

代理与信号传递视角下股权质押违约多模态风险识别体系构建

罗述君, 唐佳琪, 白 杨, 展颐华

北方工业大学经济管理学院, 北京

收稿日期: 2026年8月4日; 录用日期: 2026年8月14日; 发布日期: 2026年9月3日

摘 要

股权质押风险的跨市场传染已成为威胁资本市场稳定的重大隐患, 精准识别其关键驱动因子是防范系统性风险的重要前提。本文提出了一套融合结构化数据与非结构化文本信息的多维风险因子识别分析框架, 为后续引入机器学习模型与Shapley值归因分解提供理论支撑与变量体系基础。研究从质押行为、财务经营、市场治理三大维度构建结构化指标, 并引入年报MD&A文本的情感语调与语义风险因子两类非结构化指标, 由此形成涵盖五大维度的理论指标体系, 旨在系统揭示股权质押违约风险的潜在驱动因素。本文的边际贡献在于, 将研究视角从单一财务维度拓展至结构化与非结构化的信息融合, 丰富了股权质押风险识别的信息边界; 同时, 提出将Shapley值引入可解释的归因推断, 为金融机构贷后管理和监管部门差异化监管提供量化依据。

关键词

股权质押风险, 风险驱动机制, Shapley值

Establishing a Multi-Modal Risk Identification System for Equity Pledge Defaults from the Perspectives of Agency Theory and Signal Transmission

Shujun Luo, Jiaqi Tang, Yang Bai, Yihua Zhan

School of Economics and Management, North China University of Technology, Beijing

Received: August 4, 2026; accepted: August 14, 2026; published: September 3, 2026

Abstract

Cross-market contagion of equity pledge risks has become a major hidden threat to the stability

文章引用: 罗述君, 唐佳琪, 白杨, 展颐华. 代理与信号传递视角下股权质押违约多模态风险识别体系构建[J]. 金融, 2026, 16(5): 515-523. DOI: 10.12677/fin.2026.165051

of capital markets; accurately identifying its key drivers is a crucial prerequisite for preventing systemic risks. This paper proposes a multi-dimensional risk factor identification and analysis framework that integrates structured data with unstructured textual information, providing theoretical support and a foundation for the variable system required for the subsequent introduction of machine learning models and Shapley value attribution decomposition. The study constructs a set of structured indicators across three dimensions—pledging behaviour, financial performance, and market governance—and further introduces sentiment and semantic risk factors derived from the Management Discussion and Analysis (MD&A) sections of annual reports. Together, these components form a comprehensive theoretical indicator system spanning five dimensions in total, aimed at systematically revealing the potential drivers of equity pledge default risk. The marginal contribution of this paper lies in expanding the research perspective from a single financial dimension to the integration of structured and unstructured information, thereby broadening the informational scope for identifying equity pledge risks. Simultaneously, we propose incorporating Shapley values into interpretable attribution inference, which provides a quantitative basis for post-lending management by financial institutions and differentiated supervision by regulatory authorities.

Keywords

Risks Associated with Share Pledges, Risk-Driven Mechanisms, Shapley Value

Copyright © 2026 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

我国明确指出：“防范化解金融风险是金融工作的永恒主题”“加强金融监管”“有效防范化解金融风险”¹。金融是国民经济的血脉，资本市场作为金融体系的核心组成部分，其稳定运行直接关系金融安全与实体经济发展。当前，我国资本市场规模持续扩大、结构日益复杂，各类风险的关联性、传染性显著增强，亟需构建精准高效的风险防控体系，筑牢不发生系统性金融风险的底线。

股权质押作为一种常见的融资工具，在我国资本市场中被广泛应用。但其蕴含的风险传导机制十分复杂，高比例质押引发的控制权转移、股价崩盘等违约事件频发。2021年12月，美盛文化(002699)因控股股东长期隐瞒股权质押违约、股份司法冻结与拍卖事项，未及时履行信息披露义务，被浙江证监局出具警示函。控股股东及一致行动人质押涉诉股份合计占公司总股本 58.68%，所持股份质押比例高达 99.91%，债务违约后批量触发股份司法冻结、网络司法拍卖，公司控制权稳定性遭受根本性冲击[1]。2025年5月，瀚川智能(688022)因控股股东未及时披露股权质押合同及违约风险，被江苏证监局出具警示函，其质押股权占公司总股本比例达 14.64%，触发违约后导致股份司法冻结，公司控制权稳定性遭受重大冲击，中小投资者利益受损严重[2]。此类案例充分表明，股权质押风险已从股东个体融资风险，演变为影响公司持续经营、投资者利益保护与资本市场稳定的重大现实问题，其精准识别与防控成为落实“科学防范金融风险”重要指示的迫切实践需求。

近年来，监管部门密集出台政策强化股权质押风险精细化管控，在此背景下，本文立足于委托代理理论与信号传递理论，构建融合结构化财务数据与年报 MD&A 文本信息的多维风险因子理论框架，并在此基础上提出将机器学习模型与 Shapley 值归因方法引入该框架的分析构想，为股权质押违约风险的关

¹<http://theory.people.com.cn/n1/2025/0506/c40531-40473641.html>

键驱动因子识别提供变量选取的理论依据与后续实证研究的方法论基础。

2. 文献综述

2.1. 股权质押的动因、经济后果与风险传导

控股股东股权质押作为一种兼具融资便利与潜在风险的资本运作方式,已成为中国资本市场的重要现象。本文旨在通过对现有文献的系统梳理,构建一个从理论到方法的完整分析框架。首先,阐述股权质押的动机与经济后果;其次,探讨 Shapley 值作为一种归因分析工具在会计领域的应用潜力;最后,展望机器学习模型在股权质押风险预测与机制分析中的前沿应用。

2.1.1. 股权质押的动因

关于大股东为何进行股权质押,现有研究主要存在“融资约束”和“自利、掏空”两种观点。

从融资约束视角来看,杜丽贞(2019) [3]的研究指出,民营上市公司的大股东进行股权质押的主要动因是为了缓解上市公司的融资约束。由于民营企业面临较严重的信贷歧视,股权质押成为一种“被动”但有效的融资策略。王斌等(2013) [4]也发现,相比于国有大股东,民营大股东因融资约束更强,更倾向于采用股权质押融资,且这种融资行为带有“对赌”性质。

从自利与掏空视角看,郑国坚等(2014) [5]从大股东财务约束的视角出发,发现当大股东面临严重财务约束(股权被质押或冻结)时,更容易对上市公司进行资金占用(掏空),且这种占款行为会损害公司价值。杜丽贞(2019) [3]也证实,当大股东存在较高的资金占用时,股权质押为其掏空提供了便利条件。

2.1.2. 股权质押的经济后果

股权质押对公司业绩和价值的影响存在争议,主要取决于大股东的性质和面临的控制权转移风险。

股权质押对公司价值的影响存在“弱化激励效应”和“强化侵占效应”(郝项超和梁琪, 2009) [6]。高比例的股权质押不仅会损害公司业绩(王斌等, 2013) [4],还会降低企业风险承担水平(何威风等, 2018) [7],并显著增加股价崩盘风险(余怒涛等, 2025) [8]。为规避股价下跌触及平仓线,控股股东有强烈的动机进行市值管理。这表现为策略性的信息披露,如利用 MD&A 中的非财务信息掩盖盈余操纵行为(程新生等, 2015) [9],或进行更多的盈余管理以稳定股价(廖慧文等, 2026; 余怒涛等, 2025) [8] [10]。这些行为增加了企业的违规倾向,尤其是信息披露违规(郑维伟等, 2024) [11]。投资者对股权质押的关注呈现倒“U”形关系。在质押比例较低时,信息需求和监督动机占主导,实地调研增加;当比例过高时,风险规避心理占主导,调研活动减少(石慧等, 2025) [12]。感知到企业风险上升和信任度下降后,供应商会收紧信用政策,迫使企业更多使用应付票据等高交易成本的商业信用模式(廖慧文等, 2026) [10]。

2.1.3. 股权质押的风险传导

股权质押不仅是融资行为,更是风险的源头。何威风等(2018) [7]研究发现,大股东股权质押与企业风险承担水平呈显著负相关。大股东为了避免股价下跌触及平仓线导致控制权转移,倾向于让公司选择低风险项目,从而降低了企业的整体风险承担水平。

股权质押加剧了信息不对称和代理问题。孟庆斌等(2017) [13]指出,虽然 MD&A 披露通常有助于降低崩盘风险,但如果管理层出于掩盖负面信息的目的进行策略性披露,反而会加剧风险。股权质押带来的平仓压力可能促使大股东隐藏坏消息,一旦累积到临界点,将引发股价崩盘。

2.2. 模型可解释性与 Shapley 值在会计领域的应用

随着机器学习模型在会计与金融预测中的应用日益广泛,模型的“黑箱”特性成为一大挑战。Shapley 值作为一种基于合作博弈论的解释方法,为理解复杂模型提供了理论支撑。在会计研究中,不仅要关心

模型预测得准不准，更关心“为什么”模型会做出这样的预测。

传统的回归系数只能反映线性关系下的影响方向和大小，而 SHAP 模型可以处理非线性关系。例如，在预测股权质押违约风险时，Shapley 值可以告诉我们，究竟是“质押比例”、“现金流”还是“MD&A 语调”对违约概率的贡献最大。Shapley 值不仅能解释整个样本集上哪个变量最重要，还能解释某一个特定公司为什么被模型判定为高风险。例如，对于某家特定公司，Shapley 值可能显示其“高质押比例”推高了违约风险，但其“高 ROE”降低了违约风险，这种细颗粒度的解释对于审计师和监管者具有重要价值。

综上所述，Shapley 值作为合作博弈论框架下的归因方法，能够有效克服传统线性回归系数在非线性场景下的解释局限，为机器学习模型提供全局与局部层面的特征贡献度量化依据。就本文的研究设计而言，上述方法论梳理并非意味着本文直接实施 Shapley 值的数值计算，而是旨在论证将该方法引入股权质押违约风险因子识别研究的理论必要性与潜在价值。本文后续章节将围绕变量体系的构建展开理论分析，为后续基于实际数据的机器学习建模与 Shapley 值归因检验预设变量基础与逻辑前提。

2.3. 机器学习与文本挖掘在会计研究中的模型应用

传统的会计研究多依赖于结构化财务数据，如资产负债表、利润表数据。近年来，随着大数据技术的发展，机器学习模型结合非结构化文本数据，如年报 MD&A 成为了新的研究热点。

程新生等[9]的研究利用机器学习逻辑探讨了盈余操纵与 MD&A 非财务信息披露的关系。研究发现，存在盈余操纵的公司倾向于在 MD&A 中披露更多的非财务信息以“掩人耳目”，且这种策略性行为在有股权激励的公司中更为显著。这提示我们在构建预测模型时，不能仅看财务指标，还需结合文本特征来识别管理层的策略性动机。结合何威风等(2018) [7]和王斌等(2013) [4]的变量选择逻辑，现代机器学习模型(如随机森林、XGBoost、神经网络)可以纳入多维度的特征，主要包括两方面：

结构化特征：财务指标、治理特征、质押特征

非结构化特征：具有 MD&A 的文本向量、语调得分、可读性指标等

结合以上特征现代机器模型具有能够捕捉变量间复杂的非线性关系，并能处理高维度的文本数据的优势，从而比传统的 Logistic 回归模型提供更精准的预测。

2.4. 研究评述

综上所述，股权质押研究已从单一的动因分析转向对风险传导机制的深入探讨。在方法论上，结合 Shapley 值的可解释性机器学习模型，以及基于文本向量化的非结构化数据分析，正在成为会计与金融领域研究的新范式。这不仅提高了预测的准确性，也为理解大股东行为背后的复杂逻辑提供了更透明的视角。

3. 理论分析

立足委托代理理论、控制权私利理论与信号传递理论，结合现有股权质押风险相关文献成果，从理论层面阐释质押违约的形成逻辑，并系统搭建多维度概念指标体系。区分结构化量化信息与文本非结构化信息两大信息渠道，论证各类指标对质押违约风险的理论解释力，为后续机器学习预警模型构建、风险因子归因分析提供完整理论概念支撑。

3.1. 样本选取的理论依据与数据来源界定

从研究范畴的理论边界来看，本文选取 2012~2024 年 A 股非金融上市公司作为分析样本。金融行业的股权质押规则与监管约束与实体企业存在显著差异，剔除后可规避行业异质性对风险传导机制的干扰；该时间区间覆盖了多轮资本市场政策调整与股价周期波动，能够完整捕捉不同宏观环境下质押违约的演化规律。

在违约概念的理论界定上,本文将股权司法冻结、强制平仓、质押到期未回购三类事件统一认定为质押违约,并构建股权质押违约指标(pledge_default)。三类事件均表明控股股东质押行为已突破履约底线,实质反映控制权转移风险由潜在威胁转化为现实事件,契合现有文献对质押实质性违约的理论界定标准。

从信息维度的理论分类出发,本文将影响质押违约的信息划分为结构化量化信息与非结构化文本信息两类,二者分别具备不同的风险识别理论价值。

委托代理理论聚焦于股权分离情境下控股股东与中小股东之间的利益冲突。当控股股东通过股权质押将自身融资风险与上市公司股价挂钩时,其控制权与现金流权分离程度加剧[3],大股东以损害中小股东利益为代价进行风险转嫁的动机显著增强。结构化指标依托公司年度财务与股权交易等客观数据,直观反映企业经营、股权质押与公司治理层面的硬约束风险,划分为三大理论维度。首先,股权质押行为维度,通过质押比例、质押期限、质权主体等指标直接映射控股股东融资激进程度与到期偿债压力。依据委托代理理论,质押行为类变量直接度量代理冲突的强度:质押比例越高,控制权私利动机越强,违约风险的内生驱动力越大,质权人数量越多,反映大股东融资渠道多元化但同时也隐含更高的外部履约压力;财务经营维度则从“企业自身缓冲能力”角度回应代理问题——良好的偿债、盈利与营运能力可削弱大股东因财务困境而侵占公司资源的动机,从而降低代理成本[6],借助偿债、盈利、营运、成长四类指标刻画企业内生风险缓冲能力;市场与治理维度体现外部监督与内部制衡对代理冲突的约束作用,理论上可抑制大股东质押后的机会主义行为,利用股价波动、股权结构、机构持股等指标体现外部市场约束与内部制衡水平。

信号传递理论则解释信息不对称环境下管理层如何通过可观测信号向市场传递私有信息。财务数据是“硬信息”,反映历史经营结果;而年报 MD&A 文本是“软信息”,承载管理层对未来的预期、风险判断和策略性披露[8]。当管理层隐藏负面消息或语调异常波动时,这些文本信号可提前揭示财务指标尚未捕捉的隐性风险。据此,非结构化文本信息被纳入体系,以年报、问询函文本承载管理层私有隐性风险信息,作为结构化财务数据无法覆盖的前瞻性信号。

文本量化包括两类理论指标:一是情感语调指标,通过词汇频次量化管理层风险披露倾向;二是深层语义因子,依托金融预训练模型挖掘文本隐藏利空信号。全部文本指标与结构化指标以公司-年度为匹配标识整合,形成理论层面完整的面板概念数据集。

上述几大维度变量的选取,分别对应质押违约风险形成链条中不同环节的理论机制,而并非简单沿袭既有实证文献。

质押行为类变量(质押比例、质押期限、质权人数量等)的理论依据在于委托代理理论。控股股东通过股权质押将自身流动性风险与上市公司股价绑定,质押比例越高,控制权与现金流权分离度越大,大股东以损害中小股东利益为代价进行风险转嫁的动机越强,违约风险的内生驱动越显著。

财务经营类变量的理论依据在于财务困境理论。偿债能力、盈利能力与营运能力共同构成企业抵御外部冲击的“财务缓冲层”。当股价下跌触发补仓或平仓要求时,财务基本面健康的企业尚有追加担保或筹措资金的空间,而盈利持续弱化、负债率高企、现金流紧张的企业则缺乏风险化解的现实能力,违约从“可能”走向“必然”的阈值更低。

市场与治理类变量反映外部市场约束与内部制衡机制对控股股东质押行为的调节作用。股价波动直接决定质押物价值的安全边际,而机构投资者持股、董事会独立性等治理机制则通过监督约束功能,抑制大股东在质押后的机会主义行为,降低违约风险。

非结构化文本变量的理论依据在于信号传递理论。财务数据反映的是历史经营结果,而年报 MD&A 文本承载管理层对未来前景的预期与策略性信息披露。当管理层语调出现异常波动或文本中隐含悲观语义时,往往意味着财务指标尚未捕捉的隐性风险正在积聚,从而构成对结构化信息的增量补充。

上述几大维度变量的选取分别根植于 Jensen M C 的委托代理理论[14]与 Spence M 的信号传递理论

[15]的核心逻辑，两类理论共同构成整个指标体系的元理论框架。

3.2. 多维风险概念体系理论构建

3.2.1. 被解释变量理论内涵

被解释变量(pledge_default)对应股权质押实质性违约核心概念。从事件识别理论逻辑来看，CSMAR数据库以单笔质押合同为基础单元记录质押状态变动，能够精准区分单年度内是否出现司法冻结、强制平仓、到期未回购违约事件。理论层面，若企业当年存在任意一笔质押业务发生履约失败，代表控股股东流动性风险、控制权风险转化为现实损失，因此将该观测赋值为 1；全年质押业务均按期履约则赋值为 0，仅保留当期存在控股股东质押行为的观测作为理论分析样本，规避无质押样本带来的概念干扰。被解释变量详细定义与说明见表 1。

Table 1. Conceptual definitions of dependent variables
表 1. 被解释变量概念界定

变量名称	变量符号	变量定义	赋值说明
股权质押违约	pledge_default	衡量上市公司在特定年度是否发生股权质押违约事件	若控股股东在当年发生司法冻结、强制平仓或到期无法回购等违约事件，赋值为 1；否则赋值为 0

3.2.2. 结构化解释变量

结合委托代理理论、财务困境理论，从质押行为、财务经营、市场治理三大理论维度筛选结构化概念指标，各指标理论测算口径如表 2 所示。

Table 2. Theoretical definitions of structured indicators
表 2. 结构化指标理论界定

变量类别	变量名称	符号	变量说明
质押特征	控股股东质押比例	big_share_pledge	控股股东质押股/持股总额 $\times 100\%$
	公司股权质押比例	pledge_ratio	(公司累计质押股数/公司总股数) $\times 100\%$
	年度质押次数	pledge_times	公司当期年度内股权质押总次数(含初始质押、补充质押)
	质押平均期限	pledge_term	当期所有质押合约期限的加权平均值(年/天)，权重为对应质押股数
	质权人数量	pledge_holder_num	为公司提供股权质押的金融机构(券商、银行、信托等)总数
	是否高比例质押	is_over_pledge	0=控股股东质押比例 $< 50\%$ ；1=控股股东质押比例 $\geq 50\%$
	质押融资金额(万元)	pledge_amount	当期公司股权质押累计融资金额，按质押当日收盘价测算
	实际控制人质押占比	pledge_real_control	(实际控制人累计质押股数/实际控制人持股总数) $\times 100\%$
财务特征 - 偿债能力	流动比率	current_ratio	流动资产/流动负债
	速动比率	quick_ratio	(流动资产 - 存货)/流动负债
	资产负债率	asset_liab_ratio	(负债总额/资产总额) $\times 100\%$
	利息保障倍数 A	interest_coverA	(净利润 + 所得税费用 + 财务费用) / 财务费用
	利息保障倍数 B	interest_coverB	(净利润 + 财务费用) / 财务费用
	经营活动现金流偿债比	cash_flow_ratio	经营活动产生的现金流量净额/负债合计 $\times 100\%$
	短期负债占比	short_debt_ratio	流动负债/负债合计 $\times 100\%$

续表

财务特征 - 盈利能力	净资产收益率	roe	(净利润/平均净资产) × 100%
	总资产收益率	roa	(净利润/平均总资产) × 100%
	销售毛利率	gross_margin	(营业收入 - 营业成本)/营业收入 × 100%
	销售净利率	net_margin	(净利润/营业收入) × 100%
	EBITDA 利润率	ebitda_margin	净利润 + 所得税费用 + 财务费用 + 固定资产折旧、油气资产折耗、生产性生物资产折旧 + 无形资产摊销 + 长期待摊费用摊销
财务特征 - 营运能力	总资产周转率	asset_turn	营业收入/平均总资产
	存货周转率	inventory_turn	营业成本/平均存货
	应收账款周转率	receivable_turn	营业收入/平均应收账款
	应收账款周转天数	receivable_days	360/应收账款周转率
	营业周期(天)	operating_cycle	应收账款周转天数+存货周转天数
财务特征 - 成长能力	营业收入增长率	revenue_grow	(当期营业收入 - 上期营业收入)/上期营业收入 × 100%
	净利润增长率	profit_grow	(当期净利润 - 上期净利润)/上期净利润 × 100%
	总资产增长率	asset_grow	(当期总资产 - 上期总资产)/上期总资产 × 100%
	营业利润增长率	operating_profit_grow	(当期营业利润 - 上期营业利润)/上期营业利润 × 100%
市场与治理特征模式 - 市场特征	股票收益率波动率	stock_volatility	当期月度股票收益率标准差 × $\sqrt{12}$ (年化处理)
	股票换手率	turnover_rate	年内日换手率之和
	市净率	pb_ratio	每股市价/每股净资产
市场与治理特征模式 - 治理特征	第一大股东持股比例	top1_hold	(第一大股东持股数/公司总股数) × 100%
	董事会规模	board_size	当期公司董事会成员总人数(含独立董事)
	独立董事比例	ind_ratio	(独立董事人数/董事会总人数) × 100%
	董事长与总经理是否两职合一	is_dual	1=同一人; 2=不同一人
	机构投资者持股比例	inst_hold_ratio	机构投资者持有的上市公司股份比例

前文所述激励与掏空效应可能随质押比例变化而交替主导, 质押比例对违约风险的影响或存在非线性阈值特征, 这为后文采用机器学习模型捕捉此类复杂关系提供了理论预期。

3.2.3. 非结构化文本衍生指标理论内涵

依据信号传递理论, 年报 MD&A、问询回复文本是管理层私有信息披露载体, 财务指标仅反映历史已实现经营结果, 文本可提前释放隐性风险预期。本文从文本情感、深层语义两大理论维度构建非结构化概念指标。

1) 情感语调类概念指标

现有外文情感词典无法适配国内质押风险场景, 本文在 Loughran-McDonald 词典理论框架基础上, 结合股权冻结、平仓、流动性危机等质押特有负面词汇构建中文金融情感词典, 从词汇频次理论层面衡量管理层披露倾向。

$$\text{净语调tone} = \frac{\text{积极词汇数} - \text{消极词汇数}}{\text{文本总有效词汇数}}$$

理论上数值越高代表管理层乐观披露，利空隐藏概率更低；数值越低代表悲观预期，质押违约隐性风险更高。

语调变化率 $\text{tone_diff} = \text{本期tone} - \text{上期tone}$

该指标捕捉同一公司相邻年度间管理层态度的动态变化。若语调由乐观骤然转向悲观，可能意味着前期未充分披露的利空信息集中暴露，即存在“掩饰性信息披露”特征，此类情形下质押违约风险显著上升。

语调波动率 tone_volatility ，采用三年滚动标准差衡量披露稳定性，波动越大代表管理层判断前后矛盾，信息不对称程度越高，潜在质押风险越高。

2) 语义风险概念指标

单纯词频语调仅能捕捉表层词汇情绪，无法识别上下文组合的隐性风险语义。基于金融文本深度学习理论，采用 FinBERT 预训练模型提取 768 维文本深层语义向量；为规避高维特征理论层面的维度灾难与过拟合问题，使用主成分分析(PCA)降维，选取累计方差贡献率 85% 以上的主成分，生成 sem_risk_0 至 x 项低维语义因子，在保留文本核心风险信息的同时适配后续机器学习模型输入要求。非结构化文本指标释义见表 3。

Table 3. Definition of concepts in unstructured text
表 3. 非结构化文本概念界定

变量	指标释义
tone	文本净语调
tone_diff	年度语调变化幅度
tone_volatility	语调波动率
sem_risk_0~sem_risk_x	FinBERT + PCA 降维后语义风险特征

4. 总结

本文立足于股权质押违约风险预测中变量间非线性关系刻画不足、多维度信息融合不充分的现实背景，构建了一套融合结构化数据与非结构化文本信息的多维风险因子识别框架。

在变量体系构建层面，本文从质押特征、财务基本面、市场与治理三大维度出发，选取结构化变量，覆盖控股股东质押行为、企业偿债与盈利能力、市场波动与公司治理结构；同时引入基于金融情感词典的语调指标(tone 、 tone_diff 、 tone_volatility)和 FinBERT 语义风险因子(sem_risk_0 至 sem_risk_x)两类非结构化文本衍生变量，将年报 MD&A 中管理层态度变化与文本隐含风险信号纳入分析框架，最终构建了涵盖 5 个维度的风险因子体系。

在文献基础层面，本文系统梳理了股权质押动因、经济后果及文本信息增量价值三条研究脉络。文献分析表明，现有研究在变量选取上已形成从质押行为、财务基本面到治理环境的共识性框架，但在方法层面仍以传统线性回归为主，难以捕捉变量间的非线性关系与阈值效应；在数据层面则较少将结构化数据与非结构化文本信息纳入统一框架进行质押违约风险的预测与归因分析。此外，直接以质押违约事件为被解释变量的预警研究仍较为稀缺。上述问题的识别为本文研究设计提供了理论依据与拓展方向。

综上，本文的理论贡献有二。其一，将股权质押违约研究从单一财务指标拓展至“结构化 + 非结构化”多模态信息融合，丰富了风险识别的信息维度。其二，提出将 Shapley 值归因框架嵌入股权质押风险识别的分析构想，为后续从全局与局部层面量化特征贡献度、揭示变量间交互效应提供理论预设与方法

路径。实践层面,研究结论可为金融机构贷后管理、质权人动态保证金设定及监管部门对高比例质押公司的差异化监管提供量化参考,助力防范系统性流动性风险的跨市场传染。

本研究亦存在一定局限。文本分析中,情感词典虽已针对质押场景补充专属词汇,但中文财务文本的语义复杂性(如反讽、模糊表达)仍可能影响语调指标的精度;PCA降维虽降低了语义向量的维度,但也在一定程度上损失了可解释性。未来可进一步引入大语言模型(LLM)进行端到端的文本风险抽取,并探索动态 Shapley 值随时间演变的特征,以刻画风险驱动因子的时变规律。后续研究将基于完整样本进行实证验证,并对比不同机器学习模型的性能,持续优化预警框架。

基金项目

2026 年北京市北方工业大学大学生创业创新项目:关键风险因子识别:Shapley 值视角下股权质押风险的驱动机制研究(项目编号:10805136026XN073-108)。

参考文献

- [1] 关于对美盛控股集团有限公司、赵小强采取出具警示函措施的决定[EB/OL]. 2021-12-10. <https://www.csrc.gov.cn/zhejiang/c103952/c1653797/content.shtml>, 2026-08-13.
- [2] 江苏证监局关于对关于对苏州瀚川投资管理有限公司、蔡昌蔚采取出具警示函措施的决定[EB/OL]. 2025-05-15. <https://www.csrc.gov.cn/jiangsu/c103893/c7557926/content.shtml>, 2026-08-13.
- [3] 杜丽贞, 马越, 陆通. 中国民营上市公司股权质押动因及纾解策略研究[J]. 宏观经济研究, 2019(7): 148-160.
- [4] 王斌, 蔡安辉, 冯洋. 大股东股权质押、控制权转移风险与公司业绩[J]. 系统工程理论与实践, 2013, 33(7): 1762-1773.
- [5] 郑国坚, 林东杰, 林斌. 大股东股权质押、占款与企业价值[J]. 管理科学学报, 2014, 17(9): 72-87.
- [6] 郝项超, 梁琪. 最终控制人股权质押损害公司价值么? [J]. 会计研究, 2009(7): 57-63+96.
- [7] 何威风, 刘怡君, 吴玉宇. 大股东股权质押和企业风险承担研究[J]. 中国软科学, 2018(5): 110-122.
- [8] 余怒涛, 杨张萌. 质权人与标的物公司共享审计师能降低股价崩盘风险吗?——基于控股股东股权质押视角[J]. 现代财经(天津财经大学学报), 2025, 45(2): 71-90.
- [9] 程新生, 刘建梅, 程悦. 相得益彰抑或掩人耳目: 盈余操纵与 MD&A 中非财务信息披露[J]. 会计研究, 2015(8): 11-18+96.
- [10] 廖慧文, 童盼. 控股股东股权质押与商业信用模式——基于供应商-企业关系的视角[J]. 中央财经大学学报, 2026(1): 112-128.
- [11] 郑维伟, 瞿茜, 安玲林, 等. 控股股东股权质押与企业违规行为——基于内部控制与外部监督视角[J]. 现代金融研究, 2024, 29(12): 36-46.
- [12] 石慧, 李明辉. 投资者关注控股股东股权质押吗?——基于实地调研视角的研究[J]. 财贸研究, 2025, 36(3): 97-110.
- [13] 孟庆斌, 杨俊华, 鲁冰. 管理层讨论与分析披露的信息含量与股价崩盘风险——基于文本向量化方法的研究[J]. 中国工业经济, 2017(12): 132-150.
- [14] Jensen, M.C. and Meckling, W.H. (1976) Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, 3, 305-360. [https://doi.org/10.1016/0304-405x\(76\)90026-x](https://doi.org/10.1016/0304-405x(76)90026-x)
- [15] Spence, M. (1973) Job Market Signaling. *The Quarterly Journal of Economics*, 87, 355-374. <https://doi.org/10.2307/1882010>