Credit Risk Measure of Listed Pharmaceutical and Biological Companies Based on Genetic Algorithm KMV Model

Ruijie Liu[1*], Xinyan Zhang[2]

Department of Statistics and Mathematics, Inner Mongolia University of Finance and Economics, Hohhot, Inner Mongolia, 010070, P. R. China
*Corresponding Author

Received: 06 April 2023/ Revised: 16 April 2023/ Accepted: 22 April 2023/ Published: 30-04-2023

Abstract— *In order to measure the credit risk of listed companies in China's pharmaceutical and biological industry, a total of 28 listed companies in the A-share ST category and non-ST category were therefore selected as samples, and the model was improved using genetic algorithms, while the credit risk of 290 listed companies in the A-share pharmaceutical and biological industry in 2019-2021 was analyzed based on the improved model, and the research results showed that: the improved KMV model can effectively identify the improved KMV model can effectively identify the credit risk of listed companies in the industry, and the accuracy of the improved KMV model in determining whether an enterprise is in default reaches 78.57%; the credit risk of the pharmaceutical and biological industry decreases in the year of the outbreak of the new crown epidemic in 2020, and increases and the credit risk of enterprises appears polarized one year after the outbreak of the epidemic*.

Keywords— Credit risk measure; KMV model; GARCH model; genetic algorithm.

I. INTRODUCTION

Since the outbreak of the new crown epidemic in 2020, the public's health protection needs have increased and medical supplies for daily protection will be consumed in large quantities. For example, medical supplies such as masks, Lianhua Qingpu and nucleic acid reagents will be consumed dramatically after the epidemic, and the pharmaceutical industry is getting more and more attention. Based on the above reasons, this paper tries to measure and analyze the credit risk of listed pharmaceutical manufacturing enterprises in the western region of China, so as to promote this type of enterprises to be able to prevent their own credit risk more effectively and improve their credit quality.

Among the models for quantitative analysis of credit risk, the KMV model has the advantages of easy access to data, more intuitive calculation method, and better fit between the calculation results and the real credit risk changes. Yingchun Liu and Xiao Liu [1] (2011) used GARCH (1,1) volatility model to estimate the equity value volatility, and applied the KMV model to calculate the three-year default distance and expected default probability of 16 listed companies. The results show that the KMV model can well discriminate the difference in credit risk between ST and non-ST companies. Tang, Zhenpeng [2] and other scholars (2016) selected three industries to form the research sample, applied the modified KMV model to measure credit risk, and used factor analysis to elaborate the financial factors affecting the credit risk of listed companies in different industries. The research results show that the credit level of listed companies in the eastern coastal region is the best, followed by the central region and the western region is the worst; the credit risk of listed companies in the real estate industry is the least, followed by the pharmaceutical manufacturing industry.

Jinghai Feng[3] et al. redefined the optimal default point in the classical KMV model by combining genetic algorithm. The results show that the improved model fitting results show that the improved model fitting is more correct than the original model, i.e., the improved KMV model is more suitable for the creditworthiness assessment of listed companies in China. Shuyuan Zhou[4] used GARCH(1, 1) model to optimize the volatility and genetic algorithm to optimize the default point coefficients to modify the KMV model, and the results show that the GA-GARCH-KMV model is more accurate for the risk measure of listed companies with overcapacity industry.

II. KMV THEORETICAL MODEL AND GENETIC ALGORITHM

2.1 KMV model

The KMV model treats a firm's equity as a call option, with the strike price being the firm's liabilities and the underlying being the value of the firm's assets. When the value of the firm's assets is less than its liabilities, the firm will choose to default, otherwise it will not default. According to the Black-Scholes-Merton option pricing model, the relationship between the value of the firm's assets and the value of its equity is:

$$E = VN(d_1) - De^{-r(T-t)}N(d_2)$$

where the volatility of the firm's equity value is related to the volatility of its asset value as:

$$\begin{cases} d_1 = \dfrac{\ln(V/D) + (r + \sigma_v^2/2)(T-t)}{\sigma_v\sqrt{T-t}} \\ d_2 = d_1 - \sigma_v\sqrt{T-t} \end{cases}$$

The equation for the relationship between the volatility of equity value (σ_E) and the volatility of the firm's asset value (σ_v) is:

$$\frac{\sigma_E}{\sigma_v} = \frac{V}{E} \cdot N(d_1)$$

where D is the book value of liabilities, T is the time to maturity, t is the present time, r is the risk-free interest rate, and V is the market value of assets. σ_v is the volatility of asset value, E is the market value of equity, and σ_E is the volatility of the firm's market value of equity. The KMV model assumes that the firm's asset value follows a normal distribution and N(d) is the standard cumulative normal distribution function. From the market value of equity E and its volatility σ_E and the book value of liabilities E. The BSM option pricing model is used to find the market value of the firm's assets V and its volatility σ_v.

The default distance is the relative distance that the value of the firm's assets falls from the current level to the default point during the risk period and can be expressed as:

$$DD = \frac{E(V) - DPT}{E(V)\sigma_v}$$

where E(V) is the expected asset value and DPT is the firm's default point.

2.2 Genetic Algorithm

Genetic algorithm is a stochastic optimization search method evolved by imitating Darwin's law of natural selection and evolution in the biological evolution theory. By generating an initial population represented by strings, the poorly adapted individuals are eliminated by using the mechanism of genetic evolution in nature, thus achieving the superiority of the inferior, while the well-adapted individuals stay and become new individuals by randomly selecting and combining the same string points while exchanging them with each other. The algorithm mimics the genetic selection, crossover, and mutation of biological inheritance to generate the next generation of populations to adapt to the changing environment, and the final result is usually a more adaptive optimal solution or a local optimal solution.

III. EMPIRICAL ANALYSIS OF CREDIT RISK MEASURES OF LISTED PHARMACEUTICAL AND BIOLOGICAL COMPANIES IN CHINA

3.1 Improvement of KMV model

3.1.1 Sample selection

From the construction idea and path of GA-KMV model, it is known that the larger the number of samples, the higher the accuracy of its correction, so in this paper, in terms of sample selection, we will select all ST and *ST class enterprises that implement delisting warning from January 1, 2018 to June 30, 2021.

TABLE 1
BASIC INFORMATION OF PARAMETER CORRECTION TEST SAMPLES

Stock Name	Compare company stock names
*ST Fu Jen	Great Ginseng Forest
*ST Yunsheng	Jiangzhong Pharmaceutical
Hundred Flowers Medicine	Jichuan Pharmaceutical
ST Megadrug	National Development Co.
ST Zhongzhu	Tianyu Co.
ST Comet	Shanhe Pharmacy
*ST and Jia	Meikang Bio
*ST JIYI	Guang Sheng Tang
*ST Biocon	Ruiji Pharmaceutical
*ST Hengkang	Hysco
ST Toyo	Hansen Pharmaceuticals
*ST KH	Jiuzhi Tang
RiverChemical Co.	Yunnan Baiyao
*ST Yikang	South China Biological

A total of 14 pharmaceutical and biological companies in the ST category (including *ST) were implemented during this period. As a control for ST class companies, in order to avoid the differences in variables due to time differences and differences in enterprise size, the corresponding 14 pharmaceutical and biological companies with corresponding years and comparable market capitalization size were selected as non-ST class samples. The parameter correction test sample is shown in Table 1.

3.2 Determine the variable parameters of the KMV model

3.2.1 Value of debt D

The value of debt D for the sample data is obtained from the balance sheet of the enterprise for each year and is equal to the sum of current liabilities and long-term liabilities.

3.2.2 Debt maturity

The term of the debt is one year.

3.2.3 The risk-free rate

The risk-free rate is referenced to the Shanghai interbank 3-month interbank offered rate.

3.2.4 Equity value of the company

In this paper, the following formula is used in the model to calculate the value of the firm's equity:

Equity value = Number of shares outstanding * annual average daily closing price + Number of non-marketable shares * net assets per share

3.3 Calculate the equity value volatility σE

In the following, the equity value volatility σE is calculated by first selecting the daily closing prices of the stock trading days of the sample companies for each year from the RESET database. Since the GARCH (1,1) autoregressive conditional heteroskedasticity model requires the sample data stock returns to be stationary, the daily log returns of the sample data are subjected to ADF test and ARCH effect test. After passing the tests, the GARCH (1,1) model is applied to calculate the daily volatility of the equity value of the sample companies, which in turn yields the annual volatility of the equity value of the sample data σE. The above calculations are performed using Eviews9 econometric software.

For sample companies that are significantly free from volatility aggregation effects, equity value volatility is calculated using historical volatility.

3.3.1 Basic statistical variables

By looking at the data plot of basic statistics of Guofa (600538), it can be seen that: the mean of log return during the sample period is 0.0000892, the standard deviation is 0.022848, and the skewness is -2.120085. the kurtosis is 12.01804 The kurtosis is greater than the standard kurtosis value of 3, which indicates that the daily return of this stock has a spiky fat-tailed distribution. the J-B statistic is 1001.316 with a p-value of 0 and the series is non-normally distributed.

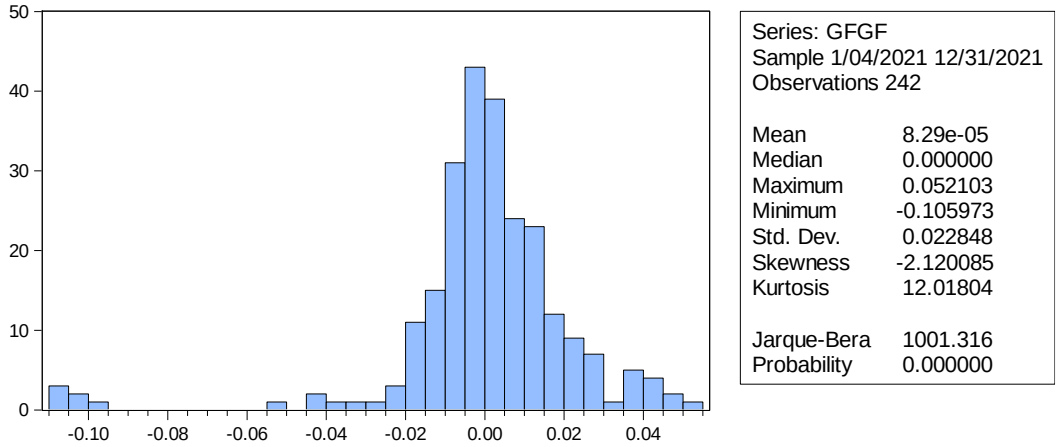


FIGURE 1: Basic Statistics of Guofa (600538)

3.3.2 Smoothness test

Time series analysis requires the data to be smooth, and the following tests whether the daily return series is smooth. According to the results of the ADF test of the log return data of Guofa (600538), we can see that the ADF value (t-Statistic) of the log return data of this stock is -12.66697, which is much smaller than the critical value of the ADF value under the 1%, 5%, and 10% criteria. This indicates that the log return data of the stock of Guofa (600538) is significantly smooth.

TABLE 2
RESULTS OF ADF TEST FOR LOG RETURN DATA OF GUOFA (600538)

Augmented Dickey-Fuller test statistic		t-Statistic	Prob.
		-10.42021	0.0000
Test critical values:	1% level	-3.4574	
	5% level	-2.87334	
	10% level	-2.57313	

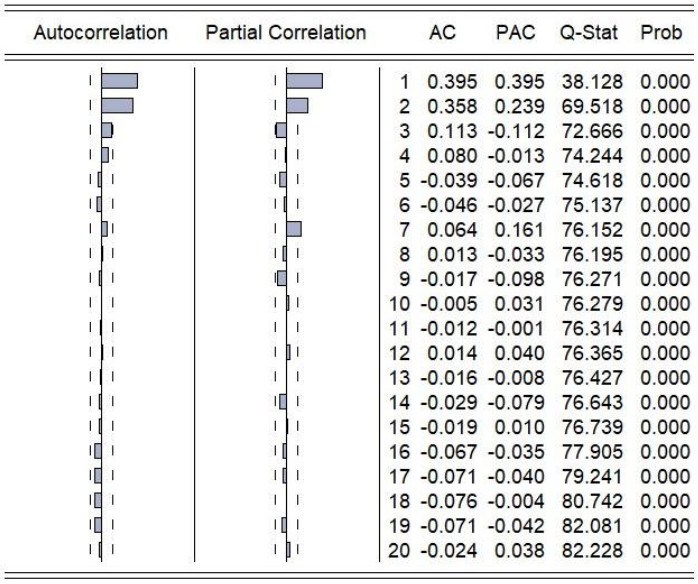
3.3.3 ARCH Effectiveness Test

Continuing with the ARCH test of conditional heteroskedasticity for Guofa (600538), Table 3 shows the results of the ARCH test with lag order of 1. The test shows that the p-value is 0.013, which is less than 0.05, and the original hypothesis is rejected, indicating the existence of ARCH effect.

TABLE 3
RESULTS OF ARCH EFFECT TEST FOR GUOFA (600538)

Heteroskedasticity Test: ARCH		p-value
F-statistic	44.11614	0.0000
Obs*R-squared	37.53019	0.0000

Also, the residuals of the daily returns are tested for AC and PAC. Through Fig. 2, it can be seen that the test results show that the coefficients of AC and PAC are not zero and the Q-Stat statistic is significant, i.e., there is a significant ARCH effect, so the GARCH (1,1) model is constructed.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.395	0.395	38.128	0.000
		2	0.358	0.239	69.518	0.000
		3	0.113	-0.112	72.666	0.000
		4	0.080	-0.013	74.244	0.000
		5	-0.039	-0.067	74.618	0.000
		6	-0.046	-0.027	75.137	0.000
		7	0.064	0.161	76.152	0.000
		8	0.013	-0.033	76.195	0.000
		9	-0.017	-0.098	76.271	0.000
		10	-0.005	0.031	76.279	0.000
		11	-0.012	-0.001	76.314	0.000
		12	0.014	0.040	76.365	0.000
		13	-0.016	-0.008	76.427	0.000
		14	-0.029	-0.079	76.643	0.000
		15	-0.019	0.010	76.739	0.000
		16	-0.067	-0.035	77.905	0.000
		17	-0.071	-0.040	79.241	0.000
		18	-0.076	-0.004	80.742	0.000
		19	-0.071	-0.042	82.081	0.000
		20	-0.024	0.038	82.228	0.000

FIGURE 2: Autocorrelation (AC) and partial autocorrelation (PAC) tests

3.3.4 Equity value volatility

After testing, it is proved that there is an ARCH effect on the log return series of the sample data and a GARCH(1,1) model can be constructed. Therefore, the GARCH(1,1) model was built in Eview9 software. As shown in Table 4, the coefficient of GARCH(1,1) model is 0.879185+0.060267 <1,which satisfies the constraint of model $\alpha+\beta<1$.

Finally, the daily volatility of the equity value of Guofa (600538) was calculated to be 0.051083, resulting in an annual volatility of the equity value of 0.797946.

TABLE 4
REGRESSION RESULTS OF GARCH(1,1) MODEL

	Variance Equation			
C	0.000158	1.69E-05	9.376304	0.0000
RESID(-1)^2	0.879185	0.141330	6.220800	0.0000
GARCH(-1)	0.060267	0.037693	1.598903	0.1098

The daily volatility and annual volatility of the equity value of the remaining sample firms were calculated as shown in Table 5.

TABLE 5
ANNUAL VOLATILITY OF EQUITY VALUE OF SELECTED SAMPLE FIRMS Σ_E

Stock Code	σE	Compare company stock code	σE
600781	0.3747	603233	0.2425
600767	0.2811	600750	0.2462
600721	0.3437	600566	0.3244
600671	0.5242	600538	0.3555
600568	0.5006	300702	0.4135
600518	0.5932	300452	0.3705
300273	0.5755	300439	0.2988

3.4 Optimization of default point parameters using genetic algorithm

The parameters of the violation points are optimized using the MATLAB toolbox. The relevant parameters of the genetic algorithm toolbox are set as shown in the Table 6.

TABLE 6
MATLAB TOOLBOX PARAMETER SETTINGS

Parameter Name	Parameter Value	Parameter Name	Parameter Value
Adaptation function	*1-(m+n)/N*	Crossover Rate	*0.8*
Population size	*50*	Mutation rate	*0.01*
Initial population range	*[0； 10]*	Maximum number of iterations	*200*
Decision Variables	$X_1 X_2$	Function Tolerance	10^{-6}

The algorithm terminates at 51 generations, when the weighted average change of the fitness function value is less than 10-6. Fig. 3 shows the iterative convergence process for one run. Fig. 4 shows the final optimization results of the genetic algorithm in MATLAB software.

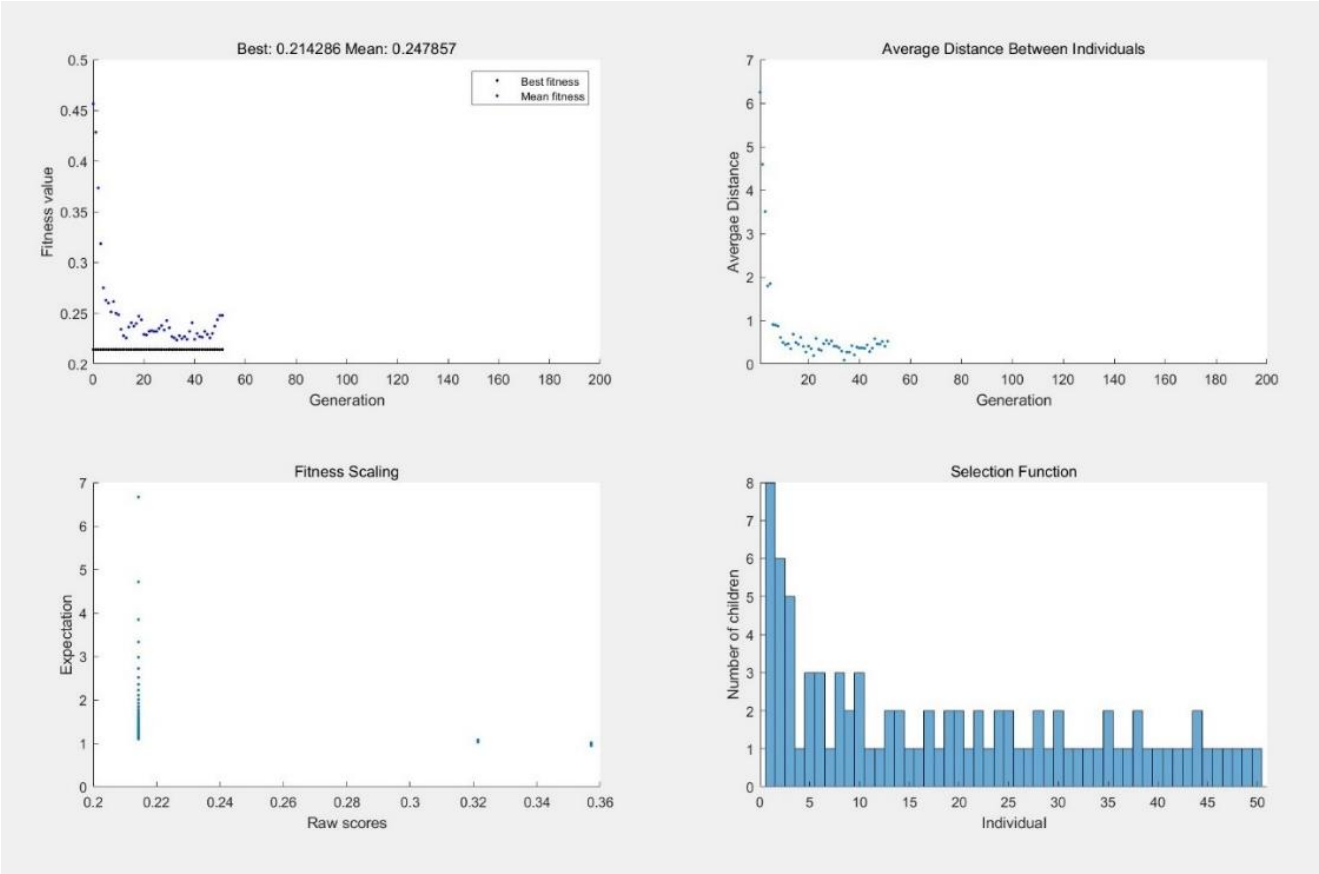


FIGURE 3: Iterative Process Diagram of Genetic Algorithm

According to the population optimization trajectory in the Fig. 3, it can be found that the value of the fitness function converges rapidly to 0.2143 from the 6th generation and reaches the algorithm termination condition in the 51st generation.

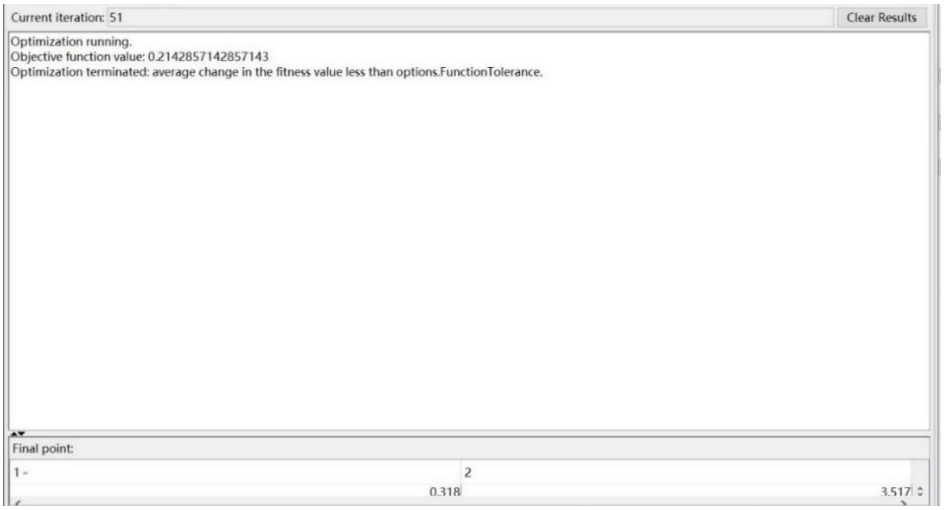


FIGURE 4: Calculation Results of Genetic Algorithm

The results of multiple runs show that the default point parameter values fluctuate up and down between 0.3 and 3.25, resulting in a formula for calculating the default point for listed pharmaceutical and biological companies, as shown in Eq:

DPT=0.318*STD+3.517*LTD

IV. EVALUATION OF GA-KMV MODEL IMPROVEMENT RESULTS

In order to verify the validity of the above parameter correction results, the α, β parameters of the modified KMV model and the parameters given in the traditional KMV model (α=1,β=0.5) were substituted into the KMV model for the two groups of samples, and the DD values of the default distance before and after the correction for the ST group and the non-ST group and the level of correctness of the judgment on the samples before and after the correction were calculated in Table 7. group and the non-ST group before and after the correction of the default distance DD value for direct comparison and analysis, and the difference in the correct rate before and after the correction is obvious. Through the direct comparison of the above data, if the traditional KMV model default point calculation is used, the comprehensive correct rate is only 53.57%, i.e., the difference between ST and non-ST classes cannot be effectively judged, while the comprehensive accuracy of the corrected model is 78.57%. Before the correction, the mean values of default distance for ST class sample and non-ST class sample are 1.66 and 2.37, respectively, and the difference between them is not significant. After the model correction, the mean values of default distance for ST class sample and non-ST class sample are -2.66 and 0.97, respectively, and the difference between them is significant.

TABLE 7
COMPARISON OF EMPIRICAL RESULTS BEFORE AND AFTER THE REVISION OF GA-KMV MODEL

Stock Name	DD before correction	Corrected DD	Stock Name	DD before correction	Corrected DD
*ST Fu Jen	1.3447	-5.7734	Great Ginseng Forest	3.9185	0.4503
*ST Yunsheng	3.4552	3.0991	Jiangzhong Pharmaceutical	2.1476	0.2754
Hundred Flowers Medicine	2.1696	-0.4834	Jichuan Pharmaceutical	2.0257	0.6908
ST Megadrug	1.9059	0.9550	National Development Co.	2.2085	1.5476
ST Zhongzhu	0.5384	-3.0365	Tianyu Co.	1.6378	-0.7915
ST Comet	0.5229	-1.2733	Shanhe Pharmacy	2.2065	0.8929
*ST and Jia	1.5422	0.8068	Meikang Bio	2.8811	0.4465
*ST JIYI	2.8281	1.5793	Guang Sheng Tang	2.2241	1.3322
*ST Biocon	0.8966	-12.7071	Ruiji Pharmaceutical	2.4024	0.8725
*ST Hengkang	2.8630	-2.1885	Hysco	1.8898	1.2146
ST Toyo	1.6662	-2.0759	Hansen Pharmaceuticals	3.1069	1.5602
*ST KH	-1.1361	-16.6567	Jiuzhi Tang	2.4109	0.3724
RiverChemical Co.	2.4758	1.2382	Yunnan Baiyao	1.6360	2.3030
*ST Yikang	2.2138	-0.7856	South China Biological	2.4895	2.4557
Number of correct	1	9	Number of correct	14	13
Correct Rate	7.14%	64.29%	Correct Rate	100%	92.86%

To further verify the validity of the above intuitive findings, the Mann Whitney U test is used for testing. Since the Mann Whitney U test has fewer assumptions and does not require that the samples meet conditions such as normal distribution and chi-squaredness, it is more widely applicable. The test results calculated by SPSS software are shown in Table 8. As can be seen from the table, before the model correction its Z value is -1.654, the asymptotic significance (two-sided) and exact significance (two-sided) are both greater than 0.05, so the original hypothesis H0 is accepted, i.e., the model before the correction cannot effectively distinguish between ST and non-ST class samples by applying it. The z-value after model correction is -2.16, the asymptotic significance (two-sided) value is 0.031 and the exact significance (two-sided) value is 0.031, both of which are less than 0.05, so the original hypothesis H0 is rejected, i.e., there is a significant difference between ST and non-ST class samples.

TABLE 8
CORRELATIONS BETWEEN PRE-CORRECTION AND POST-CORRECTION PAIRED SAMPLES

Category	Before amendment	After correction
The original hypothesis H0	There is no significant difference between ST and non-ST categories	
Mann Whitney U	62.000	51.000
Wilcoxon W	167.000	156.000
Z	-1.654	-2.160
Asymptotic saliency (two-tailed)	0.098	0.031
Exact significance [2*(one-tailed significance)	0.104[b]	0.031

Therefore, the improved KMV model based on genetic algorithm has practical value for judging the level of credit default risk of listed companies in China's A-share pharmaceutical and biological industry.

V. EMPIRICAL TEST ANALYSIS BASED ON IMPROVED KMV MODEL

This paper comprehensively counts the trading data and financial statements of all listed companies in the pharmaceutical and biological industry in A-shares from 2019-2021, excluding some companies with a long suspension period, listed for less than three years and listed separately on multiple stock exchanges, and finally recording a total of 290 listed companies that meet the conditions.The KMV model was solved by programming the MATLAB software to first find the asset market value VA and asset market value volatility σA for 290 listed pharmaceutical and biological companies for 2019-2021, and then apply the modified model to calculate the default point and finally calculate the default distance value DD for each year for the 290 sample listed companies. the descriptive statistics calculated using SPSS are shown in Table 9.

TABLE 9
DESCRIPTIVE STATISTICS OF THE DISTANCE TO DEFAULT FOR 290 LISTED PHARMACEUTICAL AND BIOLOGICAL COMPANIES, 2019-2021

	N	Minimum value	Maximum value	Average value	Standard deviation
2019 Default Distance	290	-13.88	6.83	-0.1877	2.28272
2020 Default Distance	290	-11.77	7.24	-0.0702	2.24939
2021 Default Distance	290	-63.94	7.11	-2.1643	9.15224
Number of effective cases (in columns)	290				

According to the calculation, a total of 114 companies in 2021 and 2020 and 124 listed companies in 2019 are judged to be in default, accounting for 39.31% and 42.76% of the overall total, respectively. If we apply the U.S. empirical default point formula DPT = STD + 0.5*LTD to the calculation, we will find that the formula determines the vast majority of companies as non-defaulting in their entirety, thus making it difficult to identify defaulting companies. Although more companies are judged to be in default after using the improved default point, this is acceptable because from the perspective of risk control, it can make companies take timely measures and methods for self-management to improve their credit rating and ensure their normal operation. In addition, the default distance here is only used as a relative judgment indicator of whether an enterprise is in default, and not as an absolute indicator of the probability of default.

Observing the average default distance, it can be seen that the credit risk of listed companies in pharmaceutical and biological industries decreases and then increases in the past three years, with the lowest credit risk in 2020 and the highest in 2021. Analysis of the standard deviation of default distance shows that the standard deviation of default distance in 2019 and 2020 is about 2.2, and the standard deviation in 2021 increases to 9.15, indicating that the gap of default distance among listed companies in the pharmaceutical and biological industry in 2021 is large and polarized.

VI. RESEARCH CONCLUSIONS

This paper uses the basic principles and calculation methods of KMV model and genetic algorithm to improve the default point parameters in the traditional KMV model by selecting pharmaceutical and biological listed companies marked as ST in the last three years and companies in the same industry with comparable asset value with their assets, and using MATLAB software to build a GA-KMV model to make it more consistent with the actual situation of the Chinese A-share market. Finally, the modified model is applied to empirically analyze the credit risk profile of Chinese A-share listed pharmaceutical and biological companies, and the conclusions are as follows:

Using the GA-KMV model to modify the parameters of 28 sample enterprises, the corrected default point α value is 0.318 and β value is 3.517. The comprehensive accuracy of the modified model for judging whether an enterprise is in default is 78.57%, which can effectively distinguish the credit risk of ST class enterprises and non-ST class enterprises, but if the traditional KMV model is used, its comprehensive correct rate of default or not However, if the traditional KMV model is used, the combined correct rate of default and non-default is only 53.57%, i.e., it cannot effectively distinguish the difference between ST class and non-ST class enterprises.

Using the improved KMV model to measure the credit risk of 290 listed companies in China's A-share pharmaceutical and biological industry in 2019-2021, the analysis shows that the credit risk of the pharmaceutical and biological industry decreased and then increased in the past three years, and the credit risk of the pharmaceutical and biological industry decreased in the year of the outbreak of the new crown epidemic in 2020, and increased one year after the outbreak of the epidemic, and the credit risk of the industry was polarized phenomenon. In the future, the concentration of the pharmaceutical and biological industry will probably continue to increase, and there is more room for the development of head enterprises.

REFERENCES

[1] Liu Yingchun, Liu Xiao. A study of KMV credit risk measure based on GARCH volatility model[J]. Journal of Northeast University of Finance and Economics,2011,(03):74-79.

[2] Tang ZP, Chen WEI-HONG, Huang YOU-PO. Measurement of credit risk of listed companies[J]. Statistics and decision making,2016,(24):174-179.

[3] Feng Jinghai,Tian Jing. Optimal default point determination based on genetic algorithm KMV model[J]. Journal of Dalian University of Technology,2016,56(02):181-184.

[4] Zhou Shuyuan. Research on credit risk evaluation of overcapacity industries based on GAGARCH-KMV model[J]. Modern Business,2019,(32):121-123.

[5] Zhu Ling, Wang Mingzhe. A study on default risk measurement and heterogeneity of listed enterprises' debts in Zhejiang--based on KMV model[J]. Business Development Economics,2022,(08):88-91.

[6] Yang X. Research on credit risk measure of listed technology finance companies based on KMV model[J]. China Price,2022,(04):78-80.

[7] You Zongjun,Cheng Xiaoxuan. The deductive logic of KMV correction model:A literature review[J]. Contemporary Economics,2022,39(04):17-22.

[8] Ren M,Wang C.M.,Yan H.Y.. The impact of the new crown pneumonia epidemic on the stock price of China's pharmaceutical industry--an empirical analysis based on ARIMA model[J]. Northern Economic and Trade,2022,(02):93-96.

[9] Duan Yuyuan. The impact of the new crown pneumonia epidemic on China's stock market--an empirical analysis based on the pharmaceutical industry[J]. China Business Journal, 2020,(18):28- 30.

[10] Yu Miao-Chi, Hua Si-Yu. Research on corporate bond default risk measure based on genetic algorithm KMV model[J]. Technology and Economics,2020,33(03):51-55.

[11] Zhou Xuesong. New crown pneumonia outbreak has mixed impact on pharmaceutical industry [N]. China Economic Times, 2020-02-14(002).

[12] Wang Xiu-Guo, Xie You-Huang. Extended KMV model based on CVaR and GARCH(1,1)[J]. Systems Engineering,2012,30(12):26-32.