

基于Logistic回归模型的房地产上市公司信用风险研究

杨璐菱

重庆电力高等专科学校人文素质学院, 重庆

收稿日期: 2023年10月30日; 录用日期: 2023年11月23日; 发布日期: 2024年1月8日

摘要

伴随着中国城镇化的发展, 在过去二十年里, 我国的房地产行业迅猛的发展。由于房地产行业的产业关联性极强, 已然成为我国国民经济的重要“支柱产业”之一。在此背景下, 研究房地产的企业信用风险, 对构建现代房地产企业管理机制, 以及商业银行的风险管理有着重要的现实意义。本文搜集了2020年所有房地产上市公司的财务报表中的14个有关风险管理的财务指标, 构建了房地产企业的财务信用评价指标体系。再通过因子分析法生成了对所有指标具有代表意义的4个公因子, 并挑选出具有实际意义的两个公因子: 即企业的“盈利能力” M_1 和企业的“偿债能力” M_3 。之后, 选取 M_1 和 M_3 作为自变量, 选取 Y “企业是否为ST企业”作为因变量, 建立Logistic回归模型。结果显示建立的Logistic回归模型能实际计算一个房地产企业信用风险, 且一个房地产企业的盈利能力越好, 其面临的信用风险就越低。最后根据输出结果提出了一些可行的实际建议, 例如要关注与盈利能力相关的指标的变动, 要制定相应的防范措施以避免企业出现信用风险等。

关键词

信用风险, 因子分析法, Logistic回归模型

Research on Credit Risk of Real Estate Listed Companies Based on Logistic Regression Model

Luling Yang

College of Human Quality, Chongqing Electric Power College, Chongqing

Received: Oct. 30th, 2023; accepted: Nov. 23rd, 2023; published: Jan. 8th, 2024

文章引用: 杨璐菱. 基于 Logistic 回归模型的房地产上市公司信用风险研究[J]. 金融, 2024, 14(1): 46-54.

DOI: 10.12677/fin.2024.141007

Abstract

Abstract: With the development of urbanization in China, in the past 20 years, China's real estate industry has developed rapidly. Due to the strong industrial correlation of the real estate industry, it has become one of the important "pillar industries" of our national economy. In this context, the study of real estate enterprise credit risk is of great practical significance to the construction of modern real estate enterprise management mechanism and the risk management of commercial banks. This paper collects 14 financial indicators related to risk management in the financial statements of all listed real estate companies in 2020, and constructs the financial credit evaluation index system of real estate enterprises. Then, four common factors representing all indicators are generated through factor analysis, and two common factors with practical significance are selected, namely, the "profitability" M_1 and the "solvency" M_3 . After that, the Logistic regression model was established by selecting M_1 and M_3 as independent variables and Y "whether the enterprise is ST enterprise" as dependent variable. The results show that the established Logistic regression model can actually calculate the credit risk of a real estate enterprise, and the better the profitability of a real estate enterprise, the lower the credit risk it faces. Finally, according to the output results, some practical suggestions are put forward, such as paying attention to the changes of the indicators related to profitability, and formulating corresponding preventive measures to avoid the credit risk of enterprises.

Keywords

Credit Risk, Factor Analysis, Logistic Regression Model

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

二十大报告指出, 必须坚持创新是第一动力, 加快实施创新驱动发展战略。我国经济的高质量发展离不开实体经济和虚拟经济失衡的困境[1]。随着中国城镇化的发展, 在过去二十年里, 我国的房地产行业迅猛的发展。由于房地产行业的产业关联性极强, 已然成为我国国民经济的重要“支柱产业”之一。在宽松的货币政策以及“去库存”的战略条件下, 埋下了房地产泡沫的隐患。所以, 有效的预测房地产企业的信用风险, 对抑制房地产泡沫的产生有着至关重要的作用。

众所周知, 房地产行业属于资本密集型产业, 在众多房地产企业资金回收速度放慢, 负债率高的情况下, 许多企业的资金主要来源于银行贷款。在过去的几年, 房地产的非理性繁荣, 积聚了众多风险。对于房地产企业的信用风险研究有助于商业银行在发放贷款时进行有效的风险防范。同时, 对于房地产行业的信用风险的研究, 也是能保证企业利润的一个重大措施。

信用风险又称违约风险, 是评价一笔债务由于某些因素遭受损失的可能性[2], 是信用关系中债务人用以偿还债务的资产或收入来源存在巨大的不确定性。这种不确定性可能导致债务人无力偿还债务, 从而引发违约事件[3]。

对于一般企业的信用风险研究分为三个发展阶段: 起初是传统的信用风险研究, 最具代表性的是 Altman 在 1968 年建立了传统统计模型, 通过构建判别分析对信用风险进行了研究; 而我国对于企业的

信用风险的研究则更多的依托于资产评估公司；这些方式虽从一定程度上提高了企业的风险防范水平，但实施过程艰难，评估结果精度较差，不利于推广使用。

随着互联网的发展，不同的学者通过建立不同的模型，能够更精确的预测企业的信用风险。王慧选取了 2009 年至 2016 年共计 147 家房地产上市企业的数据，通过穷举法，选用 KVM 模型预测了我国房地产上市企业的违约风险[4]；李杰辉选取了房地产上市公司的 14 个相关财务指标，采用 Logistic 模型对我国房地产企业的信用风险进行了预测[5]。Daniel C 基于问题的期望最大利润和计算利润，选取了 Logistic Regression 和 XG Boost 模型，优化了竞争性信用风险评估模型[6]。

为进一步提高信用风险的预测效果，更多的学者开始利用算法建立模型，何睿利用卷积神经网络，在构建银行间债券市场信用债知识图谱的基础上，构建了异构信息处理(R-GCN-BKG)的违约债券预测算法，其准确率可达 92% [7]。

2. 研究方法和数据来源

2.1. 研究方法

2.1.1. 因子分析法

因子分析是研究相关阵或者协方差阵的内部依赖关系，它将多个变量综合为少数几个因子，以再现原始变量与因子之间的相关关系。

2.1.2. Logistic 回归模型

Logistic 回归模型因其简单的形式和优越的性能成为了信用风险研究中较为主流的方法。Logistic 回归模型对数据的要求不严，可以不需要满足正态分布，也不需要保持自变量与因变量的线性关系。

2.2. 数据来源

本文数据来源于国泰安数据库，经过查阅资料，选择了部分能描述房地产公司信用风险的指标，最终选取了 2020 年我国房地产上市公司的 14 个财务指标。

3. 建模与分析

本文通过因子分析从 14 个财务指标中概括出能够解释原始变量的 4 个公共因子，然后根据实际意义，选择其中的 2 个公共因子作为建模的自变量，最后根据企业是否为 ST 企业(证券交易中实行“警示存在终止上市风险的特别处理”的企业，以提示该公司的股票可能有被终止上市的风险)作为因变量，建立能够描述房地产企业的风险的 Logistic 模型。

3.1. 指标筛选

通过因子分析筛选出能解释大部分原始变量的公因子，依据这几个对原始数据代表性较高的公因子作为自变量，结合因变量建立 Logistic 回归模型。

3.1.1. 碎石图

碎石图可以表示各个因子的重要程度，它将所提取的因子特征按照从大到小的顺序排列，从图中可以很直观地了解到解释变量的主要因子。

如图 1 所示，其中第一个和第二个因子的特征值较高，对解释原有变量的贡献最大；第三个和第四个因子的特征值较为明显，但第四个因子之后的其余因子则趋于平坦状态，所以提取前四个公因子作为解释原始变量的主要因子。

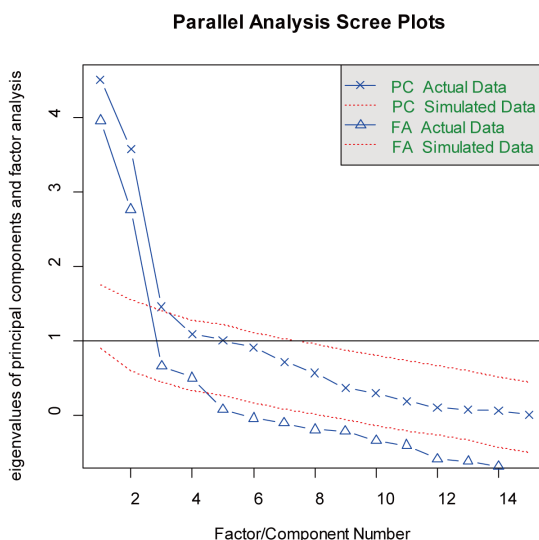


Figure 1. Scree plot
图 1. 碎石图

3.1.2. 累计贡献率

累计贡献率是因子特征值之和和所有因子特征值之和的比值。代表了抽取出的因子变异对所有变量变异的解释力，也就是抽取出的因子对所有变量的代表性。累计贡献率越高，说明这个因子的能够代表原始数据的信息越多。

Table 1. Cumulative contribution statement

表 1. 累计贡献率表

	ML1	ML2	ML3	ML4
SS loadings	3.45	2.5	2.41	0.91
Proportion Var	0.25	0.18	0.17	0.06
Cumulative Var	0.25	0.42	0.6	0.66
Proportion Explained	0.37	0.27	0.26	0.1
Cumulative Proportion	0.37	0.64	0.9	1

通过表 1 可以看到这四个因子对于变量的解释程度，第一个因子对变量的解释达到了 37%，第二个因子对变量的解释达到了 27%，第三个因子对变量的解释达到了 26%，第四个因子对变量的解释仅仅只有 10%。说明第一个因子和第三个因子对于原始数据的代表性较高。

3.1.3. 因子载荷矩阵

因子载荷矩阵是将特征向量归一化后得到的矩阵，表示了各因子对各变量的影响程度。

Table 2. Factor load matrix

表 2. 因子载荷矩阵

	ML1	ML2	ML3	ML4	h2	u2	com
资产负债率	-0.08	0.54	0.66	0.33	0.831	0.1689	2.5
净资产收益率	0.79	-0.14	0.07	-0.24	0.711	0.2887	1.3

续表

总资产净利润率	0.88	-0.47	0.05	0	0.995	0.0049	1.5
总资产回报率	0.84	-0.5	0.06	0.04	0.964	0.0364	1.6
每股收益	0.8	0.6	-0.04	0	0.995	0.005	1.9
每股净资产	0.5	0.67	-0.03	0.01	0.697	0.3034	1.8
流动比率	0.13	-0.42	-0.76	0.29	0.856	0.1445	2
产权比率	-0.11	0.34	0.53	0.66	0.843	0.1575	2.5
速动比率	0.19	-0.45	-0.76	0.33	0.912	0.0883	2.2
固定资产比率	-0.22	0.02	-0.17	-0.09	0.086	0.9138	2.3
流动负债率	0.03	0.47	0.69	0.27	0.768	0.2323	2.1
综合杠杆	0.04	0.02	0.13	0.1	0.029	0.9712	2.2
营业毛利率	0.35	-0.28	-0.08	0.05	0.211	0.7891	2.1
销售净利率	0.47	-0.34	0.08	0.16	0.366	0.6336	2.2

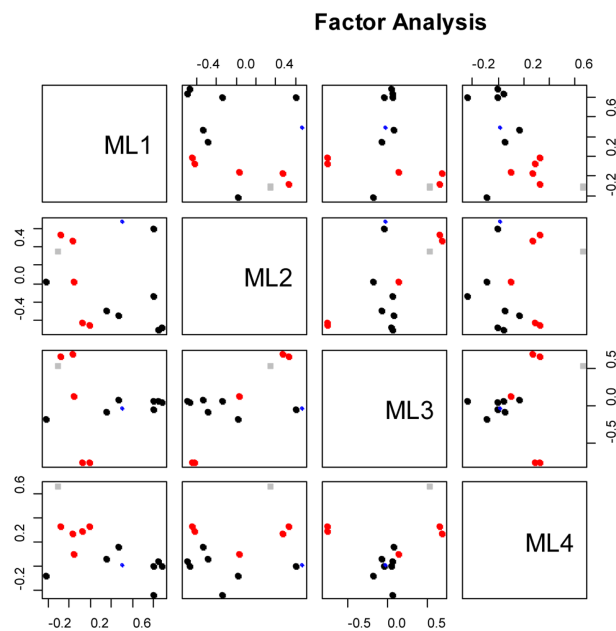


Figure 2. Point diagram of factor load matrix
图 2. 因子载荷矩阵点图

表 2 表示了因子各因子在原始变量上的载荷，图 2 可以更加清晰的看到因子载荷矩阵的分布。综上所述可以看到变量“固定资产比率”和“综合杠杆”对于任何一个变量的影响程度都很弱，说明这两个变量不具有能解释原始变量的作用，因此将其剔除。

3.1.4. 因子分析

根据因子载荷矩阵可以进一步提取出 4 个公共因子，以及归纳出每个因子所包含的变量，由此进行因子分析。

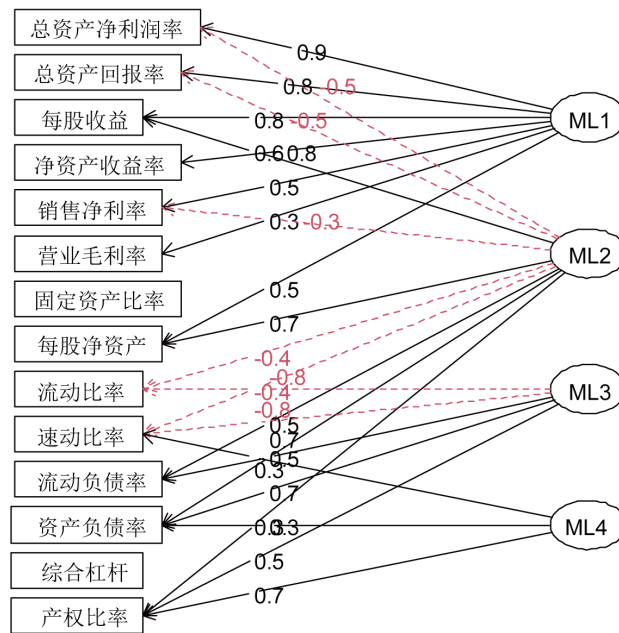


Figure 3. Factor analysis diagram
图 3. 因子分析图

Table 3. Factor interpretation table
表 3. 因子解释表

M_1 盈利能力	M_2	M_3 偿债能力	M_4
净资产收益率	每股净资产	资产负债率	产权比率
总资产回报率		流动比率	
总资产净利润率		流动负债率	
每股收益		速动比率	
营业毛利率			
销售净利率			

图 3 清晰的显示了每个因子所包含的变量，这有助于划分因子包含的不同变量。结合图 3，可以总结如表 3 的因子解释表。根据表 3 可以看到四个因子的含义以及每个因子所包含的变量，根据实际意义，将第一个因子 M_1 命名为“盈利能力”，将第二个因子 M_2 命名为“每股净资产”，将第三个因子 M_3 命名为“偿债能力”，将第四个因子 M_4 命名为“产权比率”。根据每个因子所包含的解释变量， M_2 和 M_4 这两个因子只包含了一个解释变量，不具有实际意义，所以在后续建模的过程中将其剔除。

3.1.5. 模型检验

通过卡方检验可以判断选取的房地产上市公司的财务指标是否适合建立因子分析模型，最终通过卡方检验的 P 值大小判断。

Table 4. Test table of factor analysis model
表 4. 因子分析模型检验表

因子分析模型检验输出结果
Mean item complexity = 2

续表

Test of the hypothesis that 4 factors are sufficient.
 The degrees of freedom for the null model are 105
 and the objective function was 13.9 with Chi Square of 1656.85
 The degrees of freedom for the model are 51
 and the objective function was 1.98
 The root mean square of the residuals (RMSR) is 0.03
 The df corrected root mean square of the residuals is 0.05
 The harmonic number of observations is 126 with the empirical chi square 31.06 with prob < 0.99
 The total number of observations was 126 with Likelihood Chi Square = 231.12 with prob < 9.2e-25
 Tucker Lewis Index of factoring reliability = 0.755
 RMSEA index = 0.167 and the 90 % confidence intervals are 0.146 0.19
 BIC = -15.53
 Fit based upon off diagonal values = 0.99

从表 4 可以看出模型的卡方检验结果为 9.2×10^{25} ，该值小于 0.05，说明数据通过因子分析法进行建模是可行的。

3.1.6. 因子得分系数

因子得分系数则是用来计算每个观测值在公因子上的得分。根据各个因子的得分可以写出各个因子关于各个指标的表达式。

Table 5. Factor score coefficient table
表 5. 因子得分系数表

	ML1	ML2	ML3	ML4
资产负债率	-0.0016	0.0240	0.2060	0.2838
净资产收益率	0.0090	-0.0036	0.0120	-0.1232
总资产净利润率	0.5748	-0.7083	0.5031	-0.0817
总资产回报率	0.0755	-0.1037	0.0845	0.1660
每股收益	0.5205	0.9050	-0.4694	-0.0623
每股净资产	0.0054	0.0167	-0.0047	0.0057
流动比率	0.0029	-0.0220	-0.2792	0.2950
产权比率	-0.0023	-0.0165	0.1778	0.6154
速动比率	0.0069	-0.0381	-0.4538	0.5443
流动负债率	0.0005	0.0152	0.1572	0.1710
营业毛利率	0.0014	-0.0027	-0.0051	0.0100
销售净利率	0.0024	-0.0040	0.0063	0.0371

表 5 为因子得分系数表，表中显示了不同因子对原始变量的解释程度，在剔除了前面的两个无关变量“固定资产比率”和“综合杠杆”以后，得到了 M_1 、 M_2 、 M_3 、 M_4 这四个公因子的因子得分系数。根据房地产上市公司各个财务指标的实际意义，在下一章节中将选择公因子 M_1 “盈利能力”和公因子 M_3 “偿债能力”进行 Logistic 建模，以下为 M_1 和 M_3 公因子关于房地产上市公司各个财务指标的表达式：

$$M_1 = -0.0016x_1 + 0.009x_2 + 0.5748x_3 + 0.0755x_4 + 0.5205x_5 + 0.0054x_6 \\ + 0.0029x_7 - 0.0023x_8 + 0.0069x_9 + 0.0005x_{10} + 0.0014x_{11} + 0.0024x_{12}$$

$$M_3 = 0.203x_1 + 0.012x_2 + 0.5031x_3 + 0.0845x_4 - 0.4694x_5 - 0.0047x_6 - 0.2792x_7 \\ + 0.1778x_8 - 0.4538x_9 + 0.1572x_{10} - 0.0051x_{11} + 0.0063x_{12}$$

在表达式中, x_1 为“资产负债率”, x_2 为“净资产收益率”, x_3 为“总资产收益率”, x_4 为“总资产回报率”, x_5 为“每股收益”, x_6 为“每股净资产”, x_7 为“流动比率”, x_8 为“产权比率”, x_9 为“速动比率”, x_{10} 为“流动负债率”, x_{11} 为“营业毛利率”, x_{12} 为“销售净利率”。

3.2. 建立 Logistic 回归模型

根据上述因子分析的结果,从房地产上市公司中的 14 个原始财务指标进行提取了其中 2 个具有代表性的公共因子,分别命是 M_1 “盈利能力”和 M_3 “偿债能力”。在建立的 Logistic 回归模型中, Y 表示了企业是否为 ST 企业(证券交易中实行“警示存在终止上市风险的特别处理”的企业,以提示该公司的股票可能有被终止上市的风险)。若该上市企业为 ST 企业,则反应变量取值为 1,说明该企业的信用风险较大;若该上市企业不为 ST 企业,则反应变量取值为 0,说明该企业的信用风险较小。选取 M_1 和 M_3 作为自变量, Y 作为因变量,建立 Logistic 回归模型。

3.2.1. 变量筛选

Table 6. Summary of Logistic regression results

表 6. Logistic 回归结果汇总表

	Coefficients				
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.7861	0.4838	-5.759	8.48E-09	***
M_1	-2.733	1.1053	-2.473	0.0134	*
M_3	-0.2166	0.2705	-0.801	0.4234	

注: *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, 下同。

表 6 是根据 M_1 “盈利能力”和 M_3 “偿债能力”建立的 Logistic 回归模型,可以看到自变量 M_1 的 P 值为 0.01,其结果小于 0.05,说明自变量 M_1 对回归结果的影响是显著的。而自变量 M_3 的 P 值为 0.42,其结果大于 0.05,说明自变量 M_3 对回归结果的影响不显著,故剔除该变量,选取自变量 M_1 与因变量 Y 建立 Logistic 回归模型。

3.2.2. 建立 Logistic 回归模型

Table 7. Summary of Logistic regression results

表 7. Logistic 回归结果汇总表

	Coefficients:				
	Estimate	Std. Error	Z Value	Pr (> z)	
(Intercept)	-2.6989	0.454	-5.944	2.78E-09	***
M_1	-2.3102	0.8767	-2.635	0.00841	**

剔除 M_3 后,根据变量 M_1 “盈利能力”以及因变量 Y “是否为 ST 企业”,建立二元 Logistic 回归模型。从表 7 能看 M_1 的 P 值为 0.00841,其结果小于 0.05,说明该变量对结果是显著的,代表了模型的拟

合效果很好。根据 R 语言的输出结果显示，模型的 AIC 值为 43.921，其值较小，能代表原始数据的更多信息，所以最终选择建立二元 Logistic 回归模型。根据表 7 可以写出 Logistic 模型的回归方程：

$$\log\left(\frac{p}{1-p}\right) = -2.3102M_1 - 2.6989$$

$$p = \frac{1}{(1 + e^{M_1}) - 2.6989}$$

$$M_1 = -0.0016x_1 + 0.009x_2 + 0.5748x_3 + 0.0755x_4 + 0.5205x_5 + 0.0054x_6 + 0.0029x_7 - 0.0023x_8 + 0.0069x_9 + 0.0005x_{10} + 0.0014x_{11} + 0.0024x_{12}$$

由回归方程可以看出 Logistic 模型的系数为-2.31，也就是说事件的发生比为-2.31，由于该系数小于 1，说明自变量随着事件的发生有负向作用，即房地产企业的盈利能力这一因子在企业违约的情况中占主导作用，且房地产企业的盈利能力越好，违约的概率约低。

4. 结论及建议

4.1. 结论

依据 Logistic 模型的回归结果显示，决定我国房地产企业信用风险的主要因子是企业的盈利能力，其主要具体对应财务指标是净资产收益率、总资产回报率、总资产净利润率、每股收益、营业毛利率、销售净利率。所以在房地产企业经营的过程中，要时刻关注这几个指标的波动，当指标出现过低或者过高的情况时，要及时调整经营战略，以防止出现企业的信用问题，减少企业的违约风险。

4.2. 建议

房地产企业要在新环境背景下加强内部经营与内部控制能力，提高企业盈利能力，企业就有足够利润偿还债务，企业违约风险就越小。同时也要树立优良产品理念，加强房地产企业内部管理与监督，积极通过多种途径提升员工技术专业水平、创新能力、道德水平，形成优秀企业文化，创立优质企业品牌，不断提升房地产企业竞争力，使企业盈利能力与偿债能力不断加强。

参考文献

- [1] 马昭君, 葛新权. 房地产泡沫抑制了我国技术创新吗?——来自省际面板数据的经验证据[J]. 中国流通经济, 2023, 37(9): 11-22.
- [2] 李哲, 迟国泰. 基于最大指标区分度与最优相对隶属度的上市公司信用风险研究[J]. 中国管理科学, 2021, 29(4): 1-15.
- [3] 韩立岩, 郑承利, 罗雯, 等. 中国市政债券信用风险与发债规模研究[J]. 金融研究, 2003(2): 85-94.
- [4] 王慧, 张国君. KMV 模型在我国上市房地产企业信用风险度量中的应用[J]. 经济问题, 2018(3): 36-40.
- [5] 李杰辉, 陈珊珊. 基于 Logistic 模型的供应链金融信用风险实证研究——以房地产上市公司为例[J]. 金融理论与教学, 2020(4): 21-29.
- [6] Daniel, C. and Marcin, C. (2022) The Advantage of Case-Tailored Information Metrics for the Development of Predictive Models, Calculated Profit in Credit Scoring. *Entropy*, **24**, Article 1218. <https://doi.org/10.3390/e24091218>
- [7] 何睿. 基于深度学习算法的信用债违约风险预测模型[J]. 统计与决策, 2023, 39(17): 154-158.