A Measurement and Treatment Method of Bond Default Risk Based on Monte Carlo Simulation Technology

Chunyan Xue(✉)

Xiamen University Tan Kah Kee College, Xiamen 363005, China
380040211@qq.com

Abstract. In the scenario of bank bond risk measurement, due to the scarcity of bond default sample data and feature information, the model after machine learning modeling is not enough to represent the past and predict the future, which affects the accurate measurement of bond default risk. The system uses Monte Carlo simulation based on economic and financial theory and data statistical analysis. The system realizes refined calculation and a large number of simulations of bond loss results, and effectively measures a large number of simulated loss results.

Keywords: Monte Carlo Simulation · measurement · bond risk

1 Introduction

China's corporate bond market has developed rapidly, but for the measurement of bond default risk, the traditional measurement whether based on qualitative analysis by experts or quantitative statistical analysis methods [1]. The methods have shortcomings and deficiencies in all aspects: one is based on economic and financial theoretical background models (such as Z-force model, Morton model), Failure to follow up and accurately reflect the impact of the complex and volatile economic situation on bond default risk. The other is machine learning method based on statistical analysis, which is more suitable for customer default risk prediction [2]. In the measurement of default risk of bank bonds, there are few default sample data and feature information.

Traditional risk measurement methods can no longer meet the current complex market changes and the latest regulatory requirements [3]. Therefore, in the scenario of bank bond risk measurement, the system proposes a method based on Monte Carlo simulation to achieve refined, prudent and forward-looking measurement of bond default risk.

It also quantifies the default risk based on data statistics and analysis. In customer default scenarios, early stage technologies usually rely on machine learning technologies, such as traditional logical regression, XGBoost, LightGBM and some emerging model algorithms, as well as multi model fusion technology and deep learning [4]. They can model customer default risk, predict customer credit risk, and finally quantify the risk in a numerical way.

J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 735–742, 2023.
https://doi.org/10.2991/978-94-6463-198-2_76

2 Existing Technology Analysis

For the measurement of bond credit risk, the industry has adopted or used the following algorithms or models for reference:

2.1 Z-Force Model

The default risk of bonds is predicted through financial indicators. The disadvantage is that the model only involves relevant individual financial indicators and cannot cover all other factors that may lead to bond default, such as macroeconomic indicators [5]; The model weight is too fixed and simple; Default results and financial indicators may also be non-linear.

2.2 Machine Learning/Deep Learning

In the credit system, when measuring the credit risk of PD and LGD with retail scorecards and non retail ratings, traditional logic regression, XGBoost, LightGBM and other emerging machine learning technologies, or even deep learning, are usually used for modeling [6]. However, for the bond trading business in the bank's market risk, banks generally do not obtain the financial report, credit information and other customer information of the bond issuer, and there is very little bond data for default of dependent variables, so they cannot use the credit risk modeling method in the credit business system for modeling measurement.

2.3 Advanced Measurement Model

In addition, there are other classic advanced measurement models for credit risk, but when applied to the measurement scenario of bond default risk, there are deficiencies in all aspects [7].

- KMV model: credit risk loss is measured based on Merton model framework, derived from BSM option pricing formula [8]. When it is determined that the future market value of enterprise assets is lower than the value of liabilities to be settled, the enterprise will default. The default probability from the default point of insolvency is measured by this method.

  Disadvantages: ignoring the impact of macroeconomic on bond default.

- The technology measures the deterioration of credit asset quality based on mark to market method, calculates the transfer matrix of credit risk rating through Markov chain, thus obtaining the probability distribution and fluctuation of credit rating deterioration to default level, and calculates the credit risk degree of credit assets through VaR.

Disadvantages: Most of the bonds held by banks are medium - and short-term investment grade bonds. During the duration, there are few bond rating migrations, and there is not enough sample data to model.

- CreditPortfolio View: It based on the impact of external macroeconomic factors on the probability of default, established a model relationship between macroeconomic indicators and probability of default, and used regression model to fit.

Disadvantages: Neglect the impact of asset value changes on bond default.

3 Design and Implementation of New Technology Scheme

As for bond default risk, it follows the non normal distribution characteristics of credit risk (as shown in Fig. 1), which belongs to the right fat tail form, so the measurement method is relatively complex, and generally cannot be simulated and measured through simple models. However, due to the inherent advantages and disadvantages of the traditional credit risk measurement model, there are various deficiencies and defects in the direct measurement of bond default risk. Therefore, based on the principles and characteristics of each model, the system absorbs the effective measurement of the macroeconomic loss distribution of the CreditPortfolio View model, as well as the impact of asset value changes in the KMV model on bond default, combines the advantages of the two models, and uses the model Carlo simulation to carry out simulation experiments to achieve the refined measurement of bond loss distribution. Cluster Architecture and Functional Modules of No Load Balancing Device.

The main modeling process of bond default risk is to use the CreditPortfolio View model and KMV model for reference to predict whether bonds will default. At the same time, random sampling is conducted according to the beta distribution of LGD to calculate the loss value in case of default. Based on this, a large number of repeated simulation calculations are carried out through Monte Carlo to obtain a fine loss distribution. Finally, the tail loss is measured.

3.1 Bond Default Forecast

Draw on the CreditPortfolio View model, it builds a linear relationship between macroeconomic indicators and the full price of bonds. The stepwise regression method is adopted. For each linear regression, the least significant features are removed according to Prob, and then the linear regression fitting is performed again until all characteristic variables are significant, the model R is greater than 0.75, and the model fitting degree is good.

After stepwise regression, n characteristic variables are selected for each bond, and the weight w of each characteristic variable is fitted through the linear regression model to establish a linear relationship function with the bond price-.

$$z_i = w_0 + w_1 x_{1\mathrm{i}} + w_2 x_{2i} + \cdots + w_n x_{n\dot{i}} \tag{1}$$

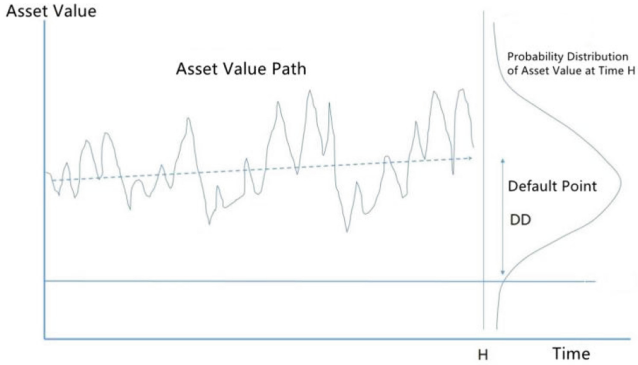

Fig. 1. Example of KMV model principle

According to the KMV model, the asset value is uncertain and fluctuates continuously over time. When H is in the future, the enterprise will default if it is insolvent (the asset value is less than the liability). According to KMV thought, the default point of each bond can be calculated by using the inverse function of normal distribution and the obtained bond PD. The value of bond assets predicted by the CreditPortfolio View model can be calculated and compared with the default point. If it is less than the default point, the bond is considered to be in default.

3.2 Simulation of Bond Default Loss Rate

Due to the lack of recovery cost, discount rate and other data and data quality issues, the LGD model measurement of loss on default has always been a difficulty in bank risk measurement. At present, the CBRC only recognizes the internal evaluation advanced LGD measurement implemented by six banks. On the other hand, LGD has considerable volatility. At present, many empirical findings show that the recovery rate (1-LGD) often follows the beta distribution. Therefore, the system calculates the beta distribution of LGD for each bond based on the characteristics of LGD and the processed LGD mean and standard deviation. During Monte Carlo simulation, LGD is randomly sampled according to the corresponding beta distribution.

Step 1. Calculate the parameters of the distribution according to the LGD mean and LGD standard deviation of the bonds: Beta

$$\alpha_i = \frac{{LGD_i}^2 \times (1 - LGD_i)}{(\sigma_{LGD_i})^2} - LGD_i \tag{2}$$

$$\beta_i = \frac{\alpha_i \times (1 - LGD_i)}{LGD_i} \tag{3}$$

Step 2. Generate the random number of the distribution according to the parameters:

$$LGD_i^{'} \sim Beta(\alpha_i, \beta_i) \tag{4}$$

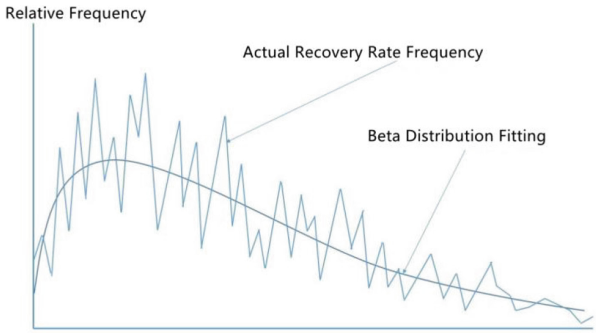

Fig. 2. Beta Distribution Fitting Chart

3.3 Monte Carlo Simulation of Loss Distribution

Based on the risk characteristic model established in the above steps, this technology generates random sampling under the random number simulation probability distribution, and calculates the loss results under the random sampling according to the risk characteristic model. Based on this, it carries out a large number of repeated random sampling calculation processes and establishes various unbiased estimators. It carries out simulation experiments close to the reality, and analyzes the loss distribution of the experimental results.

3.4 Matrix Decomposition

Both bonds and issuers have a certain degree of pairwise correlation, and the risk diversification effect of bond portfolio makes the overall risk of bond portfolio less than the sum of single bond risks. Therefore, in Monte Carlo simulation, in order to avoid uncorrelated random walk of simulation variables and generate correlated default events, Cholesky matrix decomposition method is generally used to convert independent random variables into correlated random variables. Therefore, the system adopts the bond yield matrix S, which is decomposed into and through Cholesky, and the matrix is transposed with each other.

$$S = U \times U^T \tag{5}$$

Take the decomposed matrix:

$$\mathrm{C} = U \tag{6}$$

However, when the yield matrix S is a non positive definite matrix, it cannot be decomposed by Cholesky matrix. Even though each element of matrix S can be decomposed by increasing the minimum random number and reducing the repetition rate, modifying the market data will lead to data distortion. Therefore, the system adopts singular value SVD decomposition method:

$$S = U \Sigma V^T \tag{7}$$

The decomposed matrix C is calculated according to the following formula:

$$C_{ij} = U_{ij} \times \sqrt{\Sigma_{ij}} \tag{8}$$

3.5 Monte Carlo Simulation Calculation

During each random sampling simulation, generate a normal distribution random number to replace the macro-economy in Formula (I) and calculate the predicted bond price. Then multiply the matrix with the matrix C decomposed in the previous step to obtain the bond price with correlation constraint, and substitute it into Formula (II) to judge whether it is in default.

If it is not less than the default threshold, the loss is 0.

If the default threshold is approximate, then formula (V) will generate a beta distributed random number analog LGD, and substitute it into formula (IX) below to calculate the loss value after default:

$$Loss = LGD_{\text{analog}} \times EAD \tag{9}$$

3.6 Calculation of Tail Loss Model

Through Monte Carlo simulation, we can get a large number of simulated loss data and get the overall loss distribution of bond default (as shown in Fig. 2). Generally, the VaR of the value at risk can be used to calculate the maximum loss under a certain degree of confidence for risk measurement. However, VaR's inherent ghost effect defect is not sensitive to the sudden extreme loss of the tail. In order to accurately measure extreme tail loss, the system uses tail loss weighted average algorithm ES to measure the loss distribution simulated by model Carlo.

$$ES = \sum\nolimits_{i=1}^{w} \frac{Loss_i}{w} \tag{10}$$

3.7 Model Parallel Computing

Monte Carlo simulation can simulate and analyze risks through simulation experiments, but in order to obtain higher precision and accuracy, more and more simulation calculations are needed. At present, the system is implemented in Python. In order to ensure the calculation timeliness, the traditional method uses multi thread/multi process parallel computing, but has the following disadvantages:

- To optimize the multithread/multiprocess parallel computing, the original code needs to be modified and retested.
- With the growth of business, higher requirements are constantly put forward for computing resources. It is not convenient for automatic capacity expansion through multi thread/multi process mode.

Therefore, PySpark is used in the system. With the parallel computing capability of Spark, the overall computing efficiency can be effectively improved, the intrusion

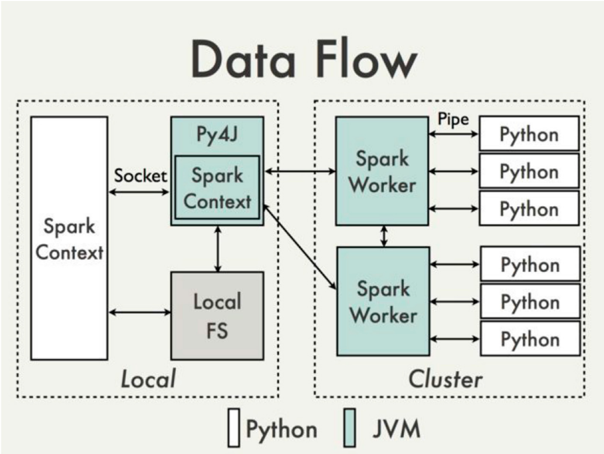

Fig. 3. Data Flow

into the original code can be avoided, and the computing resources and flexible capacity expansion can be guaranteed through the big data platform. The data flow during parallel computing is shown in Fig. 3.

3.8 Model Validation

Model risk problem will not only lead to wrong measurement results, provide wrong basis for risk management decisions, but also bring losses to banks in all aspects. Therefore, model validation is required to prove the correctness and effectiveness of the model algorithm when it is applied. At the same time, continuous tracking and monitoring of the model is required to ensure the long-term stability and effectiveness of the model after it is put into production.

Traditional machine learning can use roc curve, auc area, confusion matrix, PSI (model stability index), KS (model discrimination evaluation index) and other methods to verify the prediction ability of the model, but these methods are not suitable for the model of model Carlo simulation. Therefore, the system uses the return inspection method to calculate the number of breakthroughs under the same model confidence level on a regular basis according to the actual profit and loss, so as to judge whether the measurement results of bond default risk pass the inspection and meet the production application requirements.

4 Conclusion

Monte Carlo simulation is the most effective method to calculate credit risk VaR at present, which can effectively simulate the fat tail problem of credit risk and extreme situations. However, its biggest disadvantage is the time complexity. If 1000 assets are involved, and each asset needs to be simulated 100000 times, a total of 100 million calculations are required. It can be seen that this algorithm requires a large calculation cost. The system effectively guarantees the physical computing resources and efficient computing capability by virtue of the resource elastic expansion of the cloud platform and the parallel computing capability of the big data platform, as well as the cutting-edge technologies such as big data and cloud computing, which provides a solid and reliable foundation for the practical application of Monte Carlo simulation.

Acknowledgment. The core of risk management is risk measurement, and the core of measurement is algorithm and model. The traditional risk measurement models have their own advantages and disadvantages. When the actual business scenarios cannot meet the requirements, the system makes full use of the principles and characteristics of the traditional models, learns from each other, integrates the advantages of different models, and perfectly combines with the Monte Carlo simulation experiment to achieve accurate measurement of bond risk.

References

1. Risk Analysis on Asset Securitization of Ant Financial Services [J] Xiang Xiaohong, Wang Wenjing, Kang Fanlian Accounting study 2021(31)
2. Analysis on the Current Situation and Operation Countermeasures of Non performing Assets Securitization of Commercial Banks [J] Wang Keqin Chinese market. 2021(24)
3. Obstacles and Countermeasures to the Securitization of Non performing Assets of China's Commercial Banks [J] I wish Xinwei Productivity research 2021(06)
4. Study on the Accuracy of Credit Card Bad ABS Valuation [J] Zheng Qing, Wang Lingzhe, Wang Junying Financial market research 2021(05)
5. Feasibility analysis on securitization of non-performing assets of commercial banks under the new situation [J] Zhan Xin Management and Science and Technology of Small and Medium sized Enterprises (the first ten day issue) 2021(04)
6. Research and Development of Modern Asset Pricing Theory -- Based on the Efficient Market Hypothesis and Asset Pricing Model [J] Ginger balls Modern management science 2021(03)
7. 2020 Asset Securitization Development Report [J] Li Bo Bonds 2021(01)
8. Research on the Current Situation and Countermeasures of Non performing Assets Securitization of China's Commercial Banks [J] Yi Ping Modern commerce 2020(25)