

系统性风险与企业股价崩盘危机预警

——基于机器学习的新视角

赵丽丽, 陈瑾瑜

扬州大学商学院, 江苏 扬州

收稿日期: 2024年10月23日; 录用日期: 2024年11月14日; 发布日期: 2025年1月10日

摘要

建设“金融强国”的过程中, 系统性风险加剧, 企业发展环境险峻, 企业股价崩盘已成为动摇一国金融市场稳定的重要影响因素。本文结合Logit回归模型与LightGBM模型、XGboost模型等前沿的机器学习方法, 探究系统性风险指标在预测我国企业股价崩盘危机中的作用。结果显示, 系统性风险在预测各行业企业的股价崩盘危机时具有显著的效果, 同时, 总资产收益率较低的中游、下游及服务与支撑企业更易受到系统性风险的影响, 进而面临股价崩盘的危机, 总资产增长率较低的下游企业更易在市场波动中出现股价崩盘危机。此外, 本文进一步发现基于主成分分析构建的系统性风险指标与LightGBM模型或XGboost模型相结合可以取得较好的预测效果。在此基础上, 本文强调微观企业股价崩盘预警机制中宏观审慎视角的重要性, 建议差异化监管各行业企业, 引入机器学习方法, 为维护金融系统稳定提供一定参考。

关键词

股价崩盘, 系统性风险, 机器学习

Systemic Risk and Crisis Warning of Stock Price Collapse in Enterprises

—A New Perspective Based on Machine Learning

Lili Zhao, Jinyu Chen

Business School, Yangzhou University, Yangzhou Jiangsu

Received: Oct. 23rd, 2024; accepted: Nov. 14th, 2024; published: Jan. 10th, 2025

Abstract

In the process of building a “financial powerhouse”, the collapse of corporate stock prices has become

an important influencing factor that shakes the stability of a country's financial market. This article combines Logit regression model with cutting-edge machine learning methods such as LightGBM model and XGboost model to explore the role of systematic risk indicators in predicting the stock price collapse crisis of Chinese enterprises. The results show that systemic risk has a significant effect on predicting stock price collapse crises in various industries. At the same time, midstream, downstream, and service and support enterprises with lower total asset returns are more susceptible to the impact of systemic risk and thus face stock price collapse crises. Downstream enterprises with lower total asset growth rates are more likely to experience stock price collapse crises in market fluctuations. In addition, this article further found that the systematic risk indicators constructed based on principal component analysis combined with LightGBM model or XGboost model can achieve good predictive performance. On this basis, this article emphasizes the importance of macro prudential perspective in the micro enterprise stock price collapse warning mechanism, suggests differentiated regulation of enterprises in various industries, introduces machine learning methods, and provides certain reference for maintaining financial system stability.

Keywords

Stock Price Collapse, Systemic Risk, Machine Learning

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

2023年10月30日的中央金融工作会议全面贯彻党的二十大精神, 强调要“全面加强金融监管, 有效防范化解金融风险”, “对风险早识别、早预警、早暴露、早处置”, 为建设“金融强国”保驾护航。而在我国致力于推动金融高质量发展的背景下, 个股股价暴跌给维护金融市场稳定带来巨大压力[1]。自2008年金融危机引致的全球股市动荡事件起, 股市崩盘事件时有发生, 如2015年中国“股灾”、2020年全球范围内的股市剧烈波动。我国上市公司股价崩盘风险正处于阶段性上升中, 这不仅影响各公司自身的发展, 还会因公司间的经济活动而产生外溢效应[2], 冲击金融市场的稳定。在此情况下, 通过结合前沿的机器学习方法, 改进我国企业股价崩盘风险的预测模型, 从而实现对企业股价崩盘风险的有效预警, 具有重要现实意义。

如今, 学术界对股价崩盘问题已有较多方面的讨论, 部分研究专注于如何有效衡量股价崩盘风险。率先基于个股日收益, 以计算出的负收益偏态系数(NCSKEW)和上下波动比例(DUVOL)作为衡量股价崩盘风险的指标[3]。后续研究进一步发展, 以个股当年是否至少经历了一次股价崩盘周期构造了虚拟变量crash[4], 并基于此构建了市场崩盘模型[5]。有学者以股票未来一年内的对数收益率为基础, 将股价崩盘(Crash)的界限设定为-70%[6]。与此同时, 随着对股价崩盘风险进行有效预警的重要性逐渐显现, 部分学者已基于相关指标构建方法进行探究[7][8]。

现有研究已从多角度考虑信息集, 对企业股价崩盘进行预测, 主要影响因素包括公司财务特征[9]、外部市场表现[10]、环境与制度[11]、投资者偏好[12]等。随着相关研究的拓展, 有学者提出风险指标也能成为企业危机预测的先行指标[13]。究其原因, 首先, 居民消费会因系统性风险的冲击而受到抑制[14], 从需求端对部分企业的盈利能力造成负面影响; 其次, 宏观经济形势波动下, 金融机构往往会出现惜贷行为[15], 影响企业资金流动性与偿债能力; 此外, 经济下行时政府无法营造良好的商业环境, 会影响企业的发展[16]。企业的盈利能力、偿债能力及发展环境受到负面影响, 必然导致投资者产生悲观预期, 进

而使股价下跌。然而，目前将企业股价崩盘预测与系统性风险相结合的文献仍旧较少，本文引入系统性风险指标，为预测企业股价崩盘提供了新思路。

除了研究预测特征变量，还有部分学者致力于从模型扩展的角度优化股价崩盘危机的预测效果。以往的研究主要基于 Logit 模型等参数方法进行探究。然而，随着现代计量经济学方法的进步，将前沿的机器学习技术与其结合以预测企业的股价崩盘危机，已成为这一领域的新兴探索方向。其中，应用 LightGBM 模型，多层感知器回归器(MLPRegressor)、决策树、XGboost 等机器学习方法，探究股价崩盘的成果较多 [17][18]。为进一步探究不同模型的预测效果差异，本文基于传统的 Logit 模型以及前沿的机器学习方法分别展开测试，将企业股价崩盘预测的效果进行对比，为有效预警企业股价崩盘的模型选择提出相关建议。

有鉴于此，本文设定会计 - 经济 - 风险模型进行股价崩盘危机预测，并结合 Logit 回归模型及前沿的机器学习方法，深入考察各类模型对企业股价崩盘危机的预测能力。首先，本文采用传统的 Logit 回归模型，探究系统性风险与企业股价崩盘概率间的线性关系，并依据上市企业在产业链中位置进行划分，具体分析各行业企业的系统性风险在模型中的预测效果及系统性风险的影响机制。其次，本文将 LightGBM 模型与部分依赖分析方法相结合，进一步探究系统性风险与企业股价崩盘概率之间的非线性关系。此外，本文采用样本外测试和时间外测试两种方式对各模型的预测效果进行对比，探究系统性风险指标与预测模型的高效组合。基于上述研究，本文提出完善我国上市公司监管和企业股价崩盘有效预警的相关建议。

2. 模型设定与方法说明

2.1. 会计 - 市场 - 风险模型

股价受多种因素影响，目前国内外学者尚未建立起一致的预测指标体系。考虑到机器学习在处理复杂多变的指标时具有一定优势，本文借鉴已有的企业财务危机预警模型，在综合考虑了企业财务信息、市场表现以及风险水平的基础上，构建了会计 - 市场 - 风险模型进行股价崩盘风险预测。本文以 VaR、 Δ CoVaR 及 MES 这三个风险指标为基础，分别通过主成分分析、因子分析及标准化均值方法进行组合，构成 PCA、FACTOR 和 MEAN 这三个综合指标，同时，本文还将三个风险指标进行排序后重新组合，构建排序后的综合风险指标 PCA_RANK、FACTOR_RANK 及 MEAN_RANK，由此衡量企业的系统性风险水平 [19]。此外，考虑到实证结果中以 PCA 指标进行的一系列基准分析表现更佳，因此本文将着重以 PCA 指标展开相应探究。与此同时，为验证研究结论的稳健性，本文还进一步将 PCA 指标与其他的系统性风险指标进行对比。本文最终选用的解释变量如表 1 所示。

Table 1. Main explanatory variables
表 1. 主要解释变量

系统性风险		财务信息		市场信息	
模型指标	指标含义	模型指标	指标含义	模型指标	指标含义
PCA	基于主成分法构造的系统性风险指标	ROA	净利润/总资产	SIZE	公司市值对数
PCA_RANK	基于主成分法排序构造的系统性风险指标	TLTA	总负债/总资产	TUR	公司当月股票换手率
FACTOR	基于因子分析法构造的系统性风险指标	GROWTH	(本期末资产 - 上期末资产)/上期末资产	SZCORR	公司股票收益与 A 股各股收益的相关性均值
FACTOR_RANK	基于因子分析法排序构造的系统性风险指标			RETURN	公司股票收益与上证指数收益之差

续表

MEAN	基于标准化均值构造的系统性风险指标
MEAN_RANK	基于标准化均值排序构造的系统性风险指标

2.2. 模型设定

1) LightGBM 模型。

LightGBM 模型是基于 GBDT 提出的一种优化算法[20]。LightGBM 模型有效提升了训练效率、内存使用率及准确率, 保证高效率的同时也谨防模型过拟合。本文采用如表 1 所示的解释变量集, 被解释变量则为代表企业未来是否发生股价崩盘的二分类变量。

2) 随机森林模型。

随机森林模型通过训练随机划分的子样本获得多个弱分类器, 最终将弱分类器整合为一个强分类器。能够有效避免模型过拟合现象, 具备较强的泛化能力。

3) XGboost 模型。

XGboost 模型也是基于传统 GBDT 的一种优化算法, 该模型以 CART 为基分类器的同时, 还支持线性分类器。XGboost 模型将正则项引入损失函数中, 获得更简单的模型以及降低模型过拟合的风险。

4) 部分依赖图。

本文以机器学习方法中的 LightGBM 模型为基础, 结合部分依赖图方法, 直观描绘出系统性风险水平与企业股价崩盘概率间的非线性关联。

2.3. 模型评估指标

为全面评估模型效果, 本文选择如下评价指标: 准确率、敏感性、ROC 曲线及其面积(AUC)值[21]。下文将介绍各指标的计算方法。

在二分类的预测模型评估中, 我们可以根据模型识别的类别与样本的真实类别组合划分为四种情形, 分别是 TP(真阳性)、FP(假阳性)、TN(真阴性)及 FN(假阴性)。在本文中, TP 是指识别为会发生股价崩盘且真实情况中也发生股价崩盘的样本数; FP 是指识别为会发生股价崩盘但真实情况中未发生股价崩盘的样本数。类似地, TN 是指识别为未发生股价崩盘且真实情况中也未发生股价崩盘的样本数; FN 是指识别为未发生股价崩盘但真实情况中发生股价崩盘的样本数。ROC 曲线通过假阳性率(横坐标)与真阳性率(纵坐标)绘制而成, 模型的预测效果与 ROC 曲线接近左上方的程度成正比。同样地, 评估指标中 AUC 值越高, 说明其分类性能越佳。

其余评价指标的计算公式如下:

1) 准确率(Accuracy)表示在总样本中识别正确的样本所占的比例。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (1)$$

2) 敏感性(Sensitivity)表示预测为真的样本中真阳性样本所占的比例。

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

此外, 在二分类问题中, 合理设定模型阈值是评估模型效果的重要前提。多数二分类问题的研究中, 学者们通常将模型阈值设定为 0.5, 此时模型的泛化能力较强。然而, 当样本分类比例严重不均衡时, 模

型识别时通常会倾向于多数类, 即本文中的未发生股价崩盘情况。有鉴于此, 本文对训练集样本进行了“过采样”处理, 仍将阈值设定为 0.5。

2.4. 样本与数据说明

本文选取 2010~2024 年间我国 A 股上市公司股票日度数据作为股票交易数据样本。并对样本做如下处理: (1) 剔除所有的 ST 股票; (2) 考虑到银行、保险等金融类企业资本结构的差异, 剔除了金融类企业样本; (3) 剔除所有的缺失值; (4) 对连续性数据进行前后 1% 的缩尾处理, 以更好地刻画系统性风险与企业股价崩盘间的关联。上述数据均来源于国泰安数据库和 Wind 数据库。

3. 实证结果与分析

3.1. Logit 模型检验

基于上述准备工作, 本文采用 Logit 模型来探究模型中各项指标对企业股价崩盘的预测能力, 结果见表 2。表 2 的列(1)显示, PCA 指标能够显著预测企业股价崩盘危机。更进一步, 本文加入固定效应, 表 2 的列(2)~(4)显示系统性风险因子具有 1% 的水平下显著预测企业股价崩盘危机的能力, 这说明系统性风险是企业股价崩盘的先行指标。此外, 本文将 PCA 指标替换为系统性风险排序指标 PCA_RANK 进行稳健性分析, 列(5)~(8)的分析结果与前文一致。

究其原因, 伴随系统性风险而来的金融市场动荡会导致金融机构产生惜贷行为, 企业融资成本上升, 对企业偿债能力造成负面影响, 投资者对其丧失信心, 引致股价下跌。同时, 居民消费还会因“财富效应”而受到系统性风险的抑制作用, 给企业生产经营带来极大的压力, 企业的盈利能力受到冲击。此外, ROA 反映了企业的经营活动和未来发展的潜力, 是评估企业价值的重要参考之一, 因此, 当企业的 ROA 无法保持稳定或增长时, 投资者往往认为该企业股价将会下降, 从而使得系统性风险成为企业股价崩盘危机的先行指标。

Table 2. The impact of the collapse of corporate stock prices (based on principal components)
表 2. 企业股价崩盘的影响(基于主成分)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PCA	0.0143*** (46.0595)	0.0143*** (46.0838)	0.0144*** (46.2440)	0.0144*** (46.2299)				
PCA_RANK					0.0078*** (42.4258)	0.0078*** (42.4347)	0.0078*** (42.3042)	0.0078*** (42.4141)
ROA	-0.0030* (-1.9466)	-0.0029* (-1.8883)	-0.0030** (-1.9602)	-0.0029* (-1.8945)	-0.0008 (-0.8517)	-0.0008 (-0.8525)	-0.0008 (-0.8807)	-0.0008 (-0.8567)
TLTA	-0.0002 (-0.3072)	-0.0001 (-0.1299)	-0.0001 (-0.0789)	-0.0001 (-0.1339)	-0.0001 (-0.3024)	-0.0001 (-0.3037)	-0.0001 (-0.3193)	-0.0001 (-0.3073)
GROWTH	-0.0011*** (-2.6257)	-0.0011*** (-2.6610)	-0.0011** (-2.5709)	-0.0011*** (-2.6620)	-0.0016*** (-3.8857)	-0.0016*** (-3.8859)	-0.0016*** (-3.8074)	-0.0016*** (-3.8877)
SIZE	-0.0003* (-1.8919)	-0.0003** (-2.0465)	-0.0003* (-1.6788)	-0.0003** (-2.0330)	-0.0012*** (-7.4461)	-0.0012*** (-7.4554)	-0.0012*** (-7.0933)	-0.0012*** (-7.4201)
TUR	-0.0015*** (-4.4563)	-0.0015*** (-4.6117)	-0.0015*** (-4.4399)	-0.0015*** (-4.6214)	-0.0019*** (-5.6222)	-0.0019*** (-5.6330)	-0.0019*** (-5.3954)	-0.0019*** (-5.5873)

续表

SZCORR	0.0011 (1.5462)	0.0011 (1.5273)	0.0011 (1.4587)	0.0011 (1.5241)	0.0007 (0.9743)	0.0007 (0.9761)	0.0007 (0.9411)	0.0007 (0.9777)
RETURN	-2.3314*** (-74.1423)	-2.3305*** (-74.1202)	-2.3324*** (-74.1657)	-2.3306*** (-74.1214)	-2.5846*** (-78.5126)	-2.5846*** (-78.5149)	-2.5855*** (-78.5283)	-2.5845*** (-78.5134)
行业固定	Y	N	Y	N	Y	N	Y	N
时间固定	Y	Y	N	N	Y	Y	N	N

注：***、**和*分别代表变量在 1%、5%和 10%的显著性水平下显著，括号中为系数标准误，下同。

为探究预测效果的行业异质性，本文依据上市企业在产业链中的位置划分为四个子样本，分别是包含能源产业的上游行业，包含材料和工业产业的中游行业，包含可选消费、日常消费、医疗保健及房地产的下游行业，以及包含电信服务、信息技术与公共事业的服务与支撑行业，随后基于行业类别分别进行 Logit 回归，结果如表 3 所示。

表 3 结果表明，系统性风险能够显著预测各行业企业的股价崩盘危机。上游企业均为能源行业企业，是全球经济的重要支柱之一，系统性风险冲击下经济衰退致使能源需求下降，与此同时，投资者对于高风险行业的回避情绪增强，使得能源行业股价大跌。服务与支撑企业需要投入的资金较大，对于经济形势、政策的变化较为敏感[22]。在经济前景不乐观的情况下，该类行业面临压力时，投资者对行业和市场的悲观情绪会进一步加剧股价下跌。

对于中下游企业而言，一方面，该类企业通常是提供成品或服务的一环，其市场需求受制于居民消费意愿，经济不景气时，消费者会减少消费，降低对产品和服务的需求，影响到中下游企业的销售和收入。另一方面，该类企业通常在供应链中承担较大的生产和运营成本，系统性风险可能导致原材料价格上涨，劳动力市场不稳定，生产成本增加，对其盈利能力与资金流动性产生负面影响，使得投资者减少对该类行业企业的投资。为确保模型结果的稳健性，本文引入固定效应后发现结论依然一致，即剔除部分宏观因素并不影响结论。

Table 3. The impact of stock price crashes in different industries

表 3. 不同行业股价崩盘的影响

	上游	中游	下游	服务与支撑	上游	中游	下游	服务与支撑
PCA	0.0173*** (7.6038)	0.0156*** (33.1991)	0.0159*** (28.4407)	0.0112*** (17.2509)	0.0168*** (7.2272)	0.0155*** (33.0090)	0.0158*** (28.2077)	0.0111*** (17.0179)
ROA	-0.0079 (-0.5364)	-0.0070** (-2.5323)	-0.0010 (-0.4669)	-0.0073* (-1.7573)	-0.0069 (-0.4658)	-0.0070** (-2.5269)	-0.0009 (-0.4118)	-0.0070* (-1.6770)
TLTA	0.0096* (1.7257)	-0.0008 (-0.6464)	0.0006 (0.4960)	-0.0017 (-1.0025)	0.0096* (1.7206)	-0.0007 (-0.5560)	0.0007 (0.5912)	-0.0009 (-0.4904)
GROWTH	0.0021 (0.5483)	-0.0012 (-1.4833)	-0.0017** (-1.9928)	-0.0008 (-1.3400)	0.0022 (0.5838)	-0.0012 (-1.4758)	-0.0016* (-1.8571)	-0.0008 (-1.3792)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
行业固定	N	N	N	N	Y	Y	Y	Y
时间固定	N	N	N	N	Y	Y	Y	Y

系统性风险对企业股价崩盘危机的影响可通过多种渠道实现。总资产收益率低、总资产增长率低的企业通常对系统性风险的抵御能力较差，其中，总资产收益率低的企业通常盈利能力不足，在系统性风险冲击下，这类企业更容易面临融资成本增加和资金链断裂的风险；而总资产增长率低则意味着企业未能有效地扩大其业务规模或市场份额，在行业中的竞争力较弱。总资产收益率、总资产增长率低的情况会使得投资者对该类企业的前景感到担忧，从而影响其股价表现和投资吸引力。因此，本文引入系统性风险指标与上述企业特征的交乘项，分析系统性风险对中下游及服务与支撑企业股价崩盘危机的作用机制，结果如表 4 所示。

根据表 4，系统性风险对中下游及服务与支撑企业的影响存在差异。在下游企业中，系统性风险与总资产收益率的交乘项在 1%水平下显著为负，与总资产增长率的交乘项在 10%水平下显著为负，这表明总资产收益率低、总资产增长率低的下游企业更易受到系统性风险冲击的影响。中游及服务与支撑企业的结果显示，总资产增长率的高低不能判定其是否更易受到系统性风险冲击的影响，而系统性风险与总资产收益率的交乘项在 1%水平下显著为负，这说明总资产收益率低的中游及服务与支撑企业更易在市场波动中出现股价崩盘危机。

Table 4. The impact channels of systemic risks on enterprises
表 4. 系统性风险对企业的影响渠道

	中游企业	下游企业	服务与支撑企业
PCA	0.0176*** (35.0448)	0.0172*** (29.1828)	0.0129*** (18.6184)
PCA*ROA	-0.0459*** (-12.3796)	-0.0218*** (-6.7558)	-0.0391*** (-7.4453)
PCA*GROWTH	-0.0001 (-0.0597)	-0.0020* (-1.8707)	-0.0008 (-1.0154)
Controls	Y	Y	Y
行业固定	Y	Y	Y
时间固定	Y	Y	Y

3.2. 基于机器学习模型的系统性风险影响分析

基于 Logit 回归的研究结果可以捕捉到系统性风险与企业股价崩盘概率间的线性关系，但无法刻画出系统性风险与企业股价崩盘概率之间可能存在的非线性关系。为此，本文先基于 LightGBM 模型与随机森林模型考察企业系统性风险、财务信息以及外部市场表现等预测指标的相对重要性，结果如表 5 所示，随后将 LightGBM 模型与部分依赖图相结合，进一步探究系统性风险与企业股价崩盘之间的非线性关系。

首先，表 5 结果显示，系统性风险指标在 LightGBM 模型和随机森林模型中的相对重要性分别为 20.91%与 36.82%，位列第一。这就表明，在本文的预测模型中，系统性风险指标能够有效优化企业股价崩盘的预测。此外，在 LightGBM 模型中，企业的总资产收益率(ROA)，负债资产比(TLTA)与总资产增长率(GROWTH)相对重要性占比均超过 8%，在随机森林模型均超过 4%，与外部市场信息的重要性占比相当，进一步表明企业财务信息与外部市场信息均为我国企业股价崩盘预警体系中不可或缺的组成部分。

Table 5. Relative importance of predictor variables
表 5. 预测变量的相对重要性

	基于 LightGBM 模型		基于随机森林模型	
	相对重要性	相对重要性排序	相对重要性	相对重要性排序
PCA	20.91%	1	36.82%	1
ROA	9.89%	6	4.66%	7
TLTA	11.21%	4	5.40%	5
GROWTH	8.88%	8	4.34%	8
SIZE	9.04%	7	5.21%	6
TUR	10.16%	5	5.68%	4
SZCORR	12.20%	3	6.66%	3
RETURN	17.71%	2	31.23%	2
合计	100%		100%	

随后，本文将部分依赖图分析方法与 LightGBM 模型相结合，探究系统性风险水平与企业股价崩盘概率间的非线性关联，如图 1 所示。图 1 中，当系统性风险水平超过“-0.5”时，企业股价崩盘的概率随之升高。一方面，部分市场需求会因系统性风险冲击而下降，企业面临销售下降、产能利用率下降等问题，与此同时，就业市场竞争加剧，原材料价格上升，劳动力成本上升，这些因素会显著冲击企业经营状况；另一方面，通常情况下金融市场会因系统性风险冲击而波动加剧，市场的不确定性可能引致持悲观情绪的投资者们大规模抛售股票，造成股市中部分企业的崩盘概率上升。由此可见，系统性风险水平在预测我国企业股市健康状况方面具有关键作用。

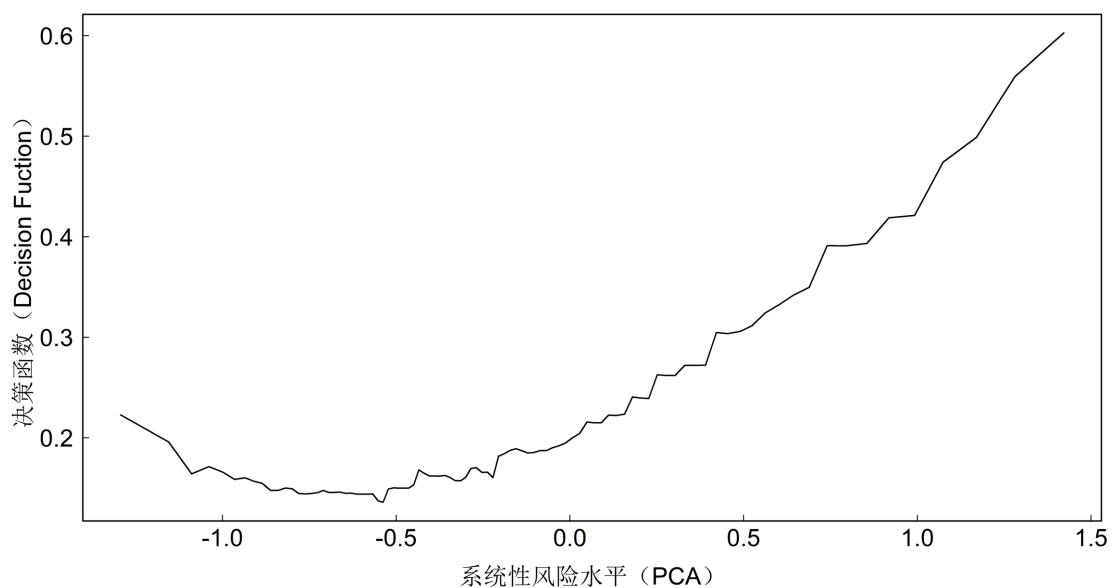
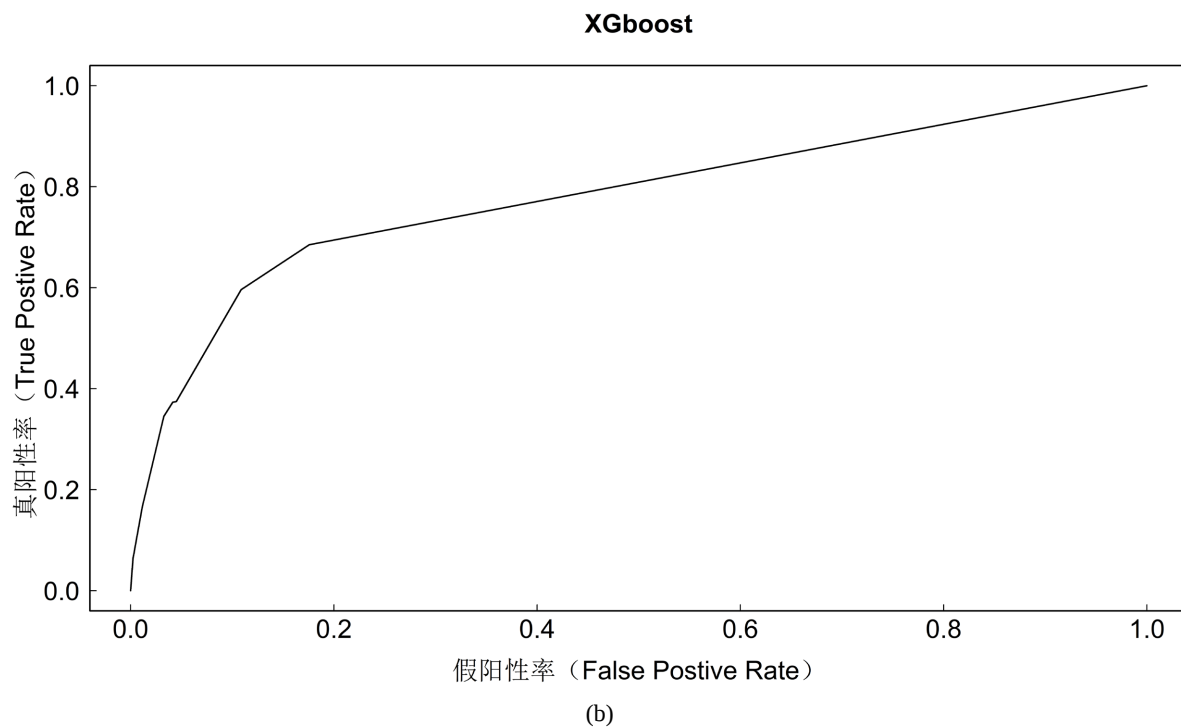
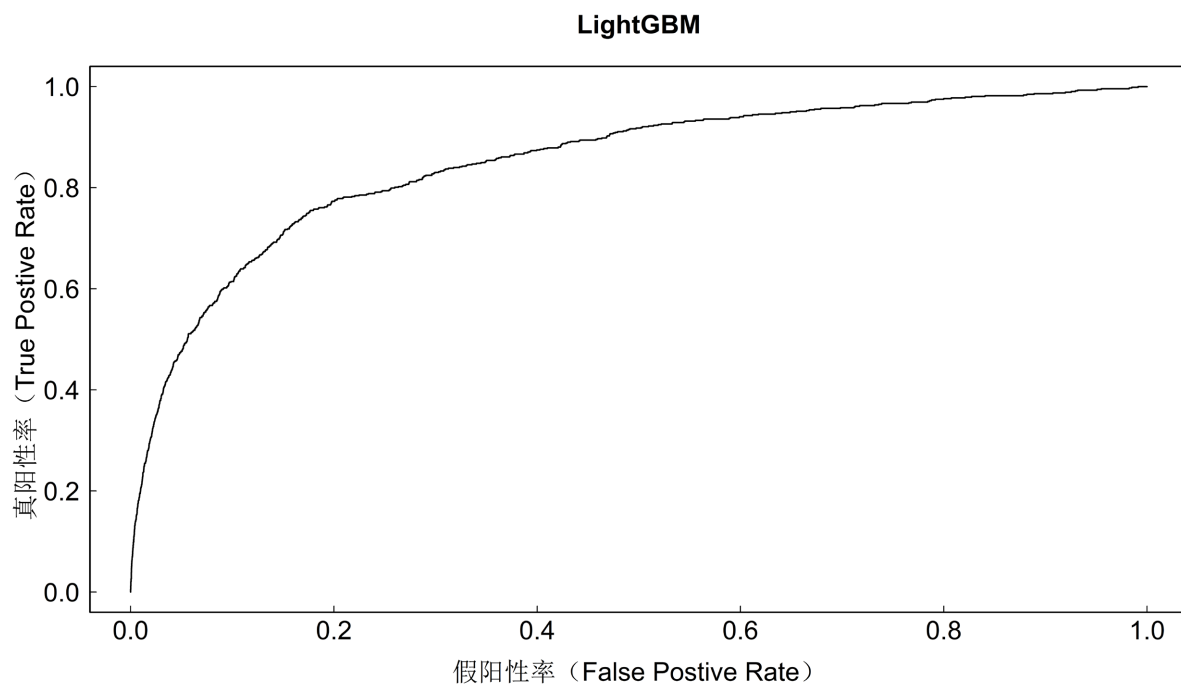


Figure 1. Partial dependency graph based on LightGBM model
图 1. 基于 LightGBM 模型的部分依赖图

3.3. 基于样本外测试的模型预测性能比较

本文从原始信息集中随机选择 80% 的数据作为训练集，训练集经“过采样”调整后构建预测模型，

测试集则余下的 20% 数据, 并基于此测试集评估模型在样本外数据上的预测性能。由于股价崩盘预测正负样本的不均衡, 本文首先采用 ROC 曲线衡量与比较各模型的性能, 结果如图 2 所示, ROC 曲线越趋于左上方说明模型的预测性能越好。由图 2 可见, 四种模型的 ROC 曲线图中, 随机森林模型与 Logit 模型的样本外测试表现出较好的预测性能, 其次为 LightGBM 模型, XGboost 模型则较弱, 说明在对企业股价崩盘危机预测时结合机器学习方法是具有实际作用的。



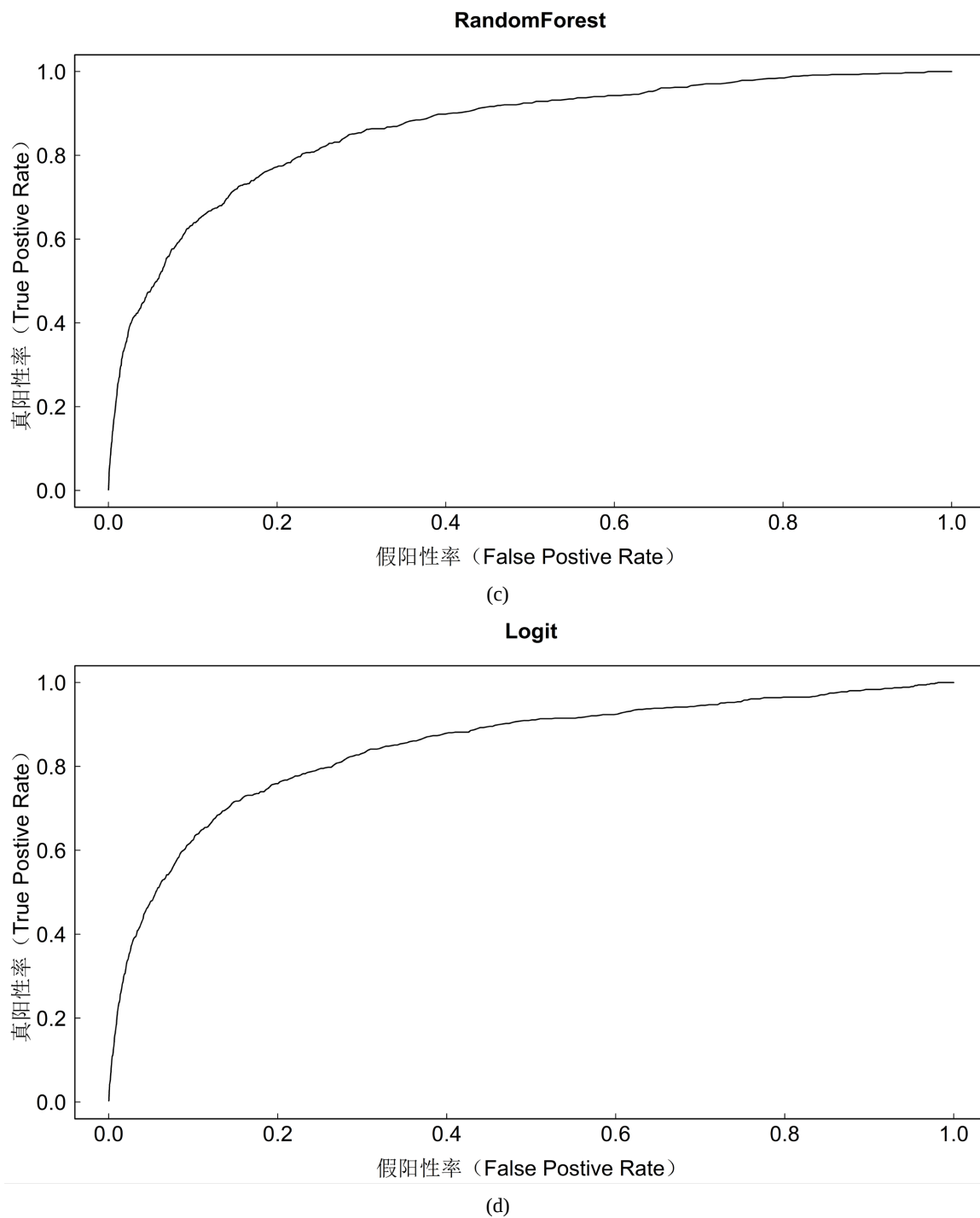


Figure 2. ROC curve of prediction model

图 2. 预测模型的 ROC 曲线图

为全面、准确地评估各模型的预测效果, 本文还考虑了模型预测的准确率、AUC 指标及敏感性, 结果如表 6 所示。表 6 结果表明, 各模型的样本外 AUC 值会因系统性风险 PCA 指标的加入而有所增加, 这就意味着系统性风险指标可以增强企业股价崩盘预警模型的性能。与此同时, 随机森林模型的准确率为 98% 优于 Logit 模型的 97.98%, LightGBM 模型、随机森林模型在 AUC 指标中均具有优于 Logit 模型

的表现，样本外 AUC 指标均达到 0.85，说明机器学习模型在企业股价崩盘预警模型的探索中具有重要地位。此外，本文是对股价崩盘危机进行预测，那么敏感性指标就极为重要。对比发现，LightGBM 和 XGboost 模型具有更高的预测能力，能够分别识别 61.28%和 68.52%的股价崩盘样本，远超 Logit 模型 25.63%的预测性能，这也进一步凸显出在企业股价崩盘预测框架中引入机器学习算法的重要性。

Table 6. Out-of-sample predictive performance (based on principal components)
表 6. 样本外预测性能(基于主成分)

	Accuracy	Sensitivity	AUC	未加入风险指标的 AUC
LightGBM	0.8998	0.6128	0.8511	0.7500
XGboost	0.8230	0.6852	0.7794	0.7423
RandomForest	0.9800	0.2911	0.8624	0.7726
Logit	0.9798	0.2563	0.8467	0.7585

此外，为探究加入不同系统性风险指标后模型的实际预测效果，本文还将 PCA 指标替换为 FACTOR 指标与 MEAN 指标，结果如表 7 所示。通过与表 6 的对比发现，系统性风险 PCA 指标在预测企业股价崩盘危机时表现更为优越。同时，系统性风险 PCA 指标与 LightGBM 模型和 XGBoost 模型结合后，其样本外敏感性水平超过 60%，这一结果优于其他组合。这也说明，主成分分析法更适用于构造我国企业股价崩盘的前瞻性系统性风险指标，并且 LightGBM 模型与 XGboost 模型均能较好地刻画系统性风险因子对企业股价崩盘概率的影响作用，两者组合可达到更佳的预警效果。

Table 7. Out-of-sample predictive performance (replacing risk indicators)
表 7. 样本外预测性能(替换风险指标)

Panel A: 基于因子分析的预测结果				
	Accuracy	Sensitivity	AUC	未加入风险指标的 AUC
LightGBM	0.8996	0.5487	0.8258	0.7500
XGboost	0.8322	0.6518	0.7749	0.7423
RandomForest	0.9823	0.2326	0.8349	0.7726
Logit	0.9790	0.2242	0.7959	0.7585
Panel B: 基于标准化均值的预测结果				
	Accuracy	Sensitivity	AUC	未加入风险指标的 AUC
LightGBM	0.8973	0.6031	0.8479	0.7500
XGboost	0.8365	0.6713	0.7766	0.7423
RandomForest	0.9826	0.2618	0.8593	0.7726
Logit	0.9785	0.2702	0.8450	0.7585

3.4. 基于时间外测试的模型预测性能比较

为验证结果的稳健性，本文还进行了时间外测试。本文以 2022 年前的数据为训练集，其余数据则为测试集，评估结果如表 8 所示。通过是否加入系统性风险指标的模型 AUC 值对比发现，除了 XGboost 模型，其余模型加入系统性风险指标后展现出更强的预测性能，这再次强调系统性风险指标在企业股价崩盘预警框架中的重要性。除此之外，还可以发现机器学习模型的样本外 AUC 指标均超过 0.80，从敏感性

指标来看，LightGBM 模型和 XGboost 模型能够对 63.22%与 69.76%发生股价崩盘的样本进行正确分类，显著优于 Logit 回归模型 33.40%的敏感性水平，也显著高于随机森林模型 34.46%的敏感性水平。可见，LightGBM 模型与 XGboost 模型均表现出稳健且准确的预判能力。

Table 8. Predictive performance over time (based on principal components)
表 8. 时间外预测性能(基于主成分)

	Accuracy	Sensitivity	AUC	未加入风险指标的 AUC
LightGBM	0.9297	0.6322	0.8820	0.8046
XGboost	0.8800	0.6976	0.8078	0.8083
RandomForest	0.9819	0.3446	0.8945	0.8284
Logit	0.9824	0.3340	0.8785	0.8124

4. 结论与启示

本文从系统性风险的视角入手，结合前沿的机器学习方法，尝试对企业股价崩盘危机展开有效预测。首先，本文采用 Logit 回归模型的分析发现，系统性风险指标在预测企业股价崩盘时表现出显著的预测效果。从行业维度分析，系统性风险冲击下，各行业企业的股价状况均易出现不利现象。其次，通过对机器学习方法预测效果的分析发现，在 LightGBM 模型和随机森林模型中，系统性风险 PCA 指标的相对重要性高于财务信息和市场表现指标，说明了系统性风险指标在企业股价崩盘预警框架中的重要性。部分依赖图进一步显示，当系统性风险指标大于“-0.5”时，企业发生股价崩盘的可能性随系统性风险的提高而上升，可见系统性风险是企业股价崩盘预测的重要前瞻性指标。最后，本文采用样本外测试与时间外测试两种方式，对比分析四种模型的预测性能。分析发现，各类模型的预测性能在引入系统性风险指标之后均有所上升，系统性风险指标在企业股价崩盘预警框架中具有重要地位。不仅如此，本文还发现，结合主成分法所构造的系统性风险指标与 LightGBM 模型和 XGboost 模型能够对企业股价崩盘进行预测时具有更加稳健及准确的预测效果。

基于以上结论，本文得出以下三点启示：第一，将系统性风险指标纳入企业股价崩盘预警框架的重要性，完善股价崩盘风险的防控长效机制。第二，通过实施差异化监管，结合多种手段来提高不同行业企业的风险抵御能力，从而有效防范风险冲击。第三，认知到前沿的机器学习监测手段在企业股价崩盘预警机制中的重要性，从技术层面帮助完善审慎监管机制。

基金项目

扬州大学商学院研究生创新项目“OBCE VXEEM 能否预测上海原油期货波动率？——基于 MSMIDAS 模型的研究”(SXYYJSKC202303)。

参考文献

[1] 朱菲菲, 吴偲立, 杨云红. ETF、股票流动性与股价崩盘风险[J]. 金融研究, 2023(6): 169-186.
[2] 许晓芳, 陆正飞. 股权质押融资存在“柠檬现象”吗?——来自股价崩盘风险的证据[J]. 金融研究, 2023(12): 56-73.
[3] Chen, J., Hong, H. and Stein, J.C. (2001) Forecasting Crashes: Trading Volume, Past Returns, and Conditional Skewness in Stock Prices. *Journal of Financial Economics*, **61**, 345-381. [https://doi.org/10.1016/s0304-405x\(01\)00066-6](https://doi.org/10.1016/s0304-405x(01)00066-6)
[4] Jin, L. and Myers, S. (2006) R2 Around the World: New Theory and New Tests. *Journal of Financial Economics*, **79**, 257-292. <https://doi.org/10.1016/j.jfineco.2004.11.003>
[5] Marin, J.M. and Olivier, J.P. (2008) The Dog That Did Not Bark: Insider Trading and Crashes. *The Journal of Finance*,

- 63, 2429-2476. <https://doi.org/10.1111/j.1540-6261.2008.01401.x>
- [6] Conrad, J., Kapadia, N. and Xing, Y. (2014) Death and Jackpot: Why Do Individual Investors Hold Overpriced Stocks? *Journal of Financial Economics*, **113**, 455-475. <https://doi.org/10.1016/j.jfineco.2014.04.001>
- [7] 胡瑾瑾, 赵雪峰, 吴德林, 等. 企业数字化转型对股价崩盘风险预测研究——基于 ED-SPCBoost 模型[J]. 管理评论, 2023, 35(8): 15-30.
- [8] 荆思寒, 王振山, 隋聪, 等. 股票间的风险传染——基于对股价崩盘风险的预测[J]. 系统工程理论与实践, 2022, 42(11): 3090-3104.
- [9] Deng, S., Zhu, Y., Duan, S., Fu, Z. and Liu, Z. (2022) Stock Price Crash Warning in the Chinese Security Market Using a Machine Learning-Based Method and Financial Indicators. *Systems*, **10**, Article 108. <https://doi.org/10.3390/systems10040108>
- [10] 史永, 许铭原, 刘永奇, 等. 审计报告信息含量对公司股价崩盘风险的影响研究[J]. 中国软科学, 2023(11): 201-212.
- [11] 褚剑, 方军雄. 中国式融资融券制度安排与股价崩盘风险的恶化[J]. 经济研究, 2016, 51(5): 143-158.
- [12] 刘圣尧, 李怡宗, 杨云红. 中国股市的崩盘系统性风险与投资者行为偏好[J]. 金融研究, 2016(2): 55-70.
- [13] 杨子晖, 张平淼, 林师涵. 系统性风险与企业财务危机预警——基于前沿机器学习的新视角[J]. 金融研究, 2022(8): 152-170.
- [14] Allen, L., Bali, T.G. and Tang, Y. (2012) Does Systemic Risk in the Financial Sector Predict Future Economic Downturns? *Review of Financial Studies*, **25**, 3000-3036. <https://doi.org/10.1093/rfs/hhs094>
- [15] 张勇, 阮培恒, 梁焱焱, 等. 宏观经济不确定性与银行流动性创造分化[J]. 数量经济技术经济研究, 2022, 39(12): 132-152.
- [16] 李小林, 司登奎. 地方政府债务治理的收入效应[J]. 财经研究, 2024, 50(7): 35-49.
- [17] Sheng, Y., Qu, Y. and Ma, D. (2024) Stock Price Crash Prediction Based on Multimodal Data Machine Learning Models. *Finance Research Letters*, **62**, Article 105195. <https://doi.org/10.1016/j.frl.2024.105195>
- [18] Li, Y., Xue, H., Wei, S., Wang, R. and Liu, F. (2024) A Machine Learning Approach for Investigating the Determinants of Stock Price Crash Risk: Exploiting Firm and CEO Characteristics. *Systems*, **12**, Article 143. <https://doi.org/10.3390/systems12050143>
- [19] Nucera, F., Schwaab, B., Koopman, S.J. and Lucas, A. (2016) The Information in Systemic Risk Rankings. *Journal of Empirical Finance*, **38**, 461-475. <https://doi.org/10.1016/j.jempfin.2016.01.002>
- [20] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 3149-3157.
- [21] 刘云菁, 伍彬, 张敏. 独立董事能识别公司财务舞弊风险吗——基于机器学习预测财务舞弊的研究[J]. 会计研究, 2023(9): 21-34.
- [22] 杨子晖, 陈雨恬, 黄卓. 国际冲击下系统性风险的影响因素与传染渠道研究[J]. 经济研究, 2023, 58(1): 90-106.