



Research on Financial Distress Diagnosis of Real Estate Listed Companies Based on PCA-Logistic Model

Yu Jiang¹, Kai Xu^{1,*}, Yan Hu², Dongyang Li¹, Dongxu Long³

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

²School of Economics and Management, Neijiang Normal University, Neijiang(641100), Sichuan, China

³School of Mechanical Engineering, Southwest Jiaotong University, Chengdu(610031), Sichuan, China

*Corresponding author. Email: xukai@cdu.edu.cn

Abstract

Financial distress diagnosis is of great significance to the risk prevention and control and sustainable operation of real estate industry. This paper takes China's a-share listed real estate companies as samples, selects financial indicators based on the characteristics of the real estate industry, and uses PCA-Logistic system to build a model of the second year before the enterprise falls into financial difficulties. The results show that the company's profitability and capital market performance factors contribute the most to the prediction, and the prediction accuracy is high. The successful construction of the model provides a basis for decision-makers to accurately and prospectively judge the financial distress of real estate companies.

Keywords: Financial Distress, Listed Real Estate Companies, Principal Component Analysis, Logistic Regression

基于 PCA-Logistic 模型的房地产上市公司财务困境诊断研究

江宇¹, 徐凯^{1,*}, 胡艳², 李东阳¹, 龙东旭³

¹成都大学商学院, 四川成都 (610106)

²内江师范学院经济与管理学院, 四川内江 (641100)

³西南交通大学机械工程学院, 四川成都 (610031)

*通讯作者. 电子邮箱: xukai@cdu.edu.cn

摘要

财务困境诊断对房地产行业风险防控与持续经营具有重要意义。文章以中国 A 股上市的房地产公司为样本, 结合房地产行业的特征选取财务指标, 运用 PCA-Logistic 系统地构建了企业陷入财务困境前第二年的模型, 结果表明公司的盈利能力和资本市场表现因子的预测作用贡献最大, 且预测准确率高。模型的成功构建为决策者准确、前瞻地判断房地产公司财务困境提供依据。

关键字: 财务困境, 房地产上市公司, 主成分分析, Logistic 回归

1. 引言

财务困境又称财务危机, 国内外学者对该概念的明确定义存在分歧, 但大多数学者都认为, 财务困境是指现金流量未能补偿现有债务的现象。该现象常导致公司产生破产和债务危机以及信用风险 (Barboza et al, 2017; Uthayakum et al, 2020)^[1,2], 公司的经营规模、流动资产和资产负债指标出现异常变化, 且这些变化常发生在违约之前。在过去的十年里, 房地产行业处于高速发展时期, 每年投资增幅稳定在 20% 左右, 房地产泡沫也在不断膨胀。房地产行业是典型的资本密集型和劳动密集型产业, 融资是房地产行业的关键环节, 资金需求大且营业周期长, 使得房地产公司的财务状况存在不稳定性 and 高风险性。近年来受新冠疫情的影响, 大部分地区的楼市成交量呈下降趋势, 资金的流动性不断收紧, 房地产公司的资金链紧张, 导致财务风险随之增大, 累积到一定程度便会造成财务困境。故本文对我国的房地产企业进行财务困境诊断, 进而进行预警具有很强的现实意义, 同时减少投资者的损失。

Logistic 回归模型属于判别式模型, 实现简单, 计算量小, 回归时通过给定的概率值进行预测, 且 Logistic 回归模型对于变量的分布无具体要求, 能很好地处理两分类问题。宋晓娜、黄业德 (2016)^[3] 分别运用主成分分析法和 Logistic 模型相结合的方法进行财务困境预警研究, 结果发现 Logistic 模型的预测效果更好。故本文采用主成分分析法 (Principal components analysis, PCA), 与 Logistic 相结合的方式, 以 2015-2021 年中国 A 股上市的房地产为样本, 构建财务困境诊断模型。文章的结构安排如下, 第二部分是构建财务困境诊断模型的研究设计; 第三部分是模型的实证分析; 第四部分是结论与建议。

2. 研究设计

2.1. 样本选取

在建立数理统计模型之前, 我们需要在数学上严格区分出陷入财务困境的企业和健康企业。根据 Altman 等人在财务困境研究过程中的判定标准, 本文将房地产上市企业中因为“财务状况异常”, 而被 ST 或者 *ST 的房地产企业选为财务困境企业样本, 未被 ST 或 *ST 的房地产企业选为健康企业样本。

根据国家统计局 2019 年修订的《国民经济行业分类》, 确定了房地产企业分类, 本文选取 2015-2021 年中国的 A 股房地产上市公司为样本, 变量 t 代表财务困境公司陷入财务困境所在年份, 相应的前第二年为 t-2。相关研究结果表明 t-2 年预测准确率最高 (宋鹏等, 2017)^[4], 故本文按照 1: 5 的比例选择财务困境企业和财务健康企业第 t-2 年的数据, 作为预测数据。共 16 家财务困境企业, 80 家健康企业作为对照, 且选取的 96 家企业样本都是独立的。数据来源于国泰安数据库, 该数据库中的指标计算根据上市公司公布的年报。

2.2. 变量定义

对于企业而言, 财务指标是用于反映公司财务状况和经营过程的统计数据。通过分析财务指标能得到公司财务与运营状况间的关系。根据我国房地产行业的特点和相关研究成果, 在传统的四大能力指标之外引入两类财务指标, 分别是资本市场表现和现金流^[5,6]。最后得到企业的偿债能力、经营能力、盈利能力、现金流、发展能力以及资本市场表现等六个维度, 共 21 个变量指标, 各变量的定义说明如表 1。

表 1 财务指标

变量类型	财务指标	符号
	企业财务状况	Y

偿债能力	流动比率	X1
	现金比率	X2
	速动比率	X3
	利息保障倍数	X4
	长期资本负债率	X5
	资产负债率	X6
经营能力	应收账款周转率	X7
	存货周转率	X8
	流动资产周转率	X9
	总资产周转率	X10
	总资产净利润率	X11
盈利能力	净资产收益率	X12
	营业净利率	X13
现金流	净利润现金净含量	X14
	营业收入现金净含量	X15
	固定资产增长率	X16
发展能力	总资产增长率	X17
	可持续增长率	X18
	每股收益	X19
资本市场表现	市盈率	X20
	市净率	X21

2.3. 模型设计

1845年,比利时科学家 Pierre François Verhulst 首次提出了 Logistic 函数^[7,8],并介绍了 Logistic 函数的一些性质。同时 Logistic 函数也被称为 sigmoid 函数,其表达式如下:

$$\gamma(s) = \frac{1}{1+e^{-s}} \quad (1)$$

二元 Logistic 回归作为一种广义的线性回归,常用于处理二分类问题,且相关研究表明该方法具有较高预测准确率(李红琨等,2011; Alifiah et al, 2014)

^[9,10],故本文将采用 Logistic 回归分析,判别房地产企业的财务困境。但是诊断财务困境过程中选取的变量是多维的,且数量众多,易出现多重共线性问题。严重影响二元 Logistic 回归模型的预测结果。手动或通过逐步回归移除共线性变量,容易剔除影响显著的变量,故采用主成分分析(Principal components analysis, PCA)解决多重共线性问题^[11]。该方法将多个指标转换为少数几个主成分,这些主成分是原始变量的线性组合,且彼此之间互不相关,其能反映出原始数据的大部分信息。且在消除共线性的影响后降低数据维度,避免了因自变量过多导致模型产生过度拟合。本文先判别财务指标,在 ST 和非 ST 样本间的差异显著性,剔除影响不显著的指标后运用 PCA 方法对财务指标降维,得到的主成分进行 Logistic 回归。结合 PCA 方法和 Logistic 回归建立财务困境诊断模型,判别模型对房地产上市企业财务困境的诊断准确率。

采用的二元 Logistic 回归方法假设自变量 X_i 和因变量 Y_i ,分别代表第 i 个房地产公司经济活动的财务指标和财务状况, $i=1, 2, \dots, n$, n 为96。公司被执行 ST 的概率为 P ,并用 $Y=1$ 表示,其表达式为:

$$P(Y = 1 | X_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_{j=1}^J \beta_j X_i \quad (2)$$

参数 β_0, \dots, β_j 的取值是通过极大似然法求得(见公式3),最终得到第 i 家公司 $Y=1$ 概率估计值 \hat{p}_i 表达式(4):

$$L(\beta_0, \dots, \beta_d) = \prod_{i=1}^n [Y_i \gamma(\beta_0 + \beta^T X_i) + (1 - Y_i) \log(1 - \gamma(\beta_0 + \beta^T X_i))] \quad (3)$$

$$\hat{p}_i(Y_i = 1 | X_i) = \frac{\exp(\beta_0 + \sum_{i=1}^n \beta_i X_i)}{1 + \exp(\beta_0 + \sum_{i=1}^n \beta_i X_i)} \quad (4)$$

3. 实证分析

3.1. 财务指标显著性检验

本文先运用柯尔莫哥洛夫-斯米尔诺夫检验(Kolmogorov-Smirnov, KS),判断样本变量是否服从正态分布。检验结果表明 X5、X6、X17 符合正态分布,故采用独立样本的参数检验(T 检验),分析 ST 样本和非 ST 样本间变量差异显著性,剩余指标采用两独立样本的非参数检验(Mann-Whitney U)。两种检验结果表明共 11 个存在显著差异的变量,分别为 X4、X6、X7、X11、X12、X13、X15、X18、X19、X20、

X₂₁。下文将采用主成分分析法对存在显著差异的指标降维，并建立二元 Logistic 模型。

3.2. 主成分分析适用性检验

为检验存在显著差异的指标是否适合主成分分析，首先运用 Bartlett 球形检验与 KMO 统计量检验样本指标（朱永忠，2012；孔宁宁，2010）^[12,13]。KMO 统计量主要用于检查变量间的相关性和偏相关性，KMO 值越接近 1 则变量间的相关程度差异越小，原有变量越适合作主成分分析，且一般认为 KMO 值>0.5 适合做主成分分析。表 2 中 KMO 值为 0.626，表明适合主成分分析。Bartlett 球状检验主要用于检验数据的分布，以及各个变量间的独立情况。若 p 值小于 0.05 则球形假设被拒绝，变量之间存在相关性，适合做主成分分析。表 2 中 Bartlett 球形检验的显著性概率为 0，表明该样本数据适合进行主成分分析。

表 2 KMO 统计量和 Bartlett 球状检验

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	Bartlett test of sphericity		
	Chi-square	Degrees of freedom	p-value
0.626	672.154	55	0.000

3.3. 主成分的提取与解释

运用 Matlab 对上述 11 个显著性指标进行主成分分析，得到如表 3 所示的总方差解释，选取前四个主成分的累计贡献率为 76.476%，在总体上能涵盖总体变量信息。根据表 4 中的成分矩阵可判别各主成分的经济含义，主成分 F₁ 代表盈利能力 (X₁₁、X₁₂、X₁₃)；发展能力 (X₁₈)；资本市场表现 (X₁₉)。主成分 F₂ 中 X₂₀、X₂₁ 负荷量比较明显，代表资本市场表现。主成分 F₃ 代表盈利能力 (X₁₃) 和现金流 (X₁₅)。主成分 F₄ 中 X₇ 负荷量明显且集中，代表经营能力。此外，计算得到成分得分系数矩阵（如表 5 所示），成分得分系数矩阵中的值作为 4 个主成分因子的系数，表达式如下：

$$F_1 = 0.124X_4 + 0.02X_6 - 0.011X_7 + 0.222X_{11} + 0.239X_{12} + 0.139X_{13} + 0.111X_{15} + 0.236X_{18} + 0.189X_{19} - 0.121X_{20} - 0.043X_{21} \quad (5)$$

$$F_2 = 0.103X_4 - 0.301X_6 - 0.023X_7 + 0.171X_{11} +$$

$$0.016X_{12} + 0.165X_{13} + 0.16X_{15} + 0.015X_{18} - 0.162X_{19} + 0.373X_{20} + 0.393X_{21} \quad (6)$$

$$F_3 = -0.153X_4 + 0.224X_6 + 0.105X_7 - 0.078X_{11} - 0.14X_{12} + 0.447X_{13} + 0.448X_{15} - 0.15X_{18} - 0.105X_{19} + 0.023X_{20} - 0.188X_{21} \quad (7)$$

$$F_4 = 0.189X_4 - 0.042X_6 + 0.977X_7 + 0.011X_{11} + 0.019X_{12} - 0.089X_{13} - 0.047X_{15} + 0.021X_{18} - 0.017X_{19} + 0.033X_{20} - 0.012X_{21} \quad (8)$$

表 3 主成分特征值和贡献率

成分	初始特征值			提取载荷平方和		
	总计	方差百分	累积 %	总计	方差百分	累积 %
1	3.815	34.681	34.681	3.815	34.681	
2	1.983	18.031	52.712	1.983	18.031	52.712
3	1.618	14.709	67.421	1.618	14.709	67.421
4	0.996	9.056	76.476	0.996	9.056	76.476
5	0.824	7.494	83.97			
6	0.769	6.992	90.962			
7	0.419	3.806	94.768			
8	0.252	2.293	97.061			
9	0.149	1.358	98.419			
10	0.139	1.262	99.68			
11	0.035	0.32	100			

表 4 成分矩阵

成分	第一主成分 F ₁	第二主成分 F ₂	第三主成分 F ₃	第四主成分 F ₄
X ₄	0.475	0.205	-0.247	0.189
X ₆	0.078	-0.597	0.363	-0.041
X ₇	-0.04	-0.046	0.17	0.973
X ₁₁	0.845	0.339	-0.127	0.01
X ₁₂	0.911	0.032	-0.227	0.019
X ₁₃	0.531	0.327	0.723	-0.088
X ₁₅	0.423	0.318	0.79	-0.047
X ₁₈	0.902	0.031	-0.243	0.021
X ₁₉	0.722	-0.321	-0.17	-0.017

X ₂₀	-0.463	0.74	0.037	0.033
X ₂₁	-0.166	0.779	-0.304	-0.012

表 5 成分得分系数矩阵

成分	第一主成分F ₁	第二主成分F ₂	第三主成分F ₃	第四主成分F ₄
X ₄	0.124	0.103	-0.153	0.189
X ₆	0.02	-0.301	0.224	-0.042
X ₇	-0.011	-0.023	0.105	0.977
X ₁₁	0.222	0.171	-0.078	0.011
X ₁₂	0.239	0.016	-0.14	0.019
X ₁₃	0.139	0.165	0.447	-0.089
X ₁₅	0.111	0.16	0.488	-0.047
X ₁₈	0.236	0.015	-0.15	0.021
X ₁₉	0.189	-0.162	-0.105	-0.017
X ₂₀	-0.121	0.373	0.023	0.033
X ₂₁	-0.043	0.393	-0.188	-0.012

3.4. Logistic 回归与模型合理性评估

使用 SPSS 软件，对执行 ST 房地产公司发生财务困境前第二年的原始数据进行二元 Logistic 回归分析，得到如表 6 中所示 Logistic 方程的变量参数。

表 6 Logistic 回归方程中的变量参数

主成分	F ₁	F ₂	F ₃	F ₄	常数
因子系数β _i	-2.1	0.472	-1.35	-0.027	-2.243

据此可以得出 t-2 期房地产上市公司的 Logistic 模型表达式：

$$P(Y = 1) = \frac{e^{-2.243 - 2.1F_1 + 0.472F_2 - 1.35F_3 - 0.027F_4}}{1 + e^{-2.243 - 2.1F_1 + 0.472F_2 - 1.35F_3 - 0.027F_4}} \quad (9)$$

财务困境诊断模型中通常将 P = 0.5，作为企业是否陷入财务困境的判定点，但是本文受限于 ST 样本数量，采用了 5:1 的非 ST 和 ST 样本配比，并且发现当判定点为 0.5 时，ST 企业的预测准确率只有 37.5%。由此我们通过枚举法^[14]，列举了判定点以 0.01

的增长率从 0.1 变化到 0.5 过程中，ST 和非 ST 公司的预测准确率以及整体预测准确率（图 1）。如图 1 所示，非 ST 公司和全体公司的预测准确率呈上升趋势，而 ST 公司呈下降趋势。当判定点为 0.36 时，总体预测准确率最大且为 89.6%，但此时 ST 公司的预测准确率仅为 62.5%。当判定点为 0.2 时，总体预测准确率不是最大，为 84.4%。但 ST 公司和非 ST 公司的预测准确率分别为 81.25%和 85%，二者差异不大且保证样本数最多的非 ST 公司预测准确率较大。本文建立财务困境诊断模型是为了进行财务困境预警，故需要保证 ST 公司预测准确率较高，同时需要保证整体预测率较高。综上所述，选择判定点为 0.2，当 P<0.2 时公司财务状况健康，当 P>0.2 时公司容易发生财务危机，具体参数如表 7 所示，此时 Logistic 模型对房地产上市公司的企业财务困境具有预测价值。

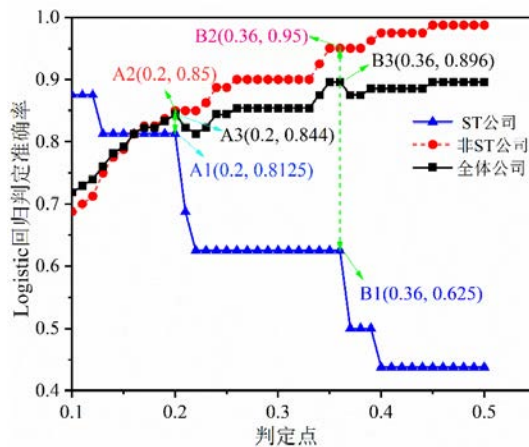


图1 预测率随判定点变化示意图

表 7 判定点 0.2 检测结果

观测公司类型	预测公司类型	预测公司数量	预测准确率	总体预测准确率
0	0	68	85.0%	84.4%
	1	12		
1	0	3	81.3%	
	1	13		

4. 结论与建议

本文通过 K-S 检验筛选，反映企业偿债能力、盈利能力、营运能力、发展能力、现金流状况以及资本

市场表现六个维度的指标,并且考虑房地产企业的关键环节是融资,增加每股收益、市盈率和市净率指标,发现陷入财务困境企业与健康企业,各维度均存在指标上的显著差异。据此,基于主成分分析结合二元 Logistic 回归分析,建立房地产上市公司的财务困境诊断模型。首先通过主成分分析对存在显著差异的指标进行降维,得到主成分因子后进行二元 Logistic 回归,最终得出房地产上市企业发生财务困境前第二年的诊断模型与预测准确率。通过枚举法选择判定点为 0.2。结果表明,判定点为 0.2 时模型具有较高的预测准确率,模型构建较为成功,企业高层、投资者以及政府官员可以通过该模型预测房地产企业未来的财务状况,防范投资风险。

上述研究过程中发现,房地产企业的盈利能力与资本市场表现指标对诊断模型有显著贡献,表明房地产企业的特征与财务困境预测显著相关。房地产企业应分析相关财务数据,并挖掘出该产业中影响显著的财务指标,了解房地产风险演变规律,从而准确判断公司的财务状况,预防与指导自身的财务风险管理工作。政府各相关部门需采取针对性措施,逐步完善房地产风险管控机制,建立完备的风险管控网络和控制中心,细化该行业的管理制度,并配备相应的专业风险监控人员以防范房地产行业风险。

致谢

本研究得到成都市哲学社会科学研究基地-成渝双城经济圈研究中心项目(CYSC21B002)、沱江流域高质量发展研究中心项目(TJGZL2022-09)的资助。

参考文献

- [1] Barboza F, Kimura H, Altman E. Machine learning models and bankruptcy prediction[J]. Expert Systems with Applications, 2017, 83: 405-417.
- [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 credit risk analysis[J]. Journal of King Saud University-Computer and Information Sciences, 2020, 32 (6): 647-657.
- [3] 宋晓娜, 黄业德, 张峰. 基于 Logistic 和主成分分析的制造业上市公司财务危机预警[J]. 财会月刊

(下), 2016(03): 67-71.

- [4] 宋鹏, 李婷婷. 中小上市企业财务危机预警研究——基于 RS—Logistic 回归的实证检验[J]. 会计之友, 2017 (9): 6.
- [5] 柴镜丹. 房地产市场与资本市场关系研究[J]. 经济师, 2015 (5): 63-64.
- [6] 刘文静, 孙利琼. 房地产企业现金流的现状分析[J]. 科技与管理, 2005, 7 (6): 38-39.
- [7] Verhulst P F. Resherches mathematiques sur la loi d'accroissement de la population[J]. Nouveaux memoires de l'academie royale des sciences, 1845, 18: 1-41.
- [8] Verhulst P F, Verhulst P F. Deuxième memoire sur la loi d'accroissement de la population[J], 1846.
- [9] 李红琨, 陈永飞, 赵根. 基于现金流的财务预警研究: 线性概率模型与 Logistic 模型之应用比较[J]. 经济问题探索, 2011 (6): 102-105.
- [10] Alifiah M N. Prediction of financial distress companies in the trading and services sector in Malaysia using macroeconomic variables[J]. Procedia-Social and Behavioral Sciences, 2014, 129: 90-98.
- [11] Wold S, Ruhe A, Wold H, et al. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses[J]. SIAM Journal on Scientific and Statistical Computing, 1984, 5 (3): 735-743.
- [12] 朱永忠, 姚焯, 张艳. 基于主成分分析和 Logistic 回归的上市公司财务困境预警模型的研究[J]. 浙江工业大学学报, 2012, 40 (6): 692-694.
- [13] 孔宁宁, 魏韶巍. 基于主成分分析和 Logistic 回归方法的财务预警模型比较——来自我国制造业上市公司的经验证据[J]. 经济问题, 2010 (6): 5.
- [14] 吴世农, 卢贤义. 我国上市公司财务困境的预测模型研究[J]. 经济研究, 2001, 6 (2008): 4.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

