huang et al. 2019), and the modeling method of government debt risk comprehensive evaluation based on multidimensional data mining proposed in this paper. The comprehensive performance of various techniques is evaluated by comparing the accuracy of government debt risk data mining, relative error rate, and time-consuming of government debt risk comprehensive assessment of different methods.

### 3.2 Analysis of experimental results

According to the above experimental scheme design, the government debt risk data mining accuracy of different methods is first compared, and the comparative results are shown in Fig. 3.

Analysis of data in Fig. 3, the evaluation method based on the pressure-state-response model of data mining accuracy between 57 and 93%, based on the center of the triangle whitenization weight function of the evaluation method of data mining accuracy between 55 and 82% of change, evaluation methods of data mining on the basis of multidimensional data mining accuracy is above 91%. It also demonstrates that, when compared to the two approaches, the proposed method’s government debt risk data mining accuracy is greater and error is lower, laying a strong basis for the correct assessment of government debt risk in the future. On the basis of the above, the relative error rates of the comprehensive evaluation of government debt risk of the three methods are compared, and the calculation results are shown in Fig. 4.

| Table 2 | Experimental environment |
|---------|--------------------------|
| **Runtime environment** | Configuration | Parameter |
| **Hardware environment** | CPU | Intel(R) Core(TM)i5-9400 |
| | Frequency | 2.90 GHz |
| | RAM | 16.0 GB |
| **Software environment** | Operating system | Windows 10 |
| | Version | 18.362.1082 pro |
| | Digits | 64bit |
| | Analog software language | APDL |
| | Simulation software | MATLAB 7.0 |
The analysis of data in Fig. 4 shows that the evaluation relative error rate of the evaluation method based on the pressure-state-response model varies between 13.5 and 28.3% and that of the evaluation method based on the central point triangular whitening weight function varies between 4.1 and 29.1%. The relative error rate of the proposed multidimensional data mining-based government debt risk comprehensive evaluation modeling method is always below 3.4 percent, indicating that when compared to the two methods, the proposed method has a lower relative error rate and higher accuracy, allowing for accurate comprehensive evaluation of government debt risk.

The time-consuming of the complete assessment of government debt risk of the three techniques is compared when combined with the aforesaid experimental scheme design. The specific results are shown in Table 3.

Based on the data in Table 3, the evaluation method based on the pressure-state-response model of evaluation takes an average of 1.42 s, the evaluation method based on the center of the triangle whitening weight function of evaluation takes an average of 1.83 s, and the government debt risk evaluation based on multidimensional data mining modeling methods of evaluation takes an average of 1.83 s.

4 Conclusion

With the rapid development of social economy, the local governments gradually build and perfect the mechanism of debt financing, but once the government facing debt risk is too big, will inevitably hurt the local economy and people, so in order to effectively solve these problems, need to research an effective government debt risk evaluation method. But the current government debt risk evaluation method of low computational accuracy evaluation indexes, the evaluation of the relative error is higher and the evaluation takes longer, so in order to solve the problems existing in the current method as the research target, this paper proposes a new government debt risk comprehensive evaluation on the basis of multidimensional data mining modeling methods. The experimental results show that the data mining accuracy of this method is always above 91%, the relative error rate of evaluation is always below 3.4%, and the average evaluation time is 0.71 s, which can realize the rapid and accurate evaluation of government debt risks. It is hoped that the relevant parts in the future will take this study as the basis to ensure the safety of government funds.
and promote the steady and rapid development of social economy by clearly dividing the boundary between state-owned enterprises and local government debts, monitoring debt risks, increasing government financing guarantee funds and other ways.

Authors' contributions LC contributed the data collection, supervision, resources, WXD contributed to the analysis, software, manuscript initial editing, ZEH contributed to the concept and design, results interpretation, manuscript final editing. LC, WXD, ZEH have read and approved the manuscript.

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest Not applicable.

Research involving Human Participants and/or Animals Not applicable.

Informed consent Not applicable.

References

ChaoYing L, Da WX, & Hui ZE (2021) Research on modeling of government debt risk comprehensive evaluation based on multidimensional data mining. Croce M, Nguyen TT, Raymond S (2021) Persistent government debt and aggregate risk distribution. J Financ Econ 140(2):347–367 Fang F (2020) Methodological developments in electron spin resonance (ESR) low-temperature thermochronometry. Holilah D, Bustamam A, & Sarwinda D (2021) Detection of Alzheimer’s disease with segmentation approach using K-Means clustering and watershed method of MRI image. Paper presented at the Journal of Physics: Conference Series.

Huang SW, W. F, Cao QW et al. (2019) Research on fast multi-objects feature point matching based on cluster algorithm. Computer Simulation, 2(36):211–214+376

Kumalasari D, Putra A, Gaffar AFO (2020) Speech classification using combination virtual center of gravity and k-means clustering based on audio feature extraction. J Inform 14(2):85

Liu (2020) Improvements on sample selection in data mining based on audit opinions and regulatory measures. Journal of Nanjing University of Science and Technology (Social Sciences Edition), 1(33):42-47

Liu H, C L, Mu L et al. (2021) A recognition method of kiwifruit flowers based on k-means clustering. Journal of Agricultural Mechanization Research.

Loey M, J. M. W, El-Bakry HM et al. (2020) Breast and colon cancer classification from gene expression profiles using data mining techniques. Symmetry, 3(12):408-415

Mengash HA (2020) Using data mining techniques to predict student performance to support decision making in university admission systems. IEEE Access 8(18):55462–55470

Moayedi H, Abdullahi MAM, Nguyen H, Rashid ASA (2021) Comparison of dragonfly algorithm and Harris hawks optimization evolutionary data mining techniques for the assessment of bearing capacity of footings over two-layer foundation soils. Eng Comput 37(1):437–447

Talha M, Sohail M, Tariq R, Ahmad MT (2021) Impact of oil prices, energy consumption and economic growth on the inflation rate in Malaysia. Cuadernos De Economı´a 44(124):26–32

Talha M, Azeem S, Sohail M, Javed A, & Tariq R (2020) Mediating effects of reflexivity of top management team between team processes and decision performance. Azerbaijan Journal of Educational Studies, 690

Talha M (2020) A history of development in brain chips in present and future. International Journal of Psychosocial Rehabilitation, 24(02).

Wenwei Z (2019) Study on the debt risk assessment of local government under the dual goal of steady economic growth and risk prevention. Contemp Econ Manage 2(41):68–76

Wu D, Z Y (2020) Design of remote monitoring system of harvester based on parallel algorithms of cloud platform data mining. Journal of Agricultural Mechanization Research, 6(42), 235–239
Xu Z, Zhu F, & Fan W (2020) Establishment of early warning index of local government debt risk—taking Hubei province of China as an example. Paper Presented at the IOP Conference Series: Materials Science and Engineering

Zhang R, Shi R, & Li C (2021) Local government debt risk measurement and early warning structural system building—Research Based on KMV and TOPSIS Models. Paper presented at the E3S Web of Conferences.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.