

上市公司违规与年报文本信息操纵

谭建华¹, 王雄元²

(1. 南京审计大学 政府审计学院, 江苏 南京 211815;

2. 中南财经政法大学 会计学院, 湖北 武汉 430073)

摘要:上市公司违规不仅使财务报告的可信度急剧下降,而且严重损害了投资者的合法权益。以2007—2019年A股上市公司为研究样本,采用文本分析法识别年报非财务信息的“距离”“难度”“情感”等可视化信息,探讨上市公司在违规后对年报文本信息的操纵行为。研究发现:上市公司在发生违规后,公司年报文本信息的相似度显著降低、复杂度显著提高、语调显著乐观。同时,上述文本特征的系统性变化主要体现在违规之心较强、机会较多、需求较强烈以及被暴露概率较小组,表明公司利用年报文本信息精心掩饰违规是一种策略性行为。此外,上市公司精心掩饰违规行为减弱了短窗口市场反应、加剧了信息漂移现象、提高了股价崩盘风险。

关键词:上市公司违规;非财务信息;文本分析;文本信息操纵

中图分类号:F272.3

文献标识码:A

文章编号:1005-0566(2022)03-0099-13

Corporate Fraud and Manipulation of Annual Report Text Information

TAN Jianhua¹, WANG Xiongyuan²

(1. School of Government Audit, Nanjing Audit University, Nanjing 211815, China;

2. School of Accounting, Zhongnan University of Economics and Law, Wuhan 430073, China)

Abstract: Fraud of listed companies not only sharply reduce the credibility of financial reports, but also seriously damage the legitimate rights and interests of investors. Using 2007 - 2019 A-share listed companies as a research sample, taking text analysis to identify visual information such as “distance”, “difficulty”, “emotion” of the annual report non-financial information, and discussing the company’s response to the annual report text after a listed company violates regulations. This study find that after a listed company violated regulations, the similarity of the company’s annual report text information decreases, the complexity increases, and the tone becomes more optimistic. At the same time, the regular changes in the above-mentioned text characteristics are mainly reflected in the groups with strong violating intentions, more opportunities, strong demand, and low exposure probability, indicating that it is a strategic behavior for the company to use annual report text information to carefully conceal violations. In addition, the meticulous concealment of violations by listed companies has weakened the short-window market response, aggravated the phenomenon of post-earnings announcement drift, and increased the stock price crash risk.

Key words: corporate fraud; non-financial information; text analysis; text manipulation

收稿日期:2021-12-20 修回日期:2022-02-20

基金项目:国家自然科学基金青年项目“基于多模态机器学习的公司财务舞弊研究:识别、侦测及风险指数构建”(72102110);教育部人文社会科学研究一般项目“人工智能应用对企业成本管理的影响研究:基于成本粘性的视角”(21YJC630123)。

作者简介:谭建华(1991—),男,土家族,湖北恩施人,南京审计大学政府审计学院讲师,博士,研究方向为文本分析、信息披露、企业创新。

基于机器学习的多模态融合技术发展,不仅促进了人工智能的进步和大数据的应用,而且会使社会科学的实证研究进入一个新拐点。针对非财务信息的文本分析已成为公司行为研究的热门话题,单独的财务定量信息已不再具有价值相关性,只有结合非财务信息才对公司行为具有解释力。但是,由于非财务信息没有固定的披露模式,其可靠性和可理解性总体上要弱于财务信息,这无疑制约了非财务信息价值的发挥。文本分析主要是通过机器学习方法提炼文本规律,增强资本市场对非财务信息的理解,有助于充分发挥非财务信息的价值相关性。如公司年报文本信息净语调越积极,越能够发挥信息心理学的框架效应机制^[1]。在预测企业未来业绩时,乐观的管理层语调有利于投资者作出积极反应^[2]。企业经济绩效对管理者的驱动作用对企业年报可读性水平存在显著性影响^[3]。年报中的特质性信息越多,公司未来股价崩盘风险越低^[4]。此外,管理层出于自利动机则会操纵年报文本信息复杂性^[5]。

上市公司违规现象历来是财务会计领域关注的焦点,近年来中国上市公司违规丑闻不断涌现,如瑞幸咖啡“财务造假”、獐子岛“扇贝事件”等,遍布各行各业。公司违规不仅使财务报告的可信度急剧下降,而且严重损害了投资者的合法权益,不利于资本市场的有序发展。上市公司是否会利用非财务信息文本操纵掩饰其违规行为?资本市场又如何识别?这些都是缺乏经验证据的重要问题。现有文献验证了非财务信息文本特征与违规密切相关,年报文本相似度可用于侦测公司违规行为^[6],语言因素结构^[7]以及文本主题内容^[8]对公司违规具有预测作用。但是,现有研究均未直接验证文本信息操纵可用于掩饰公司违规行为,均未系统考察公司违规背景下文本特征的变化,也未区分违规程度与违规动机的强弱,更未探讨资本市场能否识别公司利用文本信息操纵精心掩饰公司的违规行为。

首先,过于直观的文本信息操纵反而可能引发市场警觉并暴露上市公司违规行为。第一,降低文本相似度,提供更多合理性的解释或者用海

量信息分散或淹没投资者对公司违规的关注,比通过重复重要或不重要的信息以提高文本相似度,更符合公司掩饰违规行为的逻辑。第二,文本信息可读性较高可以向投资者传递正面信号,体现未来盈余持续性,公司股价更稳定,从而带来更加积极的市场反应^[9]。同时,较高的文本信息可读性还可以吸引投资者的关注,使分析师投入更多关注度^[10]。然而,复杂的文本对投资者的决策带来一定干扰^[11]。此外,公司违规后,某些指标可能会出现异常,管理层需要进行合理的解释,使得文本信息更加晦涩难懂,更有利于掩饰违规行为。第三,市场参与者通常对乐观语言、情绪或语调作出积极反应,对悲观语调的反应较温和^[12],因而采用乐观语调比悲观语调更有利于掩饰违规行为^[13]。上述分析表明:上市公司违规后,非财务信息文本的相似度可能降低、复杂性可能提高以及语调可能趋于乐观。

其次,上市公司违规后年报 MD&A 文本特征的变化只是相关关系,还是公司利用文本信息操纵精心掩饰公司违规的行为体现?掩饰违规动机下的文本信息操纵是公司信息披露策略性行为的体现,违规对公司的效用越大,公司越有可能利用文本信息操纵精心掩饰违规。然而,单个文本特征与违规的关系不一定具有因果逻辑的关系。因此,违规后年报 MD&A 文本特征的变化是否为公司的精心掩饰行为,需要从违规效用函数上判断。基于单个文本特征的视角探究公司违规行为,并未综合考虑企业多模态信息,因而解释力有限,不能有效识别公司违规后的年报文本信息操纵行为。本文预期公司在发生违规后,年报文本信息将发生系统性变化。因此,需要从多模态视角对年报文本信息进行情感分析、距离分析以及难度分析,以探究公司违规后对年报文本信息的操纵行为。当公司违规需求越强烈,暴露的概率越小,又有违规之心与可乘之机,公司的违规行为就会发生,而且公司会利用年报文本信息操纵加以掩饰。

最后,资本市场如何识别公司利用年报非财务信息文本操纵精心掩饰上市公司违规行为?公司利用文本信息操纵精心掩饰违规行为,会暂时性迷惑市场参与者对信息的理解,增加资本市场

对此行为的识别难度,延缓了资本市场对公司违规的识别进程。同时,公司利用文本信息操纵掩饰违规的过程也是坏消息积累的过程,而资本市场对这种行为的识别过程则是坏消息逐渐释放的过程。这些分析表明,资本市场需要较长时间才可能有效识别上市公司的精心掩饰行为。

上市公司系统性操纵年报文本特征才能达到掩饰违规行为的效果。掩饰违规动机下的文本操纵是公司信息披露策略性行为的体现,仅仅观察单个文本特征变化并不能较好理解公司的文本信息操纵策略。单个文本特征与违规的关系不一定具有因果逻辑的关系,孤立看待单个文本特征与违规的关系并不能真正理解年报文本特征变化的含义。

一、理论分析与研究假设

(一)违规前后年报非财务信息文本信息的变化规律

第一,文本距离计算是自然处理领域中的研究重点,占据基础性位置。作为一种语义空间,文本具有高维的特征,要想从数学视角出发对其相似性进行量化分析,前提是实施抽象分解。因此,文本距离通常采用文本相似度衡量。重复披露重要信息可以提供增量信息、强化促进信息使用者对公司价值的正确判断,从而引发积极的市场反应。重复不必要的信息并不能提供增量信息,其目的在于迷惑信息使用者,从而降低市场对负面消息的反应程度。当企业面临激烈的外部市场竞争时,公司管理层披露前瞻性信息的意愿越弱,以避免“言多必失”^[14]。所以,上市公司发生违规后,势必会导致公司某些指标超乎正常,需要更多能解释其合理性的说辞,因而需要大量信息披露以淹没市场信息使用者对这些指标的关注。因此,增加信息披露进而降低文本相似度更符合掩饰违规行为的逻辑,而重复重要或不重要的信息均不利于掩饰异乎寻常的违规行为。

第二,专业分析师与个人投资者通常能够从金融文本中提取有价值的信息。在对金融财务信息质量做出评价时,需要将复杂度作为重要判断标准。从年报文本复杂程度来看,增加文本复杂度更有利于公司掩饰违规行为。此外,公司在发

生违规后,某些指标可能会出现异常,管理层需要进行合理的解释,使得文本信息更加晦涩难懂,更有利于掩饰违规行为。

第三,情感分析是指采用机器学习挖掘文本信息中情感信息,探究文本信息的情感倾向性。现有研究通常利用积极、消极等情感词汇的比例与分布来度量信息的语气语调,以测定研究对象相关信息的情感指数。从市场反应来看,年报的乐观语调更有利于掩饰公司违规行为。年报的乐观语调对未来盈余、股市收益等具有预测作用^[15]。因此,上市公司利用乐观语调比悲观语调更有利于掩饰违规行为。据此,本文提出假设1:上市公司发生违规后,年报非财务信息文本的相似度显著降低、复杂度显著提高、语调显著更加乐观。

(二)公司利用年报文本信息操纵精心掩饰违规的动机分析

上市公司违规后年报 MD&A 文本特征的变化是否为公司的精心掩饰行为,可以从违规效用函数上判断,违规对公司的效用越大,公司越有可能利用文本信息操纵精心掩饰违规行为。依据经典 GONE 理论,违规需求(N)可以反映违规收益,需求越强烈违规收益越大;暴露(E)可以反映违规成本,暴露的概率越大违规成本越高;如果违规收益远大于违规成本,且公司有违规品质(G)和违规可乘之机(O),违规就会发生,而且还会用文本信息操纵精心掩饰。

具体而言:第一,G代表违规之心,主要是基于道德、品质方面的考量。如果公司面临较强的融资约束,好比“人穷志短”,更容易产生违规之心。第二,O代表违规机会。GONE理论以及 COSO 报告认为,公司治理缺陷为公司违规创造了机会。违规涉及业务的方方面面,需要组织架构中的层层授权。因此,健全的公司治理结构能有效制约公司违规。然而,当公司治理存在缺陷时,会对公司违规“大开方便之门”。第三,N代表违规需求。上市公司通过违规行为可以实现应对外界压力、财务困境的目的,此时公司违规的需求更大。第四,E代表暴露可能性。公司面临的外部监管越薄弱、法制环境越差,公司违规被查处的可能性越

小。因此,公司的违规之心越重、违规机会越多、违规需求越激烈、违规被暴露的概率越小,公司越有可能采取违规行为,违规后也越有可能利用年报 MD&A 文本信息操纵加以掩饰。据此,本文提出假设 2:公司违规后非财务信息文本特征的变化规律主要体现在违规之心较强、违规机会较多、违规需求较激烈以及违规被暴露概率较小组。

(三)资本市场能否识别公司利用文本信息操纵掩饰违规分析

信息披露策略行为之所以普遍存在在于资本市场并非完全有效,公司可以凭借信息披露策略获取短期披露收益。但是,市场又并非完全无效,也可以及时部分识别或延期完全识别信息披露策略并给予公司一定的处罚。然而公司信息披露策略行为会导致股票增发后公司业绩表现不佳,市场对公司未来业绩下降的预期会增加股票增发折价率^[16]。

如果上市公司利用文本信息操纵精心掩饰违规行为,那么资本市场对这种策略行为的识别将历经较长时期。短期内,公司精心掩饰的行为迷惑了信息使用者,有可能让市场毫无察觉,股价可能对此不产生任何反应,即短窗口期的累计超额收益率可能与违规以及文本信息操纵无关。随着时间的推移,公司精心掩饰违规的行为开始露出破绽,市场参与者开始产生质疑并慢慢接近理解事实的真相,也即市场对公司可能存在 30 或 60 或 90 天的信息漂移现象(盈余公告后漂移现象是指盈余公告后,未预期盈余为正或负,则存在正或负的超常收益,股票价格将持续向上或向下漂移,也即股价需要经过一段时间调整才能反应盈余信息)。在非完全有效市场,负面消息隐瞒能够短时间避免公司股价剧烈波动,但负面消息隐藏难度随时间推移而加大,当负面消息再也无法隐瞒并集中释放时,股价崩盘风险随即产生。公司利用文本信息操纵掩饰违规的过程其实也是坏消息积累的过程,而市场对这种行为的识别过程也是坏消息慢慢释放的过程。因此,公司利用文本信息操纵精心掩饰违规的行为可能引发未来期间的股价崩盘风险。据此,本文提出假设 3:公司利用非

财务信息文本信息操纵精心掩饰违规,降低了短窗口的市场反应、加剧了信息漂移现象、增加了未来股价崩盘风险。

二、研究设计

(一)样本选择与数据来源

本文选取 2007—2019 年 34542 个中国 A 股上市公司年度样本为初始样本,剔除 796 个金融保险业“公司—年度”样本,剔除 752 个 ST、*ST“公司—年度”样本,剔除 6134 个变量缺失的样本,最终得到 26860 个“公司—年度”样本。上市公司违规数据是根据 CSMAR 数据库中有关中国证监会、沪深交易所和财政部对上市公司的公开谴责、公开批评或公开处罚公告整理而得。公司年报非财务信息文本特征数据通过 Python 软件从年报 MD&A 部分提取所得,财务数据来源于 CSMAR 和 WIND 数据库,地区宏观经济指标取自《中国统计年鉴》。为避免异常值的影响,本文对主要变量进行了上下 1% 的 WINSORIZE 缩尾处理。

(二)模型与变量定义

为了更好地观测上市公司是否违规以及违规前后是否存在年报文本信息操纵行为,本文在控制个体与年度固定效应的基础上,设立双重差分(Difference-in-Difference)模型(1),以检验上市公司违规对年报 MD&A 文本信息披露的影响。

$$Disclosure_{i,t} = \beta_0 + \beta_1 Fraud_{i,t} + \beta_2 Controls_{i,t} + \beta_3 \sum Firm + \beta_4 \sum Year + \varepsilon_{i,t} \quad (1)$$

Disclosure 变量反映年报 MD&A 文本信息披露特征。第一,文本相似度 *Similarity*, 本文参照 Merkley (2014)^[17] 的研究,采用 Python 软件包 Gensim,使用向量空间模型计算前后两年年报 MD&A 文本向量的夹角余弦值 *Cosine* 衡量。第二,文本复杂度 *Complexity*, 本文参照 Dyer 等 (2017)^[18] 的研究,采用年报 MD&A 词汇总数的自然对数衡量。第三,文本语调变量 *Tone*, 采用公司当年积极词汇数量占年报 MD&A 词汇数量的比例衡量。*Fraud* 变量反映公司是否发生违规。若上市公司在当期发生了违规行为,则违规当年至被查处年间的样本赋值为 1,否则赋值为 0;若上市公司从未发生过违规,则该上市公司样本均赋值为

0。Controls 为系列控制变量。具体包括:国有产权性质哑变量 SOE,年末总资产的自然对数 Size,年末总负债与总资产的比值 Leverage,年末净利润与总资产的比值 Roa,企业价值 TobinQ,上市年限的自然对数 Age,研发投入的自然对数 R&D,是否四大事务所审计 Big4,行业集中度 HHI,是否增发股票 SEO,地区 GDP 增长率,年报词汇总数的自然对数 Length。

依据经典 GONE 理论中的违规之心、违规机会、违规需求以及违规暴露等特征变量对模型(1)进行分组回归,以验证本文的假设 2。其中,采用融资约束衡量公司是否具有违规之心,采用公司治理水平衡量公司是否具有违规机会,采用财务困境衡量公司是否具有违规需求,采用市场法律环境衡量违规被暴露的可能性。

此外,本文参照祝继高等(2013)^[19]的研究,设立资本市场识别模型(2):

$$CAR_{i,t} \text{ or } PEAD_{i,t} \text{ or } CRASH_{i,t} = \beta_0 + \beta_1 Fraud_{i,t} + \beta_2 Fraud_{i,t} \times Disclosure_{i,t} + \beta_3 Disclosure_{i,t} + \beta_4 Controls_{i,t} + \beta_5 \Sigma Industry + \beta_6 \Sigma Year + \varepsilon_{i,t} \quad (2)$$

CAR 反映市场对公司利用文本信息操纵精心掩饰违规的行为的短期反应。本文参照祝继高等(2013)^[19]的研究,采用[-5,+5]窗口期内的累计超额异常收益率衡量。此时,系列控制变量 Controls 包括:市场系统性风险 Beta、市净率 PB、总资产报酬率 Roa、基本每股收益 EPS、流通股股数的自然对数 LnLTA、净利润增长率 Netrate。PEAD 反映[0,+90]窗口期的信息漂移现象,即较长窗口期的累计超额收益率 CAR 对未预期盈余 UE 的反应系数^[13]。其中,解释变量未预期盈余 $UE = [Eps - E(eps)]/Price$ 。其中,E(eps)是第 t 年分析师盈利预测的中位数,Price 是年末收盘价。CRASH 变量反映一年期的股价崩盘风险。本文参考王化成等(2015)^[20]的研究,利用股价崩盘风险捕捉市场对公司利用文本信息操纵掩饰违规行为的反应。其中,股价崩盘风险变量 CRASH 用下一期的负收益偏态系数变量 FNCSKEW 和股价上下波动性变量 FNCSKEW 衡量,同时控制经市场调整的当年周收益率的标准差 Sigma,当年月平均超额换手率与上

年之差 Dturn,当年实际周收益率的平均值 Ret,过去三年可操纵性应计绝对值的总和 Opaque。

(三)描述性统计

在年报 MD&A 文本信息特征方面,年报文本相似度 Similarity 均值为 0.439,中位数为 0.430,标准差为 0.119,表明数值分布较合理、部分公司年报 MD&A 变化较大;文本复杂度 Complexity 均值为 8.010,中位数为 8.003,标准差为 0.596,表明数值分布合理;文本语调 Tone 均值为 0.608,中位数为 0.627,表明大多数年报 MD&A 文本语调总体趋于乐观。在上市公司违规方面,Fraud 变量均值为 0.114,即约有 11.4% 的样本受公司违规的影响,观测样本分布较为合理。

(四)相关性分析

Pearson 相关性分析结果显示,年报 MD&A 文本相似度 Similarity 与公司违规 Fraud 在 1% 水平上显著负相关,文本复杂度 Complexity 以及文本语调 Tone 与上市公司违规 Fraud 在 1% 水平上显著正相关。系列结果表明,相对于未发生违规的公司,违规公司年报 MD&A 文本的相似度更低、复杂度更高、语调更乐观,也即违规公司系统性地改变了其年报 MD&A 文本信息特征。同时,模型回归中变量的方差膨胀因子(VIF)均低于 5,由此多重共线性不会对研究结论产生严重干扰。

三、上市公司违规后年报文本信息发生哪些规律性变化?

(一)基准回归分析

表 1 列示了上市公司违规前后年报 MD&A 文本信息的变化规律。表 3 第(1)列的结果显示:年报 MD&A 文本相似度 Similarity 与公司违规 Fraud 在 1% 水平上显著负相关,表明公司发生违规后年报 MD&A 文本相似度显著降低,从而说明上市公司可能披露更多信息以掩饰其违规行为。表 3 第(2)列的结果显示:年报 MD&A 文本信息复杂度 Complexity 与公司违规 Fraud 在 1% 水平上显著正相关,表明公司发生违规后年报 MD&A 文本信息复杂度显著提高,说明公司可能撰写更长篇幅的 MD&A 文本以掩饰其违规行为。表 3 第(3)列的结果显示:年报 MD&A 文本信息语调 Tone 与公司

违规 *Fraud* 在 1% 水平上显著正相关,表明上市公司在违规后年报 MD&A 文本语调更加乐观,说明上市公司可能以乐观语调撰写 MD&A 文本以掩饰其违规行为。上述结果表明,上市公司违规前后年报 MD&A 文本信息发生了系统性变化,即相似度更低、复杂度更高、语调更乐观,而年报 MD&A 文本特征的这种规律性变化可能体现了公司刻意掩饰违规行为的倾向,从而验证了本文假设 1。

表 1 公司违规前后年报 MD&A 文本信息的系统性变化

变量	<i>Similarity</i> (1)	<i>Complexity</i> (2)	<i>Tone</i> (3)
<i>Fraud</i>	-0.020 *** (-3.31)	0.121 *** (4.14)	0.042 *** (3.77)
<i>SOE</i>	0.005 (0.54)	-0.041 (-0.80)	-0.021 (-1.09)
<i>Size</i>	-0.016 *** (-4.13)	0.026 (1.24)	0.023 *** (2.94)
<i>Leverage</i>	-0.005 (-0.34)	-0.121 (-1.59)	-0.058 * (-1.80)
<i>Roa</i>	-0.062 * (-1.95)	0.057 (0.49)	0.306 *** (4.23)
<i>TobinQ</i>	-0.003 *** (-3.10)	-0.003 (-0.73)	0.000 (0.14)
<i>Age</i>	0.032 *** (5.17)	-0.359 *** (-12.99)	0.011 (0.85)
<i>R&D</i>	-0.000 (-1.08)	0.000 (0.36)	0.000 (0.28)
<i>Big4</i>	0.004 (0.32)	0.012 (0.14)	0.005 (0.23)
<i>SEO</i>	-0.009 *** (-2.95)	-0.033 *** (-2.69)	0.014 *** (2.76)
<i>GDP</i>	-0.064 (-0.54)	0.610 (1.00)	-0.149 (-0.64)
<i>Length</i>	-0.029 *** (-2.99)	0.788 *** (14.87)	0.031 (1.64)
<i>Year FE</i>	Yes	Yes	Yes
<i>Firm FE</i>	Yes	Yes	Yes
<i>Constant</i>	1.018 *** (10.10)	-0.138 (-0.23)	-0.218 (-1.04)
<i>Observations</i>	26860	26860	26860
<i>Adj. R²</i>	0.235	0.302	0.075
<i>F</i>	146.64	68.57	13.27

注:括号内为 t 值;***、**和* 分别表示在 1%、5%和 10%水平上显著。

(二)稳健性检验与内生性检验

为进一步验证基准回归结果的有效性,本文还进行了如下稳健性检验与内生性检验。

第一,为佐证上述结论的可靠性,表 2 Panel A 进一步观察违规被查处前后年报 MD&A 文本特征的规律性变化。从发生违规到违规被查处,公司

具有通过改变 MD&A 文本进而掩饰其违规行为的动机。公司发生违规后会具有掩饰动机,而违规被查处后掩饰动机消失。因此,违规被查处前后 MD&A 文本特征可能呈现出与发生违规前后相反的变化规律。本文按照违规是否被查处重新对 *Punish* 进行赋值,若上市公司发生了违规,则该上市公司被查处当年及以后年度均赋值为 1,发生违规的第一年至查处年度的公司样本均赋值为 0;若上市公司从未发生过违规,则该上市公司样本均赋值为 0。表 2 第(1)列的结果显示:年报 MD&A 文本相似度 *Similarity* 与公司违规被处罚变量 *Punish* 在 1% 水平上显著正相关,说明违规被查处后公司年报 MD&A 文本相似度显著提高,从而佐证违规公司被查处前降低相似度更可能是为了掩饰其违规行为。表 2 第(2)列的结果显示:年报 MD&A 文本复杂度 *Complexity* 与违规被处罚变量 *Punish* 在 10% 水平上显著负相关,说明违规被查处后公司年报 MD&A 文本复杂度显著降低,从而佐证违规公司被查处前撰写更长篇幅的年报 MD&A 文档更可能是为了掩饰其违规行为。表 2 Panel A 第(3)列的结果显示:年报 MD&A 文本语调 *Tone* 与违规被处罚变量 *Punish* 在 1% 水平上显著负相关,说明违规被查处后公司年报 MD&A 语调更消极,从而佐证违规公司被查处前以乐观语调撰写 MD&A 文本信息更可能是为了掩饰其违规行为。上市公司违规被查处后,公司年报 MD&A 文本特征发生了规律性变化,即相似度更高、复杂度更低、语调更消极。表 1 与表 2 Panel A 的变化规律完全相反,进而验证了本文假设 1。

第二,采用倾向得分匹配法(PSM)选择控制组样本,减少实验组和控制组样本因公司特征不同可能对本文结论产生的影响。本文将公司规模 *Size*、资产负债率 *Leverage*、总资产报酬率 *Roa*、企业价值 *TobinQ*、上市年限 *Age*、研发投入 *R&D*、是否四大事务所审计 *Big4*、是否增发股票 *SEO*、地区 *GDP* 水平、年报文本长度 *Length* 作为匹配变量,并对公司是否发生违规进行回归并计算倾向得分 *Pscore*,再依据倾向得分对样本进行 1:1 无放回匹配,得到 12043 对“实验组—控制组”样本。表 2 Panel B 的

PSM + DID 回归结果显示:表 1 的回归结果依然稳健。

第三,执行安慰剂测试以消除可能存在的内生性问题。本文将公司违规发生时间向前平推 3 年,使用虚拟的违规年度重复表 1 的基准回归分析。表 4 Panel C 的安慰剂测试的结果显示:公司违规变量 *Fraud* 与 3 个年报 MD&A 文本信息指标 *Disclosure* 均无显著关系,说明因不可观测的公司固有差异导致的缺失变量偏误对本文结论的影响较小,从而验证了本文研究的因果效应。

第四,年报非财务信息披露与上一年度显著不同,可能是企业管理层审时度势并积极改变公司战略的结果。因此,本文还考虑了公司并购、公司战略的影响。为此,本文控制了公司并购变量 *M&A* 和公司战略调整变量 *Strategy*。本文参照陈爱贞等 (2019)^[21] 的研究,采用 *M&A* 度量公司是否发生并购。若公司当年发生并购,则赋值为 1,否则为 0。此外,本文又参照 Bentley 等 (2013)^[22]、孙健等 (2016)^[23] 的研究,构建一个离散变量来度量公司战略。该变量主要关注公司的研发强度、员工人数波动性、销售收入增长率、员工人数与销售收入的比值、销售费用和管理费用占销售收入的比重、固定资产占总资产的比重。首先,将上述 6 个变量取近五年的平均值。其次,对于前 5 个变量,在每一个“年度—行业”样本中从小到大平均分为 5 组,最小组赋值为 0 分,次小组赋值为 1 分,以此类推,最大组赋值为 4 分;对于第 6 个变量分组方式相反,即最小组赋值为 4 分,最大组赋值为 0 分。最后,将 6 个变量的分组得分相加,得到 0 ~ 24 分的度量变量 *Strategy*。较高的 *Strategy* 值表示公司战略较为激进,而较低的 *Strategy* 值表示公司战略较为保守。表 4 Panel D 列示了控制公司并购、公司战略影响后的回归分析结果,年报 MD&A 文本相似度 *Similarity* 与公司违规 *Fraud* 依然在 1% 水平上显著负相关,年报 MD&A 文本复杂度 *Complexity* 与公司违规 *Fraud* 依然在 1% 水平上显著正相关,年报 MD&A 文本语调 *Tone* 与公司违规 *Fraud* 依然在 1% 水平上显著正相关。上述回归结果验证了基准回归结果的稳健性。

表 2 稳健性检验与内生性检验

Panel A: 违规被查处前后年报 MD&A 文本特征的规律性变化			
变量	<i>Similarity</i>	<i>Complexity</i>	<i>Tone</i>
	(1)	(2)	(3)
<i>Punish</i>	0.013 *** (2.74)	-0.040 * (-1.77)	-0.054 *** (-5.27)
<i>Constant</i>	1.001 *** (11.12)	-1.500 *** (-2.88)	-0.633 *** (-3.17)
<i>Controls</i>	Yes	Yes	Yes
<i>Year FE</i>	Yes	Yes	Yes
<i>Firm FE</i>	Yes	Yes	Yes
<i>Observations</i>	26860	26860	26860
<i>Adj. R²</i>	0.241	0.293	0.070
<i>F</i>	187.90	82.08	14.89
Panel B: PSM + DID 方法回归			
变量	<i>Similarity</i>	<i>Complexity</i>	<i>Tone</i>
	(1)	(2)	(3)
<i>Fraud</i>	-0.021 *** (-3.43)	0.108 *** (3.68)	0.042 *** (3.76)
<i>Constant</i>	0.692 *** (7.60)	0.304 (0.45)	-0.223 (-1.06)
<i>Controls</i>	Yes	Yes	Yes
<i>Year FE</i>	Yes	Yes	Yes
<i>Firm FE</i>	Yes	Yes	Yes
<i>Observations</i>	24086	24086	24086
<i>Adj. R²</i>	0.220	0.361	0.075
<i>F</i>	89.63	59.26	14.45
Panel C: 安慰剂测试			
变量	<i>Similarity</i>	<i>Complexity</i>	<i>Tone</i>
	(1)	(2)	(3)
<i>Fraud</i>	0.003 (0.91)	-0.008 (-0.52)	-0.008 (-0.71)
<i>Constant</i>	0.601 *** (9.69)	-1.188 ** (-2.50)	-0.245 (-1.14)
<i>Controls</i>	Yes	Yes	Yes
<i>Year FE</i>	Yes	Yes	Yes
<i>Firm FE</i>	Yes	Yes	Yes
<i>Observations</i>	26860	26860	26860
<i>Adj. R²</i>	0.239	0.315	0.063
<i>F</i>	193.83	97.26	11.77
Panel D: 控制公司并购、公司战略的影响			
变量	<i>Similarity</i>	<i>Complexity</i>	<i>Tone</i>
	(1)	(2)	(3)
<i>Fraud</i>	-0.018 *** (-3.25)	0.119 *** (4.10)	0.039 *** (3.69)
<i>M&A</i>	-0.008 ** (-2.23)	-0.034 ** (-2.37)	0.014 *** (2.74)
<i>Strategy</i>	-0.004 *** (-4.04)	-0.003 (-0.71)	0.001 (0.29)
<i>Constant</i>	0.601 *** (9.69)	-1.188 ** (-2.50)	-0.245 (-1.14)
<i>Controls</i>	Yes	Yes	Yes
<i>Year FE</i>	Yes	Yes	Yes
<i>Firm FE</i>	Yes	Yes	Yes
<i>Observations</i>	26860	26860	26860
<i>Adj. R²</i>	0.228	0.309	0.086
<i>F</i>	173.85	92.11	16.85

注:括号内为 t 值;***、** 和 * 分别表示在 1%、5% 和 10% 水平上显著。

四、基于 GONE 理论对公司年报文本信息操纵的分析

经典 GONE 理论认为:公司违规由贪婪 Greed (G)、机会 Opportunity (O)、需要 Need (N)、暴露 Exposure (E)。如果上市公司存在违规之心、违规机会、违规需求且事后被发现的可能性较小,那么上市公司就更倾向发生违规行为。

(一)从贪婪(G)角度分析公司文本信息操纵行为

企业融资约束程度越高,公司利用资本市场满足融资的动机就越强。当前测度融资约束的代表性的测度方法有 KZ 指数、WW 指数和 SA 指数。其中,KZ 指数和 WW 指数均包含了很多具有内生性的金融变量。为避免内生性的干扰,本文参考海本禄等(2021)^[24]的做法,采用 Size & Age (SA) 指数衡量融资约束, $SA = -0.737 \times Size + 0.043 \times Size^2 - 0.040 \times Age$,该指数越大,表示融资约束越严重。表 3 Panel A 按照 SA 指数中位数分组的回归结果显示:公司违规变量 *Fraud* 与年报 MD&A 文本特征 *Disclosure* 的 3 个变量的显著相关关系主要存在于融资约束较高组,说明公司融资约束较高时,公司利用文本信息操纵掩饰违规的可能性越大。

(二)从违规机会(O)角度探讨公司文本信息操纵行为

公司治理效果越差,公司进行违规的机会就越多。本文参照张会丽等(2012)^[25]的研究,对第一大股东持股比例、独立董事比例、机构投资者持股比例、董事长与总经理是否兼任、董事会规模、董事持股比例、高管理持股比例以及前三名高管薪酬的自然对数进行主成分分析,构建公司治理指数 *GIndex*,如果公司治理指数大于中位数,则视为公司治理状况较好。表 3 Panel B 按照公司治理指数中位数分组的回归结果显示:公司违规变量 *Fraud* 与年报 MD&A 文本信息披露 *Disclosure* 的 3 个变量的显著关系主要存在于公司治理较差组,说明公司治理水平越差,公司进行违规并利用文本信息操纵加以掩饰的动机越强。

(三)从违规需求(N)角度探讨公司文本信息操纵行为

上市公司违规通常与公司的财务困境联系在一起,公司有可能通过违规掩饰其财务困境。本文参考林晚发等(2020)^[26]的研究,采用 *Z-score* 衡量财务状况,*Z-score* 模型是以多变量的统计方法为基础,对企业的运行状况、破产进行系统分析与判别,反映公司的财务状况。 $Z-Score = 1.2 \times X_1 + 1.4 \times X_2 + 3.3 \times X_3 + 0.68 \times X_4 + 0.99 \times X_5$ (其中, X_1 = 营运资金/资产总额, X_2 = 留存收益/资产总额, X_3 = 息税前利润/资产总额, X_4 = 权益市值/负债账面价值, X_5 = 销售额/资产总额)。*Z-score* 值越大,表示公司财务状况越差,公司发生财务困境的可能性就越大。表 5 Panel C 按照 *Z-score* 中位数分组的回归结果显示:公司违规变量 *Fraud* 与年报 MD&A 文本信息 *Disclosure* 的 3 个变量的显著关系主要存在于财务状况较差组,说明上市公司财务状况越差时,公司利用文本信息操纵掩饰违规的动机越强。

(四)从违规被暴露(E)可能性的角度探讨公司文本信息操纵行为

暴露主要是指违规被查处可能性。公司违规暴露的可能性越低,公司违规的可能性就越大。本文参照佟岩等(2021)^[27]的研究,以樊纲等(2017)^[28]披露的《中国分省份市场化指数报告(2016)》的市场化指数作为地区市场化程度的衡量变量,采用“市场中介组织的发育和法律制度环境”指数反映地区法制环境 (*Legal*)。公司注册地“市场中介组织的发育和法律制度环境”指数大于中位数,视为地区法制环境较好,违规被查处可能性较大;公司注册地“市场中介组织的发育和法律制度环境”指数小于中位数,则表示地区法制环境较差。表 3 Panel D 按照地区法制环境中位数分组的回归结果显示:公司违规 *Fraud* 变量与年报 MD&A 文本信息披露 *Disclosure* 的 3 个变量的显著关系主要存在于地区法制环境较差组,说明地区法制环境越差,公司进行违规并利用文本信息操纵掩饰的动机越强。

表 3 基于经典 GONE 理论对公司文本信息操纵行为的分析

Panel A: 从贪婪(G)角度分析公司文本信息操纵行为						
变量	Similarity		Complexity		Tone	
	融资约束高	融资约束低	融资约束高	融资约束低	融资约束高	融资约束低
	(1)	(2)	(3)	(4)	(5)	(6)
Fraud	-0.018 **	-0.021	0.112 ***	0.084	0.050 ***	0.014
	(-2.31)	(-1.47)	(3.46)	(1.37)	(3.56)	(0.69)
Constant	1.160 ***	0.569 **	-0.385	0.627	-0.378	-0.395
	(7.49)	(2.31)	(-0.53)	(0.56)	(-1.09)	(-0.94)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13441	13419	13441	13419	13441	13419
Adj. R ²	0.270	0.206	0.271	0.273	0.088	0.044
F	98.66	48.74	35.35	29.48	9.86	4.29
Panel B: 从违规机会(O)角度探讨公司文本信息操纵行为						
变量	Similarity		Complexity		Tone	
	内部公司治理较好	内部公司治理较差	内部公司治理较好	内部公司治理较差	内部公司治理较好	内部公司治理较差
	(1)	(2)	(3)	(4)	(5)	(6)
Fraud	-0.010	-0.024 **	0.072	0.115 ***	0.010	0.037 ***
	(-1.14)	(-2.38)	(1.63)	(2.68)	(1.09)	(2.83)
Constant	1.008 ***	0.974 ***	0.682	-1.101	-0.587 *	-0.261
	(6.57)	(6.64)	(0.93)	(-1.11)	(-1.68)	(-0.91)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13421	13439	13421	13439	13421	13439
Adj. R ²	0.247	0.248	0.245	0.333	0.101	0.072
F	76.19	64.15	26.77	43.37	6.19	8.89
Panel C: 从违规需求(N)角度探讨公司文本信息操纵行为						
变量	Similarity		Complexity		Tone	
	财务状况较好	财务状况较差	财务状况较好	财务状况较差	财务状况较好	财务状况较差
	(1)	(2)	(3)	(4)	(5)	(6)
Fraud	-0.018	-0.019 **	0.032	0.138 ***	0.020	0.041 ***
	(-1.63)	(-2.26)	(0.55)	(4.02)	(1.11)	(2.94)
Constant	0.620 ***	1.221 ***	0.253	0.172	-0.579	-0.280
	(2.69)	(10.07)	(0.25)	(0.29)	(-1.23)	(-1.17)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13425	13435	13425	13435	13425	13435
Adj. R ²	0.218	0.257	0.290	0.267	0.090	0.073
F	30.38	116.2	25.17	47.10	2.74	11.23
Panel D: 从违规被暴露(E)的可能性角度探讨公司文本信息操纵行为						
变量	Similarity		Complexity		Tone	
	地区法制环境较好	地区法制环境较差	地区法制环境较好	地区法制环境较差	地区法制环境较好	地区法制环境较差
	(1)	(2)	(3)	(4)	(5)	(6)
Fraud	-0.017	-0.021 **	0.014	0.127 ***	0.015	0.063 ***
	(-1.55)	(-2.55)	(1.34)	(3.24)	(0.75)	(4.27)
Constant	0.830 ***	1.061 ***	-0.341	0.222	-0.246	-0.243
	(5.32)	(7.18)	(-0.43)	(0.24)	(-0.83)	(-0.80)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13397	13463	13397	13463	13397	13463
Adj. R ²	0.258	0.219	0.274	0.310	0.091	0.061
F	83.96	65.53	29.84	33.09	9.43	5.82

注:括号内为t值;***、**和*分别表示在1%、5%和10%水平上显著。

五、资本市场能否识别公司利用文本信息操纵掩饰违规的行为

(一)5 天窗口期的累计超额异常回报率

表 4 列示了年报短窗口市场反应的回归分析结果。第(1)列不考虑年报 MD&A 文本特征的回归结果显示:公司违规变量 *Fraud* 与累计超额异常回报率变量 *CAR* 负相关但并不显著,说明资本市场对违规公司与非违规公司年报的反应不存在显著差异,也说明违规公司会精心掩饰从而不被资本市场察觉。第(2)至(5)列进一步考虑年报 MD&A 文本信息变化的掩饰动机,当文本相似度 *Similarity* 小于中位数、文本复杂度 *Complexity* 大于中位数、文本语调 *Tone* 大于中位数时,公司掩饰违

表 4 短期市场反应

变量	不考虑 掩饰行为	考虑掩饰行为		
		<i>Similarity</i>	<i>Complexity</i>	<i>Tone</i>
	<i>CAR</i> [-5, +5]	<i>CAR</i> [-5, +5]	<i>CAR</i> [-5, +5]	<i>CAR</i> [-5, +5]
	(1)	(2)	(3)	(4)
<i>Fraud</i>	-0.074 (-1.37)	0.026 (0.29)	-0.181 ** (-2.48)	-0.332 (-1.29)
<i>Fraud</i> × <i>D</i> <i>islosure_Dummy</i>		-0.130 (-1.24)	0.303 (1.11)	0.286 (1.09)
<i>Disclosure_</i> <i>Dummy</i>		0.032 (0.83)	-0.250 *** (-6.13)	0.093 (0.85)
<i>Leverage</i>	-0.038 (-0.71)	-0.012 (-0.24)	-0.011 (-0.22)	-0.012 (-0.24)
<i>Beta</i>	0.148 *** (5.73)	0.125 *** (4.34)	0.095 *** (3.26)	0.126 *** (4.36)
<i>PB</i>	0.103 (1.07)	0.062 (0.58)	0.043 (0.40)	0.075 (0.70)
<i>Roa</i>	0.019 (1.17)	0.013 (0.69)	0.029 (1.50)	0.011 (0.58)
<i>Age</i>	0.005 (1.45)	0.005 (1.48)	0.005 (1.61)	0.005 (1.48)
<i>Big4</i>	0.262 ** (2.01)	0.308 ** (2.24)	0.279 ** (2.03)	0.305 ** (2.22)
<i>EPS</i>	-0.197 *** (-5.51)	-0.185 *** (-4.27)	-0.190 *** (-4.39)	-0.187 *** (-4.30)
<i>LnLTA</i>	-0.585 (-1.21)	-0.602 (-1.28)	-0.488 (-1.04)	-0.581 (-1.24)
<i>Netrate</i>	0.652 (0.61)	0.613 (0.53)	0.631 (0.54)	0.621 (0.53)
<i>Industry FE</i>	Yes	Yes	Yes	Yes
<i>Year FE</i>	Yes	Yes	Yes	Yes
<i>Constant</i>	1.483 *** (3.25)	1.772 *** (3.27)	1.514 *** (2.79)	1.751 *** (3.20)
<i>Observations</i>	26860	26860	26860	26860
<i>Adj. R²</i>	0.098	0.084	0.086	0.084
<i>F</i>	11.87	15.25	15.37	15.46

注:括号内为 t 值;***、** 分别表示在 1%、5% 水平上显著。

规的动机更强,*Disclosure_Dummy* 赋值为 1,否则赋值为 0。回归结果显示:交乘项 *Disclosure_Dummy* × *Fraud* 与累计超额异常回报率 *CAR* 变量不存在显著关系,从而说明资本市场对违规公司与非违规公司年报的反应不存在显著差异,也即公司利用文本信息操纵精心掩饰违规的行为并未被资本市场识别。本文还检验了[-1, +1]、[-2, +2]、[-3, +3]等窗口的累计超额异常回报率,结果与此相近。

(二)90 天窗口期内的信息漂移

表 5 列示了 90 天窗口期信息漂移的回归分析结果。第(1)列不考虑文本信息操纵行为的回归结果显示:交乘项 *Fraud* × *UE* 与累计超额收益率变量 *CAR* 在 10% 水平上显著正相关,而且与 *UE* 的回归系数符号一致,说明上市公司违规加剧了

表 5 信息漂移现象

变量	不考虑 掩饰行为	考虑掩饰行为		
		<i>Similarity</i>	<i>Complexity</i>	<i>Tone</i>
	<i>CAR</i> [0, +90]	<i>CAR</i> [0, +90]	<i>CAR</i> [0, +90]	<i>CAR</i> [0, +90]
	(1)	(2)	(3)	(4)
<i>Fraud</i>	0.016 *** (3.44)	0.016 *** (4.21)	0.016 *** (4.16)	0.016 *** (4.16)
<i>UE</i>	0.013 * (1.86)	0.014 ** (2.20)	0.013 ** (2.13)	0.013 ** (2.15)
<i>Fraud</i> × <i>UE</i>	0.026 * (1.72)	0.029 ** (2.01)	0.017 (1.21)	0.007 (0.22)
<i>FraudU</i> × <i>E</i> × <i>Disclosure_</i> <i>Dummy</i>		0.018 ** (2.34)	0.013 * (1.66)	0.017 * (1.73)
<i>Disclosure_</i> <i>Dummy</i>		-0.001 (-0.08)	-0.010 (-1.41)	-0.003 (-0.19)
<i>Size</i>	0.005 (0.41)	0.004 (0.36)	0.005 (0.40)	0.005 (0.37)
<i>Leverage</i>	0.003 (0.30)	0.003 (0.27)	0.003 (0.24)	0.003 (0.25)
<i>TobinQ</i>	-0.012 ** (-2.00)	-0.011 *** (-3.67)	-0.012 *** (-3.80)	-0.012 *** (-3.74)
<i>Cash</i>	0.009 (1.33)	0.009 (1.41)	0.009 (1.43)	0.009 (1.43)
<i>Roa</i>	-0.026 (-0.43)	-0.027 (-0.67)	-0.026 (-0.64)	-0.027 (-0.66)
<i>Industry FE</i>	Yes	Yes	Yes	Yes
<i>Year FE</i>	Yes	Yes	Yes	Yes
<i>Constant</i>	0.263 ** (2.18)	0.262 *** (3.99)	0.276 *** (4.16)	0.269 *** (3.96)
<i>Observations</i>	26860	26860	26860	26860
<i>Adj. R²</i>	0.029	0.029	0.030	0.029
<i>F</i>	32.10	13.91	14.02	13.95

注:括号内为 t 值;***、** 和 * 分别表示在 1%、5% 和 10% 水平上显著。

信息漂移现象。第(2)至(5)列进一步考虑文本信息操纵行为的回归结果显示:交乘项 $Fraud \times UEDis \times closure_Dummy$ 与累计超额收益率变量 CAR 在 10% 及以上水平上显著正相关,而且与 $Fraud \times UE$ 的回归系数符号一致,说明利用年报 MD&A 文本信息操纵掩饰违规的行为加剧了信息漂移现象。上述结论表明,一方面,公司刻意进行违规并通过年报 MD&A 文本进行精心掩饰,增加了投资者对信息的理解难度,从而降低了股价对信息的反应效率;另一方面,市场能感知上市公司的刻意掩饰行为。本文同时还检验了 $[0, +30]$ 、 $[0, +60]$ 等窗口的信息漂移现象,但并未发现相同结果,从而说明公司利用文本信息操纵掩饰违规的行为迷惑市场达 90 天左右。

(三)一年期的股价崩盘风险

表 6 列示了一年期股价崩盘风险的回归分析结果。第(1)与(2)列不考虑文本信息操纵行为的回归结果显示:公司违规变量 $Fraud$ 与负收益偏态系数变量 $FNCSKEW$ 以及股价上下波动性变量 $FDUVOL$ 在 5% 水平上显著正相关,说明上市公司违规提高了股价崩盘风险。第(3)至(10)列进一步考虑文本信息操纵行为的回归结果显示:交乘项 $Fraud \times Disclosure_Dummy$ 与负收益偏态系数变量 $FNCSKEW$ 以及股价上下波动性变量 $FNCSKEW$ 具有显著正相关关系,说明公司利用文本信息操纵掩饰违规的行为加剧了公司股价崩盘风险,同时也说明资本市场能够有效识别公司的掩饰行为。

表 6 股价崩盘风险

变量	不考虑掩饰行为		考虑掩饰行为					
			Similarity	Complexity	Tone	Similarity	Complexity	Tone
	FNCSKEW	FDUVOL	FNCSKEW	FDUVOL	FNCSKEW	FDUVOL	FNCSKEW	FDUVOL
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Fraud$	0.098 **	0.079 **	0.134 *	0.116 **	0.142 **	0.133 ***	0.075	0.100
	(1.98)	(2.18)	(1.88)	(2.41)	(2.50)	(3.23)	(0.37)	(0.75)
$Fraud \times Disclosure_Dummy$			0.097 *	0.095 *	0.091 *	0.107 **	0.122 **	0.103 **
			(1.75)	(1.72)	(1.66)	(1.96)	(2.11)	(1.87)
$Disclosure_Dummy$			0.003	0.002	0.013	-0.001	0.077 *	0.092 *
			(0.12)	(0.10)	(0.45)	(-0.05)	(1.66)	(1.88)
$Ncskew$	0.015 **	0.012 **	0.016 **	0.012 **	0.016 **	0.012 **	0.016 **	0.012 **
	(2.21)	(1.99)	(2.22)	(1.99)	(2.22)	(2.00)	(2.22)	(1.99)
$Sigma$	-9.575 ***	-8.057 ***	-9.572 ***	-8.053 ***	-9.558 ***	-8.026 ***	-9.598 ***	-8.078 ***
	(-8.79)	(-10.56)	(-8.79)	(-10.56)	(-8.77)	(-10.52)	(-8.83)	(-10.59)
Ret	-77.289 ***	-78.859 ***	-77.294 ***	-78.863 ***	-77.290 ***	-78.879 ***	-77.263 ***	-78.839 ***
	(-39.02)	(-50.26)	(-38.99)	(-50.21)	(-39.02)	(-50.27)	(-39.03)	(-50.24)
$Dturn$	-0.013	-0.017	-0.013	-0.017	-0.013	-0.018	-0.013	-0.017
	(-0.40)	(-0.76)	(-0.40)	(-0.75)	(-0.41)	(-0.78)	(-0.41)	(-0.76)
$Opaque$	0.161	0.184	0.160	0.183	0.158	0.180	0.161	0.183
	(1.02)	(1.56)	(1.02)	(1.56)	(1.00)	(1.53)	(1.02)	(1.54)
$Opaque^2$	-0.113	-0.149	-0.113	-0.149	-0.109	-0.145	-0.113	-0.148
	(-0.77)	(-1.36)	(-0.77)	(-1.36)	(-0.74)	(-1.33)	(-0.77)	(-1.35)
$Size$	-0.056 *	-0.055 **	-0.056 *	-0.054 **	-0.055 *	-0.052 **	-0.056 *	-0.055 **
	(-1.90)	(-2.36)	(-1.88)	(-2.34)	(-1.85)	(-2.22)	(-1.90)	(-2.36)
$Leverage$	-0.012	-0.000	-0.012	0.000	-0.011	0.001	-0.010	0.002
	(-0.32)	(-0.00)	(-0.32)	(0.01)	(-0.30)	(0.02)	(-0.27)	(0.07)
Roa	-0.039	-0.115	-0.034	-0.110	-0.039	-0.115	-0.053	-0.128
	(-0.15)	(-0.58)	(-0.13)	(-0.55)	(-0.15)	(-0.58)	(-0.21)	(-0.64)
$Industry\ FE$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$Year\ FE$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$Constant$	1.067	1.042 **	1.055	1.030 **	1.040	0.980 *	1.025	0.999 *
	(1.61)	(2.03)	(1.59)	(2.01)	(1.56)	(1.90)	(1.53)	(1.93)
$Observations$	26860	26860	26860	26860	26860	26860	26860	26860
$Adj. R^2$	0.429	0.568	0.429	0.568	0.429	0.568	0.429	0.568
F	17.88	25.01	16.12	22.51	16.16	22.63	16.10	22.52

注:括号内为 t 值;***、** 和 * 分别表示在 1%、5% 和 10% 水平上显著。

综上所述,本文利用 5 天、90 天以及一年期三个窗口的股价变化捕捉资本市场识别公司利用文本信息操纵掩饰违规的行为的过程,并发现短期内市场无法识别公司刻意违规并精心掩饰的行为。但是,从中长期来看资本市场能逐渐察觉并反应在股价上,从而说明公司利用文本信息操纵精心掩饰违规获得了短期收益,而资本市场对这种行为的渐进识别过程也体现了其部分有效。

六、结论与启示

(一) 结论

上市公司在发生违规后,会策略性的利用年报文本信息操纵精心掩饰其违规行为。具体而言,文本相似度显著降低,文本复杂度显著提高,文本语调显著更加乐观。同时,资本市场能识别公司的信息披露策略行为,但由于公司的精心掩饰迷惑了市场参与者对信息的理解,资本市场的识别过程相对较长,表明资本市场识别是部分有效的。本文研究验证公司利用年报文本信息操纵精心掩饰违规的策略行为,有助于拓展文本分析与信息披露的相关研究。

(二) 启示

第一,本文结论有助于会计信息使用者更好地解读年报文本信息,强化监管机构对公司文本信息披露的监管。鉴于年报文本信息无统一披露范式、披露形式千差万别,因而文本信息容易被管理层粉饰,信息使用者很容易被迷惑与误导,因而不能充分发挥其信息渠道效应。本文聚焦公司年报文本信息披露的系统性变化,会计信息使用者可以充分挖掘年报文本信息的价值,对年报文本信息披露质量做出合理判断,使得年报文本信息更容易被解读。第二,早期研究主要基于财务定量信息预测上市公司违规行为,本文尝试采用文本分析深度挖掘公司年报非财务信息,建立实务界普遍认可的文本分析程序,将有助于重构基于文本分析企业年报信息的违规预测模型,进而构建全新的公司违规风险指数。

参考文献:

- [1] 李世刚,蒋尧明. 上市公司年报文本信息语调影响审计意见吗? [J]. 会计研究,2020(5):178-192.
- [2] 林乐,谢德仁. 分析师荐股更新利用管理层语调吗?——基于业绩说明会的文本分析 [J]. 管理世界,

2017(11):125-145.

- [3] 张秀敏,高云霞,高洁. 企业年报阅读难易程度的衡量与影响因素研究——基于管理者操纵视角 [J]. 审计与经济研究,2021(1):79-89.
- [4] 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.
- [5] 王克敏,王华杰,李栋栋,等. 年报文本信息复杂性与管理者自利——来自中国上市公司的证据 [J]. 管理世界,2018(12):120-132.
- [6] 钱爱民,朱大鹏. 财务报告文本相似度与违规处罚——基于文本分析的经验证据 [J]. 会计研究,2020(9):44-58.
- [7] HOBERG G, LEWIS C. Do fraudulent firms produce abnormal disclosure? [J]. Journal of corporate finance,2017,43(4):58-85.
- [8] BROWN N C, CROELEY R M, ELLIOTT W B. What are you saying? using topic to detect financial misreporting [J]. SSRN working paper,2018.
- [9] LI F. Annual report complexity, current earnings and earnings persistence [J]. Journal of accounting and economics, 2018,45(2):221-247.
- [10] RENNEKAMP K. Processing fluency and investors reactions to disclosure complexity [J]. Journal of accounting research,2012,50(5):1319-1354.
- [11] ERTUGRUL M, LEI J, QIU J P, et al. Annual report complexity, tone ambiguity and the cost of borrowing [J]. Journal of financial & quantitative analysis, 2017, 52(2):811-836.
- [12] TAN H T, ELAINE Y W, BO Z. When the use of positive language backfires: the joint effect of tone, complexity and investor sophistication on earnings judgments [J]. Journal of accounting research,2014,52(1):273-302.
- [13] ROGERS J L, BUSKIRK A V, ZECHMAN S L. Disclosure tone and shareholder litigation [J]. The accounting review,2011,86(6):2155-2183.
- [14] 雷新途,方柳卿. 言多必失:产品市场掠夺风险会影响年报前瞻性信息披露吗? [J]. 商业研究,2020(10):73-88.
- [15] ALLEE K D, DEANGELIS M D. The structure of voluntary disclosure narratives: evidence from tone dispersion [J]. Journal of accounting research,2015,53(2):241-274.
- [16] KOTHARI S P, MIZIK N, ROYCHOWDHURY S. Managing for the moment: the role of earnings management via real activities versus accruals in seo valuation [J]. The accounting review,2016,91(2):559-586.

- [17] MERKLEY K J. Narrative disclosure and earnings performance: evidence from R&D disclosures [J]. The accounting review, 2014, 89(2): 725-757.
- [18] DYE R A. An evaluation of essays on disclosure and the disclosure literature in accounting [J]. Journal of accounting and economics, 2011, 32(1): 181-235.
- [19] 祝继高, 王春飞. 金融危机对公司现金股利政策的影响研究——基于股权结构的视角 [J]. 会计研究, 2013(2): 38-44.
- [20] 王化成, 曹丰, 叶康涛. 监督还是掏空: 大股东持股比例与股价崩盘风险 [J]. 管理世界, 2015(2): 45-57.
- [21] 陈爱贞, 张鹏飞. 并购模式与企业创新 [J]. 中国工业经济, 2019(12): 115-133.
- [22] BENTLEY K A, OMER T C, SHARP N Y. Business strategy, financial reporting irregularities, and audit effort [J]. Contemporary accounting research, 2012, 30(2): 780-817.
- [23] 孙健, 王百强, 曹丰, 等. 公司战略影响盈余管理吗? [J]. 管理世界, 2016(3): 160-169.
- [24] 海本禄, 杨君笑, 尹西明, 等. 外源融资如何影响企业技术创新——基于融资约束和技术密集度视角 [J]. 中国软科学, 2021(3): 183-192.
- [25] 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.
- [26] 林晚发, 王雅炯, 幸丽霞. 企业定性信息与债券信用评级: 基于股权性质的分析 [J]. 中国软科学, 2020(4): 123-131.
- [27] 佟岩, 李鑫. 银行业竞争与企业集团母子公司对外借款决策 [J]. 中国软科学, 2021(3): 105-123.
- [28] 樊纲, 王小鲁, 余静文. 中国分省份市场化指数报告 (2016) [M]. 北京: 社会科学文献出版社, 2017.

(本文责编: 海洋)

(上接第 87 页)

- [7] FUENTES R, ALVAREZ-SUAREZ A. Productivity of travel agencies in Spain: the case of Alicante [J]. The service industries journal, 2012, 32(16): 2623-2640.
- [8] 夏赞才, 邹泉, 罗文斌, 等. 中国省域旅游业效率时空演变分析 [J]. 统计与决策, 2020, 36(8): 64-68.
- [9] GRIGGS D, STAFFORD-SMITH M, GAFFNEY O, et al. Sustainable development goals for people and planet [J]. Nature, 2013, 495(7441): 305-307.
- [10] LIU Z. Sustainable tourism development: a critique [J]. Journal of sustainable tourism, 2003, 11(6): 459-475.
- [11] GOESSLING S, PEETERS P, CERON J P, et al. The eco-efficiency of tourism [J]. Ecological economics, 2005, 54(4): 417-434.
- [12] KELLY J, HAIDER W, WILLIAMS P W, et al. Stated preferences of tourists for eco-efficient destination planning options [J]. Tourism management, 2007, 28(2): 377-390.
- [13] PENG H, ZHANG J, LU L, et al. Eco-efficiency and its determinants at a tourism destination: a case study of Huangshan National Park, China [J]. Tourism management, 2017, 60(7): 201-211.
- [14] ZHANG J J, JOGLEKAR N, HEINEKE J, et al. Eco-efficiency of service co-production connecting eco-certifications and resource efficiency in U. S. hotels [J]. Cornell hospitality quarterly, 2014, 55(3): 252-264.
- [15] 陈超凡. 中国工业绿色全要素生产率及其影响因素——基于 ML 生产率指数及动态面板模型的实证研究 [J]. 统计研究, 2016, 33(3): 53-62.
- [16] XIE H, CHEN Q, LU F, et al. Spatial-temporal disparities and influencing factors of total-factor green use efficiency of industrial land in China [J]. Journal of cleaner production, 2019, 207(10): 1047-1058.
- [17] XU X, HUANG X, HUANG J, et al. Spatial-temporal characteristics of agriculture green total factor productivity in China, 1998 – 2016: based on more sophisticated calculations of carbon emissions [J]. International journal of environmental research and public health, 2019, 16(20): 3932-3945.
- [18] 陆菊春, 钟珍, 贾倩倩. 绿色低碳视角下我国区域建筑业生产效率及演变趋势分析 [J]. 珞珈管理评论, 2017, 16(4): 189-203.
- [19] 卢飞, 宫红平. 中国旅游生态效率测度、时空特征与影响因素研究 [J]. 统计与决策, 2020, 16(8): 96-99.
- [20] 曾冰. 环境约束下中国省域旅游经济效率及其影响因素的空间计量分析 [J]. 技术经济, 2020, 39(6): 141-146, 174.
- [21] 张玉钧. 提高生态旅游产业效率的边缘性思考 [J]. 旅游学刊, 2016, 31(9): 8-11.
- [22] 王胜鹏, 乔花芳, 冯娟, 等. 黄河流域旅游生态效率时空演化及其与旅游经济互动响应 [J]. 经济地理, 2020, 40(5): 84-92.
- [23] 朱智滔, 张伟. 碳排放规制下中国主要工业行业全要素生产率研究——基于方向性距离函数与 GML 指数模型 [J]. 资源科学, 2015, 37(12): 2341-2349.

(本文责编: 王延芳)