集团特征有助于预测企业债务危机吗
——基于可解释机器学习模型的经验证据

刘 斌， 徐 舜， 李浩然， 董晋亭

[摘要] 集团型企业的内部结构与资本运作模式对其债务风险有重大影响,但现存的债务危机预警研究极少考虑相关指标。本文基于金字塔结构理论与集团内部资本市场理论,构建了一套包含8个集团特征在内的预测指标体系;以2007—2023年A股集团型上市公司为样本,利用Logit模型、RF模型与XGBoost模型构建了企业债务危机预测模型;利用SHAP可解释机器学习方法,系统比较了传统Logit模型与机器学习模型在预测和解读方面的差异,分析了集团特征指标的重要性排序及其与债务危机之间的非线性关系。研究发现,集团特征指标对预测债务危机具有重要价值,多数集团特征指标对债务危机有非线性影响。集团分权水平、向母公司分红水平、对子公司股权投资水平等是预测效力较强的指标。进一步从总部职能定位、信息沟通成本以及企业组织边界三个层面对集团进行分析发现,集团特征指标对母公司为投资型总部的集团、子公司地理位置分散的集团、执行"瘦身健体"后的国有企业集团预测效力更强。本文以集团特征为切入点,拓展了债务危机预测的指标体系,为集团型企业防范债务风险、有关部门提升监管效率提供了实践指引。

[关键词] 集团型企业； 集团特征指标； 债务危机； 可解释机器学习

[中图分类号] F272 [文献标识码] A [文章编号] 1006-480X(2025)07-0174-19

一、引言

党的二十大报告强调,依法将各类金融活动全部纳入监管,守住不发生系统性风险底线。2023年中央金融工作会议要求,全面加强金融监管,有效防范化解金融风险。2024年人民银行金融稳定工作会议指出,持续健全金融风险监测评估框架,坚持对风险早识别、早预警、早暴露、早处置,体现了对金融风险预警问题给予的高度重视。国家统计局数据显示,中国实体经济总杠杆率从2008年的157%上升至2024年的286%。与此同时,企业在追求快速增长的过程中往往忽视债务风险管

[收稿日期] 2025-01-21

[基金项目] 国家社会科学基金重点项目"企业集团债务资本野蛮生长的识别机制与有效监管研究"(批准号22AGL011);中央高校基本科研业务费项目"企业集团债务杠杆的放大效应、识别机制与有效监管"(批准号2023CDJSKJJ20)。

[作者简介] 刘斌,重庆大学经济与工商管理学院教授,博士生导师,管理学博士;徐舜,重庆大学经济与工商管理学院博士研究生;李浩然,重庆理工大学会计学院讲师,管理学博士;董晋亭,重庆大学经济与工商管理学院博士研究生。通讯作者:李浩然,电子邮箱:drlihaoran@163.com。感谢匿名评审专家和编辑部的宝贵意见,文责自负。

理,如恒大集团、海航集团、隆鑫控股等超大型集团债务暴雷的现象越来越频繁,不但严重干扰企业自身的发展,而且通过经济链条扩散甚至演化为系统性金融风险。鉴于此,如何完善并优化企业特别是集团型企业的债务危机预警模型,是一个亟须解决的现实问题。

现有基于机器学习的债务危机预警研究已经能够较为准确地预测债务危机的发生,但此类研究与经典财会理论的结合通常不够紧密,大多通过挖掘财务数据进行预测,而未深刻解释企业债务危机的成因。近年来,随着可解释人工智能的蓬勃发展,部分研究采用SHAP可解释机器学习方法对预测结果进行深入阐释(雷欣南等,2024)。这不仅提升了模型的可理解性,更为机器学习模型与专业理论的深度融合创造了有利条件。但从具体经济金融问题出发,利用财会专业知识构建指标体系,提高机器学习预测能力的研究仍然较为罕见。特别地,现有研究大多将合并后的企业视为独立的个体,以合并报表作为主要数据来源进行预测,而没有将企业及其控股子公司作为一个集团加以考虑。集团合并财务报表通过一系列会计抵销分录,将母公司与多家子公司的资产与负债进行合并,从而全面反映集团的整体财务状况。然而,这一整合过程不可避免地导致了集团特征信息的流失。例如,资产在母子公司间的分布、债务风险在母子公司间的分布、投融资行为在母子公司间的分布(王秀丽等,2017)等。为此,本文以集团金字塔结构理论、内部资本市场理论等经典理论为研究依据,构建引入集团特征指标体系的企业债务危机预警模型。

目前,中国绝大多数上市公司均采取集团化运作的模式(张会丽和陆正飞,2012)。根据《企业会计准则》的定义,集团是母公司及其控股子公司组成的经济实体。这种基于权益控股形成母子公司的关系架构,是导致集团型企业债务过度增长,进而诱发债务危机的特殊原因。一方面,集团型企业通常具有母公司控股子公司、子公司再控股孙公司,层层持股实现对末端公司控制的金字塔组织结构,这种组织结构天然具有债务的杠杆放大效应,即在母公司实施债务融资的基础上,其各级子公司仍可以同时进行债务扩张。大规模、多层次的子公司体系为母子公司的重复融资提供了可能,加剧了企业债务风险(李增泉等,2008)。同时,金字塔结构还存在所有权与现金流权不一致的治理问题,末端子公司距集团实际控制人的控制链条较长,委托—代理问题与信息交易成本也较高。此时,控股股东更可能利用其超额控制权对集团实施掏空行为,子公司管理者也更可能实施机会主义行为导致过度负债与过度投资(刘斌和袁利华,2016),进而诱发企业债务危机。另一方面,集团型企业可以通过主动实施内部资本市场操作,进一步提高资金使用效率、降低融资约束。Friedman et al.(2003)指出,集团内部母子公司间存在高水平的相互支持。资金充裕的子公司可以通过向母公司分红或关联贷款等方式转移流动性(黎文靖和严嘉怡,2021),反之,在自身资源匮乏时则从母公司或其他子公司取得融资。这一机制有助于优化集团内部资源的利用,但在经济形势下滑时,这些内部交易和担保也可能导致个别子公司的债务风险扩散(纳鹏杰等,2017)。此外,集团型企业还可以通过内部关联担保或其他手段撬动外部资金。与内部资本市场不同,企业与外部债权人间存在高水平的信息不对称,在行为经济学有限理性假设下,出于对集团规模与隐性担保的信任,银行等外部债权人倾向为其提供更多贷款。一些集团甚至通过资产证券化(ABS)、设立特殊目的实体(SPA)等特殊的资本运作方式撬动更多外部资金。这些行为同样是集团型企业运营的重要风险点,对于预测其债务危机具有重要价值。

基于金字塔结构理论与内部资本市场理论,本文剖析了集团型企业实现债务增长的特殊路径,从集团资本分布、管控模式、内部资源分配以及外部资源撬动四个方面,构建了8个能够反映集团特征的指标。随后,结合常见的企业资产质量、利润水平、高管特征等指标,利用传统的Logit模型、随机森林(Random Forest, RF)模型和极端梯度提升树(Extreme Gradient Boosting,XGBoost)模型预

测企业债务危机。在此基础上,本文系统地比较了机器学习方法与传统Logit回归在预测模式、解读方法上的差异。同时,本文还利用SHAP可解释机器学习方法,报告了集团特征指标的重要性排序、集团特征指标与债务危机之间的非线性关系等。在稳健性检验中,本文通过变换机器学习预测模型、采用时间序列滚动预测方法、检验模型远期预测效力、更换债务危机度量方式、绘制局部累积效应图等方法来提高主检验的可信度。最后,本文进一步从总部职能定位、信息沟通成本以及企业组织边界三个层面对集团进行分析,考察了模型预测效力的变化。

研究发现,集团特征指标对企业未来出现债务危机有重要预测能力,其预测效力与常见的企业内部治理、外部利益相关者治理、高管特征等传统指标相当。其中,集团分权水平、向母公司分红水平、对子公司股权投资水平等指标对企业债务危机的预测价值较高。少数股东权益水平、向子公司贷款水平等指标与企业债务危机具有明显的"U型"关系,这些指标在机器学习框架中对预测债务危机的发生有重要贡献,但在Logit框架中的重要性相对较低。其他指标与债务危机间也存在不同程度的非线性关系。进一步研究发现,集团特征对债务危机的预测能力在不同类型集团间存在差异:从总部职能定位看,对母公司为投资型总部的集团预测能力较强,而对母公司为实体型总部的集团预测效果较弱;从信息沟通成本看,对子公司地理位置分散的集团预测能力较强,而对子公司分布集中的集团预测效果较弱;从企业组织边界看,对执行国有企业"瘦身健体"后的集团预测能力较强,而对执行国有企业"瘦身健体"前的集团预测效果较弱。

本文具有以下边际贡献:①构建了集团特征指标的分析框架与指标体系。现有研究大多从单一指标角度分析集团行为(王亮亮等,2023),而系统性梳理集团特征指标的研究相对匮乏。本文以金字塔结构理论与内部资本市场理论为出发点,从集团资本分布、集团管控模式、集团内部资源分配以及集团外部资源撬动四个角度遴选了8个集团特征指标,对后续研究有借鉴意义。②加深了对集团特征指标非线性预测价值的认识。以往对集团型企业的研究通常以线性回归为主要分析手段,对金字塔结构能否提高治理效果(潘怡麟等,2018;郑志刚等,2024)、内部资本市场能否提高分配效率(陈艳利等,2014)等问题,常得出正相关或负相关的单向结论。本文讨论并实证检验了这些指标与债务危机之间的非线性关系,这对弥合现有研究争议有所帮助。此外,本文发现少数股东权益、向子公司贷款等指标在机器学习中有重要的非线性预测价值,而在Logit模型中显著性较低,表明现有文献可能忽视了这些指标对债务危机的预测效力。③丰富了集团型企业债务危机的研究内容。本文从总部职能定位、信息沟通成本以及企业组织边界三个维度对集团进行了细分,发现集团特征指标在母公司为投资型总部、子公司地理位置分散的企业、"瘦身健体"后的国有企业中对债务危机预测能力更强,进一步深化了对集团型企业运行逻辑的认识。④利用SHAP可解释机器学习方法,系统性对比了传统回归模型与机器学习模型在预测方面的差异。以往可解释机器学习领域的研究主要致力于分析、解释机器学习方法本身(雷欣南等,2024;Sun et al.,2024),而较少关注传统回归方法与机器学习方法之间的差异。本文比较并解释了三种模型在预测效果、蜂群图、SHAP依赖图、指标重要性等方面的差异,将传统计量经济学研究范式与机器学习研究范式联系起来,有助于可解释机器学习方法的进一步推广与应用。

二、文献综述与理论分析

1.文献综述

企业债务危机的预警研究是会计、金融等领域重要的研究话题之一。早期企业违约、破产风险

预测模型主要包括经典的Altman模型、Merton模型等,这些研究以回归模型为分析工具,仅考虑单个或少数几个指标对企业债务危机的影响,而对企业其他方面信息的利用较为有限。后续学者将更多宏观经济、企业财务或企业治理等指标纳入预测体系,但仍主要以回归模型为分析工具。线性识别模型通常存在"函数形式误设"问题,难以捕捉指标间的非线性关系与交互关系,且在面对高维数据时会出现"多重共线性"与"过度拟合"问题,不能充分利用数据,因此,预测效果普遍不佳,无法满足现实中对债务危机的监管、预警需求。

目前,众多学者利用机器学习模型对企业违约、破产等现象进行预测,充分挖掘企业特征大数据的潜力,提升了预测的准确率。如Bertomeu et al.(2021)发现,集成学习方法识别重大会计错报的能力显著优于Logit模型。迟国泰等(2024)遴选出一套契合于中国市场的企业特征指标组合,并基于Stacking模型开发了符合中国市场特征的企业违约预测模型。随着可解释人工智能的兴起,Lundberg and Lee(2017)以合作博弈理论为基础提出了SHAP方法,并为线性模型和树模型提供了高性能算法,攻克了SHAP方法计算复杂的难关。在此基础上,众多学者应用SHAP方法对机器学习预测结果进行解释(Sun et al.,2024;雷欣南等,2024),拓展了机器学习方法的适用范围。然而,从具体的经济金融问题着手,运用财会领域的专业知识来构建指标体系,并借助可解释的机器学习方法以深化实证解读的研究,目前仍然较为少见。

目前,已有研究借助计量经济学实证工具,讨论了集团特征与企业资本结构、债务违约之间的联系,这些成果为本文构建集团特征理论框架、训练机器学习债务危机预警模型提供了研究基础。在融资模式方面,汪玉兰等(2020)发现,相较于独立企业,集团型企业过度负债的可能性更高。在组织结构方面,李增泉等(2008)发现,企业集团金字塔持股结构具有融资优势,而且金字塔层级越多,控制链条越长,融资优势越明显,过度负债的可能性也越高(郑志刚等,2024)。在内部交易方面,陈艳利等(2014)发现,企业集团内部的关联担保交易容易产生风险传染效应,个别担保环节出现问题,可能使整个集团陷入偿债危机(纳鹏杰等,2017)。也有研究关注了母子公司的相对地位。例如,王秀丽等(2017)从集团母公司经营模式出发,发现在母公司采用经营主导型战略时,母公司报表数据的财务预警效果优于合并报表。潘怡麟等(2018)发现,集团在人事决策方面集权时,企业的市场价值更高。Beaver et al.(2019)关注了集团子公司特征与企业的债务危机问题,发现子公司破产概率高的母公司本身更容易破产,而当母公司破产概率高时,子公司也更容易出现债务危机。

在吸收与借鉴这些研究的基础上,本文进一步考察了集团在资本分布、管控模式、内部资源分配和外部资源撬动方面的特征。同时,利用先进的可解释机器学习模型,将集团特征指标引入债务危机预警模型中,分析了集团特征指标与债务危机之间的复杂非线性相关关系,对厘清集团特征对企业行为的影响机制有重要意义,是对现有研究体系的有益补充。

2.理论分析

(1)集团资本分布与企业债务危机。相比于单一法人主体,集团型企业的"独特"之处在于多层"母子公司"构成的错综复杂的金字塔结构,这种结构为集团型企业带来了更强的融资能力,但也为其债务危机埋下了隐患。一方面,金字塔结构使集团存在控制权与所有权不一致的情况,控股股东可以利用较少的股权投资"撬动"较多的资本,使集团在资本市场中具有融资优势(Almeida and Wolfenzon,2006)。同时,集团母公司与子公司能够同时实施债务融资,而利用子公司进行债务融资本身也具有杠杆放大的功能(李增泉等,2008),使整个集团的债务水平提升。在这一过程中,对子公司股权投资的额度越高,子公司扩张债务规模的本金越充足,集团内部的债务容量越高(李维安和韩忠雪,2013)。另一方面,集团型企业不仅有母子公司双层结构,还会随着母公司控股子公

司、子公司控股孙公司形成多层级的金字塔结构。随着孙公司及其下属子公司数量与层级的不断增加,集团的控制链条不断拉长,这模糊了组织的最优边界,进而导致集团的债务管控能力减弱,使其偏离了最优债务水平(汪玉兰等,2020)。同时,随着孙公司及其下属子公司的债务资本规模不断扩张,集团财务风险也进一步加剧(李焰等,2007)。因此,集团资本分布特征对企业债务危机有重要预测价值。

(2)集团管控模式与企业债务危机。在金字塔结构的集团型企业中,母公司与子公司(包括孙公司及以下子公司)间存在两类集团特有的委托—代理关系,即母公司与子公司管理者、母公司与子公司少数股东之间的代理问题。一方面,当子公司管理层的激励机制与集团整体利益发生冲突时,可能引发自利的道德风险。随着集团对末端子公司控制链条的延长,子公司的寻租空间也逐步增大(陆正飞和张会丽,2010),进而导致集团效率降低、债务风险提升。为了避免这种情况,集团需要建立有效的监督与激励体系,以确保子公司管理层的决策与母公司的战略目标保持一致。此时,集权管理的集团能够统筹协调企业资源,加强对子公司债务风险的有效管控(潘怡麟等,2018;王亮亮等,2023)。相反,分权管理模式下,子公司各自为政将损害其他子公司乃至集团整体的利益,甚至可能诱发集团债务危机。另一方面,类似的情况也出现在控股股东与少数股东之间。集团型企业存在控制权与所有权不一致的现象,大股东可能利用其控制权优势采取更激进的投融资策略,损害少数股东或债权人的利益,甚至通过关联交易、转移定价等方式直接侵吞少数股东利益(纳鹏杰等,2017)。如果少数股东话语权过高,则可能实施与企业战略目标不一致的短视行为,如过度关注股价波动、过度重视当期分红等。因此,集团管控模式特征对企业债务危机也有重要预测价值。

(3)集团内部资源分配与企业债务危机。在集团型企业中,母子公司间独特的内部资源分配行为是影响债务危机的重要因素。相较于外部资本市场,内部资本市场具有信息成本优势,能够更有效地调配自有资金,把握净现值为正的投资机会,从而缓解融资约束并降低投资不足。其中,资金在成员企业间的流转,是内部资本市场资源配置的核心环节。理论上,内部资本市场应遵循效率原则:即将资金优先配置给投资回报率较高的业务单元,同时规避低效或亏损项目(Busenbark et al.,2022)。这种配置通常表现为子公司资金归集与母公司再分配的双向流动(汤晟等,2024),子公司财务决策应服从集团整体资源配置战略(Belenzon et al.,2019)。然而,实践表明,内部资本市场的运作未必充分有效。一方面,母公司可能基于控制权优势进行非效率配置,通过高利率内部借贷转移子公司利润,或要求超额分红削弱子公司财务弹性;另一方面,总部在资源配置中可能出现“交叉补贴”现象,即对优质单元投资不足却过度支持低效单元(Lins and Servaes,2002)。这些行为导致内部资本市场功能异化,不但不能提高资源运作效率,反而助长非公允关联交易、违规担保和资金占用。此外,集团内部复杂的债务网络会形成风险传染链条(纳鹏杰等,2017),单个子公司的财务危机会通过连带责任机制向母公司及整个集团蔓延,提升集团整体违约概率。因此,集团内部资源分配行为对企业债务危机有重要预测价值。

(4)集团外部资源撬动与企业债务危机。集团型企业不仅依赖内部资本市场进行资源配置,还常通过关联担保等机制撬动外部资金。根据行为经济学的有限关注理论,出于对集团规模与隐性担保的信任,银行等外部债权人可能产生集团“大而不倒”的幻觉,降低监督水平并为其提供更多贷款,进而导致了预算软约束下的信贷扭曲膨胀(郑志刚等,2024)。随着集团内部关联担保复杂程度的增加,信息不对称进一步加剧,高风险子公司可能借助集团信用背书获得超额外部融资,推高整体债务水平(陈艳利等,2014)。母公司同样可以将其持有的子公司股权进行抵押再融资,从而形成股权投资的重复融资现象,甚至可能通过复杂的股权架构来蒙蔽债权人,最终导致债务过度扩张,

进而引发企业绩效的恶化(王琨等,2014)。此外,集团旗下的金融、财务、资产管理类子公司可以通过结构化融资工具进一步放大杠杆风险。例如,集团将内部资产打包重组后转移至表外SPE发行证券,既隐匿真实债务水平,又获取额外融资。又如,通过发行名义股权融资工具(如优先股、可转债)或合资设立子公司,配合抽屉协议承诺固定收益,实质构成隐性债务等。短期看,内部关联担保网络与金融子公司能通过获取外部融资以缓解资金压力,但若形成长期依赖则会扭曲资本结构,加剧财务脆弱性。因此,集团外部资源撬动行为对预测企业债务危机也有重要价值。①

三、研究方法与研究设计

1.预测模型

本文拟构建以下模型预测企业的债务危机情况:

$$Default_{i,t+m} = f\left(D_{i,t}\right) + \varepsilon_{i,t+m} \tag{1}$$

其中,$D_{i,t}$表示选用的变量集合;$Default_{i,t+m}$表示企业i在$t+m$年后发生债务危机的情况;$f(\cdot)$表示所采用的预测模型;$\varepsilon_{i,t+m}$表示模型的预测残差。

本文采用Logit模型、RF模型与XGBoost模型进行对比预测。其中,Logit模型是计量经济学中最常见的多元回归模型,也是具有较高可解释性的"白箱"模型,而RF与XGBoost是两种基于树模型的集成学习算法,其预测效果好、计算效率高,在机器学习方法中具有代表性且与SHAP方法适配。②

2.基于SHAP的可解释机器学习方法

SHAP方法由Lundberg and Lee(2017)提出,用来解释机器学习模型中每个变量对目标预测值的贡献大小或重要性。基于合作博弈思想,SHAP方法考虑了新变量加入模型的所有可能顺序,并计算每一顺序下加入新变量对模型预测结果造成的影响,最后用其平均值衡量新变量对预测结果的综合贡献大小。

参考肖争艳等(2022)的做法,SHAP值按以下流程计算。对于一个包含$|F|$个变量的数据集F,机器学习模型$f(\cdot)$对某个样本x_*的预测结果为$\widehat{f(x_*)}$。这一结果可以分解为预测基准值ϕ_0与每一个变量j的SHAP值ϕ_*^j之和,即:

$$\widehat{f(x_*)} = \phi_0 + \sum_{j=1}^{|F|}\phi_*^j \tag{2}$$

其中,预测基准值ϕ_0通常为样本总体的预测均值$E[\widehat{f(x)}]$,每一个变量j的SHAP值ϕ_*^j按如下方法计算:

$$\phi_*^j = \sum_{S\subseteq F\setminus\{j\}}\frac{|F|!}{|S|!\left(|F|-|S|-1\right)!}\left[v\left(S\cup\{j\}\right)-v(S)\right] \tag{3}$$

其中,S是$F\setminus\{j\}$(集合F但不包含变量j)的子集,共可以构造$\dfrac{|S|!\left(|F|-|S|-1\right)!}{|F|!}$个这样的集合。$v\left(S\cup\{j\}\right)-v(S)$表示在变量集合$S$的基础上,新增变量$j$对模型预测结果的影响。此时,SHAP值$\phi_*^j$表示在所有可能情况下向模型引入变量$j$对模型预测结果影响的均值。$\left|\phi_*^j\right|$越大,说明

① 理论框架图参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

② 预测模型详细介绍参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

该变量对目标预测值的影响越大;$\phi_*^j > 0$或$\phi_*^j < 0$则表示变量对预测结果产生正向或负向影响。

整体而言,SHAP提供了一种公平比较变量重要性的方案,具有两个重要优势:①SHAP具备模型无关性,可以应用于任何机器学习模型,而不依赖于模型$f(\cdot)$的具体结构。②SHAP具备全局的可加一致性。SHAP值ϕ_*^j能够线性地加和为模型的预测结果$\widehat{f(x_*)}$,即使模型$f(\cdot)$为非线性模型。

3.预测指标体系①

(1)债务危机。债务危机在企业中的表现形式并不完全一致,本文参考王秀丽等(2017)做法,当集团型上市公司出现以下情形时,则可认定其陷入债务危机:①破产、退市或被ST,表明企业整体经营出现重大失败,偿债能力受到严重影响;②因债务问题被提起诉讼或出现债务重组,表明企业出现实质性债务违约,无法偿还到期债务;③财务报表出现资不抵债,表明企业资产无法清偿其债务,出现会计层面的债务危机。

(2)集团特征指标。本文从集团资本分布、集团管控模式、集团内部资源分配和集团外部资源撬动四个方面构建集团特征的指标体系。①在集团资本分布方面,本文选取对子公司股权投资水平和孙公司及其以下子公司资产占比(简称孙公司及其以下占比)两个指标。其中,对子公司股权投资水平用于反映集团中子公司的体量与规模(李维安和韩忠雪,2013;刘斌和袁利华,2016);孙公司及其以下占比,即深层子公司规模占比,用于反映集团金字塔结构的深度与集团组织结构的复杂程度(李焰等,2007)。②在集团管控模式方面,本文选取集团分权水平与少数股东权益水平两个指标。其中,集团分权水平反映了上市公司作为母公司对整个集团子公司的管控水平(潘怡麟等,2018;王亮亮等,2023);少数股东权益水平用于反映少数股东参与集团治理、影响公司决策的能力。③在集团内部资源分配方面,本文选取母公司向子公司贷款水平与子公司向母公司分红水平两个指标。其中,母公司向子公司贷款水平反映了内部资本市场中资金由母公司流向子公司的水平(黎文靖和严嘉怡,2021);子公司向母公司分红水平,由于子公司富余现金通常以分红形式向母公司解缴,这一指标反映了内部资本市场中资金由子公司流向母公司的水平(汤晟等,2024)。④在集团外部资源撬动方面,本文选取集团内关联担保水平与金融类子公司数目两个指标。其中,集团内关联担保水平反映企业集团内部信用互相转移,并对外取得借款的情况(王琨等,2014);金融类子公司数目用于反映集团通过关联金融机构包装资产、取得外部融资的潜在可能。

(3)其他预测指标。除了集团特征指标外,参考姜富伟等(2023)、雷欣南等(2024)、迟国泰等(2024)的研究,本文还选取9类共计43项一般预测指标:资产质量类,包括资产规模、现金比率、有息债务率、无息债务率等指标;盈利能力类,包括总资产收益率、主营业务收入、主营业务增长率等指标;现金流结构类,包括利润收现率、投资现金率等指标;内部治理类,包括股权集中度、独立董事数量等指标;高管特征类,包括高管团队年龄、高管团队学历等指标;外部利益相关方治理类,包括媒体关注度、是否由四大会计师事务所审计等指标;宏观经济类,包括基准利率等指标;区域经济类,包括消费价格指数等指标;行业特征类,包括行业勒纳指数等指标。此外,省份虚拟变量、行业虚拟变量与年份虚拟变量也纳入预测指标体系。

4.样本与数据处理②

(1)样本选取与数据来源。由于《企业会计准则》在2007年实施了重大修订,首次规定使用成本法核算母公司对子公司长期股权投资,所以本文以2007—2023年沪深A股上市公司数据为研究

① 详细的指标定义参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

② 样本筛选过程、超参数选择和预测效果评价指标说明参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

对象。随后，样本剔除金融类企业与关键变量缺失的公司，并剔除注册资本不足5000万元或子公司数量不足5家的观测值，[①]最终获得38604个有效观测值。企业财务报告、治理、高管、行业信息来自CSMAR数据库；分析师、审计师与媒体关注数据来自CNRDS数据库；宏观经济与地方经济资料来自CCER数据库；子公司数据源自CSMAR数据库与天眼查数据。所有连续变量均在1%和99%的水平上缩尾，并按最大最小值缩放法进行归一化处理。

（2）预测方法与预测效果评价。按照机器学习研究的惯例，本文采用随机抽取方法，从样本总体中抽取90%样本作为训练数据，剩余10%样本作为测试集（雷欣南等，2024）。其中，训练集样本共34744个，用于模型的构建；测试集样本共3860个，用于模型预测性能的检验。模型所涉及超参数通过网格搜索（GridSearchCV5）方法确定。以综合反映模型预测效力的ROC曲线线下面积（AUC）作为调参目标，选取AUC值最大的参数组合。由于本文的预测目标为企业是否出现债务危机的二分类问题，本文采用准确率（Accuracy）、查准率（Precision）、查全率（Recall）、精确率与召回率的调和平均数（F-measure）、ROC曲线线下面积（AUC）来评价模型的预测性能。

四、实证分析

1.描述性统计[②]

表1报告了未经归一化处理的指标描述性统计情况。结果显示，出现债务危机的企业约占样本的7.38%。在样本中企业对子公司投资额度平均占合并报表总资产的21.51%，最高可达合并报表总资产的82.06%；孙公司及其以下子公司平均占资产总额的5.37%；集团分权水平，平均为48.96%，最高可达99.86%；少数股东权益平均占总资产的2.97%，最高可达总资产19.48%；母公司对子公司贷款余额平均占母公司资产的8.76%，最高可达57.55%；子公司每期向母公司分红平均占子公司资产的6.06%；上市公司内部关联担保平均占其净资产的15.95%；每个集团上市公司旗下平均有2.11家金融类子公司。[③]其余指标分布情况与已有研究一致。

表1　　主要变量描述性统计

指标名称	N	Min	P50	Mean	Max	S.D.
企业债务危机	38604	0.0000	0.0000	0.0738	1.0000	0.2615
对子公司股权投资水平	38604	0.0088	0.1767	0.2151	0.8206	0.1623
孙公司及其以下占比	38604	0.0000	0.0000	0.0537	0.7734	0.1402
集团分权水平	38604	0.0000	0.4793	0.4896	0.9986	0.3123
少数股东权益水平	38604	0.0000	0.0144	0.0297	0.1948	0.0393
向子公司贷款水平	38604	0.0000	0.0335	0.0876	0.5755	0.1231
向母公司分红水平	38604	0.0000	0.0000	0.0606	0.8149	0.2002
集团内关联担保水平	38604	0.0000	0.0391	0.1595	1.8345	0.2985
金融类子公司数目	38604	0.0000	0.0000	0.4801	2.8332	0.6914

2.模型预测效果

表2 Panel A展示了3种模型仅使用集团特征指标对未来1期债务危机的预测效果。当模型仅

① 参考王斌和张伟华（2014）的研究，企业集团应满足母公司注册资本在5000万元以上，并至少拥有5家子公司。

② 完整描述性统计结果参见《中国工业经济》网站（ciejournal.ajcass.com）附件。

③ 此为原值的平均而非对数值的平均。

使用集团特征指标及其平方项和两两交乘项时,XGBoost模型的查全率为24.76%,RF模型查全率为29.64%,说明超过1/4的债务危机样本在集团特征方面具有特殊性,能够被机器学习模型识别出来。RF模型与XGBoost模型的AUC值超过0.7,说明集团特征指标本身即对债务危机有良好的预测能力。Panel B展示了3种模型使用全部指标对未来1期债务危机的预测效果。由于样本中出现债务危机的正例数目较少,3种模型的预测准确率均在90%以上,但RF与XGBoost模型的预测准确率均较Logit模型有所提升。在查全率方面,机器学习模型较传统Logit模型提升非常明显。XGBoost模型能够成功识别出51.47%的债务危机样本,相对于Logit模型查全率提高了24.11%。在F-measure值与AUC值方面,两种机器学习模型预测效果也强于传统的Logit模型。其中,RF与XGBoost模型的AUC值均超过90%,说明其预测效果极佳。整体而言,XGBoost模型的预测效果优于RF模型,而RF模型又优于Logit模型。

表2　　模型预测效果

	模型	准确率	查全率	查准率	F-measure	AUC
Panel A:集团特征指标预测效果						
使用集团特征指标及其平方项和两两交乘项	Logit	0.9219	0.0391	0.5714	0.0732	0.6839
	RF	0.8821	0.2964	0.2725	0.2839	0.7103
	XGBoost	0.9021	0.2476	0.3363	0.2852	0.7356
Panel B:模型整体预测效果						
使用全部指标	Logit	0.9337	0.2736	0.7059	0.3944	0.8715
	RF	0.9386	0.4007	0.6910	0.5072	0.9012
	XGBoost	0.9417	0.5147	0.6695	0.5820	0.9360

3.模型的可解释分析

(1)Logit分析。在引入可解释机器学习方法前,本文先使用传统的Logit回归方法分析这一问题作为后续分析的基础。表3展示了传统Logit模型对样本总体的回归结果,指标按z值绝对值大小降序排列,排名前29的指标在Logit模型中均具有统计显著性。结果显示,8个集团特征指标中有7个指标对企业债务危机有显著影响,说明本文指标选择合理。其中,集团分权水平、集团内关联担保水平、对子公司股权投资水平、金融类子公司数目、向子公司贷款水平、少数股东权益水平等指标与债务危机显著正相关,向母公司分红水平与债务危机显著负相关。

(2)SHAP全局分析。[①]本文通过绘制蜂群图的方法来展示SHAP对模型的全局分析结果。若某一指标的SHAP绝对值较高,则表明该指标对模型预测的重要性较强。图1报告了3种模型对未来1期债务危机预测重要性最高的30个指标,指标按重要性(即SHAP绝对值的平均)由上至下排列。图1(a)为Logit模型的全局可解释分析。由于Logit模型属于线性模型,其SHAP值为指标值本身的大小乘以固定的回归系数。当回归系数为正时,指标的数值越大,则SHAP值越高,样本发生债务危机的可能性越大,反之亦然。Logit模型中,集团分权水平是集团特征指标中对预测结果影响最大的指标,在所有指标重要性排在第9位。结合表3可以看出,绝大部分在Logit模型中显著的指标均出现在图1(a)中,指标的重要性排序相似,系数方向也相同。图1(b)为RF模型的全局可解

① SHAP局部分析以及8个集团特征指标的完整蜂群图参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

表3 Logit模型的回归结果

排序	指标	系数	z值	排序	指标	系数	z值
1	无息债务率	4.0306***	17.7032	19	高管团队年龄	−0.0244***	−3.6930
2	主营业务收入	−0.6153***	−12.1721	20	行业勒纳指数	−2.3247***	−3.3607
3	有息债务率	2.9001***	10.8929	21	金融类子公司数目	0.1109***	2.8761
4	筹资现金率	−1.8037***	−9.1527	22	四大审计	−0.3797**	−2.5155
5	流动比率	0.4584***	8.8479	23	向子公司贷款水平	0.4634**	2.3157
6	投资现金率	−1.6293***	−8.3371	24	偿债现金率	−0.2363**	−2.3115
7	股权集中度	−0.0151***	−7.1414	25	利息保障倍数	−0.0014**	−2.1012
8	总资产收益率	−3.2922***	−6.4966	26	高管海外背景	−0.4693**	−2.0339
9	机构投资者持股	0.8976***	6.2914	27	向母公司分红水平	−0.3339*	−1.8175
10	集团分权水平	0.5952***	6.1089	28	少数股东权益水平	1.0983*	1.7570
11	集团内关联担保水平	0.4060***	5.9224	29	行业集中度	0.5771*	1.7169
12	高管金融背景	1.1358***	5.8919	30	贷款银行数	−0.0089	−1.5535
13	人均GDP	−0.0357***	−5.6881	⋮		⋮	⋮
14	营业利润率	−0.6628***	−4.5581	年份固定效应		是	
15	利润收现率	−0.0394***	−4.4907	行业固定效应		是	
16	主营业务增长率	−0.5070***	−4.4534	省份固定效应		是	
17	对子公司股权投资水平	0.8288***	4.3375	N		38507	
18	现金比率	−1.4510***	−4.3374	伪R^2		0.3040	

注：*、**、***分别表示10%、5%、1%的显著性水平。个别行业、省份的被解释变量没有变异造成了少量样本损失。

释分析。与Logit模型不同，机器学习模型考虑了指标间的交互关系与非线性关系，因此，色带中存在颜色的交杂。可以观察到，机器学习预测得到的色带大多呈“纺锤形”，绝大多数样本的SHAP值在0点周围密集，说明在机器学习中，解释变量对企业债务危机的影响是有条件的，解释变量只有在特定样本中才产生较强的预测作用。RF模型中，向母公司分红水平是集团特征指标中对预测结果影响最大的指标，在所有指标中重要性排在第4位。图1(c)为XGBoost模型的全局可解释分析。与RF模型类似，XGBoost模型的色带形状也呈“纺锤形”，但颜色的交杂情况更为明显，色带宽度更为均匀。这一定程度上说明，XGBoost模型对样本信息的利用程度更高。对于XGBoost模型，对预测结果影响最大的指标仍为向母公司分红水平，在所有指标重要性排在第7位。

在三种模型中，集团分权水平、对子公司股权投资水平始终出现在重要性排名前30的指标中，说明其对债务危机有重要的影响。①相比传统Logit模型与机器学习模型，向母公司分红水平、向子公司贷款水平与少数股东权益水平是XGBoost模型中排名前30的指标，但在Logit模型中重要性水平却偏低。这说明，其与债务危机的线性相关关系较弱，但有明显的非线性相关关系。

(3)指标组合重要性分析。前文结果仅展示了单一指标的重要性水平，但仍有必要进一步分析集团特征指标在整体上的重要性。根据前文对指标的分类，本文计算了各组指标的平均SHAP值。平均SHAP值越大，表明该类指标更为重要。表4报告了将指标按类别组合后的重要性排名。结果显示，集团特征指标对企业债务危机也有良好的预测能力，以预测效果较好的XGBoost模型为例，集团类指标的重要性居第4位，其重要性高于治理情况指标与高管特征。在Logit模型与RF模型

① 由于Logit回归中排名第29的指标具有显著性，本文大致认为重要性排名前30的指标对预测债务危机有重要价值。

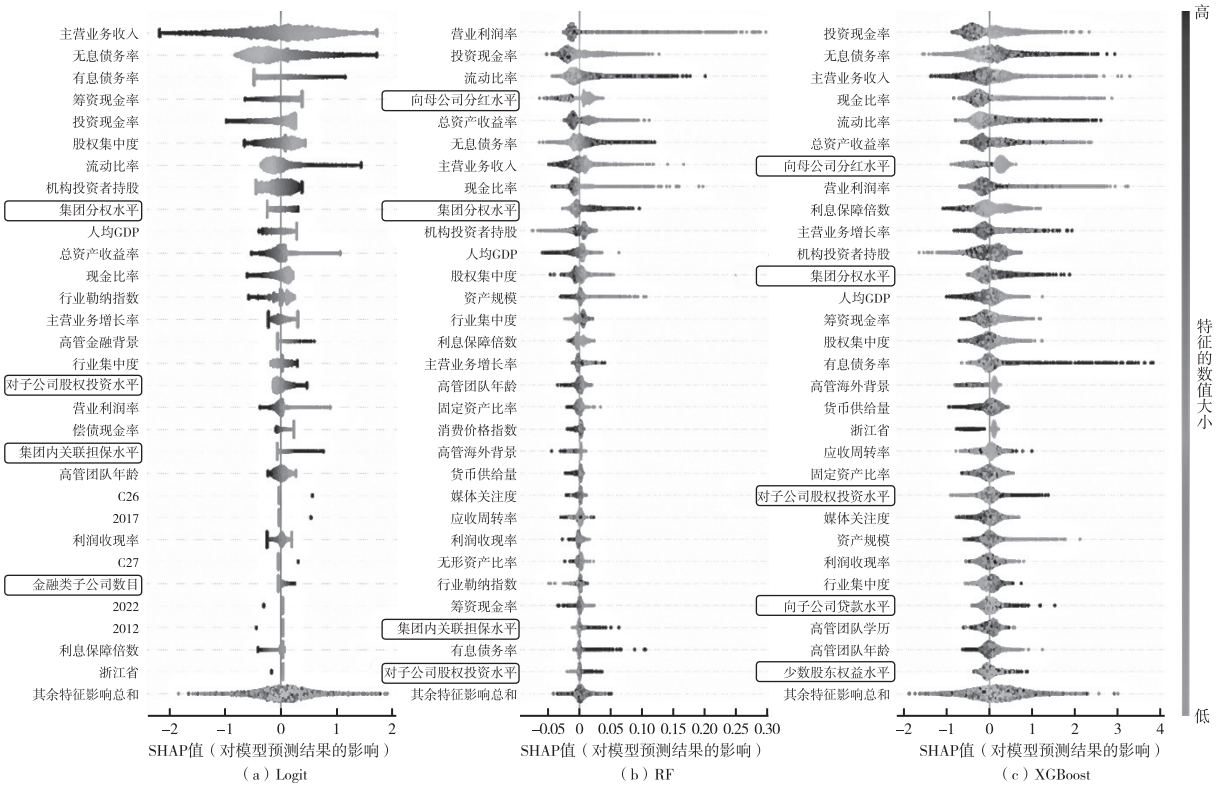

图1　指标全局可解释分析

注:纵轴为预测效果最好的30个指标,指标按其重要性由上至下排列;横轴为SHAP值。每个点代表一个样本,点的颜色越深,说明该样本中相应指标的数值越大;点的位置越偏向右侧,说明该样本中相应指标的SHAP值越大;色带的宽度反映此处堆积的样本的数量。C26、C27等指标为行业虚拟变量指标;2017、2022等指标为时间虚拟变量指标;浙江省等指标为省份虚拟变量。

表4　　指标组合重要性分析

	Logit		RF		XGBoost	
	平均SHAP值	排序	平均SHAP值	排序	平均SHAP值	排序
盈利能力	0.1822	1	0.0119	1	0.2733	1
现金流结构	0.1485	2	0.0072	3	0.2080	3
资产质量	0.1465	3	0.0081	2	0.2350	2
利益相关者	0.0712	4	0.0034	5	0.1218	5
治理情况	0.0614	5	0.0022	7	0.0855	7
集团特征	0.0603	6	0.0045	4	0.1418	4
高管特征	0.0567	7	0.0023	6	0.1175	6
宏观指标	0.0311	8	0.0003	9	0.0224	8
区域指标	0.0150	9	0.0004	8	0.0218	9
行业指标	0.0100	10	0.0001	10	0.0058	10

中，集团类指标的重要性居第6位与第4位。整体而言，集团特征指标的预测效力与治理情况指标、利益相关者指标、高管特征指标相当。后者是实证研究中常见的指标，但前者很少出现在文献中。同时，企业的资产质量类指标、盈利能力指标、现金流类指标对企业的债务危机预测能力较强，是最重要的预测指标，这与已有研究相符。

4.单一指标影响模式分析①

前文讨论了模型整体的有效性与集团特征指标的重要性，接下来将通过SHAP依赖图深入分析各集团特征指标与债务危机的具体关系。

（1）集团分权水平。图2绘制了集团分权水平的SHAP依赖图。从图中可以看出，Logit模型的SHAP依赖图为一条直线，集团分权水平越高，企业出现债务危机的概率越大。RF与XGBoost模型的SHAP依赖图显示出非线性特征。在XGBoost模型中，SHAP依赖图呈“U型”曲线，说明过度集权或过度分权均会引发集团债务危机；当集团分权水平较高时，即分权指数超过0.6（≈0.6×1.00），SHAP值则迅速上升，表明集团过度分权会提升债务危机的发生概率。上述结果与理论分析基本一致，即集团型企业内部母子公司间存在委托—代理问题，当集团分权水平较高时，母公司对子公司机会主义行为的管控水平越低，集团发生债务危机的概率越大。

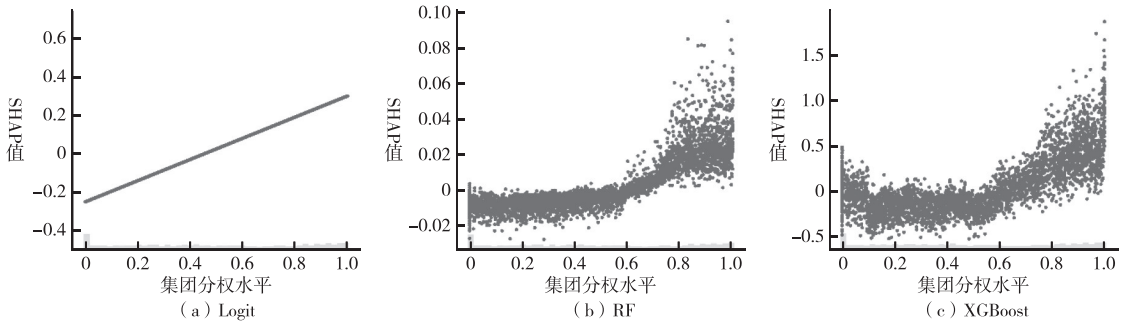

图2 集团分权水平SHAP依赖图

注：纵坐标表示样本在该指标上的SHAP值，数值越大表明样本点越有可能被判定为正例（反之为负例）；横坐标表示该指标的数值，已通过最大最小值标准化处理，取值范围为[0,1]，坐标值越大则指标的数值越大。SHAP依赖图下方的灰色柱形图表示该指标的密度分布，灰色柱越高代表此处堆积的样本数量越多。以下各图同。

（2）向母公司分红水平。图3展示了向母公司分红水平的SHAP依赖图。Logit模型表明，向母公司分红水平越高，企业出现债务危机的概率越低。RF与XGBoost模型的SHAP依赖图显示出非线性特征。当子公司不向母公司分红时，集团发生债务危机的概率显著增加；而一旦子公司开始向母公司分红，债务危机的发生概率会迅速下降。随后，曲线呈现出急剧的反转态势。随着分红水平进一步提升，债务危机的发生概率不再持续降低，反而呈现出上升的趋势。理论上，图像的前半部分表明，子公司具备基本的盈利能力，且母公司有对其现金进行统一分配的制度设计，这反映出企业内部资本市场的有效性；后半部分则表明，若母公司要求的分红水平过高，子公司的正常经营可能会受到影响，增加债务危机的风险。此外，从前文重要性分析中可知，向母公司分红水平在XGBoost模型中是集团特征指标中最重要的指标之一，但其在Logit模型中显著性较低，这表明传统线性模型严重低估了其对债务危机的非线性影响。

① 其余集团特征指标和重要控制变量的SHAP依赖图参见《中国工业经济》网站（ciejournal.ajcass.com）附件。

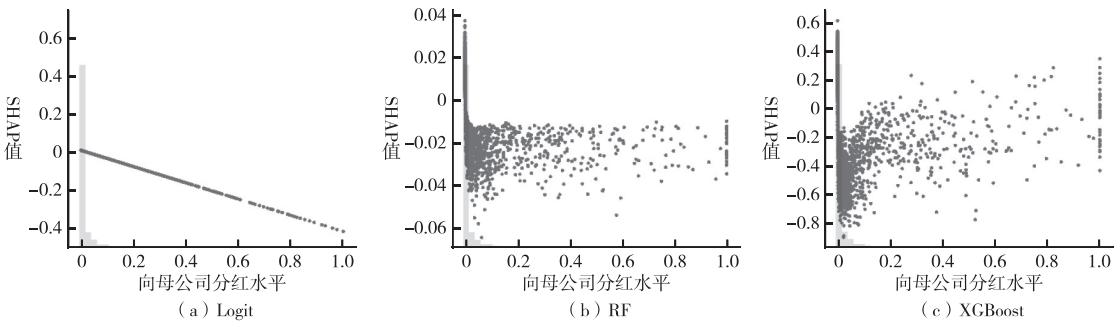

图3　向母公司分红水平SHAP依赖图

(3)少数股东权益水平。图4绘制了少数股东权益水平的SHAP依赖图。在Logit模型中,少数股东权益与债务危机呈正相关关系,但斜率较低,表明线性相关关系不强。但在RF与XGBoost模型中,少数股东权益与其SHAP值存在典型的"U型"关系。当少数股东权益水平低于1.9%(≈0.1×0.19)时,SHAP值逐步减小;当少数股东权益水平在1.9%—9.5%(≈0.5×0.19)之间时,SHAP值接近于0,此时少数股东权益水平对企业债务危机的影响不明显;当少数股东权益高于9.5%时,SHAP值逐渐增加。上述结果与理论分析基本一致,即少数股东权益水平对企业债务危机的影响具有两面性。当企业少数股东权益较低时,合理引入少数股东,能够扩大企业规模、分摊企业风险;而当少数股东的数量过多且规模过大时,这可能暗示企业存在过度扩张的现象,进而增加企业的代理成本、降低治理效率,反而加剧了企业的债务风险。

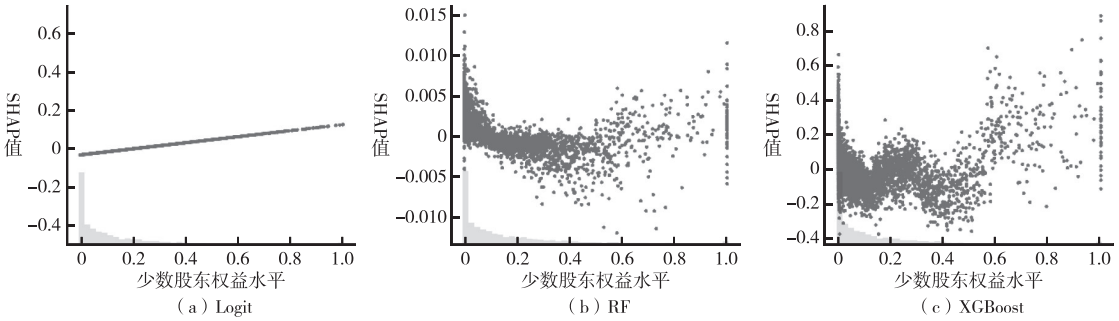

图4　少数股东权益水平SHAP依赖图

5.稳健性检验①

(1)变换预测模型。主检验使用Logit、RF与XGBoost模型进行分析。为不失一般性,本文进一步补充了决策树模型、支持向量机模型、轻度提升机模型、神经网络模型、添加L1惩罚因子的Lasso回归、60因子PCA回归与90%方差PCA回归的稳健性检验。结果显示,这些机器学习模型预测结果与原有模型一致。

(2)时间序列滚动预测。主检验从总体中随机抽取10%样本生成测试集,但这种方法可能向训练集引入未来信息。为防止这一问题,本文参考姜富伟等(2023)的做法,选择[t,t+9]期样本作为训练集,并将t+10期样本作为测试集(t=2007,…,2012),对模型进行了6组重复训练与预测。结果显

① 稳健性检验结果参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

示，模型在各滚动预测期间均有较好预测效力，其中，模型对较近年份债务危机的预测效果更佳，而对较远年份预测效果相对一般。

（3）远期预测效力。主检验仅讨论了集团特征指标对企业未来1期债务危机的预测能力。这里进一步分析了集团特征指标对未来2—5年债务危机预测的重要性水平。结果显示，集团特征指标的重要性水平与内外部治理指标的重要性水平相当，说明集团特征指标对预测远期的债务危机仍然有效。

（4）更换预测目标。主检验中债务危机被定义为企业破产、被ST或退市、债务重组、因债务被诉、因虚构利润被处罚、资不抵债等危机指标的并集，但不同口径的债务危机可能具有不同的形式与成因。因此，本文将债务危机分解为企业是否破产、被ST或退市，是否重组或被诉，是否出现资不抵债，并分别进行预测。结果显示，集团特征指标在3组检验中均有较高重要性，说明其预测能力并不局限于特定类型的债务危机。

（5）局部累积效应图。前文主要通过SHAP依赖图分析集团特征与债务危机之间的非线性关系。参考陈运森等（2024）的方法，本文补充了利用局部累积效应图（ALE图）进行单变量分析的结果。结果显示，ALE图与SHAP依赖图的图像趋势基本一致。

五、进一步研究

本文基于集团母子公司职能分工、集团子公司地理分布和国有企业集团“瘦身健体”三个场景，进一步分析了总部职能定位、信息沟通成本以及企业组织边界对模型预测效力的影响。①

1.集团母子公司职能分工

分工是社会经济系统中的基本概念。随着社会分工的不断细化和市场竞争的加剧，分工现象逐渐演变为企业分工，甚至同一集团内的内部分工。在集团发展壮大的过程中，母公司经历了生产经营主体、行政管理中心和投资控股中心等功能的演变（徐飞，2024）。目前，在西方国家的实践中，总部通常被视为集团的管理和战略核心，这符合集团企业强化分工与管理、降低内部交易成本的需求。然而，许多中国上市公司总部仍通过直属工厂和车间直接参与生产经营活动。母公司不同的分工职能对集团的行为模式和决策效率具有潜在影响。因此，在研究集团特征对企业债务危机的影响时，有必要对企业母子公司间职能分工模式进行分组讨论。

借鉴徐飞（2024）的研究，本文以母公司营业收入占合并报表营业收入比重衡量集团母公司参与实体经营的程度，如果该指标大于样本中位数，认为集团母公司倾向于实体型母公司，反之则为投资型母公司。结果显示，向母公司分红的水平是投资型母公司中最重要的预测指标，但在实体型母公司中却不显著，说明子公司盈利能力对投资型母公司债务危机的预测十分重要。集团的分权水平在实体型母公司中的重要性低于投资型母公司，说明集权或分权管理模式对投资型母公司集团债务风险有重要影响。在投资型母公司与实体型母公司两类不同的集团中，集团特征指标的重要性分别为第4位与第6位，具有较强的预测能力。相对于实体型母公司集团（平均SHAP值0.11），集团特征指标对投资型母公司集团（平均SHAP值0.18）的预测效果更佳。投资型母公司需要制定集团投融资决策并协调集团内部交易行为，这对母公司的信息整合能力和管理权威性提出了更为严苛的要求。此时，母公司对集团的集权管控效果和利润支配能力是确保其决策得以有效执行的关键。相反，实

① 进一步研究的具体结果参见《中国工业经济》网站（ciejournal.ajcass.com）附件。

体型母公司更加专注于自身的业务运营,集团整体规模通常较小,管理决策也相对直接。因此,集团特征指标对实体型母公司集团的债务危机预测能力较弱,而对投资型母公司集团的预测能力更强。

2.集团子公司地理分布

除了强有力的母公司外,集团型企业的形成还依赖于大量子公司的存在。子公司在数量、空间分布和业务类型等方面的差异会显著影响集团的管理方式和控制难度。根据2023年的数据,上市公司设立的子公司平均跨越5.42个省份,异地子公司比例超过60%。这些子公司的位置分布对企业内部沟通难度、管控难度有重要影响。因此,在研究集团特征对企业债务危机的影响时,有必要对企业子公司的位置分布进行分组讨论。本文构建基于集团所属公司省份分布的赫芬达尔—赫希曼指数(*SpaceHHI*)来衡量集团在地理位置上的集中和分散情况:

$$SpaceHHI_{i,t}=\sum_{P=1}^{31}\left(\frac{\sum_{j\in J_p}A_{i,j,t}}{TA_{i,t}}\right)^2 \tag{4}$$

其中,i为上市公司,j为子公司,P为公司所属省份,t为时间。TA为上市公司总资产,A为上市公司旗下子公司的估计资产。①随后,依据*SpaceHHI*是否大于样本中位数将样本分为2组,若观测值*SpaceHHI*较高,表明其子公司在地理位置上集中,反之则分散。

结果显示,向子公司贷款水平和集团内关联担保水平对子公司地理位置分散的集团的债务风险预测作用更强,而在地理位置集中的集团中影响较弱。这表明,即使是在集团内部,跨地区借贷仍具有较高的风险性。在子公司地理位置集中与分散的集团中,集团特征指标的重要性均为第4位。相对于子公司地理位置集中的集团(平均SHAP值0.15),集团特征指标对子公司地理位置分散集团(平均SHAP值0.17)的预测效果更佳。地理位置分散的企业通常面临更高的跨区域沟通成本,不同地区的分支机构之间信息传递效率较低,协调难度增加,从而推高运营和管理成本。此外,代理问题在分散式布局的企业中更为突出,总部对异地子公司的监管难度加大,资源分配精准度下降,资本错配风险上升,进而可能引发债务危机。因此,集团地理分布的分散性对治理模式提出了更高要求,使得集团结构和集团行为对债务风险的影响更加显著,提升了集团特征指标的预测价值。

3.国有企业集团“瘦身健体”

为推动国有企业集团“瘦身健体”,中共中央、国务院于2015年印发《关于深化国有企业改革的指导意见》,提出合理限定法人层级,有效压缩管理层级。2016年,国务院国有资产监督管理委员会出台《关于中央企业开展压缩管理层级减少法人户数工作的通知》(国资发改革〔2016〕135号),提出力争在3年内使多数中央企业管理层级控制在3—4级以内,法人层级10级以上(含10级)的企业减少3—5个层级,企业法人户数减少20%左右。国有企业“瘦身健体”对集团型企业的组织结构与管理模式有深刻的影响:一是管理层级与法人数量的大幅缩减导致组织边界明显收缩;二是组织变革带来了管理结构与控制链条的实质性改变。考察这一组织边界的动态调整过程,一方面能够揭示集团层级结构对其债务风险的作用机制,另一方面有助于侧面评估行政性政策干预对企业运营效率的实际影响效果。本文以国有企业集团样本为基础,将样本划分为“瘦身健体”前与“瘦身健体”后两个子样本。其中,对于中央企业,以2016年作为改革分界点;对于地方国有企业,根据各省份实际推进国有企业“瘦身健体”改革的具体时间进行划分。

① 具体估计方案为:若企业披露了子公司资产,直接使用该数值;若没有,则基于合并报表资产=母公司资产+子公司资产-长期股权投资+商誉的公式计算子公司资产。在此基础上,若企业披露了子公司净资产、净利润或主营业务收入,则择优按其相对于集团公司净资产、净利润或主营业务收入的比例计算。

结果显示，在国有企业“瘦身健体”改革后，孙公司及其以下占比和对子公司股权投资的预测效力显著提升。这说明，通过精简组织架构、消除冗余管理链条，“瘦身健体”改革提高了集团特征指标的信息质量，使其能更有效地预警债务危机。从指标重要性排序看，改革前集团特征指标位列第4，改革后则升至第2位，超越了资产质量和盈利能力指标。相对于“瘦身健体”前（平均SHAP值0.20），集团特征指标对“瘦身健体”后国有企业集团（平均SHAP值0.23）的预测效果更佳。相较于民营企业，国有企业通常集团化程度更高，集团特征指标对其意义也更强。但是，由于国有企业子公司数量众多、组织边界模糊不清，特征指标中常混杂大量与核心业务无关的冗余信息，这会削弱其预测能力。“瘦身健体”改革使国有企业组织结构更加精简、权责边界更为清晰，提升了集团治理效能。此时，集团特征指标的信息质量得到提升，对债务风险的预测价值也相应提高。

六、结论与启示

集团化发展是企业为适应严峻的市场竞争与多变的市场环境而采取的重要战略举措。恰当的资源部署与管控制度有助于提高集团的抗风险能力，反之则可能为集团的发展埋下隐患，导致债务危机的发生。本文从集团资本分布、集团管控模式、集团内部资源分配和集团外部资源撬动四个方面构建了集团特征指标体系，并结合Logit、RF与XGBoost模型开展中国集团型企业债务危机的预测研究。研究发现：①集团特征指标对企业债务危机的发生有重要预测价值，其指标与指标组合的预测效力优于常见的企业高管特征或企业内外部治理特征。其中，集团集权水平、向母公司分红水平、对子公司股权投资水平等指标在各类场景下均能有效预测企业债务危机的发生，具有重要价值；少数股东权益水平、向子公司贷款水平与企业债务危机间有较强的非线性相关关系，尽管在Logit模型中预测效果不佳，但在机器学习模型中是重要的预测指标；其他指标如集团内关联担保水平、金融类子公司数目等对企业债务危机也有预测价值。②相比母公司参与实体经营的企业，集团特征指标对投资型母公司集团的预测能力更强；相比子公司空间密集的企业，集团特征指标对子公司地理位置分散企业的预测能力更强；相比实施“瘦身健体”前的国有企业，集团特征指标对“瘦身健体”后的国有企业预测能力更强。③作为应用SHAP方法的一个典型案例，本文还观察到机器学习模型刻画变量非线性关系与变量交互效应方面的具体方式。一方面，在机器学习模型中，解释变量对企业债务危机的影响是非线性的，只有在少量满足特殊条件的样本中才产生较强的预测作用，绝大多数样本的SHAP值在0点周围密集；另一方面，机器学习模型在训练时考虑了变量之间的交互关系，相同取值的解释变量对被解释变量产生的影响不一定相同，解释变量的SHAP值不仅取决于解释变量的数值本身还受到其他调节因素的影响，这可能是机器学习模型预测精准性较高的原因。

本文为集团型企业相关研究和可解释机器学习在经济管理领域的应用提供了理论启示：①在集团型企业研究领域中，金字塔结构理论与内部资本市场理论具有深刻的内在联系。金字塔结构为内部资本市场的形成提供了组织基础，而内部资本市场的运作依托金字塔式的控制结构，二者分别体现了集团型企业在资本结构与资本行为方面的特征。本文构建的指标体系为理解集团型企业的行为逻辑提供了新的视角，同时提示研究者在指标选择时应注重系统性，多维度分析集团行为，以避免片面化的结论。②传统研究大多采用线性模型分析集团型企业问题，但这种方法存在函数形式误设的潜在可能。本文的实证检验表明，集团特征指标与债务危机之间存在重要非线性关系。这一发现说明，集团行为的复杂性可能超出预期，应重视非线性分析方法的应用，防止遗漏变量，从而更准确地揭示集团型企业行为的内在规律。③计量经济学与机器学习在研究手段和研究目标上

具有一致性,而可解释机器学习的发展为二者融合提供了新的可能。传统实证研究需进一步强化方法的多样性与指标的多元性,而机器学习研究应注重理论深度、指标框架的严谨性以及研究场景的多样性。将计量经济学的理论严谨性与机器学习的计算灵活性相结合,不仅能推动方法论创新和跨学科研究,还能为经济学探索复杂问题开辟新路径,从而更好地应对现实经济挑战。

同时,本文结论对金融监管部门、商业银行以及集团型企业本身也具有重要的实践启示:①企业应合理制定集团化发展路径,不盲目扩大股权投资水平及设立子公司,寻求规模效应与管理效率的平衡点;应保持合理的集团分权水平,积极加强对子公司机会主义行为的管控,充分协调与少数股东之间的关系,促进外来资本为企业健康发展服务;应理性评估子公司的债务融资行为,审慎为子公司提供担保,防止信用风险在集团内部扩散蔓延。②商业银行应充分认识到上市公司的集团属性对其融资行为的影响,在构建信贷风险预测模型过程中,应关注集团型企业母子公司重复融资的行为,切实核查集团信用风险,避免盲目信任集团内母子公司交叉担保,积极开展集团授信、联合授信等措施;应关注上市公司母子公司的管控模式、充分重视集团企业内部资本市场运作等风险事项,提高对集团违约风险的预测与防范能力;对于跨地区经营的企业与母公司投资管理职能更为典型的企业,应特别重视其集团层面的特征,从而提高自身信贷风控能力。③金融监管部门应充分重视大数据和机器学习模型在资本市场监督与管理方面的应用,充分发挥人工智能算法的优势,提高对金融风险的防范能力;应充分利用新型的可解释人工智能方法,预警系统不仅需预测危机是否发生,还需解释危机为什么会发生,才能做到对症下药、防患于未然。同时,政府部门也应积极推进国有企业集团“瘦身健体”,激活国有企业经济活力,并以集团视角加强对国有企业的管控,提高集团运作效率。

可解释机器学习方法是经济管理领域实证研究的重要工具,其对分析变量间非线性相关关系、寻找潜在的变量交互效应、防止回归函数形式误设有重要价值。后续研究可进一步利用可解释机器学习方法,寻找对被解释变量有重要非线性影响的其他影响因素。但需要说明的是,利用机器学习方法进行因果效应检验的方法论尚未成熟,如何将计量经济学领域的研究成果移植至机器学习方法中还有待进一步研究。此外,尽管本文对企业的集团特征进行了初步归纳与分类,但对指标的选择方式并不唯一,后续研究也可根据其他原则重新进行组织。

〔参考文献〕

〔1〕陈艳利,乔菲,孙鹤元.资源配置效率视角下企业集团内部交易的经济后果——来自中国资本市场的经验证据[J].会计研究,2014,(10):28-35.

〔2〕陈运森,周金泳,彭嘉续.中国上市公司分红的动因研究——基于机器学习的证据[J].中国工业经济,2024,(5):155-173.

〔3〕迟国泰,王珊珊,王逸然.基于Stacking方法的上市公司违约风险预警研究[J].系统工程理论与实践,2024,(2):1-25.

〔4〕姜富伟,林奕皓,马甜.“去刚兑”背景下的企业债券违约风险:机器学习预警和经济机制探究[J].金融研究,2023,(10):85-103.

〔5〕雷欣南,林乐凡,肖斌卿,俞红海.小微企业违约特征再探索:基于SHAP解释方法的机器学习模型[J].中国管理科学,2024,(5):1-12.

〔6〕黎文靖,严嘉怡.谁利用了内部资本市场:企业集团化程度与现金持有[J].中国工业经济,2021,(6):137-154.

〔7〕李维安,韩忠雪.民营企业金字塔结构与产品市场竞争[J].中国工业经济,2013,(1):77-89.

〔8〕李焰,陈才东,黄磊.集团化运作、融资约束与财务风险——基于上海复星集团案例研究[J].管理世界,2007,(12):

117-135.
〔9〕李增泉,辛显刚,于旭辉.金融发展、债务融资约束与金字塔结构——来自民营企业集团的证据[J].管理世界,2008,(1):123-135.
〔10〕刘斌,袁利华.土地资源获取、股权投资增加与企业投资过度[J].南开管理评论,2016,(2):75-86.
〔11〕陆正飞,张会丽.所有权安排、寻租空间与现金分布——来自中国A股市场的经验证据[J].管理世界,2010,(5):150-158.
〔12〕纳鹏杰,雨田木子,纳超洪.企业集团风险传染效应研究——来自集团控股上市公司的经验证据[J].会计研究,2017,(3):53-60.
〔13〕潘怡麟,朱凯,陈信元.决策权配置与公司价值——基于企业集团的经验证据[J].管理世界,2018,(12):111-119.
〔14〕汤晟,饶品贵,李晓溪.金融强监管与企业集团内部资本市场资源配置——来自资管新规的经验证据[J].中国工业经济,2024,(1):131-149.
〔15〕汪玉兰,窦笑晨,李井林.集团控制会导致企业过度负债吗[J].会计研究,2020,(4):76-87.
〔16〕王斌,张伟华.外部环境、公司成长与总部自营[J].管理世界,2014,(1):144-155.
〔17〕王琨,陈胜蓝,李晓雪.集团关联担保与公司融资约束[J].金融研究,2014,(9):192-206.
〔18〕王亮亮,施超,阮语,张洪辉.企业集团的决策权配置与实际税负[J].中国工业经济,2023,(6):156-173.
〔19〕王秀丽,张龙天,贺晓霞.基于合并报表与母公司报表的财务危机预警效果比较研究[J].会计研究,2017,(6):38-44.
〔20〕肖争艳,陈衍,陈小亮,陈彦斌.通货膨胀影响因素识别——基于机器学习方法的再检验[J].统计研究,2022,(6):132-147.
〔21〕徐飞.总部功能定位与集团内部职工薪酬[J].中国工业经济,2024,(12):174-192.
〔22〕张会丽,陆正飞.现金分布、公司治理与过度投资——基于我国上市公司及其子公司的现金持有状况的考察[J].管理世界,2012,(3):141-150.
〔23〕郑志刚,朱光顺,袁浩洋,郭杰."大而不倒"幻觉与金字塔控股结构下的"预算软约束"[J].世界经济,2024,(10):152-185.
〔24〕Almeida, H. V., and D. Wolfenzon. A Theory of Pyramidal Ownership and Family Business Groups[J]. Journal of Finance, 2006, 61(6):2637-2680.
〔25〕Beaver, W. H., S. Cascino, M. Correia, and M. F. McNichols. Group Affiliation and Default Prediction[J]. Management Science, 2019, 65(8):3559-3584.
〔26〕Belenzon, S., N. Hashai, and A. Patacconi. The Architecture of Attention: Group Structure and Subsidiary Autonomy[J]. Strategic Management Journal,2019,40(10):1610-1643.
〔27〕Bertomeu, J., E. Cheynel, E. Floyd, and W. Pan. Using Machine Learning to Detect Misstatements[J]. Review of Accounting Studies, 2021,26(2):468-519.
〔28〕Busenbark, J. R., M. Semadeni, M. Arrfelt, and M. C. Withers. Corporate-Level Influences on Internal Capital Allocation: The Role of Financial Analyst Performance Projections[J]. Strategic Management Journal, 2022, 43(1):180-209.
〔39〕Friedman, E., S. Johnson, and T. Mitton. Propping and Tunneling[J]. Journal of Comparative Economics, 2003, 31(4):732-750.
〔30〕Lins, K.V., and H. Servaes. Is Corporate Diversification Beneficial in Emerging Markets[J]. Financial Management, 2002, 31(2):5-31.
〔31〕Lundberg, S., and S. I. Lee. A Unified Approach to Interpreting Model Predictions [EB/OL]. https://doi.org/10.48550/arXiv.1705.07874, 2017.
〔32〕Sun, R., F. Liu, Y. Li, R. Wang, and J. Luo. Machine Learning for Predicting Corporate Violations: How Do CEO Characteristics Matter[J]. Journal of Business Ethics, 2024, 195(1):151-166.

Can Business Group Characteristics Help Predict Corporate Debt Crises: Empirical Evidence Based on Explainable Machine Learning Models

LIU Bin[1], XU Shun[1], LI Hao-ran[2], DONG Jin-ting[1]
(1. School of Economics and Business Administration, Chongqing University;
2. School of Accounting, Chongqing University of Technology)

Abstract: Business groups are economic entities composed of a parent company and its controlling subsidiaries, and most listed companies in China operate as groups. In recent years, many large business groups have experienced debt crises, making the prediction and prevention of debt risk in business groups an urgent problem to address. The internal structure and capital management of these groups significantly influence their debt risk, but few studies incorporate relevant indicators in debt crisis prediction.

Based on pyramid structure and internal capital market theories, this paper develops a framework in four dimensions: capital distribution, control mode, internal resource allocation, and external resource leverage. Additionally, it designs a predictive system with 8 group-specific indicators and 35 general enterprise indicators. Then, it develops debt crisis prediction models based on Logit, Random Forest, and XGBoost models. Using a sample of Chinese A-share listed firms from 2007 to 2023, this paper compares the predictive and interpretive differences between the conventional Logit model and machine learning models, and analyzes the importance ranking of group indicators and their nonlinear relationships with debt crises using the explainable machine learning method—SHAP.

This paper finds that group characteristic indicators have significant predictive power for corporate bankruptcy and default. The predictive effectiveness of these indicators is comparable to conventional metrics such as internal governance, external stakeholder governance, and executive characteristics. Notably, indicators such as the level of decentralization, dividend payouts to the parent company, and equity investment in subsidiaries show a strong predictive ability for debt crises. Some indicators such as minority shareholder equity and the level of loans to subsidiaries exhibit a clear U-shaped relationship with corporate debt crises. These indicators contribute significantly to predicting the debt crises within a machine learning framework, but they are less important in a Logit framework. Additionally, other indicators display varying degrees of nonlinear relationships with debt crises. The above conclusions remain valid after conducting five robustness checks: replacing the machine learning model, using rolling time series forecasting, predicting long-term debt crises, changing the forecasting target, and employing accumulated local effect graphs. Further research analyzes the group indicators from three perspectives: headquarters' functional positioning, communication costs, and organizational boundaries. It finds that indicators on group characteristics are more predictive for groups with investment-type headquarters, geographically dispersed subsidiaries, and state-owned enterprises (SOEs) that have undergone the "Lean-and-Heal Reform", compared to those with entity-type headquarters, centralized subsidiaries, or SOEs before the reform.

The main contributions are as follows. Theoretically, it constructs an analytical framework and indicator system for group characteristic metrics, enriching research on debt crises in group enterprises and deepening the understanding of the nonlinear predictive value of indicators on group characteristics. Practically, it optimizes conventional debt crisis prediction models, providing valuable insights for the selection of indicators, model construction, and methodological approaches in corporate debt crisis forecasting. This offers important empirical evidence to assist enterprises in preventing the occurrence and spread of debt crises and enhance regulatory authorities' efficiency in supervision.

Keywords: business group; indicators on group characteristics; debt crisis; explainable machine learning

JEL Classification: G33 M21 M40

〔责任编辑:崔志新〕