

数字经济 2.0 时代基于 Logistic 回归模型的我国上市公司财务预警研究

张灿灿

(青岛城市学院, 山东 青岛 266106)

摘要: 在数字经济 2.0 时代背景下, 以我国沪深两市的 A 股上市公司为研究对象, 选取 15 项具有代表性的财务指标, 去除多重共线性强的指标, 经过筛选, 最终选出 4 个财务指标进行 Logistic 回归模型的构建和检验。结果表明, 测试集样本对非 ST 企业的预测正确率超过 85%, 且评估模型优劣标准的 AUC 值超过 0.8。构建一个行之有效的财务预警模式, 能够帮助公司管理者尽早地察觉到财务问题, 并在此基础上对公司的财务和经营战略进行相应地调整, 从而提升公司的危机预警和风险管理能力; 投资者可以建立一个能对投资企业未来财务情况进行预测的财务危机预警模式, 及时调整投资策略; 债权人可以通过建立财务预警模型来评估对该公司发放贷款的安全性。

关键词: 数字经济 2.0 时代; Logistic 回归模型; 财务预警; 上市公司

随着大数据、人工智能、移动互联网、云计算、物联网等技术的迅速发展, 我们已经进入了数字经济 2.0 时代。数字经济 2.0 时代的到来使得数据变得透明化、流动化和共享化, 企业能够更好地进行决策和规划, 从而更好地把握市场机会。同时, 近年来随着上市公司数量不断增多, 每年被 ST 的企业也逐渐增多, 这将给公司的运营造成负面影响, 如股价下跌, 融资能力受限等。财务危机的产生是一个渐进的发展过程, 所以, 能否在财务危机来临前, 及早地进行预警, 并及时地采取相应的行动, 从而对公司的运营战略进行相应的调整, 这对于上市公司的良性发展具有非常重要的意义。因此, 鉴于数字经济 2.0 时代的背景和上市公司面临的挑战, 对上市公司进行财务预警, 帮助企业及时应对潜在的危机是意义且必要的。

目前, 关于财务预警的研究主要有两类, 一是定性研究, 二是定量研究。早在二十世纪三十年代初期, 国外学者就对财务预警展开了研究, 相比之下, 国内开始得较晚, 但随着统计学和机器学习等技术发展, 越来越多的学者利用数学模型进行预警, 这些模型在预测的准确性和稳定性方面各自具有优缺点, 例如, 单变量模型、多元线性模型、神经网络模型、Z 值模型以及二元 Logistic 回归模型等。其中, Logistic 回归模型具有较强的实用性, 预测准确率较高。所以, 本文利用 Logistic 回归模型, 对我国沪深两市在 A 股上市的公司进行财务预警分析, 使得上市公司能够根据预测的情况及时采取防范措施, 以在数字经济 2.0 时代保持竞争力, 实现可持续发展。

一、研究样本及指标的选取

财务预警的研究对象是财务危机, 也被称为财务困境。当前, 学界对企业财务危机的界定还没有形成共识。在此基础上, 结合我国的国情, 将那些在股市中由于“财务异常”而受到特殊处置的企业界定为“财务危机企业”, 也就是 ST 企业; 没有受到特殊对待的企业被认为是财务状况良好的企业, 即不是 ST 企业。

(一) 研究样本的选取

本论文以中国沪深两市 A 股上市公司为研究对象, 研究样本则是 2022 年被 ST 或被 ST 且 2020 年和 2021 年未被 ST 或 *ST 的 A 股上市公司, 然后在排除了金融业、公用事业类股和大量资料

缺失类股后, 我们选出了 15 只 ST 类股。由于被 ST 的公司依赖于前一年的财务数据, 所以, 如果用 ST 前一年的财务数据来进行预测, 通常会过高估计模型的预测结果, 因此选择了在 ST 之前的两年, 也就是说, 对 2020 年的财务资料进行了相关分析。此外, 采用 1:2 匹配方式, 随机筛选出 30 家 2020-2022 年均未被 ST 的公司作为配对样本, 并以随机原则及 7:3 的原则选择 31 家公司作为训练集、14 家公司作为测试集。本文样本数据均来源于 CSMAR 数据库。

(二) 指标的选取

本文在综合参考前人财务预警指标的基础上, 结合我国上市公司实际情况, 从众多的财务指标中, 初步筛选出了 15 项财务预警指标作为自变量, 通过偿债能力, 盈利能力, 营运能力, 发展能力四个层面反映公司的财务状况。包括流动比率、速动比率、资产负债率、产权比率、流动资产周转率、固定资产周转率、存货周转率、应收账款周转率、总资产周转率、资产报酬率、净资产收益率、总资产净利润率、营业收入增长率、净利润增长率、总资产增长率等 15 个财务指标, 为后续分析及建模方便, 将其命名为 X1、X2、...、X15。

上述指标即为初步选定的自变量, 对这 15 个自变量做相关性分析, 制作相关系数矩阵, 可以查看两两变量之间的相关性。为了直观地展示变量之间的相关性, 使用 Python 中的 seaborn 库中的 heatmap 函数来绘制热力图。根据相关系数矩阵制作热力图, 颜色越深说明相关性越强, 如图 1 所示, 发现有些自变量之间存在着较强的相关性, 这意味着该自变量可以被其他自变量线性表示, 从而存在多重共线性。可能会导致模型估计出的系数不准确, 进而导致模型的解释出现问题, 因此, 为了更加准确的估计模型, 需要对存在共线性的自变量进行一定的筛选。

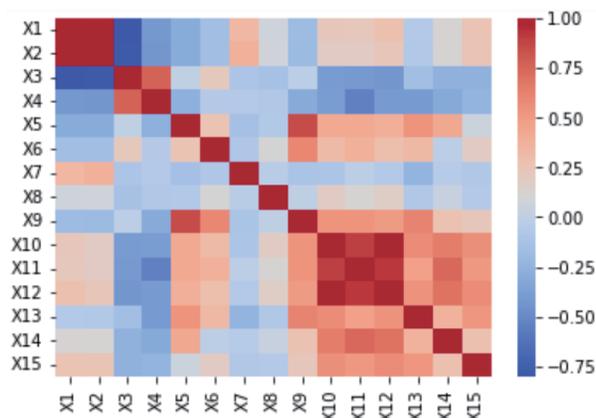


图 1 初步选取的上市公司财务预警指标相关性矩阵的热力图

经过筛选, 最终进入模型的变量是 X1、X9、X11、X13, 分别为流动比率、总资产周转率、净资产收益率和营业收入增长率。

二、Logistic 财务预警模型的估计与检验

采用 Logistic 回归模型来对我国上市公司进行财务预警。Logistic 回归模型 (逻辑回归模型) 是机器学习中的一种分类模型, 由于算法的简单和高效, 在实际中应用非常广泛。其中, 因变量

可以是二分类的，也可以是多分类的，但是二分类的更为常用，也更加容易解释。本文的因变量是二分类的，即“存在财务危机”和“不存在财务危机”，相应的模型为二项逻辑回归模型。

(一) Logistic 回归模型的原理

模型的基本形式如下：

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

在上市公司逻辑回归财务预警模型中，P(Y=1)是公司存在财务危机的概率，X₁、X₂、...、X_k是筛选出来的自变量即相应的财务指标，β₀、β₁、β₂、...、β_k是模型的参数。

可以取定一个阈值，通常以0.5为阈值，通过对比概率值与阈值的大小关系来确定是否存在财务危机，如果计算出来的概率大于等于0.5，则判断为存在财务危机，如果计算出来的概率小于0.5，则判断为不存在财务危机。

(二) 基于 Logistic 回归的财务预警模型的估计与检验

基于 Python 中 sklearn 第三方库对选定的自变量和因变量进行二元 Logistic 回归，得到 Logistic 回归财务预警模型为：

$$\hat{P} = \frac{1}{1 + e^{1.31398 + 0.2106X_1 + 0.3134X_9 + 1.2191X_{11} + 0.3919X_{13}}}$$

将训练集样本代入上述模型进行检验，得出训练集预测结果，如表 2 所示：

表 1 训练集预测结果

真实值		预测值		正确率 (%)
		是否存在财务危机		
		0 (非 ST 公司)	1 (ST 公司)	
是否存在危机	0 (非 ST 公司)	20	3	86.96
	1 (ST 公司)	3	5	62.50
总正确率				80.65

由表 1 可知，训练集包含 31 家样本，在 23 个正常的上市公司中，20 个预测准确，3 个预测错误，正确率为 86.96%。在被 ST 的企业中，3 个被错误地判定为非 ST，5 个被正确地判定，预测正确率为 62.50%。虽然异常企业的误判率较高，但这也提醒被 ST 企业加强经营管理，真正摘掉 ST 帽子。总体而言，总的正确率为 80.65%，结果较为理想，具有较高的应用价值。

为进一步验证模型的预测能力，使用训练集得到的逻辑回归模型检验测试集样本，检验的结果见表 2：

表 2 测试集预测结果

真实值		预测值		正确率 (%)
		是否存在财务危机		
		0 (非 ST 公司)	1 (ST 公司)	
是否存在危机	0 (非 ST 公司)	6	1	85.71
	1 (ST 公司)	2	5	71.43
总正确率				78.57

从预测结果中可以明显看出，7 家不存在财务危机的企业预测准确率达到 85.71%，仅有 1 家被错误地归入 ST 类。而在出现财务危机的公司中，有 2 家公司被错误地判定为不存在危机，5 家公司被准确地判定，预测正确率为 71.43%；总体预测正确率为 78.57%。

AUC 是 ROC 曲线下方的面积，通常被作为一个评估准则，用来衡量一个正向样本领先于一个负向样本的可能性。总体而言，AUC 值在 0.7 以上时，该模型具有较为显著的判别力。AUC 值较大时，其预测结果较好。利用 Python 进行 Logistic 回归分析，结果显示，训练集样本的 AUC 是 0.83，测试集样本的 AUC 是 0.82，

这表明模型的预测能力较强，效果较好，信息较为可靠。

对于财务状况较好的企业，财务预警模型的识别精度较差的企业要高。建立了预警模型，可以对我国上市公司的财务情况进行一定的预测，为公司的未来发展提供了一个很好的参考。

三、结论

经过广泛的参考，认真的研读和归纳，本文在对我国上市公司财务预警进行研究后，得到了以下结论：

将国泰安数据库得到的上市公司在被 ST 之前两个年份的 15 个财务指标数据，剔除了金融和公用事业的公司以及数据严重缺失的公司，最终筛选出 31 家训练集样本、14 家测试集样本，共 45 家上市公司作为研究对象。对 15 项财务指标进行相关性分析，对存在共线性的自变量进行一定的筛选，最终选定 4 个财务指标作为 Logistic 回归模型的自变量。

将一家上市公司在 2022 年会不会陷入财务危机，也就是被 ST 或者 *ST 当作因变量，使用 Python 中的 sklearn 第三方库，建立并验证了 Logistic 财务预警模型。通过实证分析发现，在财务状况较好的企业中，财务预警的准确率较差企业高。其中，在训练集和测试集样本中，对非 ST 企业的预测准确率都超过了 85%，训练集样本和测试集样本的 AUC 值都达到了 80% 以上，这说明该模型的预测是非常可靠的，它具备一定的实际意义和参考价值。

通过构建一种行之有效的财务预警模式，能够让公司的管理者在最短的时间内察觉到财务危机，并据此对公司的财务和经营策略进行调整，从而提升企业对危机的预判和风险控制能力，将其防患于未然。在此基础上，通过建立财务预警模型，对被投资企业的财务风险进行预测，并据此做出相应的决策，进而将投资风险降至最低。另外，在考虑对上市公司进行金融支持时，通过构建一个财务预警模型，可对其放款的安全进行评价。

当然，本文构建的财务预警模型也有一定的不足。一方面，由于本文是预测两年后的公司状况，因此预测范围还有待进一步扩大。另一方面，预测的准确性依赖于财务数据的真实有效性，而且可能受到不同行业间的影响，但本文并未考虑行业差异因素。另外，该模型还会受宏观经济条件的影响，在今后的研究中，还可以进一步考虑加入更多的变量，以提高模型的预测效果，例如，宏观经济指数，产业因子等。但是，这一模型，不论是对管理者，还是对投资人，都有很好的借鉴作用，并能在某种程度上规避或减轻企业的损失。

参考文献：

[1] 李怡雯. 基于 Logistic 回归模型的互联网服务行业上市公司财务风险预警分析 [J]. 商场现代化, 2022 (14) : 158-160.
 [2] 白加丽. 我国上市公司财务风险预警研究——基于主成分分析和 Logistic 回归的财务风险预警模型 [J]. 经营与管理, 2022 (08) : 8-15.
 [3] 刘艳华, 史培贞, 毛蓓蓓. 上市公司财务预警模型研究——基于 Logistic 回归模型的实证分析 [J]. 中国总会计师, 2022 (05) : 93-95.
 [4] 郭芷若. 基于因子分析法和 Logistic 回归的有色金属业财务预警模型研究 [D]. 成都大学, 2022.

课题项目：2021 年度青岛城市学院人文社会科学课题：数字经济 2.0 时代基于 Logistic 回归模型的我国上市公司财务预警研究 (编号 QCU21RB01)