

doi: 10.19920/j.cnki.jmsc.2026.02.007

高层团队三维综合特征视角下的财务违规倾向度量： 基于知识图谱的新指标^①

方立兵¹, 张科^{2*}, 李心丹¹, 李文中³

(1. 南京大学工程管理学院/江苏省数字金融重点实验室, 南京 210093;

2. 南京大学数字经济与管理学院, 苏州 215163; 3. 南京大学计算机科学与技术学院, 南京 210023)

摘要: 董事、监事和高级管理人员(DSE)构成的高层管理团队是公司财务违规的主要责任群体。将“知识图谱嵌入”技术应用于DSE履历综合(个人)基本特征-治理结构-交叉关联三维特征,从“整体观”视角构建新的违规倾向指标(PIFT)。基于Logit或二元Probit模型进行系列实证检验发现:PIFT越高,公司违规概率越大;与若干基于DSE履历少数特定维度的其他指标相比,PIFT具有最高的重要性。进一步检验表明:非独立董事群体对PIFT有效性起主导作用;同行业违规公司比例越高,PIFT的影响越大,但非独立董事的PIFT不受影响;公司总部所在地法治水平越高,随着PIFT增加,违规(被揭露)的可能性越大,但是否国有股权对PIFT的作用没有显著影响。希望这项工作可以为遵循标准金融实证研究范式,将知识图谱嵌入技术应用于公司金融领域建立一个新的研究起点。

关键词: 财务违规; 公司治理; 知识图谱; 机器学习

中图分类号: F832.5 **文献标识码:** A **文章编号:** 1007-9807(2026)02-0123-17

0 引言

加强公司治理、提升上市公司财务报告质量是完善现代企业治理与资本市场高质量发展的核心要求,而董事、监事和高级管理人员(DSE)是影响公司财务违规倾向的关键责任群体。近年来证监会对DSE信息披露违规的处罚力度持续加大,部分案例中主要责任人罚款已从顶格60万元提升至最高单项1000万元(如豫金刚石、宜华生活)。从现有文献来看,大多关注于DSE履历中个人特质相关的“基本特征”如性别、年龄、学历等^[1,2];反映公司“治理结构”的教育工作经历、职位信息等^[3,4];以及体现内外部“交叉关联”的社会网络信息等^[5-9]。

然而,在现代公司治理框架下,重要事项大多

是集体(如股东大会、董事会、监事会等)决策的结果,故从“整体观”视角,综合多个维度分析违规倾向十分必要。但现有理论却各有侧重,如:高层梯队理论强调个体“基本特征”及其异质性对决策的影响^[10,11];欺诈三角理论中的“压力”和“合理化”,与“基本特征”相关“机会”涉及“治理结构”^[12,13];GONE理论新增“暴露”因素,但仍为“治理结构”的体现^[14];社会资本视角关注的“交叉关联”^[5,6,9]则未被前述理论充分涵盖。整合“基本特征”、“治理结构”和“交叉关联”三个维度,测度公司违规倾向,在理论与实践层面均十分必要。理论上,可在“整体观”视角形成新的逻辑与实证结论;实践中,有助于公司避免局部优化导致的整体违规倾向上升,并为监管机构提供新的分类监管指标。近年来上市公司违规行为呈现

① 收稿日期: 2024-11-09; 修订日期: 2025-09-18。

基金项目: 国家自然科学基金资助项目(72371126; 72473063; 72495151; 72071103)。

通讯作者: 张科(1982—),男,江苏南京人,博士,助理教授。Email: kezhang@nju.edu.cn

复杂性、系统性特点,亟需从高管团队整体层面完善监管指标,以优化资源配置、提升监管效率。

鉴于此,本文先在理论层面,系统梳理高层梯队理论、欺诈三角理论以及社会资本理论,提出“三维综合特征相似的公司决策趋同”的理论观点。在具体实证中,综合DSE多维特征面临三个挑战:半结构化履历数据聚合为公司层面指标、数据维度不一致、DSE人数差异导致传统方法失效。现有研究或依赖结构化数据的机器学习模型^[2,15],或采用缺乏可解释性的图学习模型^[16],但均难以支撑标准的线性回归的实证研究范式。为此,本文创新性应用知识图谱嵌入(KGE)技术(详见2.1节),生成公司语义向量(类似自然语言处理领域的向量化,不同的是这里是基于半结构化的知识图谱数据),测度与近期违规公司DSE履历的语义相似性,从而提出“履历蕴含的违规倾向”(PIFT)指标。以2008年—2023年A股上市公司为样本,通过两组Logit基准回归验证PIFT有效性:先纳入治理、财务、市场类基础指标,再加入DSE特定维度的“重点对比指标”(如高管内部关联性^[6,9]、学术型董事占比^[17]等),并采用SHAP方法比较PIFT指标的相对重要性^[2,15]。结果显示PIFT与违规可能性显著正相关且重要性最高。稳健性检验(改用BiProbit、调整PIFT计算方式)均支持PIFT的有效性。进一步分析显示,非独立董事对PIFT有效性起主导作用,从而为现有文献关于我国公司治理的相关研究提供新的补充^[3,4]。异质性分析表明:同行业违规公司占比越高,PIFT与违规概率正相关性越强;国有股权不显著影响该关系;总部所在地法治水平高时,PIFT与违规(被揭露)可能性正相关性更强。

边际贡献可以概括为三个方面:首先,不同于现有文献仅关注DSE履历少数特定维度,在“高层团队三维综合特征”视角下,提出了一种从整体层面度量公司违规倾向的新指标;其次,遵循标准的、基于线性回归的实证研究范式,为如何将“知识图谱嵌入技术”应用于公司金融研究提供参考;最后,为基于“高层团队三维综合特征”视角的公司行为决策,提出了新的理论观点和实证内涵:高层团队综合特征较为相似的公司,其行为

决策趋于一致。

1 理论发展与相关文献评述

通过回顾相关理论与实证文献,可提炼出DSE(个人)基本特征—治理结构—交叉关联三个维度,据此提出“高层团队三维综合特征”视角下的新观点,并阐明本研究对相关领域的补充。首先,理论方面,高层梯队理论强调DSE“基本特征”及异质性对其认知与决策的影响^[10,11];欺诈三角理论指出的“压力”、“合理化”与“基本特征”相关,“机会”涉及“治理结构”(如内控薄弱)^[12,13],例如,年轻的DSE可能较为激进,女性DSE可能更重视合规^[13,18];相比之下,GONE理论进一步扩展为贪婪、机会、需求、暴露四因素,新增的“暴露”仍与“治理结构”相关,故与欺诈三角理论存在一定的重合^[14]。然而,这些理论重点关注DSE的“基本特征”,部分涉及“治理结构”,但“交叉关联”却未引起重视。近期的公司金融文献揭示了内外部社会联系的重要影响,如CEO通过任命建立内部关联增加违规风险^[6],董事网络影响治理有效性等^[5,7-9]。现代公司重要事项大多通过“集体决策”,故整合以上三个维度,从“整体观”视角描述高层团队特征显得十分必要。

拓展上述理论,涵盖“基本特征—治理结构—交叉关联”三个维度,本文提出:高层团队三维综合特征相似时,企业决策趋同。这是因为基本特征相似的管理者,认知与信息处理方式趋同^[10,11];相似的治理结构下,决策过程受到相似的约束与激励^[3,4];相似的交叉关联(如社会网络)导致信息获取与决策受到相似的外部影响^[5,7]。该观点与李维安等^[8]的发现一致,即董事会与CEO契合度高会削弱监督。由此,本文提出:与违规公司三维特征相似的公司违规倾向更高。这为PIFT指标提供了理论逻辑基础,即通过融合上述理论,利用公司层面DSE履历的语义相似性测度公司违规倾向。

实证研究中DSE履历信息在预测公司违规概率及探究违规倾向的影响因素方面备受关注,相关研究方案主要分为两类:一是采用机器学习等复杂方法考查模型对违规概率的整体预

测效力;二是运用 Logit、BiProbit 等标准的回归研究范式,从少数特定维度,分析违规倾向的影响因素。

机器学习范式下,何瑛等^[2]等基于内部治理视角选取公司财务与高管个人特征指标,借助机器学习模型及 SHAP 等工具研究其对违规行为的预测能力、重要性排序及影响模式; Xu 等^[19]比较六种机器学习模型后发现随机森林最优,且基于 GONE 理论的“贪婪”因素,选取的 CEO 年龄、学历等个人特质,显著影响违规倾向。但这类研究仅依据“经验知识”选取少数特定维度指标,无法综合 DSE 履历全维度信息。此外,对定性信息的量化处理,导致语义内涵损失较大;通过 SHAP 方法进行特征重要性分析,难以从指标含义层面给出经济学解释。针对 DSE 履历这类知识图谱数据,李建平等^[16]建议采用图学习模型,但仍然缺乏具有明确经济学含义的综合性指标,故难以为遵循标准的、基于线性回归的金融实证研究范式提供参考。

主流的金融实证文献多采用 Logit 或二元 Probit (BiProbit) 模型,针对少数特定维度检验经济逻辑,同样涉及“基本特征”、“治理结构”、“交叉关联”。“基本特征”方面,除了考虑 DSE 的年龄、性别等个人基本特征外^[1,2],具会计师事务所或学术经历的 DSE 更可能降低违规倾向^[16,17]。“治理结构”方面,Chen^[4]认为良好的董事会规模等治理结构有助于抑制违规。“交叉关联”方面, Khanna 等^[6]发现 CEO 通过任命建立内部关联会增加违规风险; Dikolli 等^[9]指出 CEO 新任命的 CFO 可能加剧盈余管理; Kuang 和 Lee^[7]认为本地社会联系较强的独立董事虽降低欺诈可能性,

但可能延长发现时间。然而,这类实证文献在“高层团队三维综合特征”视角下也面临挑战:难以从 DSE 团队整体层面得出结论,如高层团队特征多样化利于改善决策,凸显整体分析的必要性^[20];难以充分考虑不同维度特征的“协同”影响,如女性 CFO 的保守态度会因学历和涉外经历而更明显^[21];此外,基于特定维度的分析常得出冲突结果,如具政治背景的独立董事可能降低违规概率,但也可能利用政治关系保护公司,减轻监管处罚,从而激励公司从事违规行为,类似冲突在 DSE 内外部关系影响的研究中也存在^[6,22,23]。

与现有文献不同,本文应用知识图谱嵌入技术,有效聚合 DSE 履历全维度信息,从而在“整体观”视角下基于高层团队三维综合特征探究公司违规倾向的内在规律。这在方法论层面有助于克服现有文献从少数特定维度分析的局限,避免因忽视特征协同影响而导致结论冲突;在理论认知层面,可从 DSE 整体特征层面考查违规倾向,为违规行为理论机理提供新的实证内涵。

2 DSE 履历隐含的财务违规倾向

2.1 公司 DSE 履历的知识图谱与嵌入学习

知识图谱由 Google 于 2012 年提出,可有效整合样本公司及相关 DSE 的履历信息,这些信息以“〈头实体,关系,尾实体〉”的三元组形式呈现。例如,“Elon Musk 2018 年担任特斯拉 CEO”可表示为〈Elon Musk,任 CEO 于,特斯拉公司〉。类似的三元组共同构成一个由实体节点(如:人、公司等概念)和关系(如:任职、性别等)连接成如图 1 所示的知识网络。

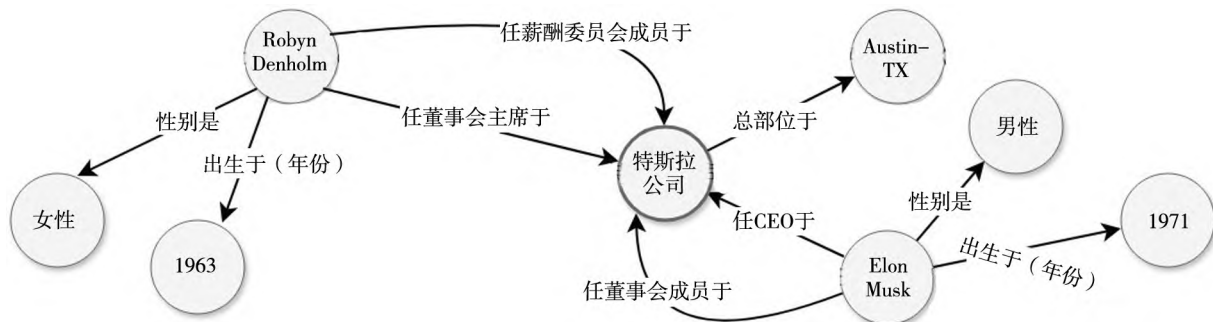


图 1 特斯拉公司部分 DSE 知识图谱 (2018 年)

Fig. 1 Partial DSE knowledge graph of Tesla Inc. (2018)

知识图谱嵌入 (KGE) 技术,也称表示学习,通过预设的语义规则,将知识图谱中所有实体和关系映射为潜空间中的向量.这与 word2vec/doc2vec 算法通过语料库将单词或文档向量化以捕捉语义的思想相似,但 KGE 以半结构化的知识图谱 (DSE 履历) 数据为上下文,生成公司的语义向量,语义相近的公司在向量空间中距离更近,从而为分析公司行为提供丰富的信息.

基于知识图谱融合 DSE 履历的多维语义构造公司层面的财务违规倾向,需完成三个步骤:构建知识图谱、利用 KGE 算法生成公司向量、通过向量相似性计算 *PIFT* 指标.知识图谱构建需先参照标准做法定义三元组模板 (本体层)^[24],再匹配样本数据 (如图 2 所示).从时间维度看,三元组记录了 DSE 的职业发展路径,如职位起止时间;从空间维度看,记录了 DSE 的交叉关联信息,如跨公司任职、校友关系等.

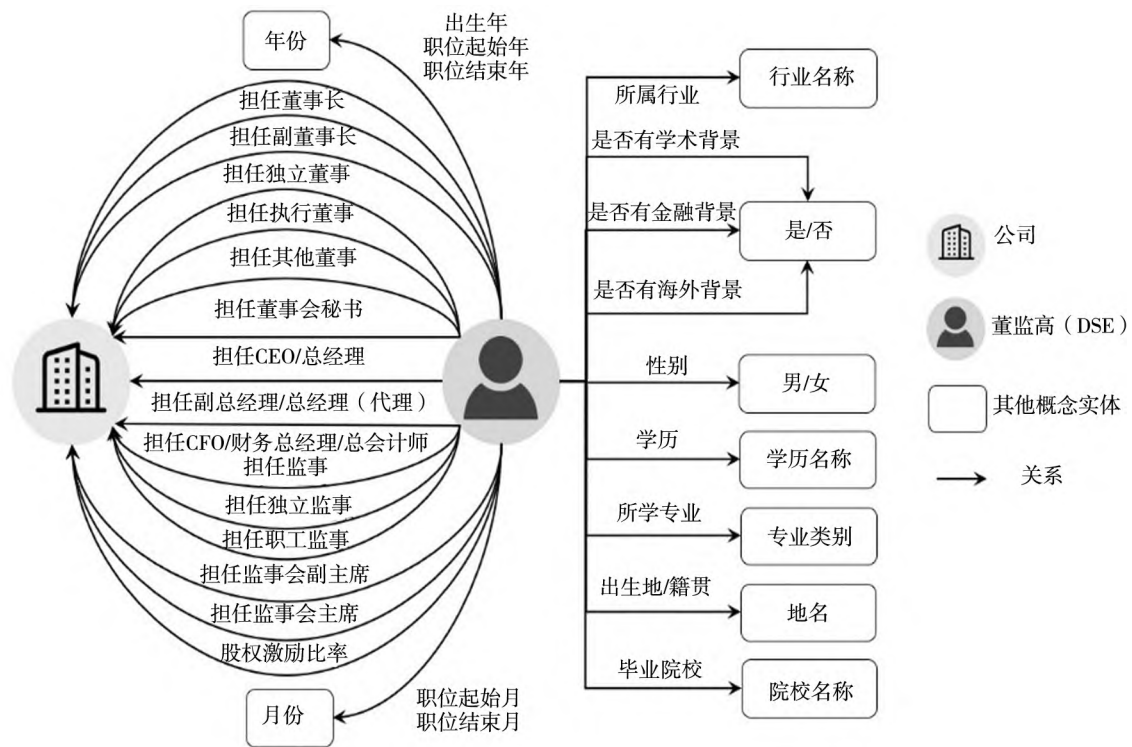


图 2 知识图谱构建的本体层(三元组模板)

Fig. 2 Ontology layer (triple template) of knowledge graph construction

KGE 技术的核心是为实体和关系生成满足语义逻辑的向量.可以采用的算法很多,Wang 等^[25]指出,最初的 TransE 算法基本思想是:对于三元组 $\langle h, r, t \rangle$,头实体向量 h 、关系向量 r 和尾实体向量 t 满足 $h + r \approx t$,其几何含义如图 3 (a) 所示.但该算法仅适用于一对一关系,无法处理多重关系 (如同一人在同一公司担任多个职位).为此,他们提出 TransH 算法:针对每个关系 r 设定对应的超平面 (相应的法向量为 w_r),投影到该超平面的关系向量为 d_r ,将头实体向量 h 和尾实体向量 t 投影到该超平面得到

h_{\perp} 和 t_{\perp} , 满足 $h_{\perp} + d_r \approx t_{\perp}$, 其中 $h_{\perp} = h - w_r^T h \times w_r$, $t_{\perp} = t - w_r^T t w_r$. 该算法通过不同超平面区分共享实体的多重关系 (如图 3 (b) 所示) 通过最小化 $\| | h_{\perp} + d_r - t_{\perp} | \|$ 生成实体和关系的语义表示向量.

就经济逻辑而言,TransH 生成的公司向量可综合三个维度的信息:基本特征维度捕捉 DSE 的个人特质,治理结构维度反映治理角色特征,交叉关联维度编码内外部关系.利用 TransH 算法,使得三维特征相似的公司,在向量空间中距离更近.

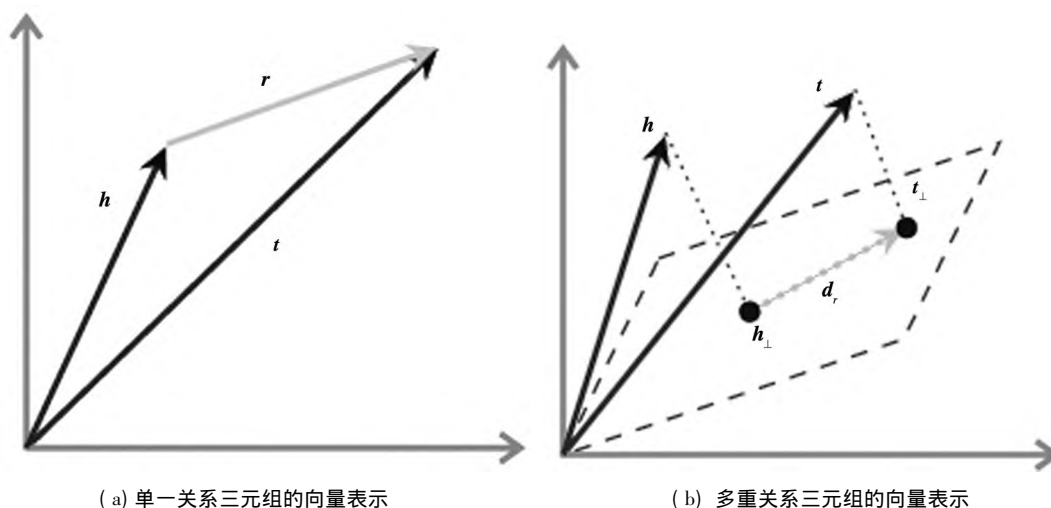


图 3 单一关系和多重关系三元组语义逻辑对应的向量表示和运算规则

Fig. 3 The vector representation and operation rules in the semantic logic of single-relation and multi-relation triples

2.2 DSE 履历蕴含的财务违规倾向

利用 TransH 算法生成的公司向量 聚合了知识图谱中该公司所有 DSE 履历三元组的语义信息. 结合“高层团队三维综合特征相似的公司决策趋同”的理论逻辑, 将 DSE 履历蕴含的财务违规倾向 ($PIFT$) 定义为: 待评估公司向量与近五年被监管确认发生财务违规的公司向量之间的平均余弦相似度. 选择近五年作为样本窗口, 一方面是因为近期的 DSE 履历信息更及时有效, 另一方面是监管机构对违规的处罚公告与实际违规时间间隔通常不超过五年(约占总样本量的 94%). 具体而言, 公司 i 在第 t 年的 $PIFT_{it}$ 定义为

$$PIFT_{it} = \frac{1}{K} \sum_{f_{jt} \in \mathcal{F}_\tau^k} sim(v_{it}, f_{jt}) \quad (1)$$

其中 v_{it} 为公司 i 在 t 年的知识图谱嵌入向量, f_{jt} 为近五年 ($\tau = t, \dots, t-4$) 确认违规公司的向量 (τ 以监管处罚公告年份为准), \mathcal{F}_τ^k 为所有违规公司集合 K 为集合中公司总数, $sim(v_{it}, f_{jt})$ 为两者的余弦相似度. 为提高 $PIFT$ 的截面可比性, 采用 z-score 方法进行标准化, 以解决数据不平衡问题.

3 研究设计与变量说明

实证分析采用 Logit 模型检验 $PIFT$ 指标的有效性. 尽管未被认定为违规的公司(白样本)可能

混杂实际违规但未受处罚的公司(不可观测黑样本), 但事实上, 在这种情况下, 基于 Logit 回归的检验更加保守: 若结果仍显著拒绝零假设 ($PIFT$ 无效), 则更能支持其有效性. 基于 Logit 回归的模型设定为

$$Prob(y_{it} = 1 | X_{i,t-1}, PIFT_{i,t-1}) = \Lambda(c + \alpha X_{i,t-1} + \beta PIFT_{i,t-1} + \varepsilon_{it}) \quad (2)$$

其中 $\Lambda(\cdot)$ 为 Logit 转换函数; 因变量 y_{it} 在公司发生财务违规时取 1, 否则取 0 (年份 t 以监管处罚公告显示的违规年份为准); 关键解释变量 $PIFT_{i,t-1}$ 基于 $t-1$ 至 $t-5$ 期间样本计算, 以规避同期双向因果的内生性; $X_{i,t-1}$ 包含后文的重点对比指标、基础控制变量及行业和年份固定效应虚拟变量.

从现有文献来看, 与 $PIFT$ 相关的关键对比指标包括 5 个: FTA 和 FDA , 分别衡量 CEO 与前四位高管(不含 CEO) 及董事会成员的关联程度(任期内任命比例), 关联越强则违规倾向越高^[6]; 虚拟变量 $CEOCFO$, CEO 任职后新任命 CFO 时取 1, 此时 CEO 对 CFO 的影响可能加剧盈余管理或违规^[9]; $Acadm$ 表示董事会中具有学术背景的董事占比, 该比例越高通常违规倾向越低^[17]; $DirTour$ 为虚拟变量, 存在董事异常离职(非退休、殉职或任期届满) 时取 1, 可能预示公司存在违规风险^[27]. 这 5 个重点对比指标将在基准回归中逐一加入, 以对比 $PIFT$ 的有效性, 并采用

SHAP方法考察其相对重要性。该方法是基于博弈论构造,为每个解释变量分配Shapley值以衡量其对目标函数的贡献,满足可加性、对称性、一致性和公平性。尽管Logit模型的相对机率(Odds ratio)具有线性化表达,但考虑到变量的取值尺度和估计误差等问题,无法直接通过系数比较重要性;而SHAP方法可基于模型整体拟合优度给出一致的全局性评价。模型还纳入文献中常见的三类基础控制变量^[2,16],包括26个基础控制变量,与上述5个重点对比指标共同构成31个控制变量,以缓解遗漏变量的内生性问题。其中,与公司治理相关的控制变量以及其他基础控制变量^②。

4 样本描述

所有样本数据均从CSMAR数据库获取,包括2008年—2023年沪深A股所有上市公司。以2008年为样本起始期,是因为我国2007年开始实行的《企业会计准则》大幅修订,使得中国的企业会计准则更加接近国际财务报告准则(IFRS),且2007年颁布的《上市公司信息披露管理办法》对企业财务违规的认定发生了重大变化。上市公司违规类型众多,参考现有文献^[22]将虚构利润、虚列资产、虚假记载(误导性陈述)、推迟披露报告、信息披露不实(其他)、重大遗漏等判定为财

务报告违规(欺诈)行为。最终共收集55 226个“公司-年度”样本,包括4 635家上市公司,其中12 943个“公司-年度”样本发生违规,涉及2 263家公司。

4.1 DSE 履历知识三元组

根据第3.1节定义的知识图谱,样本数据共整理出15种实体。按图2所述模板,在以三元组形式组织所有实体及其关系前,需进行“实体对齐”,即处理多个名称指代同一实体的情况(如“北京大学”与“PKU”指向同一所大学),故使用R软件的`adist`函数对实体名称归一化:大学名称参考中国“211工程”大学及QS世界大学排名前200的官方名称建立标准库,若数据中名称与标准名称的余弦相似度超过0.9,则替换为标准名称;地区名称依据行政区划法定术语统一。由于DSE所学专业名称无序且缺乏统一标准,故参照中国教育部2021年《普通高等学校本科专业目录》,手工整理所有专业名称并归入相应学科门类,包括哲学、经济学、法学、教育学、文学、历史学、理学、工学、农学、医学、管理学和艺术学。

4.2 变量的描述性统计

表1报告了6个主要变量的描述性统计,包括第3.2节的关键解释变量`PIFT`以及5个重点对比指标。其他与公司治理相关的控制变量以及托宾Q、杠杆率和销售增长率等财务指标,股票换手率、收益波动率等市场指标相应的描述性统计^②。

表1 主要变量的描述性统计

Table 1 Descriptive statistics of the main variables

变量名称	全样本				违规和非违规样本均值及二者之差		
	均值	标准差	最小值	最大值	违规样本	非违规样本	差异
<code>PIFT</code>	-0.006	0.979	-2.885	1.622	0.158	-0.031	0.190***
<code>FTA</code>	0.341	0.340	0.000	1.000	0.375	0.336	0.039***
<code>FDA</code>	0.468	0.321	0.000	1.000	0.494	0.464	0.030***
<code>Acdm</code>	0.208	0.119	0.000	0.571	0.194	0.211	-0.017***
<code>CEOCFO</code>	0.462	0.499	0.000	1.000	0.513	0.454	0.059***
<code>DirTovr</code>	2.799	6.523	0.000	30.769	3.647	2.675	0.972***

注:1)均值差及其t检验是双样本检验,并且考虑到了每个样本内不同的样本量和方差;2)***、**和*分别表示1%、5%和10%水平显著。

② 有需要者可向作者索要。

表1 右边三列报告违规和非违规样本均值, 以及二者之间的差异及其显著性(双样本 t -检验)。其中, 违规公司的 $PIFT$ 、 FTA 、 FDA 、 $CEOCFO$ 和 $DirTour$ 显著较高, 而 $Acdm$ 显著较低。可见, 违规公司的 CEO 与高管之间联系更紧密, CFO 在 CEO 上任后任命的比例更高, 外部董事异常离职率也显著较高; 从 DSE 学术背景来看, 违规样本 $Acdm$ 平均值在统计意义上显著较低。这些发现与现有文献的相关结果是类似的, 如参考文献 [6, 9, 17, 27]。

与虚拟变量 $CEOCFO$ 和 $DirTour$ 不同, $PIFT$ 、 FTA 、 FDA 和 $Acdm$ 为连续变量。从表1 的均值差异难以明确其与违规倾向的关系, 故进一步对这

四个变量按取值范围等距分组 (FTA 为离散变量, 直接按取值分组; FDA 和 $Acdm$ 因存在较多 0 值, 单独设为一组), 考查不同组中违规样本的分布。结果如表2 所示。可以看出, 随着 $PIFT$ 增大, 违规样本比例上升, 表明 $PIFT$ 与公司财务违规倾向显著正相关, 与理论预期一致。同时, 公司违规倾向随 FTA 或 FDA 增加而上升, 与 Khanna 等^[6] 对美国公司的研究结果吻合; 仅 $FTA = 1$ 时违规样本比率下降, Khanna 等^[6] 认为可能因 FTA 与监管稽查概率负相关, 而违规样本仅包含被曝光处罚的公司。此外, $Acdm$ 与违规倾向显著负相关, 与 Francis 等^[17] 结论一致。

表2 变量取值的样本分布

Table 2 The sample distribution of the variable values

	组内样本总量	违规样本量	违规样本占比/%
$-3 < PIFT \leq -2$	1 418	140	8.986
$-2 < PIFT \leq -1$	4 136	479	10.379
$-1 < PIFT \leq 0$	9 608	1 315	12.039
$0 < PIFT \leq 1$	13 464	2 124	13.626
$1 < PIFT < 2$	4 687	1 011	17.743
$FTA = 0.00$	16 602	1 851	11.149
$FTA = 0.25$	9 159	1 252	13.670
$FTA = 0.53$	7 587	1 040	13.708
$FTA = 0.75$	4 862	717	14.747
$FTA = 1.00$	5 524	743	13.450
$FDA = 0.00$	6 567	704	10.720
$0 < FDA \leq 0.25$	6 553	799	12.193
$0.25 < FDA \leq 0.50$	9 776	1 261	12.899
$0.50 < FDA \leq 0.75$	10 904	1 486	13.628
$0.75 < FDA \leq 1.00$	7 082	978	13.810
$Acdm = 0$	3 127	351	11.224 8
$0 < Acdm \leq 0.1$	3 055	493	16.137
$0.1 < Acdm \leq 0.2$	15 359	2 154	14.024
$0.2 < Acdm \leq 0.3$	12 465	1 620	12.996
$0.3 < Acdm \leq 0.4$	5 310	603	11.356
$0.4 < Acdm \leq 0.5$	2 064	188	9.109
$0.5 < Acdm < 0.6$	1 188	84	7.071

综合表1 和表2 的单变量统计结果可以看出, $PIFT$ 及 FTA 、 FDA 、 $Acdm$ 、 $CEOCFO$ 、 $DirTour$ 均与公司违规概率存在明确相关关系, 这说明后5 个指标有必要作为 $PIFT$ 的重点对比指标, 进而从多个角度检验 $PIFT$ 作为整体性指标的有效性。

5 实证结果与分析

5.1 基准回归结果

表3 报告了不引入重点对比变量的基准回归结果, 包括两组回归: 1) 仅含三类基础控制变量

的“模型1”；2)在模型1基础上加入关键解释变量 *PIFT* 的“模型2”。可以看出,控制所有基础变量后, *PIFT* 与违规概率显著正相关,表明 *PIFT* 越高,公司违规倾向越强。利用 SHAP 方法,将 *PIFT* 与同样源自 *DSE* 履历的治理指标(董事长兼任总经理 *CEOChair*、独立董事比例 *IndepDir*、董事会规模 *BrdSize*、审计委员会中非独立董事占比 *IndepAud*、审计委员会规模 *BrdAudit*) 进行重要性比较(图4)结果显示 *PIFT* 在这6个指标中重要性最高(平均 Shapley 绝对值最大,见图4(b)),且 *PIFT* 取值越大对模型影响越显著(见图4(a))。对数似然比(LR)检验也表明,模型2的对数似然

值显著高于模型1,说明引入 *PIFT* 能更有效评估公司违规概率。

表3中其他控制变量结果与现有文献一致^[1,2]。如:内部治理方面,董事会会议频率(*BrdMeet*)与违规概率正相关,可能因频繁会议反映决策官僚化或掩盖违规行为,董事长兼任总经理(*CEOChair*)也显著增加违规概率;外部监督方面,聘请四大会计师事务所(*Big4*)、机构投资者持股集中度(*InstOC*)及分析师覆盖(*Analyst*)均显著抑制违规;行业环境方面,行业集中度(*IndusCon*)与违规概率负相关,而公司违规比例与其所属行业违规比例(*IndusFraud*)正相关。

表3 *PIFT* 对公司违规概率的影响

Table 3 The impact of *PIFT* on the probability of corporate misbehavior

解释变量名称	模型1	模型2	解释变量名称	模型1+	模型2+
	是否违规	是否违规		是否违规	是否违规
<i>PIFT</i>		0.092*** (0.015)	<i>EbitDA</i> □	-0.017*** (0.006)	-0.016*** (0.006)
<i>CEOChair</i> ◇	0.169*** (0.034)	0.168*** (0.034)	<i>SaleG5Y</i> □	0.074 (0.055)	0.057 (0.055)
<i>IndepDir</i> ◇	0.194 (0.156)	0.233 (0.156)	<i>Larg</i> □	0.037*** (0.011)	0.036*** (0.011)
<i>InstOC</i> ◇	-0.553*** (0.090)	-0.568*** (0.090)	<i>OCFRatio</i> □	0.064 (0.049)	0.064 (0.049)
<i>BrdSize</i> ◇	0.198*** (0.047)	0.185*** (0.047)	<i>OCFVola</i> □	0.064*** (0.010)	0.066*** (0.010)
<i>BrdMeet</i> ◇	0.789*** (0.043)	0.779*** (0.043)	<i>dREC</i> □	0.201*** (0.065)	0.190*** (0.064)
<i>IndepAud</i> ◇	0.277** (0.130)	0.286** (0.130)	<i>dInvent</i> □	-0.033 (1.003)	-0.067 (0.985)
<i>BrdAudit</i> ◇	-0.187*** (0.032)	-0.179*** (0.032)	<i>dCSale</i> □	-0.027 (0.022)	-0.028 (0.022)
<i>EqIncent</i> ◇	-1.191*** (0.203)	-1.177*** (0.202)	<i>dNIBE</i> □	0.026 (0.026)	0.024 (0.026)
<i>EqCon</i> ◇	-1.249*** (0.112)	-1.183*** (0.113)	<i>IndusQ</i> □	-0.012 (0.251)	-0.007 (0.251)
<i>Big4</i> ◇	-0.762*** (0.093)	-0.763*** (0.093)	<i>IndusCon</i> □	-0.225*** (0.075)	-0.210*** (0.075)
<i>Analyst</i> ◇	-0.159*** (0.008)	-0.158*** (0.008)	<i>TOvr</i> ○	-0.063** (0.028)	-0.058** (0.028)
<i>IndusFraud</i> ◇	0.081*** (0.003)	0.081*** (0.003)	<i>Vola</i> ○	-0.232 (0.393)	-0.176 (0.385)
<i>TbnsQ</i> □	-0.091* (0.048)	-0.088* (0.048)	常数项	-5.106*** (0.229)	-5.175*** (0.230)
样本量				44 261	44 261
Pseudo- <i>R</i> ²				0.077	0.078
ln <i>L-Value</i>				-16 133	-16 115.0
LR(<i>df</i> = 1) 检验					35.952***

注: 1) 对数似然比检验(LR 检验)比较了模型1和模型2的对数似然值; 2) 变量名称上标,◇、□和○,分别标记三类基础控制变量,即第4节的公司治理指标、财务指标和市场指标; 3) ***、**和*分别表示在1%、5%和10%水平显著。因控制变量较多,表中“模型1+”列和“模型2+”列分别为模型1和模型2结果的延续,以提升可读性。

进一步将 *PIFT* 与 5 个重点对比指标 (*FTA*、*FDA*、*Acadm*、*CEO CFO*、*DirTour*) 进行比较。因 *FTA*、*FDA* 和 *CEO CFO* 相关性较高 (*FTA* 与 *FDA* 相关系数 71.3% ,*CEO CFO* 与前两者约 44%) ,故逐一加入回归; *Acadm* 和 *DirTour* 与其他变量相关性较低 (<10%) ,则始终保留在模型中。表 4 奇数列报告仅含重点对比指标及三类基础控制变量的回归结果,偶数列是进一步引入 *PIFT* (与其他四个指标相关系数约 7% ~ 8% ,没有共线性问题) 的结果,基础控制变量结果与表 3 类似。

表 4 奇数列显示,在表 3 “模型 1”基础上引入重点对比指标后,模型拟合效果显著改善 (*LR* 检验均在 1% 水平显著) ,且 *Acadm* 与违规概率显著负相关, *FTA*、*FDA* 和 *CEO CFO* 与违规概率显著正相关,均与文献结果一致^[6,9,17,27]。在此基础上引入 *PIFT* 后,模型拟合效果进一步显著提升 (*LR* 检验均在 1% 水平显著,见偶数列) ,且 *PIFT* 在所有回归中均显著为正。采用 SHAP 方法比较变量重要性(图 5) ,结果一致显示 *PIFT* 的重要性最高,与图 4 仅考虑基础治理指标的结果一致。

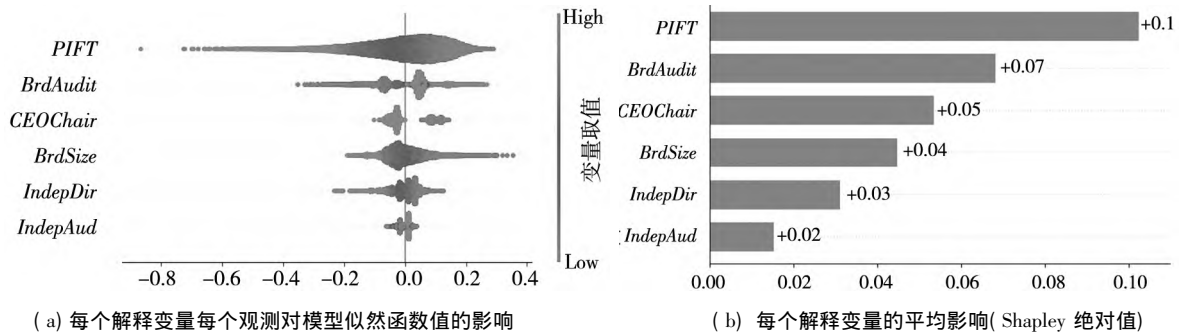


图 4 公司治理相关变量的重要性比较 (SHAP 方法)

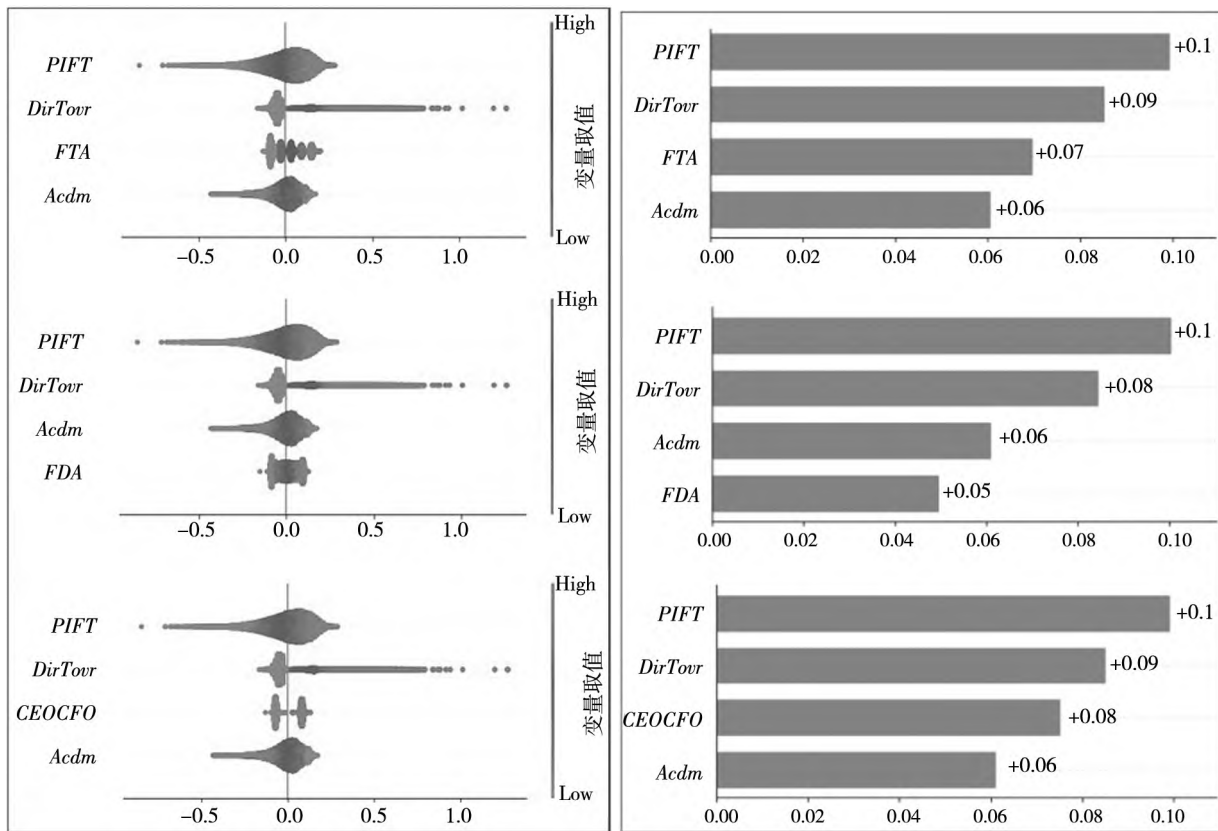
Fig. 4 Comparison of the importance of governance-related variables (using SHAP method)

表 4 *PIFT* 与重点对比指标的结果对比

Table 4 Comparison of results between *PIFT* and important comparable indicators

	(1)	(2)	(3)	(4)	(5)	(6)
<i>PIFT</i>		0.087 *** (0.016)		0.088 *** (0.016)		0.086 *** (0.016)
<i>FTA</i>	0.233 *** (0.043)	0.227 *** (0.043)				
<i>FDA</i>			0.173 *** (0.047)	0.165 *** (0.047)		
<i>CEO CFO</i>					0.119 *** (0.030)	0.110 *** (0.030)
<i>Acadm</i>	-0.445 *** (0.130)	-0.459 *** (0.130)	-0.456 *** (0.130)	-0.470 *** (0.130)	-0.446 *** (0.130)	-0.460 *** (0.130)
<i>DirTour</i>	0.017 *** (0.002)	0.017 *** (0.002)	0.017 *** (0.002)	0.017 *** (0.002)	0.018 *** (0.002)	0.017 *** (0.002)
控制变量	是	是	是	是	是	是
Pseudo. <i>R</i> ²	0.080	0.081	0.079	0.080	0.079	0.080
ln <i>L-Value</i>	-16 078.4	-16 062.4	-16 086.2	-16 070.0	-16 084.9	-16 069.3
LR (<i>df</i> = 1) 检验	109.128 ***	32.078 ***	93.501 ***	32.398 ***	96.226 ***	31.242 ***

注：该表对比 *PIFT* 与第 3 节讨论的 5 个重点对比指标的 Logit 回归结果。其中,列 (1)、列 (3)、列 (5) 的似然比 (LR) 检验以仅包含三类基础控制变量时 Logit 模型的对数似然值 (-16 133.0,见表 3 “模型 1 +”列) 为基准。列 (2)、列 (4)、列 (6) 的似然比 (LR) 检验分别以列 (1)、列 (3)、列 (5) 的对数似然值为基准。***、** 和 * 分别表示在 1%、5% 和 10% 水平显著。



(a) 每个解释变量每个观测对模型似然函数值的影响

(b) 每个解释变量的平均影响(Shapley 绝对值)

图5 PIFT 与 FTA、FDA、Acdm、CEOCFO、DirTour 的重要性比较(SHAP 方法)

Fig. 5 Comparing the importance of PIFT with FTA ,FDA ,Acdm ,CEOCFO ,and DirTour(SHAP Method)

最后,汇总表3和表4,以及图4和图5的结果,可以认为:首先 PIFT 与公司违规概率呈显著的正相关,且与其他基于 DSE 履历少数特定维度构造的公司治理指标相比,PIFT 指标的加入不仅能够显著改进模型的拟合优度(Pseudo. R^2),而且在控制其他变量的影响下,PIFT 对模型拟合优度的影响均具有最高的 SHAP 重要性.

5.2 稳健性检验

5.2.1 基于二元 Probit 模型的回归分析

第 5.1 节采用的 Logit 模型默认所有未受到监管处罚曝光的公司为非违规公司.然而,可能存在部分公司实际发生违规行为,但未受到监管稽查而未被处罚曝光,从而造成潜在的模型设定偏差.参考现有文献^[1,6],使用 BiProbit 模型可在一定程度上加以缓解.篇幅所限,具体的模型设定(有需要者可向作者索要).需要说明的是,以 Bi-Probit 模型进行检验,无法避免实证结果受到协

变量(“违规发生”和“监管稽查”的协变量集合)是否严谨的困扰^[1].同时,因为缺少“未违规”受到“监管稽查”和“发生违规”但没有被“监管稽查”的样本,故建模时,只能使用相同的变量标识“违规发生”和“监管稽查”,从而带来额外的估计偏差.因此,仅在本小节用 BiProbit 模型做稳健性检验,其他部分的实证研究仍然沿用经典的 Logit 模型.

采用极大似然法对 BiProbit 模型进行参数估计的结果如表 5 所示.与表 4 对比可见,基于 Bi-Probit 模型的回归结果中,PIFT 与公司“违规发生”概率仍然呈显著的正相关关系(如表 5 Panel A 所示),且模型拟合优度显著改进.此外,各个解释变量的 BiProbit 模型的重要性(平均的 SHAP 绝对值, $m(|SHAP|)$)显示,PIFT 指标相较于其他 5 个可比指标依然对 BiProbit 模型违规发生概率的影响最重要.

表 5 稳健性检验: 基于 BiProbit 模型的回归
Table 5 Robustness test: Regression based on the BiProbit model

	BiProbit 回归		$m(SHAP)$	BiProbit 回归		$m(SHAP)$	BiProbit 回归		$m(SHAP)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A “违规发生”概率									
<i>PIFT</i>		0.088*** (0.018)	0.084		0.101*** (0.020)	0.083		0.103*** (0.020)	0.085
<i>FTA</i>	0.392*** (0.070)	0.359*** (0.072)	0.070						
<i>FDA</i>				0.348*** (0.070)	0.319*** (0.069)	0.046			
<i>CEO CFO</i>							0.157*** (0.042)	0.127*** (0.041)	0.072
<i>Acdm</i>	-0.057 (0.221)	-0.208 (0.180)	0.044	-0.100 (0.204)	-0.103 (0.200)	0.045	-0.127 (0.209)	-0.083 (0.199)	0.046
<i>DirTour</i>	0.007** (0.003)	0.011*** (0.003)	0.065	0.007** (0.003)	0.007** (0.003)	0.069	0.010*** (0.003)	0.009*** (0.003)	0.068
控制变量	是	是		是	是		是	是	
Panel B “监管稽查”概率									
<i>PIFT</i>		0.021* (0.011)			-0.045** (0.021)			-0.049** (0.022)	
<i>FTA</i>	-0.254*** (0.065)	0.041 (0.031)							
<i>FDA</i>				-0.212*** (0.064)	-0.197*** (0.063)				
<i>CEO CFO</i>							-0.086* (0.044)	-0.052 (0.041)	
<i>Acdm</i>	-0.239 (0.223)	-0.360*** (0.102)		-0.335 (0.210)	-0.342 (0.211)		-0.244 (0.234)	-0.355* (0.214)	
<i>DirTour</i>	0.001 (0.003)	0.006*** (0.002)		0.004 (0.003)	0.004 (0.003)		0.000 (0.003)	0.003 (0.003)	
控制变量	是	是		是	是		是	是	
ln L-Value LR(2) 检验	-15 752.2	-15 745.3 13.848***		-15 686.4	-15 665.8 41.26***		-15 725.6	-15 709.5 32.274***	

注: 1) 列(2)、列(5)、列(8)的似然比(LR)检验分别以列(1)、列(4)、列(7)的对数似然值为基准,而 BiProbit 回归包括两个方程,故自由度 $df=2$;
2) ***、**和* 分别表示在 1%、5%和 10%水平显著;3) 列(3)、列(6)、列(9)报告平均的 SHAP 绝对值($m(|SHAP|)$) 用于对比各个变量的重要性。

5.2.2 调整 PIFT 计算方法

首先,违规行为可能存在“惯性”,样本中持续两年的数量占为 31.27%。这是因为一次年报的财务违规/欺诈带来的“漏洞”未来通常还需要至少一次“修补”,因此也被称为“系列违规”^[26]。在心理学层面,若管理者实施财务违规未及时受到处罚,受心理因素影响,会认为该行为相对恰当,后续更可能继续,体现出惯性违规的特征^[13]。为此,本小节将式(1)中“近五年所有违规公司”调整为:1) 待评估公司在过去任何年份没有因财务违规收到监管处罚;2) 待评估公司的高层管理团队若曾在任何已知违规公司的任职经历,则从知识图谱基础数据中剔除其对应的知识条目(三元组),并重新生成公司的向量表示。基于这

些规则尽可能避免违规行为惯性的影响后,依据式(1)计算平均余弦相似度,记为 $nPIFT$ 。

其次,式(1)利用余弦相似度的平均值计算 $PIFT$,可能受到极端值的影响,故调整为余弦相似度的中位数(记为 $mPIFT$)

$$mPIFT_{it} = \text{median}\left(\bigcup_{f_{jr} \in F_r} \text{sim}(v_{it}, f_{jr})\right) \quad (1')$$

其中 $\text{median}(\cdot)$ 取集合的中位数, \cup 表示并集;其他符号含义与式(1)相同。

$nPIFT$ 和 $mPIFT$ 的回归结果如表 6 所示。与表 4 对比可见,调整 $PIFT$ 的计算方法后,无论用哪种计算方法的 $PIFT$ 指标,仍然可以显著改进模型的拟合优度。此外,平均的 SHAP 绝对值($m(|SHAP|)$)也显示,两种方法计算的 $PIFT$ 指标相较于其他 5 个重点对比指标依然对模型具有最重要的影响。

表6 稳健性检验：排除惯性违规(*nPIFT*)、取余弦相似度中位数(*mPIFT*)
Table 6 Robustness test: Elimination of inertia misbehavior (*nPIFT*), and taking median of cosine similarity (*mPIFT*)

	Logit 回归 (1)	<i>m</i> (<i>ISHAP1</i>) (2)	Logit 回归 (3)	<i>m</i> (<i>ISHAP1</i>) (4)	Logit 回归 (5)	<i>m</i> (<i>ISHAP1</i>) (6)	Logit 回归 (7)	<i>m</i> (<i>ISHAP1</i>) (8)	Logit 回归 (9)	<i>m</i> (<i>ISHAP1</i>) (10)	Logit 回归 (11)	<i>m</i> (<i>ISHAP1</i>) (12)
<i>nPIFT</i>	0.083*** (0.022)	0.088			0.084** (0.019)	0.087			0.083** (0.019)	0.087		
<i>mPIFT</i>			0.044*** (0.013)	0.087			0.043*** (0.012)	0.085			0.042*** (0.013)	0.088
<i>FTA</i>	0.234*** (0.042)	0.058	0.233*** (0.043)	0.060								
<i>FDA</i>					0.175*** (0.046)	0.039	0.172*** (0.047)	0.034				
<i>CEOFCFO</i>									0.121*** (0.029)	0.060	0.117*** (0.030)	0.058
<i>Acadm</i>	-0.452*** (0.130)	0.023	-0.460*** (0.130)	0.019	-0.463*** (0.130)		-0.471*** (0.130)	0.019	-0.453*** (0.130)		-0.460*** (0.130)	0.016
<i>DirTour</i>	0.018*** (0.002)	0.048	0.017*** (0.002)	0.045	0.017*** (0.002)	0.044	0.017*** (0.002)	0.047	0.018*** (0.002)		0.017*** (0.002)	0.047
控制变量	是	是	是	是	是	是	是	是	是	是	是	是
Pseudo. <i>R</i> ²	0.081	0.080	0.080	0.080	0.078	0.078	0.080	0.078	0.078	0.078	0.080	0.080
<i>ln L-Value</i>	-16 067.69		-16 074.20		-16 072.68		-16 082.10		-16 076.34		-16 080.90	
LR(1) 检验	21.492***		8.528***		27.136***		8.332***		17.080***		7.982***	

注：1) 所有奇数列报告 Logit 回归的结果，似然比(LR)检验分别以表4列(1)、列(3)、列(5)的对数似然值为基准，基于 Logit 模型新增 1 个变量时的自由度 *df*=1；2) 偶数列报告平均的 SHAP 绝对值(*m*(*ISHAP1*))，用于对比各个变量的重要性；3) **、* 和 * 分别表示在 1%、5% 和 10% 水平显著。

6 进一步的实证分析

6.1 谁是主要群体?

本小节在“高层团队三维综合特征”视角下进一步考查 DSE 群体中,谁是影响 $PIFT$ 有效性的主要群体.这不仅为探讨 $PIFT$ 有效性给出进一步分析;还可以为相关主体的治理效应在“高层团队三维综合特征”的视角下带来新的认知.通常认为拥有更多独立董事的董事会监督效力更强.因此,中国证券监督管理委员会出台了关于董事会中独立董事比例和职责的相关规定.然而,独立董事的实际监督作用却不及预期^[3].主要原因在于,公司大多出于监管要求,而策略性地任命外部董事,这就削弱了公司治理的有效性.同样,公

司的监事会也面临类似困境^[4].独立董事监督效力不足、作用有限的情况并非仅存在于中国.即使在资本市场和监管体系相对发达完善的美国,独立董事的治理效果也相对有限^[6].此外,从公司的权力结构来看,内部董事(尤其是执行董事和股权董事)比高级管理人员具有更高的决策权.但是,这些结论都是基于少数特定维度构造指标进行检验的,并未综合考虑各类治理主体的自身“基本特征”及其“交叉关联”的相关因素.鉴于此,考虑三种知识图谱:仅非独立董事(执行董事和股权董事等)的履历构造 $PIFT_D$;仅高级管理人员的履历构造 $PIFT_E$;仅独立董事和监事的履历构造 $PIFT_{IDS}$.针对 $PIFT_D$ 、 $PIFT_E$ 和 $PIFT_{IDS}$ 的检验,可以在“高层团队三维综合特征”的视角下检验相关主体的治理效应.

表 7 三种群体的知识图谱构造 $PIFT_D$ 、 $PIFT_E$ 、 $PIFT_{IDS}$

Table 7 Three types of knowledge graph construction: $PIFT_D$, $PIFT_E$, $PIFT_{IDS}$

	(1)	$m(SHAP)$	(2)	$m(SHAP)$	(3)	$m(SHAP)$
$PIFT_D$	0.087 *** (0.017)	0.091	0.087 *** (0.017)	0.091	0.087 ***	0.090 (0.017)
$PIFT_E$	0.045 *** (0.017)	0.064	0.047 *** (0.017)	0.066	0.045 ***	0.063 (0.017)
$PIFT_{IDS}$	0.015 (0.016)	0.037	0.014 (0.016)	0.037	0.014 (0.016)	0.037
FTA	0.224 *** (0.043)	0.069				
FDA			0.161 *** (0.047)	0.051		
$CEOCFO$					0.108 *** (0.030)	0.072
$Acadm$	-0.493 *** (0.131)	0.061	-0.504 *** (0.131)	0.062	-0.494 *** (0.131)	0.063
$DirTour$	0.017 *** (0.002)	0.085	0.017 *** (0.002)	0.084	0.017 *** (0.002)	0.085
控制变量	是		是		是	
Pseudo- R^2	0.081		0.081		0.081	
ln L-Value	-16 051.9		-16 059.4		-16 058.6	
LR 检验	53.092 ***		53.712 ***		52.522 ***	

注: 1) $PIFT_D$ 、 $PIFT_E$ 和 $PIFT_{IDS}$ 分别表示基于以下三种知识图谱计算的 $PIFT$: 仅非独立董事(执行董事和股权董事等)、仅高级管理人员、仅独立董事和监事; 2) 列(1)、列(2)、列(3)为逐个加入 FTA 、 FDA 和 $CEOCFO$ 的模型估计结果; 3) 所有 $m(|SHAP|)$ 列报告平均的 SHAP 绝对值,用于对比各个变量的重要性; 4) 似然比(LR)检验分别以表 4 列(1)、列(3)、列(5)的对数似然值为基准,基于 Logit 模型新增 3 个变量时的自由度 $df=3$; 5) ***、** 和 * 分别表示在 1%、5% 和 10% 水平显著.

检查 $PIFT_D$ 、 $PIFT_E$ 、 $PIFT_{IDS}$ 之间的相关性发现,相关系数均在 20% ~ 30% 之间,虽然有一定

相关,但并不会带来较大的共线性问题.故可以将 $PIFT_D$ 、 $PIFT_E$ 和 $PIFT_{IDS}$ 同时放入模型(2)中进行

检验。另外,参照表 4 的做法,逐个放入 *FTA*、*FDA* 和 *CEOCFO*,而保持 *Acdm* 和 *DirTour* 始终在模型中,以便与三种 *PIFT* 指标做进一步比较。结果如表 7 所示:三种知识图谱构造的 *PIFT_D*、*PIFT_E*、*PIFT_{IDS}* 仅 *PIFT_D*、*PIFT_E* 指标显著为正,且 *PIFT_D* 具有最高的 *SHAP* 重要性。此外,5 个重点对比指标 (*FTA*、*FDA*、*CEOCFO*、*Acdm* 和 *DirTour*) 依然与表 4 保持一致。这就进一步从“高层团队三维综合特征”视角下,揭示了执行董事和股权董事等内部董事是影响公司违规倾向的主要群体。

6.2 异质性分析

本节从三个方面进行异质性分析:近五年确认违规公司中与待评估公司同行业的占比 (*Ind-Fraud5*)、公司是否为国有股权、总部所在地区的法治水平 (*LawLevel*)。首先,违规行为在同行业公司可能有“同伴效应”,即相互模仿,尤其存在共同董事时,该效应更强^[5,28]。“优势比较”理论认为,同行业违规会强化其他公司对违规的感知并合理化其行为^[13],故同行业违规占比越高,*PIFT* 与违规概率的关系可能越强。其次,国企因承担政治目标和受严格监管,违规动机较低,但国企 DSE 可能通过政治联系实施违规^[3,22],故国有股权是

潜在的异质性因素。最后,法治水平也有两种可能的结果:高法治水平可能提高违规成本、保护举报人,从而抑制违规;但在 DSE 特征相似时,违规行为在高法治地区更容易被揭露,使 *PIFT* 与违规概率的正相关性更强。鉴于此,通过以下回归考察这些因素的影响

$$Prob(y_{it} = 1) = \Lambda(\beta_0 + \beta_1 PIFT_{i,t-1} + \beta_2 PIFT_{i,t-1} \times H_{i,t-1} + \beta X_{i,t-1} + \varepsilon_{it}) \quad (3)$$

其中 $X_{i,t-1}$ 为控制变量(同式(2)); $PIFT_{i,t-1}$ 取全部 DSE 构造的 *PIFT*(式(1))和非独立董事构造的 *PIFT_D*(参考第 6.1 节结果); $H_{i,t-1}$ 依次为 *Ind-Fraud5*、国有股权虚拟变量 (*StateOwn*, 是否含国有股权)、法治水平虚拟变量 (*LawLevel*, 公司总部所在地法律环境指数是否高于均值)。法律环境指数由中国国家经济研究所 (NERI) 发布。结果如表 8 所示。可以看出, *IndFraud5* 和 *LawLevel* 对 *PIFT* (*PIFT_D*) 与违规概率的关系有显著正向影响,即:同行业违规占比高时,全 DSE 构造的 *PIFT* 与违规概率的正相关性更强 (*PIFT_D* 不受此影响,体现其独立性);高法治水平地区, *PIFT* 越高,违规越易被揭露,导致违规概率更高;而国有股权 (*StateOwn*) 无显著调节作用。

表 8 *PIFT* 对违规概率影响的异质性分析

Table 8 Heterogeneity analysis on the impact of *PIFT* on the probability of misconduct

	所有 DSE (<i>PIFT</i>)			仅非独立董事 (<i>PIFT_D</i>)		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>PIFT</i>	0.038* (0.022)	0.100*** (0.016)	0.039* (0.021)	0.082*** (0.023)	0.109*** (0.016)	0.076*** (0.022)
<i>PIFT</i> × <i>IndFraud5</i>	0.415*** (0.118)			0.164 (0.119)		
<i>PIFT</i> × <i>StateOwn</i>		-0.033 (0.024)			-0.033 (0.022)	
<i>PIFT</i> × <i>LawLevel</i>			0.123*** (0.031)			0.060* (0.031)
控制变量	是	是	是	是	是	是
Pseudo- <i>R</i> ²	0.081	0.080	0.081	0.080	0.080	0.081
ln <i>L-Value</i>	-16 048.1	-16 053.4	-16 046.2	-16 050.0	-16 050.0	-16 049.1

7 结束语

通过梳理现有文献关于高层团队特征对公

司违规倾向影响的理论观点与实证内涵,从“个人基本特征-公司治理-交叉关联”三个维度提出:高层团队三维综合特征相似的公司,违规倾向趋于一致。为检验这一观点,创新性应

用知识图谱嵌入技术,聚合公司所有 DSE 履历的多维语义信息,利用待评估公司与已知违规公司的语义相似性,构造反映“高层团队三维综合特征”的违规倾向指标(*PIFT*)。以 2010 年—2023 年我国上市公司为样本,基于 Logit 和 Bi-Probit 模型进行系列检验(含排除“违规行为惯性”的内生性处理等),结果显示:*PIFT* 越高,公司违规概率越高;与现有文献 5 个基于特定维度的重点对比指标相比,*PIFT* 的 SHAP 重要性最高。进一步检验不同 DSE 构成主体发现,非独立董事对 *PIFT* 的有效性起主导作用。异质性分析表明:同行业违规公司占比会强化 *PIFT* 与违规概率的正向关系,但对非独立董事的 *PIFT* 无显著影响;公司总部位于高法治水平地区时,*PIFT* 升高更易使违规行为被揭露,表现为更高违规概率;国有股权属性对 *PIFT* 与违规概率的关系无显著影响。

目前鲜有文献在“高层团队三维综合特征”视角下探讨公司违规倾向的理论与度量方法。理论上,DSE 的个人特征(基于高层梯队理论)、公司治理结构(基于公司治理理论)及内外部交叉关联(基于社会资本等理论),共同构成高层团队

的全面描述。实践中,*PIFT* 的价值体现在两方面:一是有助于公司在平衡高层管理团队的个人特征时,避免“按下葫芦浮起瓢”(因局部优化而导致其他问题的出现)提供一个更全面的、“整体观”视角下的预警指标;二是为监管机构提供新维度,助力精准识别高违规倾向公司,提升监管效率,减少对低违规倾向公司的不必要干预,降低社会成本。需注意的是,*PIFT* 是分类监管指标体系的补充,需与其他指标综合使用,不能作为违规事实的识别依据或法律判决依据。

PIFT 不仅为评估公司财务违规倾向提供了新的综合性指标,还为聚合 DSE 履历多维语义、探究非独立董事主导作用等结论提供了新方案。该方案可拓展至公司金融其他领域,例如,在“整体观”视角下,检验高层团队综合特征相近的公司是否具有相似的创新能力;此外,利用知识图谱嵌入技术对数据维度不一致和缺失值处理的灵活性,可在控制多维综合特征的基础上,探究特定维度(如校友联结、内部培养与外部空降等)对公司创新的影响。希望本研究能为知识图谱嵌入技术在遵循标准金融实证范式的公司金融研究建立一个新的起点。

参考文献:

- [1] Barton F J, Burnett B M, Gunny K, et al. The importance of separating the probability of committing and detecting misstatements in the restatement setting [J]. *Management Science*, 2024, 70(1): 32–53.
- [2] 何瑛,任立祺,于文蕾,等. 公司和高管特征与上市公司违规行为——基于机器学习的经验证据 [J]. *管理科学学报*, 2024, 27(6): 43–68.
He Ying, Ren Liqi, Yu Wenlei, et al. Company and executive characteristics and corporate violations: Empirical evidence based on machine learning [J]. *Journal of Management Sciences in China*, 2024, 27(6): 43–68. (in Chinese)
- [3] Jiang F, Kim K A. Corporate governance in China: A modern perspective [J]. *Journal of Corporate Finance*, 2015, (32): 190–216.
- [4] Chen T. Institutions, board structure, and corporate performance: Evidence from Chinese firms [J]. *Journal of Corporate Finance*, 2015, (32): 217–237.
- [5] 谢德仁,陈运森. 董事网络: 定义、特征和计量 [J]. *会计研究*, 2012, (3): 44–51.
Xie Deren, Chen Yunsen. Director network: Definition, characteristics and measurement [J]. *Accounting Research*, 2012, (3): 44–51. (in Chinese)
- [6] Khanna V, Kim E H, Lu Y. CEO connectedness and corporate fraud [J]. *The Journal of Finance*, 2015, 70(3): 1203–1252.
- [7] Kuang Y F, Lee G. Corporate fraud and external social connectedness of independent directors [J]. *Journal of Corporate Finance*, 2017, (45): 401–427.
- [8] 李维安,李晓琳,张耀伟. 董事会社会独立性与 CEO 变更——基于违规上市公司的研究 [J]. *管理科学*, 2017, 30(2): 94–105.

- Li Wei'an, Li Xiaolin, Zhang Yaowei. Board social independence and CEO changes: A study based on non-compliant listed companies [J]. *Management Science*, 2017, 30(2): 94–105. (in Chinese)
- [9] Dikolli S S, Heater J C, Mayew W J, et al. Chief financial officer co-option and chief executive officer compensation [J]. *Management Science*, 2021, 67(3): 1329–1992.
- [10] Hambrick D C, Mason P A. Upper echelons: The organization as a reflection of its top managers [J]. *Academy of Management Review*, 1984, 9(2): 193–206.
- [11] Carpenter M A, Geletkanycz M A, Sanders W G. Upper echelons research revisited: Antecedents, elements, and consequences of top management team composition [J]. *Journal of Management*, 2004, 30(6): 749–778.
- [12] Cressey D. *Other People's Money: A Study in the Social Psychology of Embezzlement* [M]. New York: Free Press, 1973.
- [13] Brown T J. Advantageous comparison and rationalization of earnings management [J]. *Journal of Accounting Research*, 2014, 52(4): 849–876.
- [14] Bologna G J, Lindquist R J, Wells J T. *The Accountant's Handbook of Fraud and Commercial Crime* [M]. New York: John Wiley, 1993.
- [15] 陆瑶, 张叶青, 黎波, 等. 高管个人特征与公司业绩——基于机器学习的经验证据 [J]. *管理科学学报*, 2020, 23(2): 120–140.
Lu Yao, Zhang Yeqing, Li Bo, et al. Managerial individual characteristics and corporate performance: Evidence from a machine learning approach [J]. *Journal of Management Sciences in China*, 2020, 23(2): 120–140. (in Chinese)
- [16] 李建平, 孙灏, 常闫芃, 等. 考虑审计要素多重语义关联的财务欺诈识别 [J]. *管理科学学报*, 2024, 27(3): 58–70.
Li Jianping, Sun Hao, Chang Yanpeng, et al. Financial statement fraud identification considering the multiple-dimensional semantic associations of auditing elements [J]. *Journal of Management Sciences in China*, 2024, 27(3): 58–70. (in Chinese)
- [17] Francis B, Hasan I, Wu Q. Professors in the boardroom and their impact on corporate governance and firm performance [J]. *Financial Management*, 2015, 44(3): 547–581.
- [18] Murphy P R, Dacin M T. Psychological pathways to fraud: Understanding and preventing fraud in organizations [J]. *Journal of Business Ethics*, 2011, 101(4): 601–618.
- [19] Xu X, Xiong F, An Z. Using machine learning to predict corporate fraud: Evidence based on the GONE framework [J]. *Journal of Business Ethics*, 2023, 186(1): 137–158.
- [20] Adams R B, Almeida H, Ferreira D. Powerful CEOs and their impact on corporate performance [J]. *The Review of Financial Studies*, 2005, 18(4): 1403–1432.
- [21] Luo J h, Peng C, Zhang X. The impact of CFO gender on corporate fraud: Evidence from China [J]. *Pacific-Basin Finance Journal*, 2020, (63): 101404.
- [22] Wang Y, Ashton J K, Jaafar A. Financial statement fraud, recidivism and punishment [J]. *Emerging Markets Review*, 2023, (56): 101033.
- [23] Chen D, Wang F, Xing C. Financial reporting fraud and CEO pay-performance incentives [J]. *Journal of Management Science and Engineering*, 2021, 6(2): 197–210.
- [24] Suarez-Figueroa M C, Gomez-Perez A, Motta E, et al. *Ontology Engineering in a Networked World* [M]. New York: Springer Publishing Company, 2012.
- [25] Wang Z, Zhang J, Feng J, et al. Knowledge Graph Embedding by Translating on Hyperplanes [C]//AAAI'14: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. Québec City: AAAI Press, 2014: 1112–1119.
- [26] Bao Y, Ke B, Li B, et al. Detecting accounting fraud in publicly traded U. S. firms using a machine learning approach [J]. *Journal of Accounting Research*, 2020, 58(1): 199–235.
- [27] Gao Y, Kim J B, Tsang D, et al. Go before the whistle blows: An empirical analysis of director turnover and financial fraud [J]. *Review of Accounting Studies*, 2017, 22(1): 320–360.
- [28] Tong X, Yang B. The peer effects of corporate social responsibility performance on innovation outputs: Evidence from China

[J]. Journal of Management Science and Engineering , 2025 , 10(3) : 311 – 331.

Measuring fraud tendency from the three-dimensional composite characteristics of top management teams: A new indicator based on knowledge graph

FANG Li-bing¹ , ZHANG Ke^{2*} , LI Xin-dan¹ , LI Wen-zhong³

1. School of Management and Engineering , Nanjing University / Jiangsu Key Laboratory of Digital Finance , Nanjing 210093 , China;
2. School of Digital Economics and Management , Nanjing University , Suzhou 215163 , China;
3. School of Computer Science and Technology , Nanjing University , Nanjing 210023 , China

Abstract: The top management team , comprising directors , supervisors , and executives (DSEs) , is the primary group responsible for corporate financial misconduct. By applying knowledge graph embedding (KGE) to DSE profiles , this paper constructs a novel fraud propensity indicator (*PIFT*) that captures semantic similarity across three dimensions—(personal) basic traits , governance structure , and cross ties—from a holistic perspective. Empirical tests using Logit and binary Probit models reveal that Higher *PIFTs* are associated with a greater likelihood of financial violations , and that , compared to single-dimension indicators , *PIFT* demonstrates superior predictive power. Further analyses show that non-independent directors play a dominant role in driving *PIFT*' s effectiveness , that the proportion of fraudulent firms in the same industry strengthens *PIFT*' s impact , and that higher regional legal enforcement enhances the link between *PIFT* and fraud risk , while state ownership remains insignificant. This study establishes a new starting point for applying knowledge graph embedding in corporate finance research , following the standard empirical financial research paradigm.

Key words: financial misreporting; corporate governance; knowledge graph; machine learning