

基于非平衡数据处理和多变量筛选方法的 上市公司财务困境预测研究

邢 凯¹, 盛利琴², 张 盼¹, 李 珊³

(1. 南昌大学经济管理学院, 南昌 330031; 2. 南昌大学前湖学院, 南昌 330031; 3. 中国矿业大学
经济管理学院, 徐州 221116)

摘 要 在国内供给侧结构性改革背景下, 市场环境复杂多变, 公司债务违约频发, 建立一种及时有效的财务困境预警模型十分必要. 已有的多数困境预测模型尚未有效解决数据集不平衡、关键预测指标选取不稳定、样本匹配存在随机性等问题, 且并不适应于当下中国复杂多变的市场状况. 为此, 本文采用 Bootstrap 重抽样方法构建 1000 个研究样本, 通过 LASSO (least absolute shrinkage and selection operator) 变量选择技术筛选关键预测因子构建提前 3 年预测的 Logit 模型, 并在预测阶段将样本进行 1000 次随机切割和预测以降低随机误差. 结果表明, 由 Bootstrap 样本组建方式结合 LASSO 构建的 Logit 困境预测模型相比传统应用的“同行业资产规模相近”方式所构建模型的预测能力更强. 另外, 该嵌入 Bootstrap 式 LASSO-logit 模型比主流的 O-Score 模型、Z_{China}-Score 模型预测效果更好, 准确率提高 10%, 更加适用于中国时变的市场. 本文所构建模型能帮助公司利益相关者更好地识别财务困境并及时做出调整, 以降低公司债券违约率或避免发生公司违约.

关键词 财务困境预测; LASSO-logit; bootstrap

收稿日期: 2023-08-11

基金项目: 国家自然科学基金青年项目“基于系统临界转变理论的大宗商品场外市场操纵事件识别研究”(72201120); 江西省自然科学基金面上项目“江西省大宗商品交易场所合约崩盘成因及预警研究”(20232BAB201027); 大学生创新创业训练计划国家级项目“基于机器学习的中国上市公司财务困境预测研究”(202210403051)

Supported by Youth Project of National Natural Science Foundation of China “The Study of Detecting the Events of Market Manipulation in the Commodity OTC Market Via the Theory of Critical Transitions” (72201120); Natural Science Foundation of Jiangxi, China “The Study of Detecting Crash Risk in the Commodity OTC Market Via the Theory of Critical Transitions” (20232BAB201027); National Project of the College Student Innovation and Entrepreneurship Training Program “Research on Financial Distress Prediction of Chinese Listed Companies Based on Machine Learning” (202210403051)

作者简介: 邢凯, 博士后, 副教授, 硕士生导师, 研究方向: 风险管理、金融科技、公司金融、公司治理, E-mail: xingkai@ncu.edu.cn; 盛利琴, 本科生, E-mail: shengliqin7@163.com; 张盼, 本科生, E-mail: 3544068573@qq.com; 通信作者: 李珊, 博士研究生, 研究方向: 公司金融、风险管理, E-mail: Shanlicumt@163.com.

Financial Distress Prediction of Listed Companies Based on Unbalanced Data Processing and Multivariate Selection Methodology

XING Kai¹, SHENG Liqin², ZHANG Pan¹, LI Shan³

(1. School of Economics and Management, Nanchang University, Nanchang 330031, China; 2. School of Qianhu, Nanchang University, Nanchang 330031, China; 3. School of Economics and Management, China University of Mining and Technology, Xuzhou 221116, China)

Abstract In the context of China's domestic supply-side structural reforms, the market environment has become increasingly complex and volatile, accompanied by a high frequency of corporate defaults. It is necessary to establish a timely and effective financial distress prediction model. However, most existing distress prediction models have not adequately addressed critical issues such as the imbalance of datasets, the instability in selecting key predictive indicators, and the randomness in sample matching. Moreover, these models are not well-suited for this dynamic market in China. To address these limitations, this study initially employs the Bootstrap resampling method to construct 1000 research samples and uses the LASSO (least absolute shrinkage and selection operator) variable selection technique to screen key predictors. Then, this study constructs a logit model for predicting financial distress three years in advance using LASSO-selected key predictors. To mitigate the impact of random errors, the samples are further subjected to 1000 random splits and predictions. The results demonstrate that the LASSO-logit prediction model, constructed by combining the Bootstrap sample construction method with LASSO, exhibits superior predictive ability compared to the models constructed using the traditional "similar asset size in the same industry" approach. Furthermore, this LASSO-logit model outperforms mainstream O-Score and Z_{China} -Score models, with a 10% increase in accuracy, making it more suitable for China's dynamic market. Overall, the model constructed in this study can help stakeholders of companies better identify financial distress and make timely adjustments to reduce corporate bond default rates or avoid corporate defaults.

Keywords financial distress prediction; LASSO-logit; bootstrap

1 引言

自 2014 “超日债”打破刚性兑付, 成为中国债券市场上首个违约案例后, 国内陆续出现大规模违约事件; 与此同时, 部分行业产能过剩, 债台高企, 资金链紧张, 违约风险不断上升, 导致投资者利益受到侵害, 经济下行压力增大. 近年来, 复杂多变的市场环境仍然威胁着企业的健康发展. 全国企业破产重整案件信息网数据显示, 2020 年企业破产案件为 38712 件, 相较于 2019 年同期的 21017 件, 同比增长 84.2%. 2020 年, 《国务院关于进一步提高上市公司质量》明确提出要加强资本市场基本制度建设, 大力提高上市公司质量; 2022 年中央经济工作会议和政府工作报告提出 “防范和化解重大风险”, 要求做好经济金融领域风险处置工作;

2023 年政府工作报告与二十大会议精神强调,有效防范化解经济金融风险,守住不发生系统性风险依然是政府工作的重心,其重要抓手是按照市场化、法制化的原则主动解决相关金融主体债务问题,做到早识别、早预警、早发现、早处置。债务高企往往是金融主体违约破产的主要原因。企业利益相关者规避财务困境和违约风险的需求日益上升,然而已有的多数困境预测模型并未解决关键预测指标选取不稳定、数据集不平衡、预测结果存在随机性等问题,且并不适应于时变的中国市场。有效的财务困境预测模型有助于避免困境的发生,能够帮助投资者进行科学决策,减少投资损失;能够帮助企业管理者有效评估企业发展水平,增强企业发展韧性和抗风险能力;能够有效防范行业间的风险感染,阻止市场酝酿出金融系统性风险,推动建设稳健高效的金融体系;能够为政府政策制定和市场问题分析提供数据支撑和理论依据,更能够稳就业保民生。因此,本文着力于构建一种及时有效且适用于中国市场环境的财务困境预警模型为中国企业的高质量发展保驾护航。

财务困境预测是一个经典的研究问题,但国内相关研究起步较晚,大部分理论建立在西方学者的研究基础之上。

预测指标体系方面,主要以单变量判别体系与多变量体系两种思想进行研究。Beaver (1966) 率先提出单一比率模型来预测财务困境,并发现营运资本/负债和净利润/总资产的单一预测能力很好。此后少数学者如吴世农和卢贤义 (2001) 对单变量指标效果进行了研究,证实了多指标体系的优越性。多指标体系需要分类以提高变量选择效果,更好地探究变量的经济意义 (吴世农和卢贤义, 2001)。Altman (1968)、Zavgre et al. (1985)、Shumway (2001) 的研究存在预测因子数量较少而不对其进行分类的情况。但多数学者还是基于 1968 年 Altman 提出的多元判别思想,对传统的财务比率如盈利比率、负债比率、运营比率、现金流能力等指标来分组反映公司财务状况并利用这些指标进行财务困境预测 (Altman, 1968)。将财务指标进行分组使组内变量反映同一类财务信息的不同方面,组间变量反映不同类的财务信息,在建模时加入分组变量信息可以提升指标筛选效果,能够更加全面地探究企业陷入财务困境的原因 (方匡南和杨阳, 2018; 王小燕和袁欣, 2018)。李慧等 (2019) 将 20 个盈利质量类指标分成五个维度并筛选出有显著预测作用的变量。杨贵军等 (2019) 在盈利能力、偿债能力、成长能力等组指标当中选中 15 个财务指标作为预测因子。此外,多数研究虽然采用的是多指标体系,但是选取的预测因子数量却不多,如石晓军等 (2007)、韩立岩和李蕾 (2010)、杨贵军等 (2019) 的研究,可能存在片面反映公司财务状况的问题。预测因子数量较多甚至可认为是高维数据时,学者们通常会采用收缩或降维算法来提取预测因子的主要特征,以往文献主要采用主成分分析法、最优子集法和逐步回归法等方法,这些方法在计算效率和可解释性上都存在一些缺陷 (Chen, 2011; 梁琪, 2005; 韩立岩和李蕾, 2010; Khoja et al., 2014)。后来 Tibshirani (1996) 提出的 LASSO 方法通过惩罚函数使得变量的系数为零从而达到降维目的,在处理高维数据上更具有明显特征,不仅筛选出的变量具有可解释性,还缓解了多重共线性问题。许多学者选择 LASSO 对关键指标进行筛选,并通过 LASSO 方法验证了财务指标在预测公司破产困境方面的重要性 (Tian et al., 2015)。

对于预测模型的研究,预测模型根据是否提前预测可以划分为静态与动态模型,静态模型由于只能考虑预测因子在陷入财务困境当年的作用从而人为提高了预测能力备受诟病 (Li et al., 2021)。近年来多数学者采用提前 1~5 年预测时间构建动态模型以获得实际预测能力,

其中提前3年预测时间居多(谢纪刚等, 2005; 李慧等, 2019; Li et al., 2021). Ohlson (1980) 将 Logit 方法应用于财务困境预测, 提出了 O-Score 预测模型. Logit 模型由于假设条件少、变量解释力强、预测效果好等特点, 逐渐成为财务困境预警的主流方法延用至今. 近年, 许多学者将 Logit 模型应用于中国市场财务困境预测, 预测准确率最高可达 98.6%, 但超高准确率并未预留时间窗口, 存在准确率虚高的可能性(韩立岩和李蕾, 2010; 谢纪刚等, 2005; 杨青龙等, 2016; 李慧等, 2019; Li et al., 2021). 20 世纪末到 21 世纪初是西方财务困境预测理论发展的高峰期, 国内外学者将与机器学习、人工智能有关方法引入到财务困境预测研究当中, 并获得了较高的预测准确率. 如倪志伟等 (2014) 通过流形学习和多核 SVM 的方法获得了 80% 以上的准确率; 李慧等 (2019) 通过 PCA 及 SVM 方法获得了 90% 以上的准确率; 张向荣 (2021) 采用子空间多核学习方法获得了 80% 以上的准确率; 陈艺云 (2022)、孙灏等 (2022) 通过对企业年报进行文本分析预测财务困境, 也得到了较高的预测准确率. 这类方法虽然提高了预测准确率, 但模型复杂结果不易解读. 同时, 采用这类方法的学者们鲜少在文献中提到样本非随机的问题. 财务困境预测对数据极为敏感, 随机性问题不可忽视. 比如, 在构建样本数据集阶段, 由于 ST 公司天然比非 ST 公司少, 所以在构建研究样本时, 需要对两类公司类型进行配对. 学者们主要采用两种方式: 同行业资产规模相近匹配与随机匹配 ST 与非 ST (倪志伟等, 2014; 杨贵军等, 2019; 闫达文等, 2024). 因 ST 样本数据较少, 限定行业以及资产规模这种事先人为确定的抽样标准会导致样本非随机进而产生过度选择问题 (Zmijewski, 1984; 叶佳和陈军, 2010; 董景荣和陈军, 2010). 随机匹配一次样本进行预测的效果也不具有代表性. 大多数该领域的研究均采用随机匹配样本做一次关键因子筛选, 这样可能会丧失一些可能有预测意义的预测因子. 比如倪志伟等 (2014) 对 2010–2012 年的 ST 样本按照同时期同行业原则随机选取一个正常非 ST 公司进行提前 2~3 年的预测; 杨贵军等 (2019) 将 2000–2017 年间中国 A 股市场的每一家 ST 公司按相同年度、同一行业、资产规模最相近原则来选择一家正常经营的上市公司与之匹配并得到 ST 公司和正常公司各 184 家; 闫达文等 (2022) 利用资产规模相似选取原则对 2009–2017 年合计 175 家 ST 公司抽取等数量的非 ST 公司. Bootstrap 重抽样是随机匹配 ST 公司与非 ST 公司的一种方法, 一般通过对 ST 样本和非 ST 样本重复抽样 1000 次以接近总体分布, 可以很好地解决样本片面性的问题 (Shao, 1996; Sauerbrei et al., 2015; Li et al., 2021). Shao (1996) 利用 Bootstrap 方法分别进行了 15、20、25、30、40 次随机抽样并构建五种预测模型; Li et al. (2021) 利用 Bootstrap 方法给 ST 样本随机匹配非 ST 样本并重复 100 次. 在预测阶段, 部分采用机器学习方法的学者按不同的比例划分训练集与测试集来模拟样本失衡 (Zavgre, 1985; 石晓军等, 2007; 陈艺云等, 2018). 然而训练集与测试集的划分也必然存在随机误差, 模型预测结果存在极大的不稳定性, 这可能也是导致财务困境预测研究当中预测准确率相差较大的一个原因.

综上, 在大量财务困境研究中, 普遍存在以下问题: 第一, ST 公司数据较少导致样本天然存在失衡情况, 而人为选择抽样标准会造成样本匹配非随机的问题. 第二, 该领域纳入研究的变量数量较少, 存在变量选取片面和不稳定的问题. 第三, 由于样本变动而导致的预测结果具有不稳定性, 目前还鲜有较为成熟的研究在实证中给出有力的解决方案. 因此本文紧扣三个问题, 通过更加全面细致的数据挖掘探究公司陷入财务困境的原因, 并找出相应的治理之道.

基于此,本文利用2014–2019年除金融地产行业外的所有中国上市公司财务数据,初步选取了尽可能多的91个财务指标作为候选预测因子,采用Bootstrap重抽样方法构建1000个研究样本来接近公司数据总体分布特征,通过LASSO变量选择技术并结合十折交叉验证方法筛选出六个关键预测因子:IE/NI、CHIN、CA/SALE、CFO/TL、RE/TA、TAT,利用关键预测因子来构建提前3年的Logit财务困境预测模型.为保证模型预测效果的稳定性,在预测阶段将样本进行1000次随机切割,千次的样本分割结合不同比例划分训练集与测试集有效地保证了预测结果的稳健性.通过求取千次模型评价指标结果均值的方法来评估模型,结果表明,由Bootstrap抽样方法构建的预测模型相比传统应用“同行业资产规模相近”的样本组建方式构建的模型预测能力更强,比主流的O-Score模型、Z_{China}-Score模型预测效果提升10%.

本研究的边际贡献主要包括:第一,虽然很多文献指出财务困境预测类研究结果深受样本变动影响,但鲜有作者提出针对性解决方案.本文在样本组建、变量选取、预测效果评估各阶段上有效控制了随机影响,技术上丰富了该领域研究.我们深刻地讨论了此类基于统计方法的研究应该考虑的稳定性问题:为了避免组建样本过程中的随机匹配可能导致的随机误差,我们采用了Bootstrap随机抽样匹配1000个样本(全部的ST公司与等量的非ST公司)的方法,来接近总体分布.为了避免随机划分训练集与测试集方法带来的误差,我们也将预测过程实行了1000次,以求取评价指标结果平均值的方法来评估模型效果,千次平均后的样本外预测准确率提升至71.16%.第二,预测效果上,本文构建的Bootstrap式LASSO-logit财务困境预测模型较传统主流的O-Score模型、Z_{China}-Score模型提前三年预测准确率提升约10%,不仅实现了模型简洁与预测效果的均衡,还挖掘出了两个未被重视的对中国企业财务困境有显著预测能力的因子:CA/SALE(流动资产/营业收入),提示企业适度降低流动资产中的存货比重,审慎对待宽松的流动资产投资策略;IE/NI(利息支出/净收入)提示企业应选择合适的融资渠道来增强偿债能力.第三,实际应用层面,本文选择2014–2019年作为样本时间,涵盖了一个完整的经济周期,期间见证了中国经济形势的起伏不定及信用违约的恶化.本文采用91个预测因子全面考察了在这个完整经济周期下财务变量对企业财务困境的预测作用,使得本文所构建模型具有持久适用性和生命力,可以通过定期更新预测模型来及时发现不同经济周期下风险信号,从而实现持续监测与动态风险管理. A股ST数量2014年起逐步走高,伴随百年变局加速演进、外部环境更趋复杂,该预测模型更加契合市场需求.

下文安排为:第二部分介绍数据、变量与研究设计,第三部分为实证分析,最后部分总结全文及可能的建议.

2 理论模型设定

2.1 数据来源与样本

本文使用的大部分数据来自中国北京聚源锐思数据科技有限公司开发的金融研究数据库——RESSET数据库.该数据库历史金融数据完整,涵盖股票、债券、基金、黄金、外汇、研究报告以及宏观经济和行业统计等广泛领域.在股票方面,RESSET提供了在沪深交易所上市公司的综合安全信息和新闻、财务报表和财务指标等数据,满足了本文的研究目的.如今,RESSET数据库已被有关中国金融市场的研究广泛引用(Calomiris et al., 2010).

本文采用中证指数 2016 版行业分类标准 (包括: 信息技术、工业、原材料、可选消费、主要消费、公用事业、医药卫生、能源、电信业务、金融地产), 以 2014–2019 年在沪深交易所上市的所有非金融地产行业公司的 91 个财务指标数据为研究样本。2014 “11 超日债” 首次打破刚性兑付, 成为中国债券市场违约先例, 使国内债券市场步入违约常态化, 对金融债券市场的影响深远。此外, 2014–2019 年中国经济形势起伏不定, 作为一个完整的经济周期涵盖了经济下行与回暖的不断反复, 对这一时期中国上市公司财务状态的研究契合了当期市场需求。

2.2 变量构建

2.2.1 财务困境度量

关于财务困境的界定, 由于西方发达国家破产法法律体系完善, 国外学者以企业申请破产作为其陷入财务困境标志 (Altman, 1968)。国内破产法虽然有所发展但并不充分, 并且中国有其特殊的国情, 企业在陷入财务困境后会受到多方面干预, 例如若中国国有企业陷入财务困境, 那么它就会受到来自政府的财政补贴和扶持, 以保障国有企业的经济地位, 因此中国上市公司的破产记录十分少见, 破产企业的数据也非常难获得 (Wang and Li, 2007)。基于此, 国内外大部分学者以企业被特殊处理 (Special treatment, 简称 “ST”) 作为其陷入财务困境的标志, 本文也遵循先前研究 (Zhang et al., 2010; 杨贵军等, 2019; 李慧等, 2019; Li et al., 2021), 将公司是否获得 ST 处理构建为一个二值因变量, 该值等于 0 表明该公司当年未被 ST, 财务状况正常; 等于 1 则表明该公司当年被 ST, 陷入财务困境。

ST 制度是一种危机预警机制, 旨在提醒投资者鉴别上市公司是否陷入财务困境。中国证监会于 1998 年 4 月 22 日推出 ST 制度, 符合以下任一标准的公司将被归类为 ST 公司: ①连续两年盈利为负; ②最近一年盈利和权益为负; ③最近一年经审计的财务报表存在重大疑问; ④证券交易所认定为异常活动或者退市风险较高的其他情形。另外, 若上市公司已经连续 3 年以上净利润为负, 公司经营持续亏损则会被标记为 *ST, 表明该股票有退市风险。已有研究发现, 前二者是触发 ST 警告最常见的原因, 且 ST 企业兼具高杠杆和负收益的特征, 由此可见, ST 可以较好地代表公司财务困境 (Zhang, 2010)。因此, 本文将上市公司是否被标记为 ST 构建为因变量, 其中, ST 与 *ST 均归为 ST 公司, 其余处理类型为非 ST。本文所选 ST 与 *ST 公司皆由《上海证券交易所股票上市规则》与《深圳证券交易所股票上市规则》认定。

2.2.2 预测因子

关于财务预警指标的选择, 前人的研究当中已提供了多种参考方法。Altman (1968) 采用统计过滤器从 22 个财务比率中最终筛选出 5 个关键预测指标。Brédart et al. (2021) 总结反映公司财务状况的指标包括流动性、财务结构、偿付能力、盈利能力、活跃度和周转率等 8 类; Klietnik et al. (2020) 收集了 400 多个困境预测模型, 总结出相关研究中最常用的 3 种财务比率: 流动比率、总负债比总资产和总销售比总资产。虽然已有研究有详尽的财务预警指标选择方法, 但在当今社会, 财务指标不断细化, 尽管采用相同的方法选择了相同的指标, 也会因样本长度、数量和模型效应的不同而展现出不同的性能 (倪志伟等, 2014)。因此, 根据 RESSET 提供的财会信息, 本文初步选取了 91 个预测因子。选取尽可能多的候选预测因子不仅可以解决因样本差异带来的指标性能差异, 而且能更加全面地反映公司财务状况, 提高模型预测性能。参考 Beaver (1966)、Altman (1968)、Ohlson (1980)、Zhang et al. (2010) 的

研究, 本文将这 91 个预测因子进行分组, 包括活动率、流动性比率、盈利能力比率和负债率, 具体如表 1 所示. 其中, SIZE (总资产账面价值的对数转换) 主要反映公司的资产规模, 不同于前面四组反映公司财务状况的指标, 故将其归为“企业规模”(吴世农和卢贤义, 2001).

A. 活动率. 一般而言, 企业会尽快将生产转化为现金或销售额以获得更多的收入. 活动率可以衡量公司将资产负债表中的各种账户转换为现金的能力, 资产变现能力越强公司财务状况越好, 陷入财务困境的可能就越低 (杨贵军等, 2019). 现金流、营业收入和库存通常用于构建活动率. 在这些变量中, 营业收入指标因其与企业获得收入能力密切相关而使用最频繁.

B. 流动性比率. 公司资产的流动性可以用公司的现金、总资产和流动负债来衡量. 流动比率代表了将流动资产转化为现金的能力, 也表明了偿还短期负债的能力. 如果流动比率过低, 公司将难以支付短期负债, 相反, 高比率则表示闲置资金过多 (闫达文等, 2022).

C. 盈利能力比率. 盈利能力指标可以从销售和运营角度来评估公司获利的能力, 同时显示了公司为股东创造利润和价值的效率, 一般来说, 盈利能力比率越高代表公司财务状况越好 (李慧等, 2019).

D. 负债率. 负债率可以衡量负债和资产的规模, 以反映公司的健康状况和长期偿债能力. 如果负债水平高于一般水平, 那么公司无法及时偿还债务造成违约的可能性将更大 (杨贵军等, 2019; 闫达文等, 2022). 因此, 债权人通常会着重考虑公司的负债率以判断其财务状况.

表 1 变量描述性统计

变量	描述	变量	描述
A 组: 活动率			
CF/SALE	营业收入现金含量	CF/OI	现金流/营业收入
EBITDA/SALE	息税折旧摊销前利润率	SALE/WC	营业收入/营运资金
CF/TA	总资产现金回收率	SALE/CA	营业收入/流动资产
NI/SALE	营业净利润率	SALE/TL	营业收入/总负债
INV/OI	净利润/营业收入	VA/SALE	营业总收入增长率
CA/SALE	流动资产/营业收入	NI/VA	净利润增长率
INV/SALE	存货/营业收入	VA/TA	总资产增长率
REC/SALE	应收账款/营业收入	AC/SALE	应付账款与收入比
SALE/TA	营业收入/总资产		
B 组: 流动性比率			
CLCH/TA	(流动负债 - 现金)/总资产	CL/SALE	流动负债/营业收入
CH/TA	现金/总资产	WC/SALE	营运资金/营业收入
CH/CL	现金/流动负债	CH/CA	现金流动负债比率
QA/CL	速动资产/流动负债	CA/TA	流动资产/总资产
CA/CL	流动资产/流动负债	SA/TA	短期资产/总资产
WC/TA	营运资金/总资产	INV/TA	存货/总资产
CL/TL	流动负债/总负债	EBIT/SL	息税前收益/短期负债
CFO/TL	经营活动产生的现金流/总负债	QA/TA	速动资产/总资产
RE/CL	留存收益/流动负债	CL/TA	流动负债/总资产
C 组: 盈利能力比率			

表 1 (续)

变量	描述	变量	描述
OI/SALE	营业利润/营业总收入	ROIS	营业利润率
ln(SALE)	营业总收入的自然对数	IT	存货周转率
ln(TA)	总资产的自然对数	DSO	应收账款周转天数
EBIT/TA	息税前利润/总资产	DPO	应付账款周转天数
EBIT/SALE	息税前利润/营业总收入	PT	应付账款周转率
PROFIT	利润总额	CAT	流动资产周转率
dSALE	营业收入增长率	FAT	固定资产周转率
NI/TA	净利润/资产总计	TAT	总资产周转率
RE/TA	留存收益/总资产	CFM	现金流利润率
ROIC	投入资本回报率	OP/TA	营业收入/资产总计
ROA	资产净利率	CF/SF	净现金流量/净资产
ROE	净资产收益率	NI/SF	净利润/所有者权益
ROATA	平均净资产收益率	EBIT/(TA-CL)	息税前利润/(总资产-流动负债)
ROTC	资产报酬率	EBIT/IE	息税前利润/利息支出
GPM	毛利润率	SALE/INV	营业收入/存货
RT	应收账款周转率	EBIT/SF	息税前利润/股东权益
CHIN	$\frac{NI_t - NI_{t-1}}{ NI_t - NI_{t-1} }$	INTWO	若 NI>0, 则为 1; 若 NI<0, 则为 0
D 组: 负债率			
D/TA	资产负债率	LD/WC	长期负债/营运资金
E/TL	权益对负债比率	D/E	负债/权益
IE/SALE	利息支出/营业收入	FE/EBITDA	财务费用/息税折旧摊销前利润
IE/NI	利息支出/净利润	FE/NI	财务费用/净利润
FD/CF	金融负债/现金流量	FE/TA	财务费用/总资产
TL/TA	总负债/总资产	TA/TL	总资产/总负债
E/TA	所有者权益/总资产	CF/TL	现金流/负债总计
TL/E	产权比率	CL/CA	流动负债/流动资产
QR	速动比率	TD/TA	总债务/总资产
IE/EBITDA	利息支出/息税折旧摊销前利润	IE/TA	利息支出/总资产
OI/TA	营业收入/总资产	CF/TD	现金流/总负债
企业规模			
SIZE	总资产账面价值的对数转换		

2.3 研究方法

2.3.1 研究设计

首先, 本文选取了沪深交易所除金融地产行业外的所有上市公司的 91 个财务指标作为候选预测变量, 将上市公司是否陷入财务困境构建为因变量。

而后, 通过 Bootstrap 抽样组建样本、结合十折交叉验证的 LASSO 降维法和共线性检验筛选出关键预测因子, 并用于构建提前三年的 Logit 财务困境预测模型。同时, 我们也将这类研究中常采用的“同行业资产规模相近”原则来组建样本, 重复以上操作得到的变量筛选

结果与 Bootstrap 方式对比. 随机将训练集样本中的 ST 公司与非 ST 公司进行 1:1 配对, 将整体样本的其余非 ST 公司放入测试集样本中, 并对 Logit 模型进行训练和测试.

通过 McFadden's pseudo R^2 , AUC, 平衡准确率, F_β 值这四个指标对样本内和样本外的模型预测效果进行评估, 通过评价指标结果将本文所构建 Bootstrap 式 LASSO-logit 模型与传统主流的 Totass (采用同行业资产规模相近原则匹配样本) 式 LASSO-logit 模型、O-Score 模型、Z_{China}-Score 模型进行对比.

最后进行稳健性检验, 在组建研究样本数据阶段采用 ST: 非 ST = 1:1、1:2、1:3 三种比例来模拟 ST 与非 ST 公司的数据失衡情况; 构建 logit 模型时, 按照训练集 ST 公司数: 测试集 ST 公司数 = 1:1、2:1、3:1 的标准重复训练 logit 模型并进行预测. 将此划分数据集方式重复 1000 次, 以减少随机切分样本内外数据带来的误差. 图 1 展示了本文的设计思路.

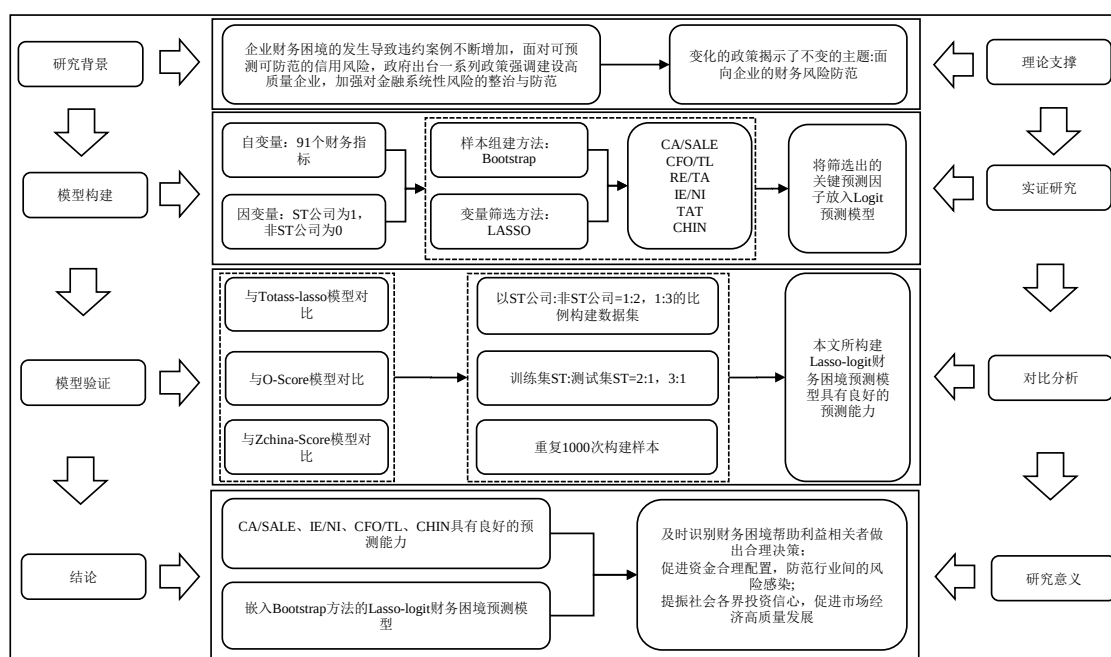


图 1 研究思路图

2.3.2 技术方法

1) 数据预处理. 本文采用与大部分研究相同的数据清洗方法, 对样本数据进行百分之一缩尾处理, 缺失值采用均值填补, 最后进行归一化处理, 使数据更加合理、准确 (杨贵军等, 2019; 李慧等, 2019; 闫达文等, 2022). 处理后最终得到了 3786 家公司样本, 其中 ST 公司 158 家, 非 ST 公司 3628 家. 表 2 显示了 ST 公司与非 ST 公司的各年度分布状况.

表2 ST与非ST公司数量分布状况

年份	ST 公司	非 ST 公司
2014	32	3509
2015	39	3531
2016	48	3722
2017	48	3817
2018	44	3820
2019	71	3794

2) 样本组建方法. Bootstrap 抽样是由 Efron (1979)、Efron and Tibshirani (1986) 首先提出的一种广泛适用且极其强大的统计工具, 它从数据源中重复抽取样本数据进行替换以估计总体参数. 本文采用 Bootstrap 方法来组建样本使样本选择更具随机性, 构建的预测模型更具普适性. 抽样步骤:

a) 构建 ST 公司数据集 S , $S = \text{matrixes}(D_{it}, CP_{it}^1, \dots, CP_{it}^m)$, 其中 D_{it} 表示第 i ($i = 1, 2, \dots, n$) 公司在第 t 年陷入财务困境, $CP_{it}^1, \dots, CP_{it}^m$ 表示第 i ($i = 1, 2, \dots, n$) 家公司在第 t 年的第 m 个候选预测变量.

b) 构建非 ST 公司数据集 NS , $NS = \text{matrixes}(D_{it}, CP_{it}^1, \dots, CP_{it}^m)$, 其中 D_{it} 表示第 i ($i = 1, 2, \dots, n$) 家公司在第 t 年未陷入财务困境, $CP_{it}^1, \dots, CP_{it}^m$ 表示第 i ($i = 1, 2, \dots, n$) 家公司在第 t 年的第 m 个候选预测变量.

c) 实现 Bootstrap 抽样 1000 次, 得到 ST 公司数据集 S_{pn_1} 和非 ST 公司数据集 NS_{pn_2} . 其中 p ($p = 1, 2, \dots, 1000$) 表示数据集个数; n_1 表示 ST 公司数量, n_2 表示非 ST 公司数量.

杨青龙等 (2016) 提到大部分文献会按照 1:1 匹配 ST 与非 ST 公司, 由于实际中健康企业的数目要明显大于被 ST 企业的数目, 这可能导致建模分析的性能不佳. 张向荣 (2021) 因此考虑了健康企业与非健康企业数 2:1 来观察不均衡配比情况下预测方法随着样本变化的预测结果. 本文将 ST 公司与随机选择的非 ST 公司按 1:1、1:2、1:3 进行配对, 设置这三种比例匹配样本是为了保证预测的稳健性.

3) 变量筛选方法. 我们采用由 Tibshirani (1996) 提出的惩罚收缩方法 LASSO 作为变量选择技术, 能够有效地降低拟合方差. 当候选预测变量过多或相关时, LASSO 方法可以很好地处理多重共线性以最大限度地减少由于过拟合导致的数值不稳定性 (Van Rijsbergen, 1979).

LASSO 的主要思想是最小化以下公式来估计每个自变量的系数:

$$\beta = \arg \min \left\{ \sum_i^n \left[y_{i,t} - \sum_j^p \beta_j x_{i,j,t-3} + \lambda \sum_j^p |\beta_j| \right] \right\}, \quad i = 1, 2, 3, \dots, n; j = 1, 2, \dots, p, \quad (1)$$

其中 $i = 1, 2, 3, \dots, n$ 代表样本中的 n 家公司, $j = 1, 2, \dots, p$ 代表个变量, $x_{i,j,t-3}$ 表示第 i 家公司在第 $t-3$ 年的第 j 个解释变量. $y_{i,t}$ 表示第 i 家公司在第 t 年的财务状况. 当企业陷入财务困境时, $y_{i,t}$ 等于 1, 否则等于 0. 公式中第一项表示预测模型的拟合优度, 第二项是带有调整参数的惩罚函数, 用于防止过拟合. 如果 λ 逐渐增加, 预测变量的估计系数将相应缩

小,直到部分预测变量的估计系数严格等于0时,关键预测变量的选择才算完成.在本研究中, λ 由 K 折交叉验证程序确定,其中 K 设置为10.

4) 共线性检验.多重共线性是指自变量之间存在高度线性相关关系,即一个自变量可以是其他一个或几个自变量的线性组合.多次的LASSO筛选可能导致变量筛选结果存在共线性问题(Li et al., 2021).本文采用方差膨胀系数(VIF)和容忍度(TOL)检验所构建的模型是否存在共线性问题.VIF是衡量多元线性回归模型中多重共线性严重程度的一种度量,该值小于10表明共线性程度较低.容忍度是方差膨胀系数的倒数,TOL值介于0至1之间,该值小于0.1时,表示此自变量与其他自变量之间存在共线性.若VIF值太大或者TOL值太小,即VIF高于10或TOL值低于0.1,都表明模型存在严重的共线性问题(Zhang et al., 2010; Serrano-Cinca, 2019).

5) 预测模型构建.已有大量研究使用Logit模型预测破产、违约和财务困境,证明该模型在财务困境预测中的可行性(Ohlson, 1980; Pereira et al., 2016; Li et al., 2021).Logit模型的预测准确率较高,杨青龙等(2016),杨贵军等(2019)和李慧等(2019)的研究中Logit模型的财务困境预测准确率均达到了70%以上,且李慧等(2019)的实证表明lasso-logit对财务困境的预测准确率达到88.50%.Logit模型公式如下:

$$\text{Logit}(P_D)_{i,t} = \ln \left(\frac{P_{D,i,t}}{1 - P_{D,i,t}} \right) = \beta^T \mathbf{X}_{i,t-3}, \quad (2)$$

其中, $P_{D,i,t}$ 表示第*i*家公司在第*t*年陷入财务困境的概率, $\text{Logit}(P_D)_{i,t}$ 是 $P_{D,i,t}$ 的Logit变换. β^T 表示系数矩阵的转置, $\mathbf{X}_{i,t-3}$ 表示第*i*家公司*t*-3年的关键预测变量矩阵.根据上文可知,ST公司的主要特征是连续两年净利润为负,因此在公司被标记为ST的前3年是预测公司困境的关键,而且Geng et al. (2015)分别运用3年、4年和5年的时间窗口去预测公司财务困境,发现3年时间窗口的财务数据的预测效果是最准确的,这跟公司被标记ST的特征相关,同时也说明越靠近公司获得ST处理的时间财务变量预测效果也更加准确(Geng et al., 2015; Li et al., 2021).因此,我们使用ST前三年的财务数据来探索模型的性能,由此来判断提前三年数据的预测能力和公司财务状况的变化.考虑到在两年之内很有可能可以通过人为因素来使公司的财务困境逆转,所以采用提前3年的数据以避免人为干扰来进行财务预测是合理的,并且这样才使得预测具有实际意义.且已有多数研究采用提前3年的财务数据进行预测且效果较佳,其预测可靠性得到了验证(Altman, 1968; 倪志伟等, 2014; Geng et al., 2015).

6) 样本分割.根据以往关于财务困境预测的研究,本文应用样本内和样本外分析来评估本研究中的模型性能(Tian et al., 2015; Serrano-Cinca, 2019; Li et al., 2021).结合Geng et al. (2015)、Li et al. (2021)研究中的样本分割方法,我们将数据随机划分为训练样本和测试样本,采用不同的训练集与测试集中的ST公司数量比率如1:1、2:1、3:1,我们随机将训练样本中的ST公司与同等数量的非ST公司进行配对,并将整体样本的其余非ST公司放入测试样本.

7) 模型评价指标. a) McFadden's pseudo R^2 该指标可以衡量广义线性模型的拟合优度,

值越大表示预测模型更适合拟合训练数据, 模型效果越好 (Mcfadden, 1974).

$$R^2_{\text{McF}} = 1 - \frac{\ln(L_M)}{\ln(L_0)}. \quad (3)$$

在该公式中, $\ln(L_M)$ 表示被估计模型的对数似然, 而 $\ln(L_0)$ 表示没有预测变量的模型的对数似然.

b) AUC. AUC (area under curve) 为 ROC 曲线 (receiver operating characteristic curve) 下与坐标轴围成的面积, 是模型辨别率的测量值, 取值范围为 0.5~1. ROC 曲线以真阳性率 (即正确地将 ST 公司归类为 ST) 为 y 轴, 假阳性率 (即错误地将非 ST 公司归类为 ST) 为 x 轴, AUC 值越高表明分类器效果越好, 即财务困境预测模型的效果越好 (Hosmer et al., 2013).

c) 平衡准确率. 由于本文 ST 公司与非 ST 公司数量存在巨大差异, 我们使用平衡准确率来评估样本内数据和样本外数据的预测精度. 准确率是一种常见的财务困境预测衡量指标 (Altman, 1968; Ohlson, 1980; Zhang et al., 2010). 一般来说, 该指标用来衡量预测模型准确将 ST 公司分类为财务困境公司的比例和将非 ST 公司分类为财务正常公司的比例. 当处理非常不平衡的数据集时, 平均准确率很容易得出有偏分类性能的结论, 平衡准确率可以很好地克服这个问题, 该值越大表明模型效果越好 (Brédart, 2021). 平衡准确率公式如下:

$$\text{平衡准确率} = \frac{\text{真阳性率} + \text{真阴性率}}{2}, \quad (4)$$

其中, 真阳性率表示准确分类为财务困境公司的 ST 公司的比例. 真阴性率表示被准确归类为非财务困境公司的非 ST 公司的比例.

d) F_β 值. F_β 值是考虑错误分类成本的模型性能测量方法 (Van Rijsbergen, 1979). Norton and Smith (1979) 认为, 对于投资者而言, 错误地将 ST 公司分类为非 ST (假阴性) 要比将非 ST 公司分为 ST 公司 (假阳性) 的所带来的投资成本高得多, 具体来说, 将非 ST 公司错误分类的成本等于投资的机会成本, 而将 ST 公司错误归类的成本等于原始投资 (Norton and Smith, 1979). 由于困境公司和非困境公司的数量存在显著差异, 如果两种类型的错误分类的权重相同, 那么将会人为拉低假阴性率, 导致预测结果将过于乐观. 在这种情况下, 为了将错误分类引入模型评估的成本, 我们使用 F_β 值来检验各种模型的准确性. F_β 值主要通过精确度和召回率计算, 公式如下所示:

$$F_{\beta\text{-score}} = (1 + \beta^2) * \frac{\text{精确度} * \text{召回率}}{\beta^2 * \text{精确度} + \text{召回率}}, \quad (5)$$

$$\text{精确度} = \frac{\text{真阳率}}{\text{真阳率} + \text{假阳率}}, \quad (6)$$

$$\text{召回率} = \frac{\text{真阳率}}{\text{真阳率} + \text{假阴率}}, \quad (7)$$

其中, β 值设置为 35, $F_{\beta\text{-score}}$ 表示衡量错误分类 ST 公司为非 ST 公司的重要性是错误分类非 ST 公司为 ST 公司重要性的 35 倍, 即更强调保护投资者准备的原始投资金额, 降低误分类成本.

3 实证结果分析

3.1 统计分析

表3显示了本文所使用四组候选预测因子的描述性统计数据. 前两栏提供了ST公司和非ST公司的汇总统计信息, 本文还进行了Mann-Whitney U检验, 以检查ST公司和非ST公司之间是否存在显著的统计差异, 检验结果展示在(3)列(Li et al., 2021). ***表示ST公司与非ST公司在给定的显著性水平下各数据存在本质差别, 由此表明本文选取的大部分预测指标可以用来鉴别ST公司与非ST公司.

表3 变量描述性统计

	ST 公司 (1)		非 ST 公司 (2)		Mann-Whitney U Test (3)
	平均值	标准差	平均值	标准差	Z 值
A 组: 活动率					
CF/SALE	0.357	0.156	0.375	0.116	-2.003***
EBITDA/SALE	0.633	0.116	0.632	0.1	-1.256
CF/TA	0.347	0.129	0.373	0.136	-2.350***
NI/SALE	0.723	0.099	0.748	0.096	-7.922***
INV/OI	0.392	0.205	0.374	0.112	-1.45
CA/SALE	0.242	0.226	0.161	0.15	-5.261***
INV/SALE	0.231	0.261	0.139	0.156	-4.460***
REC/SALE	0.27	0.255	0.226	0.207	-1.694
SALE/TA	0.164	0.172	0.235	0.171	-7.388***
CF/OI	0.503	0.182	0.505	0.111	-0.591
SALE/WC	0.456	0.088	0.464	0.087	-5.017***
SALE/CA	0.17	0.182	0.212	0.172	-5.262***
SALE/TL	0.118	0.132	0.211	0.177	-9.259***
VA/SALE	0.269	0.19	0.297	0.134	-5.379***
NI/VA	0.746	0.072	0.731	0.098	-1.153
VA/TA	0.204	0.171	0.235	0.146	-5.822***
AC/SALE	0.13	0.207	0.058	0.112	-7.219***
B 组: 流动性比率					
CLCH/TA	0.598	0.185	0.522	0.177	-5.203***
CH/TA	0.213	0.185	0.256	0.195	-3.438***
CH/CL	0.056	0.081	0.091	0.14	-4.846***
QA/CL	0.072	0.078	0.118	0.142	-5.650***
CA/CL	0.101	0.109	0.146	0.155	-5.479***
WC/TA	0.468	0.213	0.564	0.194	-5.742***
CL/TL	0.799	0.209	0.786	0.233	-0.418
CFO/TL	0.218	0.104	0.301	0.155	-9.379***
CL/SALE	0.2	0.188	0.099	0.117	-10.044***
WC/SALE	0.418	0.183	0.442	0.13	-3.304***

表 3 (续)

	ST 公司 (1)		非 ST 公司 (2)		Mann-Whitney U Test (3)
	平均值	标准差	平均值	标准差	Z 值
CH/CA	0.322	0.232	0.346	0.221	-1.654
CA/TA	0.502	0.251	0.554	0.236	-2.790***
SA/TA	0.332	0.217	0.219	0.186	-7.204***
INV/TA	0.274	0.244	0.256	0.2	-0.162
EBIT/SL	0.021	0.11	0.036	0.119	-10.578***
QA/TA	0.306	0.213	0.373	0.222	-4.027***
RE/CL	0.376	0.14	0.47	0.125	-10.817***
TL/TA	0.439	0.239	0.35	0.206	-4.817***
C 组: 盈利能力比率比率					
OI/SALE	0.691	0.132	0.747	0.099	-9.429***
ln(SALE)	0.451	0.211	0.465	0.185	-0.094
ln(TA)	0.484	0.188	0.441	0.193	-3.454***
EBIT/TA	0.49	0.096	0.56	0.124	-9.995***
EBIT/SALE	0.68	0.105	0.696	0.099	-4.766***
PROFIT	0.91	0.281	0.93	0.252	-0.877
dSALE	0.28	0.188	0.283	0.124	-4.749***
NI/TA	0.524	0.096	0.594	0.119	-11.217***
RE/TA	0.376	0.14	0.47	0.125	-10.817***
ROIC	0.458	0.103	0.522	0.127	-9.640***
ROA	0.46	0.092	0.54	0.127	-11.279***
ROE	0.65	0.129	0.693	0.103	-9.729***
ROATA	0.554	0.128	0.617	0.114	-10.538***
ROTC	0.428	0.092	0.506	0.132	-10.179***
GPM	0.293	0.209	0.383	0.207	-6.326***
ROIS	0.679	0.135	0.737	0.102	-9.431***
IT	0.031	0.083	0.042	0.134	-2.245***
DSO	0.034	0.137	0.036	0.127	-1.812
DPO	0.262	0.239	0.183	0.158	-4.047***
PT	0.066	0.123	0.084	0.131	-4.082***
CAT	0.17	0.169	0.223	0.171	-5.892***
FAT	0.059	0.181	0.062	0.139	-5.667***
TAT	0.159	0.162	0.237	0.17	-7.927***
CFM	0.373	0.242	0.34	0.103	-1.24
OP/TA	0.475	0.114	0.572	0.128	-11.811***
CF/SF	0.299	0.062	0.313	0.1	-3.417***
NI/SF	0.022	0.11	0.036	0.118	-10.984***
EBIT/(TA-CL)	0.542	0.127	0.576	0.108	-7.271***
EBIT/IE	0.546	0.068	0.541	0.1	-1.228

表 3 (续)

	ST 公司 (1)		非 ST 公司 (2)		Mann-Whitney U Test (3)
	平均值	标准差	平均值	标准差	Z 值
SALE/INV	0.017	0.044	0.033	0.128	-3.182***
RT	0.034	0.137	0.036	0.127	-1.812
EBIT/SF	0.021	0.11	0.036	0.119	-10.578***
CHIN	0.612	0.332	0.519	0.212	-3.697***
INTWO	0.04	0.189	0.01	0.117	-2.449***
D 组: 负债率					
D/TA	0.209	0.244	0.158	0.203	-2.111***
E/TL	0.105	0.144	0.139	0.163	-5.299***
IE/SALE	0.275	0.153	0.203	0.12	-7.717***
IE/NI	0.283	0.287	0.138	0.127	-7.431***
FD/CF	0.465	0.044	0.47	0.035	-0.27
TL/TA	0.439	0.239	0.35	0.206	-4.817***
E/TA	0.507	0.259	0.608	0.216	-5.147***
TL/E	0.199	0.235	0.119	0.144	-4.903***
QR	0.08	0.081	0.13	0.154	-5.597***
LD/WC	0.07	0.121	0.055	0.132	-4.143***
D/E	0.163	0.241	0.09	0.162	-3.040***
FE/EBITDA	0.582	0.164	0.529	0.115	-6.605***
FE/NI	0.17	0.259	0.093	0.132	-5.064***
FE/TA	0.452	0.225	0.347	0.178	-5.899***
TA/TL	0.11	0.119	0.152	0.155	-5.013***
CF/TL	0.295	0.075	0.315	0.104	-2.754***
CL/CA	0.237	0.195	0.17	0.149	-5.479***
IE/EBITDA	0.582	0.164	0.529	0.115	-6.605***
IE/TA	0.452	0.225	0.347	0.178	-5.899***
OI/TA	0.475	0.114	0.572	0.128	-11.811***
CF/TL	0.295	0.075	0.315	0.104	-2.754***
企业规模					
SIZE	0.471	0.193	0.421	0.196	-3.874***

3.2 关键预测因子筛选

为了保证所选变量的稳健性, 本文采用两种方式将 158 家 ST 公司按 1:1 比例与非 ST 公司配对, 一种是 Bootstrap 有放回抽样, 按照其原理将样本组建 1000 次; 另外一种是按“同行业且资产规模相近”原则进行匹配, 将这两种样本匹配方式应用于 LASSO 变量选择技术. 为了便于说明, 将这两种基于不同样本组建方式的建模过程指代为 Bootstrap 和 Totass. 为了减小在抽样阶段产生的随机误差以提高模型的实际应用性, 本文对 Bootstrap 的交叉验

证程序重复 1000 次¹, 并采用 λ_{1se} 的标准得到最精简的筛选结果. 表 4 列 (1)、列 (2) 展示了两种 LASSO 方法的筛选结果以及每个预测变量的筛选频率, 筛选频率是 1000 次 LASSO 程序运行后预测变量被判断为有意义预测因子的次数, 末行是预测变量被选中次数的平均值, * 表示筛选次数高于平均值. Li et al. (2021) 运用 LASSO 方法进行 100 次的关键预测变量筛选, 并将各指标筛选次数划分为大于 50 次, 30~50 次, 20~29 次, 10~19 次四个层级, 将筛选次数超过一半即大于 50 次作为有显著预测作用关键因子用于构建 Logit 模型. 这样的做法拥有其合理性, 因为筛选次数越高则意味着模型正确估计该指标系数的次数越高, 模型的拟合度越高, 那么用于困境预测的准确率就会越高. 本文参考 Li et al. (2021) 的做法, 通过十折交叉验证法并以最小均方误 (MSE) 为标准确定最优调参数 λ , 其对应的 Bootstrap-LASSO 模型筛选出 23 个变量. 为保证模型的简洁性和效率性, 本文以筛选次数平均值作为临界值, 从初步筛选出的 23 个变量中继续提炼更加有效的预测变量, 将高于平均值的变量作为构建预测模型的关键因子. 结果表明, IE/Ni (利息支出/净收入)、CHIN (净收入的增量形式) 总是被这两种选择方法选出. 同时, Bootstrap 相较于 Totass 多筛选出了四个预测变量, 分别是: CA/SALE (流动资产/营业收入)、CFO/TL (经营活动产生的现金流/总负债)、RE/TA (留存收益/总资产)、TAT (总资产周转率). LASSO 进行 1000 次变量选择后易人为造成变量之间的多重共线性, 因此本文参照 Li et al. (2021)、闫达文等 (2022) 的研究, 对筛选出的关键预测因子进行多重共线性检验. 这两个模型的关键预测因子的 VIF 值都小于 10, 且 TOL 值都大于 0.1, 表明筛选出的指标之间不存在完全共线性.

在 Bootstrap 方法筛选出的六个关键预测因子之中, CFO/TL、CHIN 也曾被筛选出用来构建 O-Score 模型, 再次验证上市公司可通过提高现金周转率、减少通过借贷实现盈利的行为来降低陷入财务困境的概率; RE/TA 用于构建 Z_{China} -Score 模型, 表明可以通过提高留存收益, 用自身积累的资本来扩大公司业务, 降低财务风险 (Ohlson, 1980; Zhang et al., 2010). 本文筛选出与这两个主流困境预测模型相同的预测因子, 在回顾过去研究的同时表明了本文所用筛选方法的可靠性, 启示我们这三个变量在对中国市场的财务困境预测方面依旧能做出重要贡献.

基于 Bootstrap 样本的 LASSO 筛选结果筛选出三个未受 O-Score、 Z_{China} -Score 模型关注的预测因子, CA/SALE (流动资产/营业收入)、IE/Ni (利息支出/净收入)、TAT (总资产周转率). 前有学者苗培熙等 (2020) 发现流动资产周转率 (营业收入/流动资产) 在财务预警方面的显著作用. 尹斌 (2012) 认为流动资产周转率反映公司的营运能力, 而公司的营运能力往往影响公司的盈利能力和获利能力, 营运能力的高低对上市公司财务困境的恢复有着重大的影响. CA/SALE 是流动资产周转率的倒数形式, 田宝新和王建琼 (2017) 实证出流动资产周转率对财务困境有显著负向作用, 流动资产周转越快, 企业越不容易陷入财务困境, 可

¹Bootstrap 进行 1000 抽样是为了让数据更加接近变量总体分布, 而按“同行业资产规模相近”原则进行一次抽样是为了匹配相近的固定样本.

² λ_{1se} 是指在 λ_{min} 一个方差范围内得到最简单模型的那一个 λ 值. 十折交叉验证程序中, λ 值到达一定大小之后, 继续增加模型自变量个数即缩小 λ 值, 并不能很显著的提高模型性能. λ_{1se} 给出的是一个具备优良性能但是自变量个数最少的模型. λ_{min} 是指在所有的 λ 值中, 得到最小目标参量均值的那一个.

表 4 变量选择结果

	Bootstrap (1)	Totass (2)		Bootstrap (1)	Totass (2)
A 组: 活动率			ROE	0	0
CF/SALE	27	0	ROATA	0	0
EBITDA/SALE	132	0	ROTC	0	0
CF/TA	0	0	GPM	3	0
NI/SALE	0	0	ROIS	0	0
INV/OI	0	0	IT	0	0
CA/SALE	243*	0	DSO	0	0
INV/SALE	46	0	DPO	188	0
REC/SALE	13	0	PT	0	0
SALE/TA	0	0	CAT	20	0
CF/OI	0	0	FAT	0	0
SALE/WC	0	0	TAT	214*	0
SALE/CA	0	0	CFM	0	0
SALE/TL	26	0	OP/TA	36	0
VA/SALE	0	0	CF/SF	0	0
NI/VA	1	0	NI/SF	0	0
VA/TA	0	0	EBIT/(TA-CL)	0	0
AC/SALE	0	0	EBIT/IE	0	0
B 组: 流动性比率			SALE/INV	2	0
CLCH/TA	0	0	RT	0	0
CH/TA	0	0	EBIT/SF	0	0
CH/CL	0	0	CHIN	986*	1*
QA/CL	0	0	INTWO	0	0
CA/CL	0	0	D 组: 负债率		
WC/TA	0	0	D/TA	0	0
CL/TL	5	0	E/TL	0	0
CFO/TL	629*	0	IE/SALE	0	0
CL/SALE	58	0	IE/NI	1000*	1*
WC/SALE	0	0	FD/CF	1	0
CH/CA	0	0	TL/TA	0	0
CA/TA	0	0	E/TA	0	0
SA/TA	180	0	TL/E	0	0
INV/TA	0	0	QR	0	0
EBIT/SL	0	0	LD/WC	0	0
QA/TA	0	0	D/E	0	0
RE/CL	0	0	FE/EBITDA	0	0
CL/TA	0	0	FE/NI	23	0
C 组: 盈利能力比率			FE/TA	0	0
OI/SALE	0	0	TA/TL	0	0
ln(SALE)	0	0	CF/TL	0	0
ln(TA)	0	0	CL/CA	0	0
EBIT/TA	0	0	IE/EBITDA	0	0
EBIT/SALE	0	0	IE/TA	0	0
PROFIT	0	0	OI/TA	36	0
dSALE	0	0	CF/TD	0	0
NI/TA	0	0	企业规模		
RE/TA	509*	0	SIZE	0	0
ROIC	0	0			
ROA	0	0	平均值	199	1

以推测 CA/SALE 对财务困境预测可能存在显著预测作用. 通过理论分析提示周转率与财务费用对财务风险可能存在的解释作用, 如杨贵军等 (2019) 认为总资产周转率反映了公司的营运能力, 公司由于运作不善易增加财务风险, 这提示本文的关键预测因子 TAT (总资产周转率) 可能存在有效预测作用. 孙宝民 (2004) 认为应收账款过多企业不得不增加贷款以弥补资金不足, 从而增大企业的财务费用, 其后果是企业的大量利润转为利息支出, 企业经营可能陷入支付困境, 甚至使企业资不抵债, 最终导致破产. 吕长江等 (2004) 认为提高偿债保障、降低财务费用有利于缓解公司财务困境. 企业持有过多债务会加重企业利息支出, 这提示关键预测因子 IE/Ni (利息支出/净收入) 可能存在有效预测作用.

3.3 预测模型构建

利用 LASSO 筛选的关键预测因子构建 Logit 模型, 公式如下:

$$\begin{aligned} \text{Logit}(\text{Bootstrap} - \text{lasso}_{D,i,t}) = & \beta_0 + \beta_1(\text{CA}/\text{SALE})_{i,t-3} + \beta_2(\text{CFO}/\text{TL})_{i,t-3} + \\ & \beta_3(\text{RE}/\text{TA})_{i,t-3} + \beta_4(\text{IE}/\text{NI})_{i,t-3} + \\ & \beta_5(\text{TAT})_{i,t-3} + \beta_6(\text{CHIN})_{i,t-3}, \end{aligned} \quad (8)$$

$$\text{Logit}(\text{Totass} - \text{lasso}_{D,i,t}) = \beta_0 + \beta_1(\text{IE}/\text{NI})_{i,t-3} + \beta_2(\text{CHIN})_{i,t-3}. \quad (9)$$

在这两个模型中, 本文采用 $t-3$ 年的公司年度数据来预测其在 t 年是否被 ST 处理, 若是则因变量等于 1, 否则等于 0. 本文记录模型评价指标的样本内和样本外测试结果以及训练样本的逻辑回归系数, 以此来判断 Bootstrap 模型与 Totass 模型预测效果. 表 5 展示了模型对比结果, A 组结果表明 IE/Ni、CHIN 在这两个模型当中均具有统计学意义. IE/Ni (利息支出/净收入) 在 1‰ 的水平上显著为正说明该比率值越大企业越容易陷入财务困境, 企业的利息费用增加时, 会导致其负债成本增加, 若企业的营业收入不足以支付利息, 则其陷入财务困境的风险就越大. CHIN 表示企业在 t 时刻的净收入的增量, 它在 1‰ 的水平上显著为正说明净收入的增量越大, 公司陷入财务困境的概率就越大. 在 Bootstrap 模型当中, CA/SALE、CFO/TL 在 10% 水平上显著. 其中, CA/SALE (流动资产/营业收入) 的正系数说明流动资产占用越多越容易导致公司陷入财务困境. 高流动资产水平 (主要是持有大量存货) 会增加企业的流动资产成本, 从而降低企业的收益水平, 进而恶化公司的财务状况. CFO/TL (经营活动中的现金流/总负债) 的负系数表明公司现金流越少其偿还短期负债的能力就越弱, 那么企业陷入财务困境的概率也就越大.

A. 样本内预测结果. 将筛选出的关键预测因子纳入 Logit 模型, 并在训练集上估计变量系数, 模型评价指标结果展示在表 5 中. 显然, Bootstrap 模型的平衡准确率、AUC 值、 F_β 值、真阳率均优于 Totass 模型结果, 表明 Bootstrap 模型性能更好. 同时, pseudo R^2 值更高表明 Bootstrap 模型的数据拟合度优于后者, 即 Bootstrap 抽样方法较按“同行业资产规模相近”抽样原则能够训练出更好的预测模型.

B. 样本外预测结果. 为了测试训练集模型的实际预测效果, 现进行样本外检验. 根据样本内测试建立的这两个 logit 模型结果, 本文对 ST 公告前三年的测试集样本中的 79 家 ST 公司与 3549 家非 ST 公司进行预测, 结果展示在 C 组. 其中, Bootstrap 模型的平衡准确率、

表 5 纵向模型对比结果

	Bootstrap	Totass		Bootstrap	Totass
A 组: 参数估计			B 组: 样本内结果		
(intercept)	-1.812	-9.601	真阳率	0.6667	0.5679
CA/SALE	0.134*		真阴率	0.7769	0.8026
CFO/TL	-1.883*		平衡准确率	0.7218	0.6852
RE/TA	-1.202		AUC	0.775	0.719
IE/NI	4.670***	5.445***	F_β 值	0.666	0.568
TAT	-1.542		C 组: 样本外结果		
CHIN	3.801***	3.619**	真阳率	0.622	0.622
pseudo R^2	0.172	0.131	真阴率	0.7944	0.7881
			平衡准确率	0.7082	0.7051
			AUC	0.759	0.713
			F_β 值	0.609	0.609

注: *, **, *** 分别表示自变量在 10%, 1%, 1% 水平上显著。

AUC 值、 F_β 值、真阳率均高于或持平 Totass 模型, 这表明 Bootstrap 模型在测试集数据的预测能力依旧略好于 Totass 模型。值得一提的是, Bootstrap 模型虽较后者多筛选出四个关键预测因子, 但其模型预测能力并没有高出很多, 这是由于构建的预测样本具有随机性, 预测结果也就会具有随机误差 (这将在稳健性检验部分深入讨论)。为了降低随机误差, 本文将对训练集和测试集的样本进行 1000 次分割, 并进行 1000 次的训练与测试以验证 Bootstrap 模型预测效果的提升。

3.4 预测效果分析

本节将嵌入 Bootstrap 式 LASSO-logit 模型与主流财务困境预测模型——O-Score 模型与 Z_{China}-Score 模型进行对比 (Ohlson, 1980; Zhang et al., 2010)。O-Score 模型是采用 20 世纪 70 年代美国破产公司数据进行财务困境预测的 Logit 模型; 而 Z_{China}-Score 模型是基于 Altman 的 Z-Score 模型, 使用中国市场数据构建的。本文构建嵌入 Bootstrap 式 LASSO-logit 模型在此基础上有所更新, 希望验证该模型更加适用于中国市场。我们采用与前文相同的样本分割方法来对两个模型进行样本内和样本外测试, 并采用相同的模型评价指标。O-Score 模型与 Z_{China}-Score 模型公式如下:

$$\begin{aligned} \text{Logit}(\text{O-Score}_{D,i,t}) = & \beta_0 + \beta_1(\text{WC/TA})_{i,t-3} + \beta_2(\text{NI/TA})_{i,t-3} + \beta_3(\text{TL/TA})_{i,t-3} + \\ & \beta_4(\text{CFO/TL})_{i,t-3} + \beta_5(\text{CHIN})_{i,t-3} + \beta_6(\text{SIZE})_{i,t-3} + \\ & \beta_7(\text{INTWO})_{i,t-3} + \beta_8(\text{CL/CA})_{i,t-3}, \end{aligned} \quad (10)$$

$$\begin{aligned} \text{Logit}(\text{Z}_{\text{China-Score}}_{D,i,t}) = & \beta_0 + \beta_1(\text{E/TL})_{i,t-3} + \beta_2(\text{EBIT/TA})_{i,t-3} + \\ & \beta_3(\text{WC/TA})_{i,t-3} + \beta_4(\text{RE/TA})_{i,t-3}. \end{aligned} \quad (11)$$

表 6 记录了模型对比结果。表 7 汇总了三个模型关键预测因子的经济信息。从关键预测因子经济含义角度, LASSO-logit 模型中四个显著因子 CA/SALE、IE/NI、CFO/TL、

表 6 横向模型对比结果

	Bootstrap	O-Score	Z _{China} -Score
A 组: 参数估计			
(intercept)	-1.812	-1.089	-1.129
CA/SALE	0.134*		
IE/NI	4.670***		
TAT	-1.542		
CFO/TL	-1.883*	-2.879**	
CHIN	3.801***	4.834***	
NI/TA		-1.278	
TL/TA		-0.025	
CL/CA		0.933	
SIZE		1.634	
INTWO		-1.663	
WC/TA		0.401	-0.682
RE/TA	-1.202		-3.161***
EBIT/TA			1.197
E/TL			-0.413
pseudo R^2	0.172	0.083	0.031
B 组: 样本内结果			
真阳率	0.6667	0.7037	0.6543
真阴率	0.7769	0.6179	0.5538
平衡准确率	0.7218	0.6608	0.6041
AUC	0.775	0.711	0.629
F_β 值	0.666	0.703	0.653
C 组: 样本外结果			
真阳率	0.622	0.6585	0.6463
真阴率	0.7944	0.5947	0.634
平衡准确率	0.7082	0.6266	0.6402
AUC	0.759	0.721	0.673
F_β 值	0.609	0.632	0.623

注: *, **, *** 分别表示自变量在 10%, 1%, 1‰ 水平上显著.

CHIN 分属于活动率、负债率、流动性比率与盈利能力比率, 而 O-Score 模型的两个显著预测因子只体现在流动性与盈利能力, Z_{China}-Score 模型中的一个显著预测因子只体现盈利能力, LASSO-logit 模型能够更加全面地反映公司财务状况, 为公司治理提供更加具体的方向. 从模型评价指标预测能力角度, LASSO-logit 模型中 IE/NI、CHIN 在 1‰ 水平上显著, CA/SALE、CFO/TL 在 10% 水平上显著, 关键预测因子对模型的强贡献让评价指标样本外预测的能力更佳. 在平衡准确率上, LASSO-logit 模型高出 O-Score 约 8%, Z_{China}-Score 模型约 6%. 在 AUC 方面, LASSO 模型的分类效果也要优于另外两个模型, 这说明 LASSO-logit 模型较主流的 O-Score 模型和 Z_{China}-Score 模型的数据拟合度更高. 在显著预测因子的时空变换层面, LASSO-logit 模型较 O-Score 模型多筛选出了两个显著预测因子,

分别是 CA/SALE (流动资产/营业收入)、IE/Ni (利息支出/净收入), 它们分别衡量公司活动率和负债率. 而 Z_{China}-Score 模型中的 1% 显著预测因子 RE/TA 在我们的模型中并不显著, 说明留存收益/总资产对当下中国的财务困境预警效果有所减弱. 相应地, 我们更应重视 CA/SALE (流动资产/营业收入)、IE/Ni (利息支出/净收入) 对当下中国上市公司财务困境预警的指示作用.

表 7 关键预测变量信息

变量	经济含义	分组	Bootstrap	O-Score	Z _{China} -Score
CA/SALE	流动资产/营业收入	活动率	0.134*		
IE/Ni	利息支出/净利润	负债率	4.670***		
TAT	总资产周转率	盈利能力比率	-1.542		
CFO/TL	经营活动产生的现金流/总负债	流动性比率	-1.883*	-2.879**	
CHIN	$\frac{NI_t - NI_{t-1}}{ NI_t + NI_{t-1} }$	盈利能力比率	3.801***	4.834***	
RE/TA	留存收益/总资产	盈利能力比率	-1.202		-3.161***
WC/TA	营运资金/总资产	流动性比率		0.401	-0.682
NI/TA	净利润/资产总计	盈利能力比率		-1.278	
TL/TA	总负债/总资产	负债率		-0.025	
CL/CA	流动负债/流动资产	负债率		0.933	
SIZE	总资产账面价值的对数转换	企业规模		1.634	
INTWO	若 NI < 0, 则为 1; 若 NI > 0, 则为 0	盈利能力比率		-1.663	
EBIT/TA	息税前利润/总资产	盈利能力比率			1.197
E/TL	权益对负债比率	负债率			-0.413

注: *, **, *** 分别表示自变量在 10%, 1%, 1% 水平上显著.

3.5 稳健性检验

为了尽可能地修正统计过程中的随机性以及进一步验证模型性能, 我们对嵌入 Bootstrap 式 LASSO-logit 模型进行稳健性检验³, 具体如下:

首先, 由于财务困境研究中 ST 公司与非 ST 公司天然的样本失衡情况, 可能会导致模型在实际应用过程中的不稳定 (Zavgre, 1985; 韩立岩和李蕾, 2010; 石晓军等, 2005; 石晓军, 2006). 我们在进行数据集构建时将所有 ST 公司都用于变量选择, 而非 ST 公司的数量采用 1:1、1:2、1:3 的比例来模拟样本失衡的情况. 前文已经展示了 1:1 的结果, 由于篇幅限制, 本文不予展示后两种比例结果, 但是其结论与前者一致, 嵌入 Bootstrap 式 LASSO-logit 模型依旧优于其他模型.

其次, 采用以上三种比例筛选出关键预测因子后, 预测阶段使训练集数据 ST 公司数量: 测试集 ST 公司数量 = 1:1, 2:1, 3:1 以不断训练模型系数, 以挖掘数据噪声影响. 同时为了适应样本失衡情况, 也在预测阶段对 ST 公司数量: 非 ST 公司数量运用 1:1, 1:2, 1:3 的失衡比例. 结果均显示嵌入 Bootstrap 式 LASSO-logit 模型预测效果更好.

³ 本文所使用的数据和代码请参见科学数据银行 (ScienceDB) 期刊社区, DOI: 10.57760/sciencedb.j00214.00026 和 CSTR: 31253.11.sciencedb.j00214.00026. 若使用文中数据信息, 请注明引文和数据出处.

最后, 为了检验本文构建的 Bootstrap 式 LASSO-logit 模型具有更强的预测能力, 现将预测样本重复构建 1000 次, 以减少随机切分样本内外数据带来的误差, 并通过计算比较各模型评价指标结果的千次均值来验证模型性能, 表 8 展示了重复 1000 次预测的结果. 由样本内和样本外预测结果可知, Bootstrap 模型的平衡准确率和 AUC 值均高于另外三个模型, 且 Bootstrap 模型的五个模型评价指标样本外结果都好于 Totass 模型并具有显著优势. 另外, 大部分研究中存在的过拟合问题导致样本外真阳率较样本内结果有所损失但差距很小, 在正常范围之内.

由以上指标结果可知, 按 Bootstrap 方法来组建样本的 LASSO-logit 模型较按“同行业资产规模相近”原则构建的 LASSO-logit 模型具有更强的预测能力, 由于按“同行业资产规模相近”原则来组建样本是已有的大部分研究都会采用的方法, 所以本文的发现可以给其他学者做出新的启示. 嵌入 Bootstrap 式 LASSO-logit 模型较广泛应用的 O-Score 模型和 Z_{China}-Score 挖掘出了新的对中国上市公司财务困境预测有显著作用的关键预测因子, 可以更加准确地预测财务困境, 这为公司利益相关者的相关决策提供了重要依据.

表 8 重复 1000 次的预测结果

	Bootstrap	Totass	O-Score	Z _{China} -Score
样本内结果				
真阳率	0.6793	0.6322	0.6768	0.7211
真阴率	0.6965	0.7041	0.6111	0.4372
平衡准确率	0.6879	0.6682	0.644	0.5792
AUC	0.75	0.706	0.712	0.617
F_{β} 值	0.679	0.632	0.676	0.72
样本外结果				
真阳率	0.6622	0.6291	0.6497	0.7013
真阴率	0.7611	0.7051	0.6424	0.5151
平衡准确率	0.7116	0.6671	0.6361	0.6082
AUC	0.768	0.685	0.699	0.665
F_{β} 值	0.646	0.61	0.625	0.668

4 结论和建议

基于 2014–2019 年间沪深交易所除金融地产行业外的所有上市公司共 91 个财务指标的数据, 本文采用 Bootstrap 重抽样方法组成 1000 个数据集, 通过 LASSO 变量筛选方法, 结合 10 折交叉验证程序确定 LASSO 程序中最佳惩罚系数 λ 以得到最精简有效的预测因子, 通过筛选次数高于均值的标准来提取关键预测因子以构建提前 3 年的 Logit 预测模型. 本文采用机器学习方法预测财务困境, 过程中解决了财务困境预测研究中的数据集不平衡、关键指标选取不稳定、样本匹配随机性等问题. 实证结果表明: 一是嵌入 Bootstrap 抽样方法的 LASSO-logit 模型各方面预测结果都要比“同行业资产规模相近”(Totass) 抽样方法构建的 LASSO-logit 模型的预测效果好; 二是较主流的 O-Score 模型、Z_{China}-Score 模型, 基于本土数据 Bootstrap 抽样的 LASSO-logit 模型的预测准确率提升约 10%. 三是该模型还挖掘出了

新的对财务困境预测有显著贡献的预测因子: CA/SALE (流动资产/营业收入)、IE/NI (利息支出/净收入), 这两个预测因子值越高越容易导致公司陷入财务困境. 要维持宽松的流动资产投资策略, 企业通常由于补给原材料或不愿意因为产成品不足而失去销售来保持高水平存货. 高水平存货管理费用高、利息费用过高导致偿债成本过高, 这都增加了企业陷入财务困境的可能性. 这启示公司可以通过降低流动资产中存货比重来降低 CA/SALE 从而降低企业陷入财务困境的可能性 (杨贵军, 2019; 闫文达, 2022). 通过选择合适的融资渠道来降低利息费用等手段增强偿债能力以降低 IE/NI, 从而降低企业陷入财务困境的可能性 (吕长江等, 2004; Li et al., 2021). 此外, 本文所构建模型不仅仅是提供一次性的预测, 还可以作为一种持续监测工具. 通过定期更新预测模型并与实际表现进行比较, 可以及时发现变化趋势和风险信号, 从而实现动态的风险管理. 本文所构建的财务困境预测模型具有多重价值. 微观层面上, 有助于提高企业财务报告的透明度和质量, 促进信息披露的规范化和有效性, 帮助公司利益相关者更早识别财务困境, 引导投资者向更加具备价值性的项目集中目标, 优化决策方向, 实现投资收益升级; 该模型具备对于公司前瞻性风险预测的反应能力, 使得公司财务监管提高应对水平, 拓宽应对空间, 预防财务困境, 避免债券违约, 提高公司发展韧性和抗风险能力. 中观层面上, 该模型将显著优化企业决策的风险控制性, 推动决策结构向产业利好方向调整, 进而夯实金融产业运行基石, 降低资金链式反应对产业区间内的负面影响, 提高产业市场内部稳定度, 赋能金融市场投资热情, 拉动金融产业内部结构性升级, 同时能够防范行业间的风险感染, 引导资金在行业间有效配置, 有利于构建金融稳定保障体系, 推动信贷市场稳健高效运行. 宏观层面上, 成功的财务困境预测为困境避免及治理提供了先决条件, 将有效为政策制定提供信息参考, 提升金融政策风险应对能力, 从而稳定金融民生市场, 提高社会对产业前景信心, 对于国家经济高质量发展具备积极意义, 有利于国家双循环市场发展定位的推进, 既能够稳定并拓展国内资金市场, 对于海外国际金融市场引流、提振投资信心也具备重要意义.

进一步研究内容有: Chava and Jarrow (2004) 指出财务困境预测存在明显的行业效应, 但某些有重要研究意义的分行业如房地产业由于研究样本较小, 不适合该模型. 财务困境研究存在明显的时变特点, 即不同的研究时间区间与时间窗口会导致模型效果差别较大, 可以考虑在传统的财务困境预测因子外添加具有时变作用的因子, 或者更换具有自适应时间特征的模式.

参 考 文 献

- 陈艺云, (2022). 基于文本信息的上市中小企业财务困境预测研究 [J]. 运筹与管理, 31(4): 136–143.
- Chen Y Y, (2022). Research on Financial Distress Prediction of Listed Smes Based on Text Information[J]. Operations Research and Management Science, 31(4): 136–143.
- 陈艺云, 贺建风, 覃福东, (2018). 基于中文年报管理层讨论与分析文本特征的上市公司财务困境预测研究[J]. 预测, 37(4): 53–59.
- Chen Y Y, He J F, Qin F D, (2018). Research on Financial Distress Prediction of Listed Companies Based on the Text Discussion and Analysis of Management in Chinese Annual Report[J]. Forecasting, 37(4): 53–59.
- 董景荣, 陈军, (2010). 论经典统计财务困境预测模型的理论误区 [J]. 统计与决策, (4): 11–15.

- Dong J R, Chen J, (2010). On the Theoretical Misunderstanding of the Classical Statistical Financial Dilemma Prediction Model[J]. Statistics and Decision, (4): 11–15.
- 方匡南, 杨阳, (2018). SGL-SVM 方法研究及其在财务困境预测中的应用 [J]. 统计研究, 35(8): 104–115.
- Fang K N, Yang Y, (2018). SGL-SVM Method Study and Its Application in Financial Distress Prediction[J]. Statistical Research, 35(8): 104–115.
- 韩立岩, 李蕾, (2010). 中小上市公司财务危机判别模型研究 [J]. 数量经济技术经济研究, 27(8): 102–115.
- Han L Y, Li L, (2010). Research on the Financial Crisis Discrimination Model of Small and Medium-sized Listed Companies[J]. The Journal of Quantitative & Technical Economics, 27(8): 102–115.
- 梁琪, (2005). 企业经营管理预警: 主成分分析在 logistic 回归方法中的应用 [J]. 管理工程学报, 19(1): 100–103.
- Liang Q, (2005). Early Warning of Business Management: The Application of Principal Component Analysis in Logistic Regression Method[J]. Journal of Industrial Engineering and Engineering Management, 19(1): 100–103.
- 李慧, 温素彬, 焦然, (2019). 基于盈利质量的 DANP 变权财务预警模型 [J]. 系统工程理论与实践, 39(7): 1651–1668.
- Li H, Wen S B, Jiao R, (2019). Danp Variable Weight Financial Early-Warning Model Based on the Earnings Quality Pyramid[J]. Systems Engineering — Theory & Practice, 39(7): 1651–1668.
- 吕长江, 徐丽莉, 周琳, (2004). 上市公司财务困境与财务破产的比较分析 [J]. 经济研究, (8): 64–73.
- Lü C J, Xu L L, Zhou L, (2004). Comparative Analysis of the Financial Difficulties and Financial Insolvency of Listed Companies[J]. Economic Research, (8): 64–73.
- 苗培熙, 孙浩瑜, 顾意刚, (2020). 基于 Logistic 模型的物流业上市公司财务预警 [J]. 技术与方法, 39(2): 71–75.
- Miao P X, Sun H Y, Gu Y G, (2020). Financial Early Warning of Listed Logistics Companies Based on Logistic Model[J]. Technology and Method, 39(2): 71–75.
- 倪志伟, 薛永坚, 倪丽萍, 肖宏旺, (2014). 基于流形学习的多核 SVM 财务预警方法研究 [J]. 系统工程理论与实践, 34(10): 2666–2674.
- Ni Z W, Xue Y J, Ni L P, Xiao H W, (2014). Research on the Financial Early Warning Method of Multicore SVM Based on Manifold Learning[J]. Systems Engineering — Theory & Practice, 34(10): 2666–2674.
- 石晓军, (2006). Logistic 违约率模型最优样本配比与分界点的模拟分析 [J]. 数理统计与管理, 25(6): 675–682.
- Shi X J, (2006). Logistic Simulation Analysis of the Optimal Sample Ratio and Cut-off Point of the Default Rate Model[J]. Journal of Applied Statistics and Management, 25(6): 675–682.
- 石晓军, 任若恩, 肖远文, (2007). 边界 Logistic 违约率模型及实证研究 [J]. 管理科学学报, 10(3): 44–51.
- Shi X J, Ren R E, Xiao Y W, (2007). Boundary Logistic Default Rate Model and an Empirical Study[J]. Journal of Management Sciences in China, 10(3): 44–51.
- 石晓军, 肖远文, 任若恩. (2005). Logistic 违约率模型的最优样本配比与分界点研究 [J]. 财经研究, 31(9): 38–48.
- Shi X J, Xiao Y W, Ren R E, (2005). Logistic Study on the Optimal Sample Ratio and Cut-Off Point of the Default Rate Model[J]. Journal of Finance and Economics, 31(9): 38–48.
- 孙宝民. (2004). 应收账款风险及其对策刍议 [J]. 中国农垦经济, (10): 48–49.
- Sun B M, (2004). Account Receivable Risk and Its Countermeasures[J]. China's Agricultural Reclamation Economy, (10): 48–49.

- 孙灏, 朱晓谦, 李建平, (2022). 考虑财务报告中文本风险信息的财务困境预测 [J]. 系统管理学报, 31(6): 1204–1215.
- Sun H, Zhu X Q, Li J P, (2022). Financial Distress Prediction Considering Text Risk Information in Financial Reports[J]. Journal of Systems & Management, 31(6): 1204–1215.
- 田宝新, 王建琼, (2017). 基于财务与非财务要素的上市公司财务困境预警实证研究 [J]. 金融评论, 9(5): 103–115.
- Tian B X, Wang J Q, (2017). Empirical Study on Financial Distress Warning of Listed Companies Based on Financial and Non-Financial Factors[J]. Chinese Review of Financial Studies: 9(5): 103–115.
- 王小燕, 袁欣, (2018). 基于惩罚组变量选择的 COX 财务危机预警模型 [J]. 系统工程, 36(3): 113–121.
- Wang X Y, Yuan X, (2018). A Cox Financial Crisis Early Warning Model Based on the Selection of Penalty Group Variables[J]. Systems Engineering, 36(3): 113–121.
- 吴世农, 卢贤义, (2001). 我国上市公司财务困境的预测模型研究 [J]. 经济研究, (6): 46–55.
- Wu S N, Lu X Y, (2001). Research on the Prediction Model of the Financial Difficulties of Listed Companies in China[J]. Economic Research Journal, (6): 46–55.
- 谢纪刚, 裘正定, 韩彦俊, 莫莉, (2005). 上市公司财务困境预测模型比较研究 [J]. 系统工程理论与实践, 25(9): 29–35.
- Xie J G, Qiu Z D, Han Y J, Mo L, (2005). Comparative Study on the Prediction Model of Financial Difficulties of Listed Companies[J]. Systems Engineering — Theory & Practice, 25(9): 29–35.
- 闫达文, 李存, 迟国泰, (2024). 基于混频数据的中国上市公司财务困境动态预测研究 [J/OL]. 中国管理科学: 1–14. Doi: 10.16381/i.cnki.issn1003-207x.2021.1570.
- Yan D W, Li C, Chi G T, (2024). Research on Dynamic Prediction of Financial Difficulties of Chinese Listed Companies Based on Hybrid Data[J/OL]. Chinese Journal of Management Science: 1–14. Doi: 10.16381/i.cnki.issn1003-207x.2021.1570.
- 杨贵军, 周亚梦, 孙玲莉, (2019). 基于 Benford-Logistic 模型的企业财务风险预警方法 [J]. 数量经济技术经济研究, 36(10): 149–165.
- Yang G J, Zhou Y M, Sun L L, (2019). Enterprise Financial Risk Early Warning Method Based on Benford-Logistic Model[J]. The Journal of Quantitative & Technical Economics, 36(10): 149–165.
- 杨青龙, 田晓春, 胡佩媛, (2016). 基于 LASSO 方法的企业财务困境预测 [J]. 统计与决策, (23): 170–173.
- Yang Q L, Tian X C, Hu P Y, (2016). Forecast of Enterprise Financial Distress Based on the LASSO Method[J]. Statistics & Decision, (23): 170–173.
- 叶佳, 陈军, (2010). 经典统计方法在财务困境预测中的应用综述 [J]. 经济研究导刊, (14): 115–116.
- Ye J, Chen J, (2010). Review of the Application of Classical Statistical Methods Infinancial Distress Prediction[J]. Economic Research Guide, (14): 115–116.
- 尹斌, (2012). 我国上市公司财务困境恢复的影响因素研究 [J]. 会计之友, (17): 74–77.
- Yin B, (2012). Research on Influencing Factors of Financial Distress Recovery of Chinese Listed Companies[J]. Friends of Accounting, (17): 74–77.
- 张向荣, (2021). 基于子空间多核学习的企业财务困境预测方法 [J]. 运筹与管理, 30(1): 184–191.
- Zhang X R, (2021). Enterprise Financial Distress Prediction Method Based on Multikernel Learning of Subspace[J]. Operations Research and Management Science, 30(1): 184–191.
- Altman E I, (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy[J]. The Journal of Finance, 23(4): 589–609.
- Beaver W H, (1966). Financial Ratios as Predictors of Failure, Empirical Research in Accounting: Selected Studies[J]. Journal of Accounting Research, 4: 71–111.

- Brédart X, Séverin E, Veganzones D, (2021). Human Resources and Corporate Failure Prediction Modeling: Evidence from Belgium[J]. *Journal of Forecasting*, 40(7): 1325–1341.
- Calomiris C W, Fisman R, Wang Y X, (2010). Profiting from Government Stakes in a Command Economy: Evidence from Chinese Asset Sales[J]. *Journal of Financial Economics*, 96(3): 399–412.
- Chava S, Jarrow R A, (2004). Bankruptcy Prediction with Industry Effects[J]. *Review of Finance*, 8(4): 537–569.
- Chen M Y, (2011). Bankruptcy Prediction in Firms with Statistical and Intelligent Techniques and a Comparison of Evolutionary Computation Approaches[J]. *Computers & Mathematics With Applications*, 62(12): 4514–4524.
- Efron B, (1979). Bootstrap Methods: Another Look at the Jackknife[J]. *The Annals of Statistics*, 7(1): 1–26.
- Efron B, Tibshirani R, (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy[J]. *Statistical Science*, 1(1): 54–75.
- Geng R B, Bose I, Chen X, (2015). Prediction of Financial Distress: An Empirical Study of Listed Chinese Companies Using Data Mining[J]. *European Journal of Operational Research*, 241(1): 236–247.
- Hosmer D W, Lemeshow S, Sturdivant R X, (2013). *Applied Logistic Regression*[M]. 3rd ed. New Jersey: John Wiley & Sons.
- Khoja L, Chipulu M, Jayasekera R, (2014). Analysing Corporate Insolvency in the Gulf Cooperation Council Using Logistic Regression and Multidimensional Scaling[J]. *Review of Quantitative Finance and Accounting*, 46(3): 483–518.
- Kliestik T, Valaskova K, Lazaroiu G, Kovacova M, Vrbka J, (2020). Remaining Financially Healthy and Competitive: the Role of Financial Predictors[J]. *Journal of Competitiveness*, 12(1): 74–92.
- Li C Y, Lou C X, Luo D, Xing K, (2021). Chinese Corporate Distress Prediction Using LASSO: The Role of Earnings Management[J]. *International Review of Financial Analysis*, 76: 101776.
- McFadden D, (1974). Conditional logit analysis of qualitative choice behavior[M]. *Frontiers in Econometrics*. New York: Academic Press.
- Norton C L, Smith R E, (1979). A Comparison of General Price Level and Historical Cost Financial Statements in the Prediction of Bankruptcy[J]. *The Accounting Review*, 54(1): 72–87.
- Ohlson J A, (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy[J]. *Journal of Accounting Research*, 18(1): 109–131.
- Pereira J M, Basto M, Silva A F D, (2016). The Logistic LASSO and Ridge Regression in Predicting Corporate Failure[J]. *Procedia Economics and Finance*, 39: 634–641. Doi: 10.1016/s2212-5671(16)30310-0.
- Sauerbrei W, Buchholz A, Boulesteix A L, Binder H, (2015). An Stability Issues in Deriving Multivariable Regression Models: On Stability Issues in Deriving Multivariable Regression Models[J]. *Biometrical Journal*, 57(4): 531–555.
- Serrano-Cinca C, Gutiérrez-Nieto B, Bernate-Valbuena M, (2019). The Use of Accounting Anomalies Indicators to Predict Business Failure[J]. *European Management Journal*, 37(3): 353–375.
- Shao J, (1996). Bootstrap Model Selection[J]. *Journal of the American Statistical Association*, 91(434): 655–665.
- Shumway T, (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model[J]. *The Journal of Business*, 74(1): 101–124.
- Tian S, Yu Y, Guo H, (2015). Variable Selection and Corporate Bankruptcy Forecasts[J]. *Journal of Banking & Finance*, 52: 89–100.

- Tibshirani R, (1996). Regression Shrinkage and Selection Via the LASSO[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1): 267–288.
- Van Rijsbergen C J, (1979). *Information Retrieval*[M]. 2nd ed. London: Butterworths.
- Wang Z, Li H, (2007). Financial Distress Prediction of Chinese Listed Companies: A Rough Set Methodology[J]. *Chinese Management Studies*, 1(2): 93–110.
- Zavgre C V, (1985). Assessing the Vulnerability to Failure of American Firms: A Logistic Analysis[J]. *Journal of Business Finance & Accounting*, 12(1): 19–45.
- Zhang L, Altman E I, Yen J, (2010). Corporate Financial Distress Diagnosis Model and Application in Credit Rating for Listing Firms in China[J]. *Frontiers of Computer Science in China*, 4(2): 220–236.
- Zmijewski M E, (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models[J]. *Journal of Accounting Research*, 22: 59–82.