Study on credit risk of real estate industry based on genetic algorithm KMV model

Xusheng Cheng¹, Ziran Sun² and Weiqi Bao³

¹Southwest University of Finance and Economics, Sichuan, China
²Southwest University of Finance and Economics, Sichuan, China
³Southwest University of Finance and Economics, Sichuan, China

Corresponding author and e-mail: Ziran Sun, szr4655@126.com

Abstract. This paper firstly analyzes the current situation of credit risk in China's real estate industry, and then compares the traditional and modern credit risk measurement models. On this basis, the KMV model is selected and the artificial intelligence model genetic algorithm (GA) is introduced to improve the accuracy of KMV model. Secondly, annual financial data and stock trading data of 108 real estate listed companies from 2010 to 2019 are selected for empirical research. The analysis of the total default distance between the 108 companies and the actual economic development in China proves that the results of the GA-KMV model are in good agreement with the economic development trend, indicating that the model has good applicability. Finally, some suggestions are put forward according to the empirical results.

1. Introduction
Real estate not only has a huge role in promoting China’s GDP growth, but also improves people's lives, promotes the urbanization process, and the common development of upstream and downstream industries, which show the important role of the real estate industry in the national economy, and also, the necessity of researches on the real estate industry. Since 2004, China fully implemented the "bidding, auction, and listing" system, China's real estate market has experienced rapid development for 15 years. While China’s real estate industry is developing rapidly, the bank's credit funds invested in the real estate industry are also growing rapidly. However, due to the relatively short development time of China’s real estate marketization, negative factors emerge during the development process, China's real estate industry has problems, such as unbalanced market supply and demand, unbalanced regional development, and unsustainable development due to excessively high profit margins. Therefore, it is particularly important to analyze and study the credit risk of the real estate industry.

2. Literature review

2.1. Domestic literature review

2.1.1. Domestic research on the credit risk of the real estate industry. Huiping Li, Yingjie Chen (2010) [1] classified risk probability, risk loss, loss expectation, etc. as direct indicators of real estate investment risk, along with investment returns, investment return periods and other indicators as indirect indicators by studying the evaluation indicators of real estate investment risk, providing a
reference for measuring the risk of real estate investment projects. Bing Xiao and Chunhong Li (2010) [2] used the Logistic model to evaluate and predict the credit default risk of the real estate industry. Besides, it is concluded that there are special credit risk characteristics in China’s real estate industry, and in addition to financial indicators, macroeconomic indicators are also important factors affecting the credit risk of China's real estate companies. Wei Li (2013) [3] selected the KMV model to conduct an empirical analysis of China's A-share market and verified the effectiveness of KMV model in China. Deliang Pan, Jiehua Xie, Chengqing Li (2018) [4] analyzed the credit risk of the real estate industry under strong financial supervision, and carried out innovative development thinking, proposing that commercial banks should not only focus on preventing real estate risks, but also should further adjusting the real estate financing structures.

2.1.2. Domestic research on credit risk measurement models. Since China's real estate finance started late, and the credit data is incomplete, research on real estate credit risk measurement is very limited. In limited research in China, a common method to measure credit risk in the real estate industry is KMV, but the classic KMV model is derived from historical data in the United States, so the traditional KMV model is not applicable to China's real estate market. Besides, there are few studies on the applicability improvement of the KMV model in China. Some scholars are subjectively set a fixed DPT coefficient, usually higher than the classic KMV model default point, for example, in 2017, Zhang Zhimei changed the coefficient of long-debt to 0.75[6], however, this method is quite subjective and does not fully demonstrate the theoretical basis for setting the coefficient, so it's slightly rough. Some scholars are subjectively set up multiple fixed default points, and select appropriate DPT through empirical research, for example, in 2004, Zhang Ling and her group set up multiple coefficient of long-term debt: 0, 0.5, 0.75 [7], their results show that the model's ability to identify credit risk is the strongest when the coefficient is set at the highest 0.75, and is the worst when coefficient is set at the lowest 0, which empirically supports that classic KMV is not applicable to China's real estate market. Xiuyun Yang, Yuanyuan Jiang, Zhenzhen Duan (2016) [8] took Chinese listed companies during 2013 to 2014 as a sample, to compare and analyze the applicability and limitations of KMV model, Credit Metrics model, Credit Risk model and Credit Portfolio View model, and finally concluded that KMV makes the measurement of credit risk more reliable and is most suitable for China's national conditions.

Therefore, after discovering the lack of relevant research in China, we decided to use genetic algorithm, which is suitable for solving complex optimization problems, to optimize the traditional KMV model to get a GA-KMV model, and then we use the GA-KMV model to study the credit risk of China’s real estate industry. And after our experiments, we concluded that the credit risk of the China’s real estate industry is relatively high, besides, compared with the classic KMV model, the overall accuracy of the GA-KMV model for measuring credit risk has increased by 2%, and the success rate of identifying ST companies has increased 19%.

2.2. Foreign literature review

2.2.1. Foreign research on the credit risk of the real estate industry. The foreign scholars’ researches on the credit risk of the real estate industry are ahead of Chinese researchers both in theory and practice. Tien Foo Sing, Seow Eng Ong, GangZhi Fan & C. F. Sirmans (2004) [12] applied a theoretical default-risky swaps valuation model to evaluate credit risks in ABS bonds in Singapore. The Monte-Carlo simulation results, based on the Century Square shopping mall ABS case, show significant effects of the changes in rental volatility and default-free interest rate volatility on the default-risk premium of swap. Their results suggest that the rental dynamics of the securitized real estate are critical in determining the default risks of ABS deals. The fixed-rate (coupon yield) and floating-rate (rental cash flows) should therefore be adequately determined to reflect the default risks [12]. Kanak Patel & Prodromos Vlamis(2006)[13] estimated the distance to default and the “risk neutral” default probabilities for a sample of 112 real estate companies in UK over the period 1980 to
2001. The empirical results support the theoretical underpinnings of the BSM-type structural model in that the two driving forces of default are high leverage and high asset volatility [13].

2.2.2. Foreign research on credit risk measurement models. Foreign scholars' research on credit risk measurement models ranges from the initial expert assessment method to later single-variable and multivariable credit discrimination models, to credit risk measurement models and artificial intelligence models. These models make the measurement of credit risk gradually accurate.

In modern times, with the increasing importance of risk measurement models, some international financial and consulting companies have developed credit risk measurement models from a non-traditional perspective. One innovative forecasting model which has been widely applied in both practice and academic research is a particular application of Merton’s model (Merton, 1974) that was developed by the KMV Corporation. Korablev & Dwyer (2008) compared the results of the KMV model with the traditional Z score and Logit model by collecting company data from three regions in Asia, Western Europe and North America from 1996 to 2006. It is found that the KMV model has a forward-looking advantage in pre-company credit risk [15].

3. Methodology

3.1. The classic KMV model

The classic KMV model is used to estimate the default probability of the borrowing company. It regards corporate liabilities as buying a European call option, that is, the owner of the corporate holds a European call option with an exercise price at the face value of the corporate’s debt and the market value of the corporate’s assets. The operation principles of classic KMV model are first to predict future changes in the value of corporate assets according to option pricing theory. Second, the enterprise's default point (denoted as DPT) is calculated based on a relationship between long-term and short-term liabilities of the enterprise shown as below: In classic KMV model: 

\[ DPT = SD + 0.5LD \]

After getting DPT, default distance (denoted as DD) of the borrowing corporate is calculated based on the formula:

\[ DD = \frac{\ln \frac{VA}{DPT} + \left( r - \frac{\sigma_A^2}{2} \right) \sqrt{T}}{\sigma_A \sqrt{T}} \]

The enterprise’s default situation is judged based on the following principles:

- DD > 0 Enterprise is judged NOT to default
- DD < 0 Enterprise is judged to be in default

Finally, expected default rate (denoted as EDF) of the enterprise is calculated.

3.2. Genetic algorithms

3.2.1. The concept of genetic algorithm. Genetic algorithm, as the name implies, is an algorithm that simulates the process of species reproduction. He draws on Darwin's idea of "survival of the fittest, survival of the fittest" to find the global optimal solution to the problem through the algorithm.

3.2.2. The general idea of genetic algorithm. Species have phenotypes and genes that determine their phenotypes, and problems have objective equations and decision variables for objective equations. Genetic algorithm is to compare the objective equation of the problem to the phenotype, and the decision variable to the gene.

In nature, different phenotypes of the same species will have different adaptability to the environment, such as the full wings and residual wings of butterflies. The residual wings butterflies will gradually be eliminated if they can’t fly, resulting in most of the butterflies are intact wing.
The genetic algorithm also introduces the concept of fitness (specific definition of specific problems), and sorts the individuals in the virtual population according to the fitness. Individuals with high fitness are more likely to propagate genes into the next generation.

The simple use of fitness to calculate is very limited, because the simple use of fitness is only to sort the individuals in the initial population, and it is easy to find the local optimal solution. Genetic algorithms will use mutation, crossover, etc. to solve this problem. For example, mutation means that the algorithm will adjust the decision parameter values of individuals to be passed to the next generation in a certain range, so as to always keep new individuals appearing in the algorithm and know to find the global optimal solution.

3.3. Optimize the classic KMV Model though genetic algorithms

3.3.1. Parameter selection. Because the calculation at the default point is related to long-term and short-term liabilities, we have two decision variables a and b in this problem.

\[ \text{DPT} = a \times SD + b \times LD \]

Objective equation selection:

\[ DD = \ln \frac{v_A}{DPT} + \left( r - \frac{\sigma_A^2}{2} \right) T \]

We determine the company attributes by judging the default distance DD, so there are four possible situations:

1) ST company was judged as ST company (judgment is correct);
2) The ST company is judged to be a non-ST company (wrong)
3) Non-ST companies are judged as ST companies (judgment is correct);
4) Non-ST company was judged as ST company (wrong)

ST: company who continuously loss
Non-ST: company with good financial status

Our goal is to improve the accuracy of the KMV model in predicting the types of listed companies in China. Therefore, letting cases 1 and 3 happen as much as possible is our goal. Assuming that our data set uses X companies in total, of which ST and non-ST each account for half, and the number of occurrences of cases 1 and 3 is m and n, respectively, then our fitness function is F.

3.3.2. Specific actual operation. Initialized population. Create a matrix to represent the initialized population, where each row of the matrix represents an individual in the population, and the number of matrix columns represents the number of decision parameters. In this problem, we construct a (100, 2) matrix. Each individual is a group (a, b) value, which represents a possible default point.

Fitness and Objective function value calculation. We use Geatpy to complete the genetic algorithm, which uses the Numpy array type matrix to store the target function value of the population. Generally named ObjV, each row corresponds to each individual, because for a single objective function, ObjV will have only 1 column; and for the objective function, ObjV will have multiple columns. In this question

\[ \text{ObjV} = \begin{pmatrix}
    f_1(x_{1,1}, x_{1,2}, \ldots x_{1,Nvar}), f_2(x_{1,1}, x_{1,2}, \ldots x_{1,Nvar}) \\
    f_1(x_{2,1}, x_{2,2}, \ldots x_{2,Nvar}), f_2(x_{2,1}, x_{2,2}, \ldots x_{2,Nvar}) \\
    f_1(x_{3,1}, x_{3,2}, \ldots x_{3,Nvar}), f_2(x_{3,1}, x_{3,2}, \ldots x_{3,Nvar}) \\
    \vdots \\
    f_1(x_{Nind,1}, x_{Nind,2}, \ldots x_{Nind,Nvar}), f_2(x_{Nind,1}, x_{Nind,2}, \ldots x_{Nind,Nvar})
\end{pmatrix} \]

Geatpy uses column vectors to store individual fitness of populations. Generally named FitnV, it is also a numpy array type, each row corresponds to each individual of the population matrix.
FitnV = \begin{pmatrix} 
fit_1 \\
fit_2 \\
fit_3 \\
\vdots \\
fit_{N_{ind}} 
\end{pmatrix}

Here it is not difficult to see that in fact ObjV and FitnV are similar data formats, this helps us in the next operation.

Because it is different from the general idea of sorting individuals by fitness to achieve the selection effect.

In this problem, the function value directly related to the decision variable is not the goal we want to optimize. We do not want to know the maximum or minimum value of the DD (default distance) parameter.

The value we want is the accuracy of the kmv model defined by ourselves, which is our fitness function:

\[ F = 1 - \frac{m + n}{x} \]

So we do not calculate the population as a complete matrix, in contrast, we calculate each individual separately. For every individual who represent a group of (a,b), we let them get a “F” (value of the function shown above), We turn all the obtained “F” into a list. We let the FitnV represent the ObjV to complete the calculation.

Calculate the default distance DD. For each default point, use the formula to calculate the default distance of each sample company.

Determine the type of company. For each default point, determine the type of the ST and non-ST sample companies, and obtain the values of m and n

Check whether the termination condition is reached. If the termination condition is reached, stop the algorithm and obtain the optimal solution, that is, the optimal coefficient of short-term and long-term debt, otherwise return to step (3).

4. Data analysis

4.1. Data description

We selected two kinds of dataset from CSMAR and official website of SHIBOR as shown in figure 1. Based on the previous introduction of our methodology, we mainly focus on the real estate industry in China’s A-share market [4].

All the selected companies should not be cross listed in B-share market or H-share market, and the data of all selected companies in the first year of listing will be eliminated by SAS. According to the methodology, the dataset is separated into ST/non-ST, while figure 1 illustrates the transaction status of selected companies from 2010 to 2015.
Figure 1. Transaction Status of Selected Companies in the Real Estate Industry in China's A-share Market.

Table 1. Data Used in Research.

| Variables                        | Time range            | Frequency               | Source            |
|----------------------------------|-----------------------|-------------------------|-------------------|
| individual stock                 |                       |                         |                   |
| trading days                     |                       | per year                |                   |
| daily return                     |                       | per trading day         |                   |
| non-current share capital        |                       |                         |                   |
| current share capital            |                       |                         |                   |
| net asset value per share        |                       |                         |                   |
| annual closing price             | 2010.01.01-2019.12.31 | per year                | CSMAR             |
| annual return                    |                       |                         |                   |
| total liability                  |                       |                         |                   |
| current liability (SD)           |                       |                         |                   |
| non-current liability (LD)       |                       |                         |                   |
| risk-free rate                   | one year fixed deposit rate | 2010-2015 |                      |
|                                  | SHIBOR                 | 2016-2019               | http://www.shibor.org/ |

From the perspective of individual stock, we use daily return to calculate the standard deviation of equity value for each stock per year, and the value of equity \( V_E \) per company per year is calculated from the following formula:

\[
V_E = \text{non - current share capital} \times \text{net asset value per share} + \text{current share capital} \times \text{annual closing price}
\]

For the risk-free rate, because of the interest rate reform from 2014 to 2015, we use one-year fixed deposit rate to represent the risk-free rate before 2015, and use the market rate which is generally regarded as Shanghai interbank offered rate (SHIBOR) in China to represent the risk-free rate.

4.2 Results

Based on the data set of the decade from 2010 to 2019, the model coefficients calculated by the GA-KMV model are \( \alpha = 1.940359 \) and \( \beta = 1.078522 \) separately.

Compared with the original KMV model, the overall accuracy of the GA-KMV model is improved by 2%, but the success rate of recognizing ST-company is increased by 19%. The detailed result is shown below in the table 4-2:
Table 2. Final result.

|        | Total # | # of nst recognized successfully | # of st recognized successfully | Total success rate | % of nst recognized successfully | % of st recognized successfully |
|--------|---------|----------------------------------|----------------------------------|--------------------|----------------------------------|----------------------------------|
| KMV    | 108     | 34                               | 15                               | 0.45               | 62.70%                           | 27.00%                           |
| GA-KMV | 108     | 26                               | 25                               | 0.47               | 47.00%                           | 46.00%                           |

Firstly, the new GA-KMV model is shown as $DPT = 1.94SD + 1.08LD$, compared with the original KMV model, the GA-KMV model has larger coefficients on short-term debt (SD) and long-term debt (LD). However, the coefficient of short-term debt is still about twice that of long-term debt. This indicates that China's real estate industry has low default points and high default risk, but the main ingredient of the risk that China's real estate industry need to consider is still the short-term debt.

Secondly, as can be seen from Table 4-3, the GA-KMV model has greatly improved the accuracy of recognizing the ST companies compared to the original model while the accuracy of recognizing non-ST companies is slightly reduced. This shows that the GA-KMV model is stricter in determining the default risk than the original KMV model. This leads to the prediction accuracy of the GA-KMV model is only 2% higher than the original KMV model, but it is undeniable that the GA-KMV model is more accurate in predicting companies that are prone to default.

5. Conclusions

This article uses the genetic algorithm KMV model to conduct an empirical analysis of the real estate listed companies in China from 2010 to 2019. The conclusions are as follow:

1) The new GA-KMV model is shown as $DPT = 1.94SD + 1.08LD$ which has larger coefficients on short-term debt (SD) and long-term debt (LD) than the original KMV model.

This shows that the credit risk of China's real estate industry is relatively high. Since 2010, real estate companies have begun to finance non-standard debt asset investments. According to the actual situation of China's real estate industry, real estate companies will only invest a relatively low proportion of their own funds in order to develop multiple projects at the same time. The remaining part mainly depends on external financing, in which bank loans are the main source of external financing. In 2013, non-standard investments accounted for 30% of the newly increased social financing throughout the year. Most of their underlying assets are connected to real estate. In 2014, with the economic growth, listed companies obtained a steep increase in financing scale through the issuance of private placements. After 2015, real estate companies have significantly increased their investments in issuing bonds abroad because of the gradual tightening of domestic financing. Therefore, a large amount of debt financing has increased the default risk of real estate companies.

2) Compared with the original KMV model, the overall accuracy of the GA-KMV model is improved by 2%, and the success rate of recognizing ST companies is increased by 19%.

It is acceptable that the success rate of recognizing non-ST companies is low (26%). From the perspective of risk control, it can enable enterprises to take timely measures and methods for self-management, improve the credit rating of the enterprise, and ensure the normal operation of the enterprise. This shows that the optimal default point calculated by genetic algorithm is better than the original model formula.

The sample data is a single industry data which is relatively small, but the KMV model requires a lot of historical data to measure, so the results obtained in this article may be different from the real situation of the China's real estate industry.

Acknowledgement

Xusheng Cheng, Ziran Sun and Weiqi Bao contributed equally to the article.

References

[1] Huiping Li, Yingjie Chen. Analysis on the Evaluation Index of Real Estate Investment Risk [J]. Market Week (Theoretical Studies), 2010(10):41-42 + 76. (In Chinese)
[2] Bing Xiao, Chunhong Li. Research on Credit Risk of Real Estate Industry Based on Logistic Model [J]. Technical economy, 2010, 29(03):65-68 +78. (In Chinese)

[3] Wei Li. Research on Credit Risk Measurement of Listed Companies Based on KMV Model [D]. Southwest University of Finance and Economics, 2013. (In Chinese)

[4] Deliang Pan, Jiehua Xie, Chengqing Li. Considerations on Credit Risk and Innovation Development of Real Estate Industry under Strong Financial Supervision [J]. China Urban Finance, 2018(04):25-28. (In Chinese)

[5] Yong Zhang. Research on Real Estate Credit Risk Management Based on CPV Credit Risk Measurement Model [D]. Harbin University of Technology, 2011. (In Chinese)

[6] Zhimei Zhang, Renjun Zhang. Improvement of KMV Model and Measurement of Credit Risk of Listed Companies [J]. Statistics and decision-making, 2006(18):157-160. (In Chinese)

[7] Ling Zhang, Zhenshi Yang, Chen Guan. Study on the Application of KMV Model in Credit Risk Evaluation of Listed Companies [J]. Systems Engineering, 2004(11):84-89. (In Chinese)

[8] Xiuyun Yang, Yuanyuan Jiang, Zhenzhen Duan. Applicability Analysis and Empirical Test of KMV Model in Credit Risk Management of Commercial Banks in China [J]. Financial Theory and Practice, 2016(1):34-40. (In Chinese)

[9] Chunfeng Wang. Application of Neural Network Technology in Credit Risk Assessment of Commercial Banks [J]. Systems Engineering Theory and Practice, 1999,19(9):24-32. (In Chinese)

[10] Chunfeng Wang, Li Kang. Credit Risk Assessment Model [J]. Assessment Model of Commercial Banks Based on Genetic Planning Method [J]. Systems Engineering Theory and Practice, 2001(02):73-79. (In Chinese)

[11] Lili Zhou. Research on IDSS Modeling of Real Estate Credit Risk Based on Complex Multiple Projects [D]. Ocean University of China, 2012. (In Chinese)

[12] Sing, T.F., Ong, S.E., Fan, G. et al. ‘Analysis of Credit Risks in Asset-Backed Securitization Transactions in Singapore’. The Journal of Real Estate Finance and Economics 2004(28):235–253.

[13] Patel, K., Vlamis, P. ‘An Empirical Estimation of Default Risk of the UK Real Estate Companies’. J Real Estate Finan Econ 2006 (32):21–40.

[14] Altman E I, Haldeman R, Narayanan P. ‘Zeta analysis: a new model to identify bankruptcy risk of corporations’ [J]. Journal of Banking & Finance, 1977:1-29+ 54.

[15] Korablev & Dwyer. ‘Power and Level Validation of Moody’s KMV EDF Credit Measures In North America, Europe, and Asia’ [Z]. Moody’s KMV Corporation, 2007.

[16] Härdle, Wolfgang K., Prastyo, Dedy & Hafner, Christian, Support Vector Machines with Evolutionary Feature Selection for Default Prediction. SFB 649 Discussion Paper 2012,030.