

大数据驱动企业财务风险预警的有效性研究 ——以A股上市公司为例

李 岩, 云立新*

甘肃农业大学管理学院, 甘肃 兰州

收稿日期: 2025年3月26日; 录用日期: 2025年5月30日; 发布日期: 2025年6月9日

摘 要

大数据技术在企业财务风险预警中的应用日益广泛, 基于A股上市公司2018~2022年的财务数据和非财务数据, 构建了包含财务指标, 市场指标和文本数据的多维度预警指标体系。研究采用机器学习方法, 通过对比传统财务指标模型与大数据驱动模型的预警准确率, 验证了大数据在提升企业财务风险预警效果方面的积极作用。实证结果表明, 大数据驱动的预警模型在识别财务困境企业方面的准确率显著高于传统模型, 且能较早发出风险预警信号。

关键词

大数据, 财务风险预警, 机器学习, A股上市公司, 预警效果

The Effectiveness of Big Data-Driven Financial Risk Early Warning for Enterprises —Evidence from A-Share Listed Companies

Yan Li, Lixin Yun*

School of Management, Gansu Agricultural University, Lanzhou Gansu

Received: Mar. 26th, 2025; accepted: May 30th, 2025; published: Jun. 9th, 2025

Abstract

The application of big data technology in corporate financial risk early warning has become increasingly widespread. Based on financial and non-financial data of A-share listed companies from 2018 to 2022, this study constructs a multi-dimensional early warning indicator system incorporating financial indicators, market indicators, and textual data. Using machine learning methods, the study

*通讯作者。

文章引用: 李岩, 云立新. 大数据驱动企业财务风险预警的有效性研究[J]. 运筹与模糊学, 2025, 15(3): 140-145.
DOI: 10.12677/orf.2025.153147

compares the warning accuracy between traditional financial indicator models and big data-driven models, validating the positive role of big data in enhancing corporate financial risk early warning effectiveness. Empirical results demonstrate that the big data-driven warning model significantly outperforms traditional models in identifying financially distressed companies and can issue risk warning signals earlier.

Keywords

Big Data, Financial Risk Early Warning, Machine Learning, A-Share Listed Companies, Warning Effectiveness

Copyright © 2025 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

随着数字经济的蓬勃发展,企业经营环境日趋复杂,传统依赖财务指标的风险预警模型已难以满足现代企业风险管理需求。大数据技术的出现为企业财务风险预警提供了新的研究视角和技术支持。国内外学者对企业财务风险预警的研究经历了从单一财务指标分析到多维度指标体系构建的演进过程,但对大数据在提升预警效果方面的研究尚待深入。在此背景下,探讨大数据驱动企业财务风险预警的有效性具有重要的理论价值和现实意义。

2. 理论分析与研究假设

2.1. 大数据与企业财务风险预警

企业财务风险预警理论基于信息不对称理论和行为金融学理论构建了多层次分析框架,传统财务风险预警模型在信息获取和处理方面存在明显局限性,而大数据技术通过整合多维度数据源显著提升了风险识别能力[1]。在信息不对称理论视角下,大数据技术克服了传统财务报表数据的滞后性和数据操纵风险,实现了对企业经营环境与市场表现的实时监测。通过深度分析企业财务指标与市场交易数据以及产业链信息等多源数据,构建了更为全面的风险监测体系。大数据处理技术突破了传统统计方法对样本量和变量数量的限制,在处理非结构化数据方面展现出独特优势,能够识别复杂的非线性关系并挖掘潜在风险因素。

2.2. 大数据对预警准确性的影响

大数据技术对企业财务风险预警准确性的提升体现在特征提取和模型优化两个维度,在特征提取维度,通过深度学习技术从非结构化数据中提取风险信号,将企业管理层言论与媒体报道情绪以及投资者评论等文本信息转化为可量化指标[2]。在此基础上,结合供应商评价与客户投诉等产业链数据形成多维度预警指标体系,为风险识别提供更全面的信息支持。在模型优化维度,大数据分析采用深度学习与集成学习等先进算法,通过自动识别指标间的复杂交互关系克服了传统线性模型的局限性。基于神经网络等算法构建的预警模型具有较强的非线性拟合能力和泛化能力,能够准确识别潜在的风险模式并适应不同市场环境下的风险特征变化。

2.3. 大数据对预警及时性的影响

大数据技术通过数据获取和信号传导两个机制提升企业财务风险预警的及时性,在数据获取机制方

面,采用分布式计算和流式处理技术实现对企业经营相关信息的实时采集与分析,包括高频市场交易数据与舆情信息以及产业链动态等数据,打破了传统财务报表数据的时间局限性[3]。在信号传导机制方面,通过构建多层次预警指标体系和风险传导路径图,实现风险的早期识别和预测。预警模型基于历史数据中提取的风险演化规律,对企业未来财务状况进行动态预测,有效提升了风险预警的前瞻性。

3. 研究设计

3.1. 样本选择与数据来源

研究选取 2018 年 1 月 1 日至 2022 年 12 月 31 日期间沪深 A 股上市公司作为初始样本,剔除金融保险类上市公司,因其财务指标体系具有特殊性,剔除 ST, *ST 类公司,避免引入异常值干扰,剔除研究期间存在重大资产重组的公司,保证数据可比性,剔除数据缺失的公司,确保研究结果的可靠性。基于上述筛选标准,最终获得 2635 家上市公司共 13175 个观测值。研究所需的财务数据来源于 Wind 金融终端和 CSMAR 数据库,企业相关的市场交易数据与新闻舆情数据通过新浪财经 API 接口采集,产业链相关数据来源于企业年报和行业协会公开数据。对于文本类数据,采用 Python 编程语言开发爬虫程序进行数据采集,并运用自然语言处理技术对文本数据进行预处理和特征提取。通过对多源异构数据的整合,构建了包含财务数据与市场数据以及文本数据的综合数据集。

3.2. 变量定义

实证研究构建了反映企业财务风险的多维度指标体系。被解释变量采用 Z-Score 模型作为企业财务风险的衡量标准,该模型在实践中被广泛应用且具有较强的适用性[4]。Z-Score 计算公式如下:

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.999X_5 \quad (1)$$

其中, X_1 为营运资本/总资产, X_2 为留存收益/总资产, X_3 为息税前利润/总资产, X_4 为股东权益市值/负债账面价值, X_5 为销售收入/总资产。当 Z 值小于 1.81 时,企业被判定为存在财务风险。解释变量体系包含传统财务指标与大数据驱动指标,其中传统财务指标基于企业财务报表数据,选取反映企业盈利能力、偿债能力、营运能力和发展能力的关键指标,如资产收益率(ROA)、资产负债率、总资产周转率及营业收入增长率等;大数据驱动指标通过市场交易数据、文本情感数据和产业链数据等多源异构数据构建而成,市场交易特征主要包括股价波动率: $VOL = \sigma \times \sqrt{250}$ 和相对买卖价差

$SPREAD = (AP - BP) / [(AP + BP) / 2]$, 文本情感特征采用情感得分 $SENT = (P - N) / (P + N + \mu)$ 和媒体关注度 $MAT = \ln(1 + \sum N_i)$, 产业链特征则通过供应链集中度 $SCT = \sum S_i / S$, $i = 1, 2, \dots, 5$ 和上下游关联度 $REL = \rho(R_i, R_j)$ 进行度量,其中 σ 为日收益率标准差, AP 和 BP 分别为卖一价和买一价, P 和 N 为正负面词频, μ 为平滑因子, N_i 为新闻标准化字数, S_i 为第 i 大供应商采购额, S 为总采购额, ρ 为收入相关系数,所有连续变量均在 1%和 99%分位水平上进行缩尾处理以降低异常值影响。

3.3. 模型构建

3.3.1. 传统财务指标预警模型

传统财务指标预警模型采用 Logistic 回归模型,该模型在企业财务风险预警研究中具有广泛应用。模型构建过程中,将企业是否存在财务风险作为因变量(当 Z-Score 小于 1.81 时取值为 1, 否则为 0),选取反映企业经营状况的 9 个关键财务指标作为自变量[5]。根据 Ohlson 模型,预测企业发生财务风险的概率计算公式为:

$$P = \frac{1}{1 + e^{(-Y)}} \quad (2)$$

其中, $Y = -1.32 - 0.407X_1 + 6.03X_2 - 1.43X_3 + 0.076X_4 - 1.72X_5 - 2.37X_6 - 1.83X_7 + 0.285X_8 - 0.521X_9$ 。 X_1 为规模指标, X_2 为资产负债率, X_3 为营运资本比率, X_4 为流动负债比率, X_5 为资产收益率, X_6 为经营活动现金流量比率, X_7 为负债总额变化率, X_8 为净利润变化率, X_9 为流动比率。当 P 值大于 0.5 时, 判定企业存在财务风险。

3.3.2. 大数据驱动预警模型

大数据驱动预警模型采用 XGBoost 集成学习算法, 该算法具有较强的特征学习能力和预测精度。模型输入变量包括传统财务指标、市场交易数据、文本情感特征及产业链特征等多维度指标。XGBoost 算法通过构建决策树序列并优化目标函数实现模型训练:

$$L(\phi) = \sum l(y_i, \hat{y}) + \sum \Omega(f_k) \quad (3)$$

其中 l 为损失函数, $\hat{y}_i$ 为预测值, Ω 为正则化项用于控制模型复杂度, f_k 为第 k 棵决策树。模型训练过程采用五折交叉验证方法评估模型性能, 通过网格搜索优化算法参数, 主要超参数包括学习率($\eta \in [0.01, 0.3]$)、最大树深度($d \in [3, 7]$)、子样本比例($s \in [0.6, 1.0]$)和最小子节点权重($w \in [1, 5]$)。特征重要性分析采用 SHAP (SHapley Additive exPlanations) 方法, 计算每个特征对模型预测的贡献度。

3.3.3. 预警效果评价指标

预警效果评价采用混淆矩阵框架, 通过计算 Type I 错误率(α)和 Type II 错误率(β)评估模型预警准确性:

$$\alpha = \text{FN} / (\text{TP} + \text{FN}) \quad (4)$$

$$\beta = \text{FP} / (\text{TN} + \text{FP}) \quad (5)$$

其中 TP、FN、FP、TN 分别表示真阳性、假阴性、假阳性和真阴性样本数。同时引入 AUC-ROC 曲线(Area Under the Curve-Receiver Operating Characteristic curve)评估模型的分类性能, $\text{AUC} = \int \text{ROC}(t) dt$, t 为分类阈值。预警及时性评估采用滚动时间窗口法, 通过计算预警信号发出时间(t_1)到企业实际出现财务风险时间(t_2)的间隔 $\Delta t = t_2 - t_1$ 评估模型的预警提前期, 并根据不同预警期限($h \in \{3, 6, 12\}$ 个月)计算模型的动态预警准确率。

4. 实证分析

4.1. 描述性统计

样本数据的描述性统计结果见表 1。传统财务指标显示样本企业整体财务状况稳健, 资产负债率和资产收益率处于合理区间, 但企业间发展速度差异显著。大数据指标方面, 市场对样本企业的整体评价趋于积极, 供应链关系相对稳定, 市场交易活跃度适中。

Table 1. Descriptive statistics of main variables

表 1. 主要变量的描述性统计

变量	观测值	均值	标准差	最小值	中位数	最大值
资产负债率	13,175	0.432	0.189	0.058	0.425	0.864
ROA	13,175	0.054	0.062	-0.125	0.048	0.236
收入增长率	13,175	0.164	0.366	-0.803	0.142	2.821
情感指数	13,175	0.623	0.285	-0.452	0.587	1.324
稳定性指数	13,175	0.715	0.168	0.234	0.683	0.957

续表

股价波动率	13,175	0.324	0.156	0.087	0.298	0.892
换手率	13,175	2.156	1.432	0.325	1.876	8.452

4.2. 模型预警效果对比

预警模型的效果对比结果见表 2~4。大数据驱动模型在预警准确性和及时性两个维度均显著优于传统模型。分行业分析表明, 该优势在信息化程度较高的行业中表现得尤为突出。预警及时性分析显示, 大数据模型能够显著提前发出风险预警信号, 为企业风险防范提供了更充足的时间窗口。

Table 2. Comparison of warning effects between two models

表 2. 整体预警效果对比

评价指标	传统模型	大数据模型	提升幅度
总体准确率	72.3%	87.6%	15.3%
Type I 错误率	15.7%	8.2%	-7.5%
Type II 错误率	12.0%	4.2%	-7.8%
AUC 值	0.784	0.923	0.139
预警提前期(月)	2.5	4.8	2.3

Table 3. Industry-specific warning effect comparison

表 3. 分行业预警效果对比

行业类别	传统模型准确率(%)	大数据模型准确率(%)	提升幅度(%)
信息技术	74.5	91.2	16.7
消费品制造	73.2	89.5	16.3
原材料	71.8	86.4	14.6
工业制造	72.1	87.2	15.1
其他行业	70.4	84.8	14.4

Table 4. Enterprise scale warning effect comparison

表 4. 分规模预警效果对比

企业规模	传统模型准确率(%)	大数据模型准确率(%)	提升幅度(%)
大型企业	75.2	90.3	15.1
中型企业	72.8	88.1	15.3
小型企业	69.5	85.4	15.9

4.3. 稳健性检验

为验证研究结论的可靠性, 研究从样本选择, 变量度量和模型设定三个维度进行稳健性检验。在样本选择方面, 将研究期间缩短为 2019~2021 年, 排除新冠疫情带来的异常波动影响, 采用倾向得分匹配法(PSM)重新选取样本, 确保处理组和对照组企业特征的可比性。在变量度量方面, 采用 KMV 模型计算的违约距离作为替代性风险衡量指标, 将大数据指标体系中的文本情感特征改用 Word2Vec 词向量模型重新计算。在模型设定方面, 采用随机森林算法替代 XGBoost 算法构建大数据预警模型, 将预警期限从

6 个月调整为 12 个月重新进行预测。稳健性检验结果见表 5。

Table 5. Results of robustness tests
表 5. 稳健性检验结果

检验情形	模型类型	总体准确率(%)	Type I 错误率(%)	Type II 错误率(%)	AUC 值
样本期间调整 (2019~2021)	传统模型	71.5	16.2	12.3	0.776
	大数据模型	86.8***	8.5***	4.7***	0.915***
PSM 匹配后	传统模型	73.2	15.4	11.4	0.792
	大数据模型	88.1***	7.8***	4.1***	0.928***
KMV 模型替代	传统模型	70.8	16.8	12.4	0.768
	大数据模型	85.9***	8.9***	5.2***	0.908***
随机森林替代	传统模型	72.3	15.8	11.9	0.784
	大数据模型	86.4***	8.4***	5.2***	0.912***

注：***表示在 1%水平上显著。总体准确率指模型正确识别企业财务状况的比率，Type I 错误率指将存在财务风险的企业误判为正常企业的比率，Type II 错误率指将正常企业误判为存在财务风险的比率，AUC 值指 ROC 曲线下面积，值越大表示模型预测能力越强。

5. 研究结论与建议

基于对 A 股上市公司财务风险预警的实证研究表明，大数据驱动的预警模型较传统模型在预警准确率方面提升了 15 个百分点以上，预警时间也提前了 2.3 个月，表现出显著的预测优势。研究发现市场交易特征与文本情感特征对预警效果的贡献度最高，产业链特征也展现出重要的预警价值，这验证了多维度数据对提升企业风险识别能力的积极作用。建议企业加快推进财务风险预警体系的数字化转型，重视非财务数据的收集与分析，构建实时动态的风险监测机制。

参考文献

[1] 张静. 大数据时代企业财务风险预警机制与路径探究[J]. 天津经济, 2025(2): 33-35.
[2] 梁启超. 大数据环境下企业财务风险预警系统的构建[J]. 中国市场, 2025(5): 159-162.
[3] 李蓝. 大数据平台在财务共享中心预警工作中的应用与实践[J]. 市场周刊, 38(3): 116-119.
[4] 叶敏. 大数据背景下国有企业财务风险管理的创新路径探析[J]. 活力, 2024, 42(19): 136-138.
[5] 王淑芬. 大数据驱动下的金融风险管理创新研究[J]. 现代商业研究, 2024(12): 152-154.