

Article

Financial Distress Prediction for Digital Economy Firms: Based on PCA-Logistic

Dongyang Li ¹, Kai Xu ^{1*}, Yun Li ¹, Yu Jiang ¹, Ming Tang ¹, Yangdan Lu ¹, Chun Cheng ¹, Chunxiao Wang ¹ and Guanbing Mo ¹

¹ Business School, Chengdu University, Chengdu (610106), Sichuan, China

* Correspondence: xukai@cdu.edu.cn

Received: February 16, 2022; Accepted: March 30, 2022; Published: April 15, 2022

Abstract: Financial distress prediction is important for risk prevention and control of digital economy firms, as well as going concern guarantee. This paper takes 100 Chinese A-share listed digital economy firms from 2017 and 2021 as samples, obtains financial indicators by combining the characteristics of digital economy firms, the first three periods of financial distress are systematically modeled employs Logistic regression, while we use the Principal Component Analysis method to deal with the problem of multicollinearity. The results show that the profitability factor has the greatest contribution to the predictive role; the closer to the year in which the financial distress occurred, the higher the prediction accuracy rate. Finally, this model achieves 86% prediction accuracy. The successful modelling provides a basis for information users to determine the financial distress of firms accurately and prospectively in the digital economy.

Keywords: Financial Distress; Digital Economy; Principal Component Analysis; Logistic Regression

1. Introduction

In the economic globalization and the new economy era, the financial situation and financial distress of digital economy firms are under attention of the society. Compared with other industries, the digital economy industry has typical characteristics of high growth, high profitability, focus on innovation, and highly competitive intensity in the field, which also makes the financial situation of digital economy firms dynamic and high-risk. When financial risks are not controlled and accumulate to a certain level to produce qualitative changes, it will cause financial distress. Financial distress can be considered as a business failure, bankruptcy crisis [1], debt crisis or credit risk [2]. During a firm's decline, the indicators reflecting the firm's operations, assets, and liabilities often show abnormal changes from those of a normally operating firm, followed by difficulties and eventual bankruptcy. Timely predicting of corporate financial distress purposes to reduce investors' losses and help management make strategic adjustments to avoid facing operating crisis.

The so-called financial distress prediction, that is based on risk management, according to the combination of theory and methodology, analysis of the macro environment, micro governance, business conditions and financial management faced by firms [3]. PCA (Principal Component Analysis) constructs a few representative principal factor indicators from many indicators of the firms, which provide a method to avoid the multicollinearity caused by high correlation among indicators and achieve the purpose of reducing and simplifying variables. The advantages of PCA-Logistic are, firstly, using the symmetric feature of PCA covariance matrix to weaken the sensitivity

of Logistic model to covariance. Secondly, replacing most indicators with few indicators to realize the initial data dimensionality reduction of Logistic model. Thirdly, reducing the workload and errors of manual screening indicators and improving the overall prediction accuracy of the model.

In view of this, we aim to contribute to research into the accounting-based distress model, and specifically in two ways. Firstly, by adding R&D investment, a core characteristic indicator of digital economy firms. Secondly, PCA-Logistic method is used to make the model more fit for digital economy firms while having high accuracy. Therefore, this paper utilizes a combination of PCA and Logistic to build a financial distress prediction model with a sample of Chinese A-share listed digital economy firms from 2017-2021 to improve the accuracy of prediction. The innovations of this paper are, firstly, the combination of PCA method and logistic model to build a financial distress prediction model compared to the single prediction model used in the past. Secondly, digital economy firms are emerging industries and growing rapidly in China, previous studies lacked a financial predicting methodology for the digital economy industry, and this paper complements this.

The remainder of this paper is organized as follows. Section 2 gives overview of previous literature in the field of distress prediction. Section 3 describes the research design for building the distress prediction model. Section 4 presents the empirical analysis, which contains PCA and logistic models. Section 5 concludes.

2. Literature Review

The study of corporate financial distress has been carried out since the early 20th century, and the first research scholar Fitzpatrick (1932) defined financial distress as the inability of a firm to pay its debts as they fall due [4]. Foster (1985) considered the existence of financial distress when a firm needs to face a reorganization situation [5]. Laitinen (1991) in his study classifies firms with different levels of distress and thus identifies the firms that are in financial distress [6]. Following behind, Chinese researchers have also conducted in-depth studies on financial distress. Wu and Huang (1987) argued that a firm is found to be in financial distress when it cannot pay its debts as they fall due and loses its legal personality [7]. Wu (2011) defines financial distress as a situation in which an enterprise's ability to pay is insufficient after weighing the force of financial distress occurrence against its resisted force in the context of embedded stakeholder behavior [8]. Based on previous researchers' definitions, this paper defines a financially distressed firm as a firm that has been given special treatment and is at risk of delisting due to consecutive years of losses.

The construction of financial distress prediction models is the key to ensure the efficiency of forecasting, which has been fruitfully studied in the academic. Fitzpatrick (1932) pioneered the univariate analysis of financial indicators of listed companies; subsequently, Altman (1968) used multivariate linear discriminant methods to build multivariate financial prediction model [9]. Since multivariate discriminant models require strict distribution and covariance of independent variables, which limit the selection of samples, in response Maitin (1977) explored and used logistic model to construct a bank warning system [10]. Based on this, Ohlson (1980) introduced four non-financial variables and used logistic regression for predictive analysis [11], and since then more researchers have used logistic model for financial distress prediction. Chinese researcher Jiang (2001) proved that logistic model has high predictive accuracy through model testing [12]. Li et al. (2011) concluded that logistic model is more suitable for analyzing the actual situation of financial distress compared to linear models [13]. Li (2018) predicted financial risk with high accuracy through logistic regression

[14]. Based on the comparison of previous studies, it is found that the use of logistic model is more representative, and given that machine learning is still developing, this paper uses the widely used logistic regression to construct a financial distress prediction model.

3. Data and Methodology

3.1. Sample Selection

According to the research of Altman et al (2017) and Sun et al (2014), firms with two consecutive years of losses, insolvency [15], and audit negativity are defined as firms in financial distress [16]. In accordance with the rules of China Shanghai and Shenzhen A-share markets, this paper uses firms that have been ST or *ST for the above reasons as a sample of firms in financial distress, and firms that have not been ST or *ST as a sample of healthy firms.

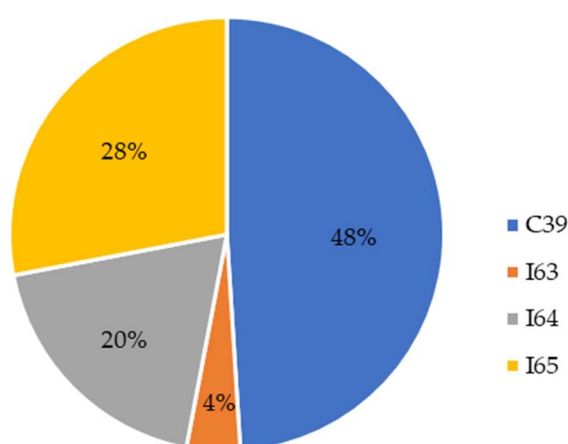


Figure 1. Industry Distribution of Digital Economy Firms.

Previous researchers and institutions have not been consistent in defining the digital economy, but their definitions have some common points, which emphasize digital technologies, networks, industrial convergence and their impact on the economy [17]. Hence, based on the characteristics of digital economy, this paper defines digital economy firms as relying on digital platforms and applications, etc., providing products or services with digital technology innovation as the core, at the same time promoting the industrial integration of digital technology and real economy. According to the "Statistical Classification of Digital Economy and its Core Industries" released by China Statistics Bureau in 2021 to determine the classification of digital economy firms, this paper obtains A-share listed firms in China comes from between 2017 and 2021 as the sample, the year in which the financial distress occurred is taken as period t . Referring to the method of Wang et al. (2017), data from the previous three years, i.e. period $t-1$, period $t-2$ and period $t-3$ are selected as forecast data according to the ratio of 1:1 between financially distress firms and financially healthy firms [18]. A total of 50 firms in financial distress and 50 healthy firms are used as controls. According to the industry classification of China Securities Regulatory Commission (CSRC), the sample can be divided into four industries, namely, Telecommunications, radio and television and satellite transmission services (I63); Internet and related services (I64); Software and information technology services (I65); Computer, communication and other electronic equipment manufacturing (C39). And the industry

distribution of the sample is shown in Figure 1. The data are obtained from a database in which the indicators are calculated according to the annual reports by listed firms.

3.2. Variable Definition

Financial indicators are an intuitive indicator to reflect the financial and operational status of firms. In this paper, variables are selected according to five dimensions: solvency, profitability, operational capacity and development capacity as well as cash flow status of each firm. In addition, a total of 20 independent variables are selected considering the characteristics of digital economy firms focusing on innovation and R&D. The definition of each variable is described in Table 1.

Table 1. Variable Definition.

| Variables | Abbrev | Formulas |
|--------------------------------------|-----------------|---|
| Financial Distress | FAIL | financial distress using 1, financial health using 0 |
| Current Ratio | X ₁ | Current assets / Current liabilities |
| Quick Ratio | X ₂ | (Current assets - Inventories) / Current liabilities |
| Cash Ratio | X ₃ | Ending balance of cash and cash equivalents / Current liabilities |
| Debt-to-Asset Ratio | X ₄ | Total liabilities / Total assets |
| Accounts Receivable Turnover Ratio | X ₅ | Operating income / Accounts receivable |
| Inventory Turnover | X ₆ | Operating costs / Inventory |
| Current Asset Turnover | X ₇ | Operating income / Current assets |
| Total Assets Turnover | X ₈ | Revenue from main business / Total assets |
| Return on Assets | X ₉ | (Total profit + Finance costs) / Total assets |
| Total Net Asset Margin | X ₁₀ | Net profit / Total assets |
| Return on Net Assets | X ₁₁ | Net income / Shareholders' equity balance |
| Gross Operating Margin | X ₁₂ | (Operating revenues - Operating costs) / Operating revenues |
| Total Assets Growth Rate | X ₁₃ | (Closing assets - Opening assets) / Opening assets |
| Operating Income Growth Rate | X ₁₄ | (Ending operating revenue - Opening operating revenue) / Opening operating revenue |
| Sustainable Growth Rate | X ₁₅ | Return on net assets * Earnings retention rate / (1 - Return on net assets * Earnings retention rate) |
| Net Assets per Share Growth Rate | X ₁₆ | (Closing net assets per share - Opening net assets per share) / Opening net assets per share |
| Net Profit Growth Rate | X ₁₇ | (Closing net income - Opening net income) / Opening net income |
| Net Cash Content of Operating Income | X ₁₈ | Net cash flow from operating activities / Total operating income |
| Operating Index | X ₁₉ | Net cash flows from operating activities / Cash generated from operations |
| R&D Investment Ratio | X ₂₀ | R&D investment / Total assets |

3.3. Model Design

Due to the limited number of digital economy firms in A-shares, but the variables for discriminating financial distress are more and multidimensional, using a large number of financial

variables will have multicollinearity and overfitting problems. Meanwhile, Logistic regression is sensitive to the covariance of variables, and manual elimination of variables is easy to ignore multidimensional discrimination. To deal with this issue, the advantage of PCA lies in the dimensionality reduction of variables by factor coefficients to avoid over-fitting problems caused by too many independent variables.

The concept of PCA is based on the maximum variance theory, which maps the original features to feature vectors, and the feature value corresponding to each feature vector is the variance after the projection of the original features to the projection surface of the feature vector. In order to ensure that the information is not lost as much as possible, PCA will select the projection plane with larger variance as the projection plane of the original features to get the feature value with the maximum information, so that the loss of information after dimensionality reduction is minimized.

In this study, we first discriminate the significance of the difference of variables between the two types of samples, use PCA method to reduce the dimensionality of variables when there are more variables, and use logistic regression for the obtained principal component factors, finally discriminate the accuracy of the model for financial distress prediction of digital economy firms.

The expression about the Logistic model is show in equation (1):

$$\text{Logistic}(FAIL_{i,t}) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_{j=1}^J \beta_j X_{i,j} \quad (1)$$

Among other things, $i=1, \dots, n$, $j=1, \dots, J$, $X_{i,j}$ is the j^{th} variable of the i^{th} firm, variable j has a total of 20 variables, and t is the prediction period.

Logistic regression, a linear model commonly used to deal with dichotomous problems, is widely applicable in the field of predicting financial distress, besides Alifiah and Norfian (2014) showed a high prediction accuracy [19]. In this study, the explanatory variables are dichotomous {0, 1} variables to measure whether a firm is in financial distress. X is the financial index of the firm, and the coefficient β is obtained by the method of great likelihood estimation, which ultimately leads to the estimated $FAIL=1$ probability $\hat{p}_{i,t}$ for the i^{th} firms in period t . The expression is as in equation (2):

$$\hat{p}_{i,t}(FAIL_{i,t} = 1 | X_{i,t}) = \frac{\exp(\hat{\beta}_{0,t} + \sum_{j=1}^J \hat{\beta}_{j,t} X_{i,t})}{1 + \exp(\hat{\beta}_{0,t} + \sum_{j=1}^J \hat{\beta}_{j,t} X_{i,t})} \quad (2)$$

If $\hat{p}_{i,t}$ is greater than 0.5, there is a financial distress signal. Conversely, if $\hat{p}_{i,t}$ is less than 0.5, there is no financial distress signal [20].

4. Empirical Results

4.1. K-S and Mann-Whiteoy U Test

We first apply Kolmogorov-Smirnov(K-S) to test whether the variables of the sample overall are normally distributed, and the results show that period t-3, period t-2 X_4 and period t-1 X_8 , X_{13} are in line with normal distribution using independent samples T-test, and the remaining indicators are subjected to Mann-Whiteoy U test for significance of differences in variables between the financial distress sample and the financial health sample. The indicators find to be significantly different in

period t-3 are $X_1, X_3, X_5, X_8, X_{20}$. The indicators with significant differences in period t-2 are $X_3, X_5, X_7, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{19}, X_{20}$. In period t-1, the indicators with significant differences are $X_1, X_2, X_3, X_4, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{18}, X_{19}$. Since there are only five significantly different indicators in period t-3, which is not suitable for principal component factor construction, this paper uses Logistic regression after testing the covariance to construct the model for this period. For the remaining two periods, the indicators with significant differences are first downscaled using Principal Component Analysis, and then a Logistic model will build.

4.2. Principal Component Analysis

The Kaiser-Meyer-Olkin test (KMO) and the Bartlett's test for sample indicators is first used to determine the suitability for principal component analysis. The results in Table 2 show that the KMO is greater than 0.5 in both periods and the Bartlett's test of sphericity is significant at the 1% level of the mean, indicating that the data in both periods are suitable for principal component analysis.

Table 2. KMO and Bartlett's Test Results.

| | t-2 | t-1 |
|---|-------|-------|
| KMO measure of sampling adequacy quantity | 0.554 | 0.563 |
| Bartlett's test of sphericity sig | 0.000 | 0.000 |

Table 3. Component matrix.

| | t-2 | | | | | | t-1 | | | | | |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | F_1 | F_2 | F_3 | F_4 | F_5 | F_6 | F_1 | F_2 | F_3 | F_4 | F_5 | F_6 |
| X_1 | | | | | | | -0.006 | 0.944 | 0.132 | 0.186 | 0.060 | 0.008 |
| X_2 | | | | | | | -0.029 | 0.943 | 0.131 | 0.189 | 0.071 | 0.004 |
| X_3 | -0.019 | 0.003 | 0.235 | 0.618 | 0.660 | 0.198 | 0.135 | 0.871 | 0.038 | 0.151 | -0.047 | -0.038 |
| X_4 | | | | | | | -0.569 | -0.444 | -0.209 | 0.038 | 0.379 | -0.084 |
| X_5 | 0.099 | 0.797 | 0.094 | 0.081 | 0.295 | 0.004 | | | | | | |
| X_6 | | | | | | | | | | | | |
| X_7 | 0.038 | 0.916 | 0.075 | -0.134 | -0.117 | -0.060 | 0.226 | -0.433 | 0.647 | 0.427 | -0.001 | 0.012 |
| X_8 | 0.049 | 0.919 | 0.094 | -0.044 | -0.059 | -0.105 | 0.244 | -0.383 | 0.668 | 0.431 | 0.013 | -0.005 |
| X_9 | 0.851 | 0.067 | -0.341 | 0.083 | -0.130 | -0.021 | 0.921 | -0.134 | -0.168 | -0.038 | -0.114 | -0.011 |
| X_{10} | 0.894 | 0.082 | -0.300 | 0.098 | -0.100 | -0.023 | 0.929 | -0.114 | -0.168 | -0.022 | -0.128 | -0.012 |
| X_{11} | 0.721 | 0.072 | -0.257 | 0.164 | 0.095 | -0.008 | 0.226 | 0.064 | 0.027 | 0.026 | -0.141 | 0.826 |
| X_{12} | | | | | | | 0.382 | 0.189 | -0.144 | -0.134 | 0.718 | 0.038 |
| X_{13} | 0.470 | -0.141 | 0.639 | -0.317 | -0.138 | -0.064 | 0.782 | -0.062 | -0.312 | -0.156 | -0.178 | 0.045 |
| X_{14} | 0.065 | -0.174 | -0.329 | -0.053 | 0.206 | -0.572 | 0.524 | -0.007 | -0.124 | 0.031 | -0.019 | -0.401 |
| X_{15} | 0.787 | -0.145 | 0.114 | 0.036 | 0.135 | 0.082 | 0.397 | 0.139 | 0.540 | -0.659 | -0.124 | -0.010 |
| X_{16} | 0.736 | -0.188 | 0.526 | -0.082 | 0.025 | -0.009 | 0.101 | -0.065 | -0.492 | 0.748 | -0.209 | 0.083 |
| X_{17} | 0.213 | -0.054 | -0.139 | -0.542 | 0.403 | 0.214 | 0.499 | 0.082 | 0.104 | 0.222 | 0.402 | -0.004 |
| X_{18} | | | | | | | 0.546 | -0.107 | 0.028 | 0.135 | 0.469 | -0.016 |
| X_{19} | 0.028 | 0.084 | -0.215 | -0.054 | -0.209 | 0.762 | 0.102 | 0.170 | 0.076 | 0.144 | -0.343 | -0.412 |
| X_{20} | 0.063 | -0.031 | 0.169 | 0.582 | -0.471 | -0.046 | | | | | | |

In extracting the common factors, the paper takes the eigenvalues with a cumulative contribution greater than 70% as the main components. A total of 6 principal component factors are extracted in period t-2, with a cumulative contribution of 75.01%, and a total of 6 principal component factors are extracted in period t-1, with a cumulative contribution of 75.34%. According to Table 3 of the component matrix, the economic meaning of each principal factor can be determined. In period t-2, the significance of principal factor F_1 is profitability (X_9, X_{10}, X_{11}), development capacity (X_{15}, X_{16}); the significance of principal factor F_2 is operating capacity (X_5, X_7, X_8); the significance of principal factor F_3 is development capacity (X_{13}, X_{16}); the significance of principal factor F_4 is solvency (X_3), innovation capacity (X_{20}); the significance of the main factor F_5 is solvency (X_3), development capacity (X_{17}), innovation capacity (X_{20}); the significance of the main factor F_6 is development capacity (X_{14}), cash flow (X_{19}). In period t-1, the significance of principal factor F_1 is profitability (X_9, X_{10}), development capacity (X_{13}, X_{14}); the significance of principal factor F_2 is solvency (X_1, X_2, X_3, X_4); the significance of principal factor F_3 is operating capacity (X_7, X_8); the significance of principal factor F_4 is development capacity (X_{15}, X_{16}); the significance of principal factor F_5 is profitability (X_{12}), development capacity (X_{17}), cash flow (X_{18}, X_{19}); the significance of the main factor F_6 is profitability (X_{11}), cash flow (X_{19}).

The calculated two-period principal component factor score matrices, which are omitted from the table to save space, and the two-period score matrices as coefficients of the six principal component factors, are expressed as follows.

Each principal factor score in period t-2 equations:

$$F_1 = -0.006X_3 + 0.028X_5 + 0.011X_7 + 0.014X_8 + 0.244X_9 + 0.256X_{10} + 0.207X_{11} + 0.135X_{13} + 0.018X_{14} + 0.255X_{15} + 0.211X_{16} + 0.061X_{17} + 0.008X_{19} + 0.018X_{20} \quad (3)$$

$$F_2 = 0.001X_3 + 0.325X_5 + 0.373X_7 + 0.375X_8 + 0.027X_9 + 0.033X_{10} + 0.029X_{11} - 0.057X_{13} - 0.071X_{14} - 0.059X_{15} - 0.077X_{16} - 0.022X_{17} + 0.034X_{19} - 0.012X_{20} \quad (4)$$

$$F_3 = 0.188X_3 + 0.075X_5 + 0.060X_7 + 0.075X_8 - 0.273X_9 - 0.240X_{10} - 0.205X_{11} + 0.511X_{13} - 0.263X_{14} + 0.091X_{15} + 0.421X_{16} - 0.112X_{17} - 0.172X_{19} + 0.135X_{20} \quad (5)$$

$$F_4 = 0.515X_3 + 0.068X_5 - 0.112X_7 - 0.037X_8 + 0.069X_9 + 0.082X_{10} + 0.137X_{11} - 0.265X_{13} - 0.044X_{14} + 0.030X_{15} - 0.069X_{16} - 0.452X_{17} - 0.045X_{19} + 0.485X_{20} \quad (6)$$

$$F_5 = 0.609X_3 + 0.272X_5 - 0.108X_7 - 0.055X_8 - 0.120X_9 - 0.093X_{10} + 0.087X_{11} - 0.127X_{13} + 0.190X_{14} + 0.124X_{15} + 0.023X_{16} + 0.371X_{17} - 0.193X_{19} - 0.434X_{20} \quad (7)$$

$$F_6 = 0.194X_3 + 0.004X_5 - 0.059X_7 - 0.103X_8 - 0.021X_9 - 0.023X_{10} - 0.008X_{11} - 0.062X_{13} - 0.560X_{14} + 0.081X_{15} - 0.008X_{16} + 0.209X_{17} + 0.745X_{19} - 0.045X_{20} \quad (8)$$

Each principal factor score in period t-1 equations:

$$F_1 = -0.001X_1 - 0.007X_2 + 0.034X_3 - 0.143X_4 + 0.057X_7 + 0.061X_8 + 0.232X_9 + 0.234X_{10} + 0.057X_{11} + 0.096X_{12} + 0.197X_{13} + 0.132X_{14} + 0.100X_{15} + 0.025X_{16} + 0.126X_{17} + 0.138X_{18} + 0.026X_{19} \quad (9)$$

$$F_2 = 0.294X_1 + 0.293X_2 + 0.271X_3 - 0.138X_4 - 0.135X_7 - 0.119X_8 - 0.042X_9 - 0.035X_{10} + 0.020X_{11} + 0.059X_{12} - 0.019X_{13} - 0.002X_{14} + 0.043X_{15} - 0.020X_{16} + 0.026X_{17} - 0.033X_{18} + 0.053X_{19} \quad (10)$$

$$F_3 = 0.078X_1 + 0.077X_2 + 0.023X_3 - 0.124X_4 + 0.384X_7 + 0.396X_8 - 0.099X_9 - 0.100X_{10} + 0.016X_{11} - 0.085X_{12} - 0.185X_{13} - 0.074X_{14} + 0.320X_{15} - 0.292X_{16} + 0.062X_{17} + 0.017X_{18} + 0.045X_{19} \quad (11)$$

$$F_4 = 0.117X_1 + 0.119X_2 + 0.095X_3 + 0.024X_4 + 0.268X_7 + 0.271X_8 - 0.024X_9 - 0.014X_{10} + 0.016X_{11} - 0.084X_{12} - 0.098X_{13} + 0.019X_{14} - 0.414X_{15} + 0.470X_{16} + 0.139X_{17} + 0.085X_{18} + 0.090X_{19} \quad (12)$$

$$F_5 = 0.046X_1 + 0.054X_2 - 0.036X_3 + 0.289X_4 - 0.001X_7 + 0.010X_8 - 0.087X_9 - 0.098X_{10} - 0.107X_{11} + 0.548X_{12} - 0.136X_{13} - 0.014X_{14} - 0.095X_{15} - 0.159X_{16} + 0.307X_{17} + 0.358X_{18} - 0.262X_{19} \quad (13)$$

$$F_6 = 0.008X_1 + 0.004X_2 - 0.037X_3 - 0.081X_4 + 0.012X_7 - 0.005X_8 - 0.011X_9 - 0.012X_{10} + 0.800X_{11} + 0.037X_{12} + 0.043X_{13} - 0.389X_{14} - 0.009X_{15} + 0.081X_{16} - 0.004X_{17} - 0.016X_{18} - 0.399X_{19} \quad (14)$$

4.3. Logistic Regression

In this paper, the t-3 period variables through the presence of differences and the t-2 and t-1 period principal factors derived through principal component analysis are used as independent variables in the Logistic regression. The significant difference indicators for t-3 periods are X_1 , X_3 , X_5 , X_8 , X_{20} . After the cointegration test the variance inflation factor (VIF) of X_1 and X_3 is greater than 5, which indicates that there may be cointegration. Considering that the current ratio (X_1) includes cash-like assets, the cash ratio (X_3) is excluded. The Logistic regression coefficients for the three periods are shown in Table 4. From the regression results, it can be seen that the R&D investment ratio (X_{20}) in period t-3 is a significant predictor of financial distress, and the lower the R&D investment ratio is, the higher the possibility of financial distress. Additionally, the principal factor F_1 in both period t-2 and period t-1 is negatively and significantly correlated with financial risk at the 1% level, and F_1 in both periods includes corporate profitability, indicating that the profitability factor has the largest contribution to the financial distress prediction of digital economy firms.

Table 4. Logistic model estimated coefficients.

| Period/Factors | t-3 | t-2 | t-1 |
|---------------------------|-----------|-----------|-----------|
| X_1 | -0.100 | | |
| X_3 | -0.016 | | |
| X_8 | 0.294 | | |
| X_{20} | -25.827** | | |
| F_1 | | -1.234*** | -3.158*** |
| F_2 | | -0.079 | -0.099 |
| F_3 | | 0.078 | -1.625 |
| F_4 | | -0.578** | 1.699 |
| F_5 | | -0.005 | 0.624 |
| F_6 | | -0.331 | 0.033 |
| Constant | 0.740 | 0.130 | 0.901** |
| Observations | 100 | 100 | 100 |
| -2 Log Likelihood | 130.227 | 117.612 | 85.595 |
| Nagelkerke R ² | 0.107 | 0.253 | 0.549 |

Note: *, **, *** indicate statistically significant at 10%, 5%, 1% level.

Accordingly, the equation of the Logistic model for each period can be derived

$$\text{Logistic}(FAIL_{i,t-3}) = 0.740 - 0.100X_1 - 0.016X_5 + 0.294X_8 - 25.827X_{20} \tag{15}$$

$$\begin{aligned} \text{Logistic}(FAIL_{i,t-2}) = & 0.130 - 1.234F_1 - 0.079F_2 + 0.078F_3 - 0.578F_4 - 0.005F_5 \\ & - 0.331F_6 \end{aligned} \tag{16}$$

$$\begin{aligned} \text{Logistic}(FAIL_{i,t-1}) = & 0.901 - 3.158F_1 - 0.099F_2 - 1.625F_3 + 1.699F_4 + 0.624F_5 \\ & + 0.033F_6 \end{aligned} \tag{17}$$

4.4. Model Accuracy

Table 5. Distress prediction accuracy.

| Period | Firms | Financial Health | Financial Distress | Prediction Accuracy |
|--------|--------------------|------------------|--------------------|---------------------|
| t-3 | Financial Health | 25 | 25 | 50% |
| | Financial Distress | 14 | 36 | 72% |
| | Total | | | 61% |
| t-2 | Financial Health | 43 | 7 | 86% |
| | Financial Distress | 16 | 34 | 68% |
| | Total | | | 77% |
| t-1 | Financial Health | 45 | 5 | 90% |
| | Financial Distress | 9 | 41 | 82% |
| | Total | | | 86% |

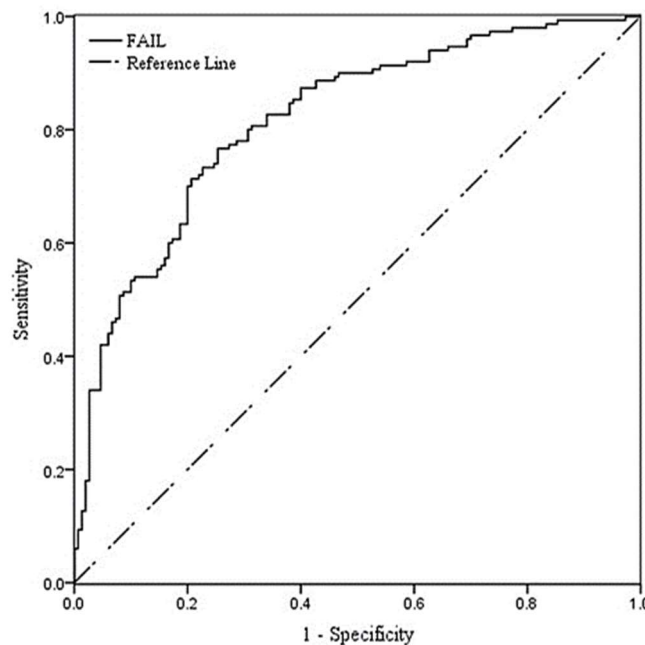


Figure 2. Financial Distress Prediction ROC curve.

The model is used to further calculate the accuracy of financial distress prediction for each period, and the results in Table 5 show that the prediction accuracy for firms in financial distress in period t-3 reached 72%, and the total prediction accuracy is 61%. The total prediction accuracy in period t-2 is 77%, and the total prediction accuracy in period t-1 is 86%. This result shows that the

closer the year of financial distress, the higher the accuracy of the prediction. In conclusion, the model has a high accuracy rate and a strong prediction function for the financial distress situation of digital economy firms. In addition, according to the ROC curve in Figure 2, the AUC (Area Under the ROC Curve) is 0.818, which imply that the model has predictive value for the financial distress of digital economy firms.

5. Conclusion and Discussion

In this paper, we screen indicators reflecting five dimensions of firms solvency, profitability, operating capacity and development capacity as well as cash flow status by K-S test, and has increased the proportion of R&D investment considering the innovation characteristics of digital economy firms, then we find that there are significant differences in indicators between financial distress firms and healthy firms in each dimension. Based on the analysis above, PCA-Logistic regression is applied to predict the financial distress situation of digital economy firms, and the indicators with differences are downscaled by PCA to obtain the principal component factors and then Logistic regression, finally the prediction model and prediction accuracy of digital economy firms three years before the occurrence of financial distress are obtained. The results show that the model achieve high accuracy and the closer to the year of financial distress the higher the prediction accuracy, indicating that the three distress prediction models are good enough to predict the financial distress of digital economy firms, and the model is constructed successfully and can provide a basis for decision making for management and information users such as investors and government.

It is not difficult to find that the significant contribution of profitability and innovation capacity of digital economy firms to the prediction model, which reflects that the characteristics of digital economy firms are significantly related to the prediction of financial distress and the deterioration of related indicators is the direct cause of the distress of firms. The financial data of digital economy firms listed in China A-shares have become easily accessible, and firms in the digital economy should make full use of these data to dig deeper into the risk patterns of various factors in the industry in order to obtain risk evolution patterns to prevent and guide their own financial risk management. In this context, the regulator is responsible for ensuring that firms operating have efficiency and sustainability incentives and can avert financial distress. At the same time, it is also able to promote the healthy development of the industry and improve risk identification and resilience.

Funding: This research was funded by the Chengdu Research Base of Philosophy and Social Sciences--Research Center Project of Chengdu Chongqing Twin City Economic Circle (No. CYSC21B002), the Youth Foundation Project of Social Science and Humanity, China Ministry of Education (No. 17YJC790083), and the Sichuan County Economic Development Research Center (No. xy2018016).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- [1] Barboza F, Kimura H, Altman E. Machine Learning Models and Bankruptcy Prediction[J]. *Expert Systems with Applications*, 2017,83:405-417. DOI: <https://doi.org/10.1016/j.eswa.2017.04.006>.
- [2] Uthayakumar J, Vengattaraman T, Dhavachelvan P. Swarm intelligence-based classification rule induction (CRI) framework for qualitative and quantitative approach: An application of bankruptcy prediction and DOI: <https://doi.org/10.54560/jracr.v12i1.319>

- credit risk analysis[J]. Journal of King Saud University - Computer and Information Sciences, 2020, 32(6):647-657. DOI: <https://doi.org/10.1016/j.jksuci.2017.10.007>.
- [3] Ashraf S, Félix G S, Elisabete, Serrasqueiro, Zélia. Do Traditional Financial Distress Prediction Models Predict the Early Warning Signs of Financial Distress?[J]. Journal of Risk and Financial Management, 2019, 12(2):138-143. DOI: <https://doi.org/10.3390/jrfm12020055>.
- [4] Fitzpatrick P J. A Comparison of the Ratios of Successful Industrial Enterprises with Those of Failed Companies[J]. Certified Public Accountant, 1932,2:589-605.
- [5] Foster, George. Financial Statement Analysis[M]. Prentice Hall, 1985,2:625.
- [6] Laitinen E K. Financial Ratios and Different Failure Processes[J]. Journal of Business Finance & Accounting, 1991,18(5):649-673. DOI: <https://doi.org/10.1111/j.1468-5957.1991.tb00231.x>.
- [7] Wu S, Huang S. Analytical indicators and prediction models of enterprise bankruptcy[J]. China Economic Issues, 1987(06):8-15. DOI: <https://doi.org/10.19365/j.issn1000-4181.1987.06.002>.
- [8] Wu X. Financial Crisis Early Warning Research: Problems and Framework Reconstruction[J]. Accounting Research, 2011(02):59-65+97. DOI: <https://doi.org/10.3969/j.issn.1003-2886.2011.02.009>.
- [9] Altman E I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy[J]. Journal of Finance, 1968,23(4):589-609. DOI: <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>.
- [10] Martin D. Early warning of bank failure: A logit regression approach[J]. Journal of Banking & Finance, 1977,1(3):249-276. DOI: [https://doi.org/10.1016/0378-4266\(77\)90022-X](https://doi.org/10.1016/0378-4266(77)90022-X).
- [11] Ohlson J. A. Financial Ratios and the Probabilistic Prediction of Bankruptcy[J]. Journal of Accounting Research, 1980,18(1):109130. DOI: <https://doi.org/10.2307/2490395>.
- [12] Jiang X, Sun Z. Weak governance and financial crisis: A predictive model[J]. Nankai Management Review, 2001(05):19-25. DOI: <https://doi.org/10.3969/j.issn.1008-3448.2001.05.005>.
- [13] Li H, Chen Y, Zhao G. Research on cash flow-based financial early warning: A comparison of linear probability model and logistic model applications[J]. Exploration of Economic Issues, 2011(06):102-105+111. DOI: <https://doi.org/10.3969/j.issn.1006-2912.2011.06.022>.
- [14] Li C. Construction of prediction model of enterprise financial risk based on logistic regression method[J]. Statistics and Decision, 2018,34(06):185-188. DOI: <https://doi.org/10.13546/j.cnki.tjyj.2018.06.045>.
- [15] Altman E I, Małgorzata Iwanicz-Drozdowska, Laitinen E K, et al. Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model[J]. Journal of International Financial Management & Accounting, 2017,28(2):131-171. DOI: <https://doi.org/10.1111/jifm.12053>.
- [16] Sun J, Li H, Huang Q H, et al. Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches[J]. Knowledge-Based Systems, 2014, 57(feb.):41-56. DOI: <https://doi.org/10.1016/j.knosys.2013.12.006>.
- [17] Zeng Y. Research on The Trend and Social Effects of Digital Economy[M]. China Social Science Press, 2021:11-12.
- [18] Wang X, Zhang L, He X. A comparative study on the effectiveness of financial distress prediction based on consolidated statements and parent company statements[J]. Accounting Research, 2017(06):38-44+96. DOI: <https://doi.org/10.3969/j.issn.1003-2886.2017.06.007>.
- [19] Alifiah M N, Norfian M. Prediction of Financial Distress Firms in the Trading and Services Sector in Malaysia Using Macroeconomic Variables[J]. Procedia - Social and Behavioral Sciences, 2014,129(6):90-98. DOI: <https://doi.org/10.1016/j.sbspro.2014.03.652>.
- [20] He P, Lan W, Ding Y. Is China's stock market predictable? -- A perspective based on a combined LASSO - logistic approach[J]. Statistical Research, 2021(05):82-96. DOI: <https://doi.org/10.19343/j.cnki.11-1302/c.2021.05.007>.



Copyright © 2022 by the authors. This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).