引用格式:杨肃昌, 曹益宁. 企业社会责任信息披露与股价同步性——基于NER和TF-BIDF文本技术的分析[J]. 技术经济, 2025, 44(11): 1-17.

Yang Suchang, Cao Yining. Corporate social responsibility disclosure and stock price synchronicity: An analysis based on NER and TF-BIDF textual techniques[J]. Journal of Technology Economics, 2025, 44(11): 1-17.

数字经济驱动高质量发展:融合、创新与共享

企业社会责任信息披露与股价同步性

——基于NER和TF-BIDF文本技术的分析

杨肃昌, 曹益宁

(兰州大学经济学院, 兰州 730000)

摘 要:基于2008—2022年上市企业公开发布的社会责任报告,利用命名实体识别(named entity recognition,NER)和词频-前逆文档频率(term frequency-back Inverse document frequency,TF-BIDF)技术对文本内容进行分析,探讨企业社会责任信息披露所传递信号对股价同步性的影响机制。企业通过社会责任报告向外界传递与经营相关的信号,这些信号会对企业股价同步性产生重要影响。研究表明,社会责任报告中反映出的两类信号——传递企业特异信息的异质性信号,以及体现企业传达一致性的跨期一致性信号——均有助于减轻企业与市场间的信息不对称,从而降低股价同步性。进一步分析显示,非国有企业的信号异质性对股价同步性的降低作用更加显著;企业实际履行社会责任的水平越高,异质性信息含量和跨期一致性指数越显著;而信息披露水平的提高可能在一定程度上削弱信号效应。此外,通过更换样本及采用PSM-DID和Heckman模型进行了稳健性检验,以确保结论的可靠性。从信号性质的不同维度出发,对社会责任报告进行了内部内容分析与外部纵向比较,为研究企业社会责任报告对市场表现的影响提供了新视角。研究成果不仅为提高中国资本市场的信息效率提供了理论支持,也为当前社会责任信息披露制度的完善提供了政策建议。

关键词:社会责任报告;股价同步性;机器学习;文本分析;信息披露

中图分类号:F832.5 **文献标志码**:A **文章编号**:1002-980X(2025)11-0001-17

DOI:10.12404/j.issn.1002-980X.25021503

一、引言

股价同步性作为衡量单个公司股票价格变动与市场平均价格变动关联性的指标,有效反映了市场对信息的反应程度和速度,因此一直备受学界和资本市场关注。从现有学界研究来看,在中国这样一个新兴市场国家中,股价更多受市场层面影响,因此股价同步性往往较高[1-2]。2023年的中央金融工作会议指出,高质量发展是全面建设社会主义现代化国家的首要任务,金融要为经济社会发展提供高质量服务。金融如何推动经济社会高质量发展?关键在于提升股价变动对市场信息的敏感性和反应速度,使金融市场能够更高效地配置资源,最终服务于新时代社会主义建设。因此,研究中国资本市场上的信息披露如何影响股价同步性,探讨提升市场信息配置效率和降低股价同步性的路径,具有重要的理论与实践意义。

企业信息披露是减少与相关利益主体信息不对称的重要手段[3]。上市公司通常通过定期报告和临时报告进行信息披露。然而,临时报告通常仅针对特定事件进行披露,缺乏周期性,不利于系统性研究;定期

收稿日期:2025-02-17

作者简介:杨肃昌(1964—),博士,兰州大学经济学院教授,博士研究生导师,研究方向:审计体制、内部控制与风险管理、环境审计与区域经济发展;(通信作者)曹益宁(2001—),兰州大学经济学院硕士研究生,研究方向:公司金融、环境经济学。

① 参见中国证监会《上市公司治理准则》,国家生态环境部《企业环境信息依法披露管理办法》、上海、深圳证券交易所《上市公司环境信息披露指引》等文件。

报告虽有较长的时间跨度,但主要提供财务数据,文本信息有限,因此选择企业社会责任(corporate social responsibility,CSR)报告作为研究对象。同企业公布的其他报告相比,其有如下优势:第一,同企业的临时报告相比,企业社会责任报告受到相关规章制度规定,具有确定的披露时间①,并且自2008年起,沪深两市A股上市公司开始陆续发布社会责任报告,为研究提供了长时间跨度的文本数据;第二,同企业其他类型的定期报告相比,社会责任报告没有固定格式要求,因此企业之间的内容差异性为研究进行深入文本分析提供了可能性。

社会责任报告与环境、社会与治理(environment、society and governance,ESG)报告都是企业信息披露的重要形式,旨在减少信息不对称,提升市场透明度,帮助投资者更好地理解企业的社会责任履行和治理水平。尽管两者在具体内容和评估标准上有所不同:CSR报告侧重于企业社会责任的实践,而ESG报告更注重可持续发展指标。但它们共同提供了企业行为和绩效的关键信息。本文将从这两类报告的信息披露角度同股价同步性关系分析当前研究存在的争议:一是当前文献对于二者关系的研究存在不同的方法,但均存在缺陷。李新丽等[4]从2008年部分企业开始被强制要求披露企业社会责任报告入手,利用倾向得分匹配-双重差分(PSM-DID)模型研究得出了强制披露企业社会责任报告会对股价同步性产生积极影响。还有一批学者选择用文本分析方法来解决这个问题。例如,Xu等[5]选择A股上市公司的ESG报告作为研究对象,利用词频-逆文档频率(term frequency-inverse document frequency,TF-IDF)方式来计算企业的ESG报告和同行业内其他企业的ESG报告之间的相似度,最终得出相似度同股价同步性存在正相关的关系。以上文献虽然或是利用政策冲击,或是利用行业内比较等方法进行研究,但实际上仍是比较不同群体之间的社会责任报告差异带来的影响,忽视了不同企业的社会责任报告对于个体股价同步性的影响。二是当前部分文献选择采用第三方机构数据衡量企业社会责任行为,但评估标准之间存在差异性。不少文献选择采用ESG评级、分析师和政治身份等[6-8]来衡量企业的社会责任表现,但实际上已有越来越多的证据表明第三方机构因评估标准不同,其差异性可能进一步放大股价同步性[9-10]。三是当前文献对于企业信息披露和股价同步性之间的关系仍然存在争议。Xu等[5]通过计算行业间ESG报告的文本相似度证明了,行业间的ESG报告文本相似度越高,说明传达出来的独特有效信息越少,继而股价同步性也会越高。但Hu等[11]的研究却得出了相反结论,即企业的ESG分数越高,说明信息不对称程度得到了有效缓解,企业更加融入市场,企业的股价同步性因此提高,以这两篇文献为代表,说明当前学界对于企业信息披露对于股价同步性的影响仍然具有不同的观点。

从已有文献看出,学界在企业信息披露对股价同步性的影响的研究中已经取得了一定的成果,但既有文献仍然存在一定不足。一是对于二者关系的研究方法存在缺陷,其往往将企业划分为不同群体进行分析却忽视企业个体差异;二是常常使用第三方机构数据,但却因不同机构评估标准存在差异导致对信息披露实际影响产生差异化结果;三是当前文献仍然对企业信息披露和股价同步性之间的关系存在相反的结论。因此,近年来越来越多学者尝试从企业社会责任报告本身入手,通过文本分析和机器学习技术探索报告所传递的信号。

综上所述,本文研究认为从企业发布社会责任信息报告到市场根据报告所传达的信号做出反应,再到企业的股票市场价格表现发生变动的过程中,对于社会责任报告信号的多方面性质分析仍然是有必要的,也值得对这方面展开进一步详细的实证分析。现基于巨潮资讯网获得的2008—2022年的8936份企业社会责任报告,利用命名实体识别(named entity recognition,NER)技术和词频-前逆文档频率(term frequency-back inverse document frequency,TF-BIDF),通过对企业社会责任报告文档中的命名实体关键词和相同企业不同年份之间的文档相似度对比,分别从社会责任报告文本实际内容出发和从纵向时间维度对企业不同年份之间文本进行相似度比较,对企业社会责任报告之间的信息异质性含量和跨期一致性指数进行分析,来尝试得出企业社会责任报告中的多种性质对于企业自身的股价同步性的影响作用。

本文研究可能存在如下贡献:一是突破传统文本分析框架,首次将NER与TF-BIDF技术相结合构建信息披露度量体系。相较于既有研究采用的词频统计、可读性分析等表层指标,通过深度挖掘企业社会责任报告中的实体信息与跨期一致性,为信息披露研究提供了更为精准的技术路径。二是创新性地构建了信号

传递的双维度分析范式。基于信号理论的内在逻辑，将企业释放的信息解构为“异质性信号”与“一致性信号”两大类别。通过两种文本分析方式验证发现，两类信号对股价同步性分别产生差异化影响，这不仅完善了信号分类理论在公司治理领域的应用框架，更揭示了信息类型与市场反应间的传导机制。三是系统厘清了企业信息披露与股价同步性的作用关系。针对既有文献中“信息透明降低股价同步性”与“信号趋同加剧股价联动”的理论争议，通过实证表明，异质性信号通过提升企业信息透明度有效降低股价同步性，一致性信号通过传达企业信息连贯性也有效降低股价同步性。这一发现不仅为当前学术分歧提供证据，更为监管机构引导企业信息披露实践提供了理论依据。

二、文献综述和研究假设

（一）信息披露异质性与股价同步性

股价同步性反映了个股价格对市场或行业公共信息的依赖程度，是衡量资本市场信息效率的核心指标之一。当企业披露更多异质性信息，市场参与者能够更准确地识别公司真实价值，股票价格将更多反映企业自身基本面，而非市场或行业噪声，进而表现出较低的股价同步性[12]。因此，异质性信息的有效传递，成为改善市场定价效率的重要路径。

在宏观制度环境层面，研究指出制度缺陷将阻碍信息传播效率，导致企业特质信息难以传导至市场[1-2]。例如，巫岑等[13]发现，注册制下的首次公开发行会因为更充分的信息披露而改善同行其他已上市公司的信息环境，从而降低后者的股价同步性；孙凡和孙泽宇[14]基于各省份逐步上线公共数据开放平台的准自然实验发现，公共数据开放有助于提升资本市场信息效率，显著降低企业股价同步性；王雪和丁亚楠[15]发现投服中心行权通过改善公司治理和影响审计师决策，继而降低股价同步性。

在微观行为层面，不同外部主体的参与也影响信息异质性的传播。郑祖玄等[16]通过 ChatGPT 识别分析师深度报告指出，在投资者主动关注的情况下，分析师对公司数字化水平的评价能够影响资本市场表现，呈现出股价同步性的降低；于连超等[17]发现环境法庭的设立能降低股价同步性和提高股票定价效率；王明涛和李茜[7]利用分析师跟踪网络发现，这一网络中同样蕴含着公司异质性信息对于股价同步性产生积极影响；付强等[18]进一步指出，审计师数字化专长能带来更低的股价同步性且使分析师预测更精准；刘智宇等[19]发现投资者实地调研分别通过降低公司盈余管理程度和分析师盈余预测分歧两种机制发挥治理效应和信息效应，从而降低股价同步性。

企业内部结构也决定信息披露异质性的质量与路径。王建琼等[20]指出，高管连锁网络会提高公司的股价同步性，处于高管连锁网络越中心位置的上市公司，其股价同步性越高；化兵和孙武军[21]研究发现，多个大股东股权结构能够降低股价同步性，且通过提高信息披露质量和提升公司治理水平降低了股价同步性；赵彦锋[22]揭示了国有股东参股通过长期投资导向机制和市场关注机制能够显著降低民营企业股价同步性。

（二）信息披露一致性与股价同步性

同信息异质性相比，信息披露的一致性更强调跨期内容的延续性、可比较性与逻辑连贯性，是衡量企业信息披露质量的重要维度。信息经济学理论指出，高一致性披露行为可以有效降低信息不对称与市场摩擦，增强市场对企业未来预期的理解与判断力，从而提高价格对企业异质性信息的响应能力[23]。

从信号可信度角度进行分析，持续一致的信息披露有助于增强市场信任。Wu[24]关注中国强制性企业社会责任披露政策对股价同步性的影响，研究发现强制性社会责任披露使得企业持续向社会传递企业信息，影响投资者的分类学习行为，进而对股价同步性产生正向影响。李子健等[25]以管理层讨论与分析（management’s discussion and analysis，MD&A）文本为研究对象发现，MD&A 中前瞻性内容越多，股价同步性越低。该前瞻性信息减少了分析师预测误差与分歧，抑制了噪声交易者的参与，为一致性信息提升市场信息效率提供了另一角度的验证。

一致性信息的传递还有助于降低信息处理负担。毕晓方等[26]发现，稳定型机构投资者通过促进信息属性盈余平滑，显著降低了股价同步性，特别是在内部控制薄弱或市场竞争激烈的情形下效果更显著。李玉敏等[27]进一步指出，共同机构投资者通过“用脚投票”治理机制提高企业信息透明度，在降低噪声交易和提

升信息效率方面发挥关键作用,从而实现股价同步性的下降。此外,杨科等[28]在其研究中指出,投资者的有限关注和情绪渠道可能导致对信息的处理不充分。而一致性的信息披露可以缓解这一问题,使投资者更容易获取和处理企业信息,从而降低股价同步性。

(三)研究假设

基于信号传递理论[29],CSR报告中的异质性信息具有不可模仿性与专属性特征[30],其本质是管理层向市场传递企业独特的可持续发展战略与资源禀赋的专属信号。李四海和李震[31]的实证研究表明,由于CSR实践需要长期组织资源投入与治理结构适配,低社会责任履行度的企业难以通过短期模仿行为复制高CSR企业的异质性信息,这使得异质性信息成为缓解信息不对称的有效载体。具体而言,当企业披露ESG投资倾向、利益相关者管理创新或绿色技术专利等异质性信息时,将吸引专业信息中介(如分析师与机构投资者)的解读与验证[32],促使私有信息通过市场交易转化为公共信息[23],从而提升股价对企业异质性冲击的敏感性。

此外,根据Akerlof[33]的"柠檬市场"理论,信息不对称可能引发投资者对低质量企业的逆向选择。而高异质性信息含量的CSR报告通过传递可验证的专有知识(如经第三方审计的减排数据),形成"自我选择信号"[29],将高CSR履行度企业与低质量模仿者分离。市场通过信号甄别强化对优质企业的定价,减少因信息模糊导致的股价同步性较高的现象。

据此,本文提出假设1:

企业CSR报告异质性信息含量与企业股价同步性负相关,即异质性信息含量越高,股价同步性越低(H1)。

信息经济学理论指出,信息披露一致性是降低市场信息摩擦的关键机制[12]。一方面,跨期稳定的CSR报告通过"声誉抵押"效应增强信号可信度:当企业持续披露逻辑自洽、数据可比的可持续发展目标(如碳排放强度下降路径或供应链合规承诺)时,投资者可依据历史披露验证企业承诺的兑现能力,进而形成对管理层诚信度的贝叶斯更新[34],减少因信息不确定性引发的估值分歧[35]。另一方面,一致性信息通过构建"叙事连贯性"降低投资者认知成本[28]。例如,持续被披露积极履行ESG责任的企业,其市场表现和股票收益更易被市场识别[36],从而削弱由短期噪声交易引发的股价共移现象。实证研究表明,信息披露一致性高的企业,其经营稳定性和市场反应更为良好[37],分析师预测离散度也显著下降[38],上述结果为一致性降低股价同步性提供了直接证据。

据此,本文提出假设2:

企业CSR报告信息一致性与企业股价同步性负相关,即信息一致性程度越高,股价同步性越低(H2)。

三、研究设计

(一)变量定义与模型建立

1. 异质性信息含量的度量

在信息量较高的社会责任报告中,通常包含更多具有特定语义指向的命名实体,这些实体包括具体的地名、机构名称、人物姓名等。通过这些具有明确语义边界的实体词汇,高信息含量的社会责任报告能够显著区别于那些输出低质、内容同质化的报告文本。以平安银行2008年和2022年的社会责任报告为例②:2008年的报告中提到,"深圳发展银行致力于持续推动区域经济发展。华东地区,大力扶持民营企业的发展,鼓励江浙私营经济的快速发展,并培育了一大批忠诚客户……"。而2022年的报告则写道,"平安银行等三方合作研发全国首个智慧破产审理系统,并推出智慧破产'芯'服务,为办理破产提供信息化、智能化服务。2022年10月,本行智慧破产'芯'服务荣获广东省金融科技进步奖二等奖。"两段文字均涉及企业在报告期内促进区域经济发展的相关内容,但从信息输出质量来看,后者显著更高。其原因在于,2008年的报告

② 数据来源于巨潮资讯网,报告内容分别节选于《深圳发展银行股份有限公司2008年度社会责任报告》和《平安银行股份有限公司2022年可持续发展报告》,其中深圳发展银行股份有限公司于2012年更名为平安银行股份有限公司。

主要使用“推动区域经济快速发展”“扶持民营企业”等泛化表达，而2022年的报告则提供了“智慧破产审理系统”“广东省金融科技进步奖二等奖”等具体的专业实体。这种内容层次的差异正是本文采用NER技术来度量社会责任报告信号异质性的逻辑基础。本文研究在技术层面的度量方式具体如下：在巨潮资讯网以“社会责任”为关键词对文本进行下载，并摒弃掉含有关键词“英文”“已取消”“公告”“摘要”“委员会”“意见”“指导”“说明”“制度”的下载文件，以确保分析内容都为企业社会责任报告原文而非企业公布的其他制度性文件（人员安排、公司规章制度说明等）。随后，针对部分识别不完整或格式混乱的报告文本，本文使用百度“飞桨”平台的字符识别技术进行文字提取与修正。接着，利用搭建在TensorFlow机器学习平台上的NER技术进行抽取，借鉴张忠良等[39]、李四海和李震[31]及Huang等[40]通过BERT语言处理预训练模型来识别命名实体。最后，再将借此分类出来的实体词统计出词频，得出了对于社会责任报告的异质性信息含量如式（1）所示。

$$Specificity_{i,t} = \sum_{k=1}^{n} NER_{k,i,t} \times NERTimes_{k,i,t} / Reports_{k,i,t} \tag{1}$$

其中：$Specificity_{i,t}$为企业i在t年的异质性信息含量，是指在企业i在t年发布的拥有n个命名实体词的企业CSR报告中所有命名实体词的总词频；$NER_{k,i,t}$为企业i在t年的CSR报告利用NER技术识别出来的第k个命名实体词的具体字数；$NERTimes_{k,i,t}$为企业i在t年的CSR报告第k个命名实体词出现的词频；$Reports_{k,i,t}$为第k个命名实体词所在的企业i在t年的CSR报告总词数。

2. 跨期一致性指数的度量

本文中定义的“跨期一致性”是指同一家企业在不同年度发布的社会责任报告之间的文本相似程度，即报告内容相较于前一年度报告的相似度。为衡量这种跨期相似性，参考Kelly等[41]、吕璐成和赵亚娟[42]及Li等[43]的研究方法，采用TF-BIDF方法对报告文本的分词结果进行加权处理，并基于余弦相似度计算不同年度报告之间的文本相似度。TF-BIDF方法最初多用于衡量专利文本之间的差异性，现将其引入企业社会责任报告文本分析中，以实现对报告跨期一致性的度量。具体而言，TF-BIDF是在传统的TF-IDF算法基础上的一种改进形式。为便于理解，首先阐述TF-IDF算法的基本计算原理，随后介绍TF-BIDF相较于TF-IDF在权重计算方面的改进机制，最后说明如何利用余弦相似度对经加权处理的文本向量进行计算，从而获得企业社会责任报告间的跨期文本相似度指标。

在以中文为研究对象的文本处理中，文档间相似度的计算一直是一项具有挑战性的任务。由于中文语言中存在大量虚词与功能词，如“的”“地”“得”等，这些词在文本中出现频率较高但缺乏实质语义信息，容易对相似度计算结果造成干扰，从而影响模型的准确性。为降低无意义高频词的影响，学界普遍采用利用TF-IDF对文本词项进行恰当的加权，该方法会使词组的重要性随着其在文件中出现的次数成正比增加，但同时会随着其在语料库中出现的频率成反比下降。*TF-IDF*具体计算方式如式（2）~式（4）所示。

$$TF_{i,j} = \frac{n_{i,j}}{\sum_{k} n_{k,j}} \tag{2}$$

$$IDF_i = \ln \frac{|D|}{|1 + \{j : t_i \in d_j\}|} \tag{3}$$

$$TF\text{-}IDF_{i,j} = TF_{ij} \times IDF_i \tag{4}$$

其中：$TF_{i,j}$为指词语i在文档j中的词频；$n_{i,j}$为指词语i在文档j中出现的次数；$\sum_{k} n_{k,j}$为文档j共k个词语出现的次数总和；IDF_i为指词语i的逆文档频率；$|D|$为语料库中的文档总数；$\{j : t_i \in d_j\}$为语料库d内包含词语i的文件总数，如果词语i不在语料库内会导致分母为0，所以一般会加上1。以上两者相乘便是词语i在文档j的*TF-IDF*指数。

尽管TF-IDF算法在文本加权中具有较好的适用性，但在处理长时序文本时仍存在一定局限。其主要问题在于：某些词语的出现具有显著的时间阶段性，即这些词语可能仅在特定时期流行。如果将这类

词语在后续年份零散出现的文档全部纳入IDF的计算范围,则会导致其IDF值被低估,从而造成权重分配失衡。以本文研究的社会责任报告为例,2008年的社会责任报告中出现的高频率词有“驰名商标”“名牌产品”“劳资关系”等,2015年的社会责任报告中的高频率词有“生态建设”“非营利性”“风险管理”等,2022年的社会责任报告中的高频率词有“芯片级”“高技术”“十四五”等。这些变化反映了不同时期社会环境与政策导向对企业社会责任话语风格的影响。显然,某一时期的流行词并不意味着其在后续年份完全消失,但若将其零星出现的文本继续计入IDF统计,将使权重偏向低估该词在特定时期内的重要性,进而影响TF-IDF的准确性。为解决上述问题,本文在TF-IDF的基础上引入时间修正机制,提出TF-BIDF。该方法通过设定时间截点,仅统计具有明显时代特征的词语在该时间点之前出现的文档数,从而修正传统IDF在时间维度上的累积偏差。随后,根据报告年份对分词结果重新分类,计算各年份词组的出现频率。对于在当年词频排名位于前5%且在前三年未出现过的词组,将其定义为“当年流行词”。对于这些流行词,本文只计算在该年(含当年)之前的逆文档频率,因此命名该种方式计算出来的IDF为BIDF,具体计算公式如式(5)所示。

$$BIDF_i = \ln \frac{|D_t|}{|1 + \{j_t : t_i \in d_j\}|} \tag{5}$$

其中:$|D_t|$为指流行词i在流行年t及其之前语料库中的文档总数;$\{j_t : t_i \in d_j\}$为指流行词i在流行年t及其之前在语料库d内包含流行词i的文件总数。

本文采用余弦相似度进行文本相似度的计算,余弦相似度是指文本分词后用向量空间表示,之后计算两个向量夹角的余弦值,用这个余弦值来衡量两个个体间差异的大小。余弦值越接近1,就表明夹角越接近0度,亦即两个向量越相似,余弦相似度越高。具体公式如式(6)所示。

$$\rho_{it,it-1} = \frac{V_{it}}{\| V_{it} \|} \frac{V_{it-1}}{\| V_{it-1} \|} \tag{6}$$

其中:$\rho_{it,it-1}$为同一家企业i在t年的企业社会责任报告同前一年($t-1$年)的报告相似度之比;V_{it}指用TF-BIDF将企业i在t年的社会责任报告构词组构成的向量空间;V_{it-1}指用TF-BIDF将企业i在$t-1$年的社会责任报告构词组构成的向量空间。

3. 股价同步性的度量

通过借鉴已有文献[1,44-45],本文分别计算出三种股价同步性,计算公式如(7)~式(9)所示。

$$R_{i,w,t} = \beta_0 + \beta_1 R_{m,w,t} + \varepsilon_{i,w,t} \tag{7}$$

$$R_{i,w,t} = \beta_0 + \beta_1 R_{m,w,t} + \beta_2 R_{I,w,t} + \varepsilon_{i,w,t} \tag{8}$$

$$R_{i,w,t} = \beta_0 + \beta_1 R_{m,w,t} + \beta_2 R_{I,w,t} + \beta_3 R_{m,w-1,t} + \beta_4 R_{I,w,-1,t} + \varepsilon_{i,w,t} \tag{9}$$

其中:$R_{i,w,t}$为企业i在t年w周考虑现金红利再投资的收益率;$R_{m,w,t}$为企业i在t年w周考虑现金红利再投资的流通市值加权综合周市场回报率;$R_{m,w-1,t}$为企业i在t年$w-1$周考虑现金红利再投资的流通市值加权综合周市场回报率;$R_{I,w,t}$为企业i在t年w周所在行业剔除自身后其他股票考虑现金红利再投资的流通市值加权综合周市场回报率;$R_{I,w-1,t}$为企业i在t年$w-1$周所在行业剔除自身后其他股票考虑现金红利再投资的流通市值加权综合周市场回报率(行业划分参照证监会2012年行业分类标准)。

在对式(7)~式(9)分别计算得出拟合优度R^2后,对拟合优度进行对数化处理,得到的$SYNCH_{i,t}$为企业i在t年的股价同步性指标。

$$SYN_{i,t} = \ln \frac{R^2_{i,t}}{1-R^2_{i,t}} \tag{10}$$

4. 控制变量说明与具体定义

参照Hu等[11]的研究,选择以下变量作为控制变量:总资产收益率(*ROA*)、营业收入增长率(*Grow*)、托宾Q(*TQ*)、控股股东持股比率(*ContrashProportion*)、董事长和总经理是否为同一人(*ConcurrentPosition*)、董事会人数对数值(ln*Board*)、独立董事占比对数值(*IndDirectorRatio*)等。相关变量的具体定义见表1。

表 1　变量说明

变量符号	变量含义	定义说明
CSRSP	基于企业社会责任报告的异质性信息含量	异质性词组总频数占全报告字数
CSRSim	基于不同年份企业社会责任报告的跨期一致性指数	报告相较于前一年度相似值
ROA	总资产收益率	净利润/总资产
Grow	营业收入增长率	营业收入增加值/营业收入值
TQ	托宾 *Q*	企业市值/企业总资产
SYN_1	基于式(7)计算出的 R^2 计算出	对式(7)R^2 对数化处理
SYN_2	基于式(8)计算出的 R^2 计算出	对式(8)R^2 对数化处理
SYN_3	基于式(9)计算出的 R^2 计算出	对式(9)R^2 对数化处理
ContrashProportion	控股股东持股比率	控股股东或第一大股东占总公司持股
ConcurrentPosition	董事长和总经理是否为同一人	两职合一则为 1，否则为 0
ln*Board*	董事会人数	对董事会人数进行对数化处理
IndDirectorRatio	独立董事占比	独立董事占董事会总人数占比

5. 模型建立

本文选取利用式(7)~式(9)算出的股价同步性 SYN_1、SYN_2、SYN_3 作为因变量，将 *CSRSP* 和 *CSRSim* 作为自变量，具体模型建立如下：

$$SYN_{it} = \beta_0 + \beta_1 CSRSP_{it} + \beta_2 Controls_{it} + \varepsilon_{it} + \lambda_t + \mu_I \tag{11}$$

$$SYN_{it} = \beta_0 + \beta_1 CSRSim_{it} + \beta_2 Controls_{it} + \varepsilon_{it} + \lambda_t + \mu_I \tag{12}$$

其中：本文分别使用 $CSRSP_{it}$（异质性含量）和 $CSRSim_{it}$（跨期一致性）来当作对于股价同步性的主要解释自变量；$Controls_{it}$ 为控制变量，包括总资产收益率（*ROA*）、营业收入增长率（*Grow*）、托宾 *Q*（*TQ*）、控股股东持股比率（*ContrashProportion*）、董事长和总经理是否为同一人（*ConcurrentPosition*）、董事会人数对数值（ln*Board*）、独立董事占比对数值（*IndDirectorRatio*）等变量。考虑到企业社会责任报告在同一行业内可能存在的差异性等因素，本文参照李四海和李震[31]及伊志宏[45]采用行业和年份固定效应，并使用对于企业个体的聚类标准误。

（二）数据来源和描述性统计

本文以 2008—2022 年的上市企业 A 股企业社会责任报告为研究样本，并对部分利用光学字符识别（OCR）技术仍然无法处理的文本做出了删除处理。此外，在进行文本信号一致性的分析时，因为处理方式是利用当前年份文本同前一年做出对比，因此所有上市企业的社会责任报告的第一年数据都无法被统计在内。

表 2 是本文主要变量的描述性统计。结果表明，企业社会责任报告的异质性信息含量均值为 0.031，最大值却为 0.572，体现出企业整体社会责任报告输出有效信息量的含量较低，但也存在一部分企业具有较高的社会责任报告异质性。对于社会责任报告的跨期一致性指数来说，均值为 0.390，体现出企业整体倾向于向社会表达出较高一致程度的信号。

表 2　描述性统计

变量名	观测值	均值	标准差	最小值	最大值
CSRSP	8392	0.031	0.023	0.000	0.572
CSRSim	6946	0.390	0.200	0.000	0.999
SYN_1	8392	−0.067	1.438	−17.320	1.873
SYN_2	8392	−0.351	1.033	−6.675	2.374
SYN_3	8392	−0.202	0.935	−5.752	2.426
ROA	8392	0.048	0.949	−0.540	4.248
Grow	8392	0.286	4.111	−0.972	251.200
TQ	8392	2.317	1.334	0.705	25.430
ContrashProportion	8392	37.010	15.690	0.004	89.410
ConcurrentPosition	8392	0.189	0.392	0.000	1.000
ln*Board*	8392	2.196	0.224	1.386	3.045
IndDirectorRatio	8392	37.550	5.773	0.000	80.000

四、实证结果分析

(一)社会责任报告中异质性信息含量对于股价同步性影响的检验

表 3 为利用式(11)求出的股价同步性和社会责任报告异质性信号之间的关系。其中(1)列~(3)列分别指用式(7)~式(9)计算得出的三种股价同步性作为因变量的解释结果。可见,随着对于拟合优度的求值公式的不断细化,常数项绝对值随指标精确性提高而减小。

表 3 的(2)列和(3)列显示,信号异质性的系数分别是-1.532 和-1.368,均在 1%的水平显著,表明社会责任报告中的异质性信息含量越高,对于企业的股价同步性指标就呈现出越强的负相关作用,验证了本文假设 H1。

表 3　异质性信息含量检验

变量	(1)	(2)	(3)
	SYN_1	SYN_2	SYN_3
CSRSP	-0.630(0.501)	-1.532***(0.459)	-1.368***(0.410)
ROA	0.007(0.01)	0.011***(0.00)	0.008*(0.00)
Grow	-0.004(0.00)	-0.006*(0.00)	-0.005*(0.00)
TQ	-0.110***(0.02)	-0.072***(0.01)	-0.061***(0.01)
ContrashProportion	-0.003***(0.00)	-0.002*(0.00)	-0.001*(0.00)
ConcurrentPosition	-0.064*(0.04)	-0.052*(0.03)	-0.053**(0.03)
ln*Board*	0.103(0.08)	0.222***(0.07)	0.195***(0.06)
IndDirectorRatio	0.004*(0.00)	0.008***(0.00)	0.007***(0.00)
Constant	-1.062***(0.229)	-0.844***(0.202)	-0.665***(0.185)
观测值	8392	8392	8392
R^2	0.425	0.366	0.370
行业固定效应	是	是	是
年份固定效应	是	是	是

注:***、**、*分别表示估计系数在 1%、5%、10%的水平上显著;括号内为聚类到企业层面的标准误。

(二)社会责任报告中跨期一致性指数对于股价同步性影响的检验

表 4 为利用式(12)求出的股价同步性和企业社会责任报告的信号一致性的检验。同样地,随着股价同步性指标的精确性增加,本文计算出的常数项也在不断变小,反映出本文对于跨期一致性指数和股价同步性关系判断的愈发准确。

表 4 的(2)列和(3)列表明,利用跨期一致性指数作为解释变量的系数分别为-0.138 和-0.129,且二者均在 5%水平显著。这证明企业通过数年来连续一致的社会责任报告的发布,可以向社会传达出自己经营稳定的信号,继而降低企业自身的股价同步性,验证本文假设 H2。

表 4　信号一致性检验

变量	(1)	(2)	(3)
	SYN_1	SYN_2	SYN_3
CSRSim	-0.099(0.071)	-0.138**(0.063)	-0.129**(0.057)
ROA	0.008(0.01)	0.012***(0.00)	0.008*(0.00)
Grow	-0.020***(0.01)	-0.022***(0.01)	-0.018***(0.01)
TQ	-0.108***(0.02)	-0.071***(0.02)	-0.061***(0.01)
ContrashProportion	-0.002**(0.00)	-0.001(0.00)	-0.001(0.00)
ConcurrentPosition	-0.050(0.04)	-0.053(0.03)	-0.053*(0.03)
ln*Board*	0.119(0.08)	0.263***(0.07)	0.240***(0.06)
IndDirectorRatio	0.005*(0.00)	0.008***(0.00)	0.008***(0.00)
Constant	-1.095***(0.250)	-0.952***(0.218)	-0.778***(0.200)
观测值	6946	6946	6946
R^2	0.414	0.348	0.352
行业固定效应	是	是	是
年份固定效应	是	是	是

注:***、**、*分别表示估计系数在 1%、5%、10%的水平上显著;括号内为聚类到企业层面的标准误。

（三）稳健性检验——基于信号可靠性

1. 利用企业年报文本和分析师报告摘要进行替代性解释排除

基于企业社会责任报告的研究一直具有争议，企业社会责任报告不强制要求发布，同时也不要求具体格式，因此存在潜在的漂绿和自选择问题[46-47]。基于企业社会责任报告的文本分析得出的结论也会因此受到质疑。

本文参照 Hollweck 等[48]的做法，利用“排除替代性解释”来验证本文结论的稳健性，即通过另一个更客观且更具说服力的文本也采取同样的处理手段，如果得出的结论仍与前文相同，则有理由认为前文所做的实证措施是具有可信度的。

表 5　基于年报和分析师报告的稳健性检验

变量	(1)	(2)	(3)	(4)	(5)	(6)
	SYN_1	SYN_2	SYN_3	SYN_1	SYN_2	SYN_3
年报异质性信息	−2.067***	−1.734***	−1.401***			
	(0.45)	(0.38)	(0.32)			
分析师异质性信息				−0.185**	−0.460***	−0.396***
				(0.09)	(0.07)	(0.07)
ROA	0.017***	0.019***	0.013***	0.014***	0.019***	0.013***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Grow	−0.000***	−0.000***	−0.000***	−0.000***	−0.000***	−0.000***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
TQ	−0.040***	−0.034***	−0.029***	−0.055***	−0.038***	−0.034***
	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
ContrashProportion	−0.002***	−0.000	−0.000	−0.003***	−0.002***	−0.002***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ConcurrentPosition	−0.093***	−0.076***	−0.069***	−0.120***	−0.102***	−0.094***
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
ln*Board*	0.362***	0.391***	0.343***	0.341***	0.384***	0.338***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
IndDirectorRatio	0.004***	0.006***	0.006***	0.006***	0.008***	0.008***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	−1.939***	−1.501***	−1.241***	−1.781***	−1.364***	−1.113***
	(0.14)	(0.13)	(0.12)	(0.16)	(0.14)	(0.12)
观测值	37586	37586	37586	25738	25738	25738
R^2 值	0.427	0.369	0.381	0.412	0.355	0.362
行业固定效应	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是

注：***、**、* 分别表示估计系数在 1%、5%、10%的水平上显著；括号内为聚类到企业层面的标准误。

2. 利用 Word2Vec 模型检验 BERT 模型准确性

与 Word2Vec 模型相比，BERT 模型在强调动态上下文联系的同时，可能削弱了对静态词向量的识别精度[39]。因此，采用 Word2Vec 模型替代 BERT 模型对社会责任报告文本进行处理，对企业异质性信息含量进行重新计算，以检验前述结论的稳健性。具体而言，本文以腾讯 AI 实验室发布的 Word2Vec 预训练模型所得的静态词向量为基础，构建企业社会责任报告的文本表示，并按照原模型设定重新回归。结果如表 6 所示，Word2Vec 处理下的企业社会责任报告异质性含量（*Word2Vec CSRSP*）系数虽然受静态词向量识别影响被放大，但在所有模型中均为负且在 1%水平上显著，表明社会责任报告的异质性信息含量越高，企业股价同步性显著降低，与基于 BERT 模型的主回归结果保持一致，结论依然稳健。这进一步验证了企业通过社会责任报告传递的异质性信息确实能有效减少市场信息不对称，降低股价同步性。

3. 利用 PSM-DID 缓解内生性问题

为更加稳健地评估企业社会责任报告所传递的公司信号对股价同步性的影响,本文引入政策性外生冲击作为检验情境。具体而言,《关于构建绿色金融体系的指导意见》及其后续配套文件提出,重点排污上市公司应强制披露环境信息,其他上市公司实行“不披露则解释”原则,到 2020 年实现所有上市公司强制披露环境信息。正如前文所说,企业社会责任报告通过传递出异质性信号和一致性信号,继而降低了股价同步性。而上述文件的出台,通过半强制式要求企业进行相关环境信息的披露,使得上市企业信息披露变多,继而稀释了企业中社会责任报告所传达出的企业独特信号。本文将发布了企业社会责任报告的企业定为处理组,取值为 1,将那些未发布企业社会责任报告的企业定为对照组,取值为 0。对处理组企业进行了近邻匹配,并按照式(11)和式(12)的控制变量选取了协变量,回归结果见表 7 的(1)列。结果显示,政策效应在 10%水平显著,显示政策效应稀释社会责任报告的信号效应,与主要假设相一致。

4. 利用 Heckman 两阶段回归缓解内生性问题

需要指出的是,鉴于目前我国企业社会责任报告的披露仍处于自愿阶段,企业是否选择发布报告往往与其公司治理水平、社会形象或信息披露动机等内生特征相关。因此,在以发布社会责任报告的企业为研究对象时,可能面临样本选择偏误的问题。若未能有效控制这种内生性,可能导致实证结果对总体企业群体的外推性受到一定限制。为了尝试解决这一问题,选择使用 Heckman 两阶段模型进行检验。在第一阶段构建虚拟变量企业是否出具 CSR 报告的选择模型,之后使用 Probit 模型将其作为解释变量进行回归,同时选择基础回归中的所有控制变量,利用这次模型的拟合值计算出逆米尔斯比率(*IMR*)。在第二阶段,将第一阶段算出的 *IMR* 加入原基础回归中,来纠正可能潜在的选择性偏差对结论的干扰。回归结果见表 7 的(2)列,由于异质性信息含量指标在样本覆盖范围上较为完整,而跨期一致性指数受计算方法限制,需对样本进行删减,故本文仅选取信号异质性指标进行回归。结果显示,在控制样本选择偏误问题后,企业社会责任报告的异质性信息含量与股价同步性之间依然呈现显著的负相关关系,且在 1%的显著性水平上通过统计检验。该结果进一步验证了基础回归的结论,表明社会责任报告中所蕴含的异质性信息确实能够降低企业股价同步性。

表 6　基于 Word2Vec 模型的稳健性检验

变量	(1)	(2)	(3)
	SYN_1	SYN_2	SYN_3
Word2Vec CSRSP	−20.722***	−24.940***	−21.871***
	(7.44)	(6.45)	(5.76)
ROA	0.007	0.011***	0.008*
	(0.01)	(0.00)	(0.00)
Grow	−0.004	−0.006*	−0.005*
	(0.00)	(0.00)	(0.00)
TQ	−0.110***	−0.072***	−0.061***
	(0.02)	(0.01)	(0.01)
ContrashProportion	−0.003***	−0.002*	−0.001*
	(0.00)	(0.00)	(0.00)
ConcurrentPosition	−0.064*	−0.053*	−0.054**
	(0.04)	(0.03)	(0.03)
ln*Board*	0.087	0.207***	0.183***
	(0.08)	(0.07)	(0.06)
IndDirectorRatio	0.004	0.007***	0.007***
	(0.00)	(0.00)	(0.00)
Constant	−0.987***	−0.789***	−0.618***
	(0.23)	(0.20)	(0.19)
观测值	8392	8392	8392
R^2	0.426	0.366	0.371
行业固定效应	是	是	是
年份固定效应	是	是	是

注:***、**、*分别表示估计系数在 1%、5%、10%的水平上显著;括号内为聚类到企业层面的标准误。

表 7　利用 PSM-DID 和 Heckman 两阶段缓解内生性问题

变量	(1)	(2)
	SYN_3	
政策效应	0.095***	
	(0.02)	
CSRSP		−1.357***
		(0.41)
ROA	0.012***	0.031
	(0.00)	(0.03)
Grow	−0.000***	−0.007**
	(0.00)	(0.00)
TQ	−0.027***	−0.093**
	(0.01)	(0.04)
ContrashProportion	−0.000	0.000
	(0.00)	(0.00)
ConcurrentPosition	−0.054***	−0.266
	(0.01)	(0.23)
ln*Board*	0.287***	1.091
	(0.03)	(0.97)
IndDirectorRatio	0.006***	0.025
	(0.00)	(0.02)
Constant	−1.184***	−4.575
	(0.11)	(4.24)
观测值	39699	8392
R^2	0.388	0.370
行业固定效应	是	是
年份固定效应	是	是

注:***、**、*分别表示估计系数在 1%、5%、10%的水平上显著;括号内为聚类到企业层面的标准误。

五、异质性分析

在下文中的异质性检验中，出于简洁性的需要，本文统一选择由式(9)计算出的 *SYN*3 作为被解释变量。

(一)企业的所有权(产权)性质对于企业信号的影响

企业产权性质对信息披露具有显著影响，本文选择将样本分为国有和非国有两类，继而尝试探寻产权性质是否会对企业社会责任报告的信号传递产生影响。如表 8 所示，分别将研究样本分为国有和非国有两类企业，结果显示，将样本数据按照企业所有权性质划分后，无论是异质性信息含量还是跨期一致性指数都对股价同步性有着负向影响，但是只有非国有企业中的信号异质性对于股价同步性的影响在 1%的水平显著。

具体机理推测可能如下：非国有企业同国有企业相比，本身市场就对其信任程度较低，其信息披露透明度也相对不足，因此对其社会责任报告中的异质性信息，市场反应更为敏感，因此结果更为显著。至于信号一致性方面，结果与前文跨期一致性指数对股价同步性影响较弱的结论相一致。在异质性分组分析中，一致性指标在国有与非国有企业间的差异并未表现出显著性。这表明无论是国有企业还是非国有企业，对社会责任报告的跨期一致性均缺乏足够重视，企业在不同年度间对报告内容的延续性与一致性管理尚未形成稳定机制。

表 8　基于企业产权性质的异质性检验

变量	(1)	(2)	(3)	(4)
	SYN_3			
	国有	非国有	国有	非国有
CSRSP	−0.592 (0.555)	−2.018*** (0.660)		
CSRSim			−0.085(0.082)	−0.106 (0.081)
ROA	0.005 (0.00)	−0.022*** (0.01)	0.006 (0.00)	−0.016 (0.01)
Grow	−0.018*** (0.00)	−0.002 (0.00)	−0.016*** (0.00)	−0.015 (0.01)
TQ	−0.084*** (0.02)	−0.040*** (0.01)	−0.079*** (0.02)	−0.044*** (0.02)
ContrashProportion	−0.002* (0.00)	−0.003** (0.00)	−0.002* (0.00)	−0.002 (0.00)
ConcurrentPosition	−0.033 (0.05)	−0.036 (0.03)	−0.043 (0.05)	−0.037 (0.04)
ln*Board*	0.143* (0.08)	0.078 (0.10)	0.231*** (0.08)	0.095 (0.11)
IndDirectorRatio	0.007** (0.00)	0.003 (0.00)	0.008*** (0.00)	0.003 (0.00)
Constant	−0.358 (0.24)	−0.420 (0.30)	−0.604** (0.25)	−0.452 (0.34)
观测值	4834	3533	4113	2808
R^2	0.359	0.375	0.345	0.351
控制变量	是	是	是	是
行业固定效应	是	是	是	是
年份固定效应	是	是	是	是

注：***、**、*分别表示估计系数在 1%、5%、10%的水平上显著；括号内为聚类到企业层面的标准误。

(二)实际责任履行程度对于企业信号的影响

企业社会责任报告是企业自身基于报告前一年自身表现进行的总结，而对于企业 ESG 评级，是第三方机构基于企业在评级前一年 ESG 领域的实际表现打出。第三方机构具有相对独立性，其出具的评估结果在一定程度上能够客观反映企业的社会责任履行水平。因此，选取 ESG 评级中的“S”(社会)维度独立评级作

为衡量企业实际社会责任履行状况的指标。

本文选择彭博ESG评级中的社会评级作为区分企业内实际责任履行程度高低的依据，理由如下：①彭博ESG评级相对覆盖时间较长，从2011年以后都有，且覆盖企业范围较广；②彭博ESG评级不仅具有ESG评级，且对它评价的企业还都有独立的E、S、G评级，这符合本文基于企业社会责任报告的研究。实证处理中，本文将彭博S评级按照对于中位数高低进行区分，高于中位数的划分为高S组，低于中位数的划分为低S组。

实证结果如表9所示，高S组无论是异质性信息含量还是跨期一致性指数，对于股价同步性的影响均在统计结果上显著，系数分别为-1.358和-0.237，分别在5%和1%的水平上显著。这说明企业社会责任报告不足以削减市场上的信息不对称现象，只有在独立第三方都认可的行为加持下，市场才会予以反应，股价同步性才会发生显著变动。此外，需要注意低S组中的跨期一致性同股价同步性呈现正相关关系，说明在低社会责任实际履行情况下，重复度高低质量的社会责任报告反而加重企业同市场之间的信息不对称情况，提升了股价同步性。

表9　基于实际责任履行情况的异质性检验

变量	(1)	(2)	(3)	(4)
	SYN_3			
	高S	低S	高S	低S
CSRSP	-1.358**	-1.052		
	(0.555)	(0.702)		
CSRSim			-0.237***	0.028
			(0.081)	(0.089)
ROA	0.001	0.011	0.001	0.008
	(0.00)	(0.01)	(0.00)	(0.01)
Grow	-0.002	-0.016**	-0.016***	-0.013
	(0.00)	(0.01)	(0.01)	(0.01)
TQ	-0.040***	-0.068***	-0.036**	-0.079***
	(0.01)	(0.02)	(0.02)	(0.03)
ContrashProportion	0.000	-0.003**	0.001	-0.002
	(0.00)	(0.00)	(0.00)	(0.00)
ConcurrentPosition	-0.022	-0.063	-0.014	-0.077*
	(0.04)	(0.04)	(0.05)	(0.04)
ln*Board*	0.075	0.208**	0.128	0.205**
	(0.09)	(0.09)	(0.09)	(0.09)
IndDirectorRatio	0.007**	0.009***	0.007**	0.008**
	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-0.457*	-0.619**	-0.554*	-0.635**
	(0.28)	(0.26)	(0.29)	(0.27)
观测值	2951	3028	2608	2667
R^2	0.376	0.354	0.383	0.357
控制变量	是	是	是	是
行业固定效应	是	是	是	是
年份固定效应	是	是	是	是

注：***、**、*分别表示估计系数在1%、5%、10%的水平上显著；括号内为聚类到企业层面的标准误。

(三)实际信息披露程度对于企业信号的影响

企业社会责任报告属于企业信息披露中的重要组成部分，因此考察不同信息披露程度下社会责任报告信号效应的差异，具有重要研究意义。对于信息披露情况完善的企业来说，可能存在企业信息披露的“信息过载”现象，即过度的信息披露会在一定程度上稀释社会责任报告中蕴含的异质性信息含量与跨期一致性

信号，使市场难以准确识别其中的核心信息。相反地，对于信息披露程度相对有限的企业，社会责任报告所传递的信号可能更为集中与显著，从而放大其在资本市场中的边际影响。

本文选取上交所、深交所的信息披露评级作为衡量企业实际信息披露程度的指标。上交所和深交所每年都会对本所上市公司进行信息披露等级的打分，分为 1~4 级，1 级最好（优秀），4 级最差（不合格），将两个交易所的信息披露等级评分同前文样本数据相匹配。本文将 1 级评级的分为高信息披露组，非 1 级的为低信息披露组。

实证结果如表 10 所示，对于低信息披露组来说，异质性信息含量的系数达到了-1.555，跨期一致性指数的系数达到了-0.218，系数绝对值高于表 3 和表 4 的基础回归结论，且都在 5%及以下的水平上显著。表明在低披露情形下，企业社会责任报告中的异质性信号更易被市场识别。而对于高信息披露组来说，他们自身相对完整的信息披露使得他们社会责任报告中呈现出的跨期一致性指数受到了影响，这体现系数较小（仅为-0.001）和统计显著性不高的特征上。

表 10　基于实际信息披露情况的异质性检验

变量	（1）	（2）	（3）	（4）
	*SYN*3			
	高信息披露	低信息披露	高信息披露	低信息披露
CSRSP	-1.292** （0.588）	-1.555** （0.619）		
CSRSim			-0.001 （0.105）	-0.218*** （0.072）
ROA	-0.015 （0.24）	-0.004 （0.01）	0.029 （0.25）	-0.005 （0.01）
Grow	-0.090** （0.04）	-0.016** （0.01）	-0.069 （0.05）	-0.014** （0.01）
TQ	-0.031* （0.02）	-0.057*** （0.02）	-0.052** （0.02）	-0.049*** （0.02）
ContrashProportion	-0.005*** （0.00）	-0.001 （0.00）	-0.005*** （0.00）	-0.002 （0.00）
ConcurrentPosition	-0.066 （0.05）	-0.020 （0.03）	-0.046 （0.06）	-0.034 （0.04）
ln*Board*	-0.044 （0.10）	0.252*** （0.08）	0.030 （0.11）	0.258*** （0.09）
IndDirectorRatio	0.007* （0.00）	0.008*** （0.00）	0.007* （0.00）	0.008** （0.00）
Constant	0.035 （0.29）	-0.944*** （0.26）	-0.129 （0.33）	-0.891*** （0.29）
观测值	2098	4077	1,869	3379
R^2	0.414	0.393	0.394	0.371
控制变量	是	是	是	是
行业固定效应	是	是	是	是
年份固定效应	是	是	是	是

注：***、**、*分别表示估计系数在 1%、5%、10%的水平上显著；括号内为聚类到企业层面的标准误。

（四）基于报告信号的超额水平做出的分析

本文借鉴李四海和李震[31]的做法，利用式（13）的模型分别分离出企业社会责任报告的异质性信息含量和跨期一致性指数的超额水平及残差值，来同样作为回归模型的解释变量。一般而言，企业通过社会责任报告传达出来的信号往往会受到企业多方面的影响，包括企业基本面指标、时间和企业个体因素等，如果本文通过模型进行回归分解，因此得出的残差，即可能代表了企业管理对于企业经营的未来预期信号。正如前文所证明的信号理论，超额的未来预期信号与基准回归中的两类信号对股价同步性具有一致的方向性影响。

$$CSRSim_{it}/CSRSP_{it} = \beta_0 + \beta_1 Controls_{it} + CSRSim_e_{it}/CSRSP_e_{it} + \lambda_t + \mu_I \quad (13)$$

其中:本文选取了包括总资产收益率(*ROA*)、营业收入增长率(*Grow*)、托宾 *Q*(*TQ*)、控股股东持股比率(*ContrashProportion*)、董事长和总经理是否为同一人(*ConcurrentPosition*)、董事会人数对数值(ln*Board*)、独立董事占比对数值(*IndDirectorRatio*)等变量分别对企业的财务方面和管理层方面进行控制。此外,考虑到企业信息披露可能还受到披露指引的影响,此外也衡量了企业是否加入全球报告倡议(Global Reporting Initiative,GRI),以 GRI 二元变量表示。最后,还加入了年份和行业固定效应来消除不同行业和年份之间潜在的但是难以直接观测潜在的道德因素。利用上述公式计算得出异质性信息含量的超额水平(*CSRSP_e*)和跨期一致性指数的超额水平(*CSRSim_e*)之后,代入式(11)和式(12)中,得出实证结果见表11。

表11的结果显示,对于企业异质性信息含量和跨期一致性指数的超额水平都呈现比较显著的负相关关系,*CSRSP_e* 的系数为-1.261,在1%水平上显著;*CSRSim_e* 的系数为-0.109,在10%的水平上显著。这说明企业管理层对于企业的未来预期信号也能向资本市场传达出有效信号,继而有助于降低股价同步性。

表11　基于信号超额水平的分析

变量	(1)	(2)
	SYN3	
CSRSP_e	-1.261*** (0.407)	
CSRSim_e		-0.109* (0.058)
ROA	0.008* (0.00)	0.009* (0.00)
Grow	-0.005* (0.00)	-0.017*** (0.01)
TQ	-0.062*** (0.01)	-0.061*** (0.01)
ContrashProportion	-0.001* (0.00)	-0.001 (0.00)
ConcurrentPosition	-0.054** (0.03)	-0.055* (0.03)
ln*Board*	0.203*** (0.06)	0.252*** (0.06)
IndDirectorRatio	0.007*** (0.00)	0.008*** (0.00)
Constant	-0.727*** (0.18)	-0.864*** (0.20)
观测值	8392	6946
R^2	0.370	0.352
控制变量	是	是
行业固定效应	是	是
年份固定效应	是	是

注:***、**、*分别表示估计系数在1%、5%、10%的水平上显著;括号内为聚类到企业层面的标准误。

六、结论与启示

本文以沪深两市上市公司社会责任报告为研究对象,最终搜集并整理8936份报告进行分析,考察社会责任报告所蕴含的文本信号与股价同步性之间的关系。在方法上,本文基于机器学习平台的命名实体识别与TF-BIDF技术,从"异质性信息含量"和"跨期一致性指数"两个维度识别并量化报告中的关键信号。实证结果表明:第一,从报告内容视角看,社会责任报告中的异质性信息越多(信号异质性越强),企业股价同步性显著降低。第二,从纵向比较视角看,同一企业跨年度报告之间的相似度越高(跨期一致性越强),同样显著降低股价同步性。第三,将命名实体技术扩展至年报文本得到一致结论,即年报的异质性信息对股价同步性亦存在显著负向影响,在一定程度上削弱了潜在的替代性解释,从而进一步支撑本文主要结论。第四,异质性分析表明,相较于国有企业,非国有企业的报告异质性对股价同步性的抑制作用更强;社会责任履行水平较高的企业,其报告更受市场关注,表现出更强的信号特性与更显著的负相关性;在信息披露不足的企业中,报告信号的"放大效应"更为明显,回归系数较基准结果更大。

在此基础上进一步拓展,学界多关注企业社会责任报告、ESG评级等相关信息披露对于市场流动性的影响,却较少探讨过度披露可能引发的信息过载效应,最终可能导致信号效应减弱或是消失。因此,对于潜在信息过载的研究及不同信号之间的削弱或强化作用或许是未来的一个重要研究方向。

本文对于监管机构有着重要的启示作用。监管机构对社会责任报告与财务报告的监管强度与技术规范长期不对称,是造成社会责任报告质量参差与"漂绿"风险的制度根源;然而,本文的证据表明,高质量CSR文本能够通过提升信息异质性与跨期一致性来降低股价同步性、缓解信息不对称。鉴于此,监管层宜在制度与技术两个层面同步推进:一是尝试建立行业可比披露框架,明确量化指标、核算口径与基期并要求对历年变动作出充分的口径变更说明和跨期可追溯披露;二是推动机器可读格式与版本管理,便于第三方

复核与自动化一致性检验，同时针对高度模板化与复刻文本引入文本自动审阅与抽查，从源头减少“形式化披露”现象；三是完善配套激励与约束机制，将披露质量纳入再融资、债券注册、指数纳入与费用减免等市场化安排，对严重失真与漂绿行为实施更正与惩戒；四是开展先行试点，在高环境敏感行业与非国有企业占比较高的领域先行试点差异化监管与第三方鉴证，以形成可复制推广的制度范式。

本文对于市场第三方机构有着重要的启示作用。评级机构应以方法透明与证据链可核验为原则，将基于命名实体识别与 TF-BIDF 获得的“文本异质性指数”“跨期一致性指数”等特征纳入模型，并与年报、行政执法记录、舆情与供应链数据开展多源交叉验证，降低主观打分带来的误差；同时建立年度方法学审计与回溯检验机制，公开关键参数与稳定性检验结果，以保证评级对投资决策的可用性。对投资机构而言，本文结果提示应将 CSR 文本中的异质性信息纳入研究与决策框架、与基本面结合，并围绕披露窗口实施事件研究与持仓调整；在执行层面，需要设置信息预警机制，避免因冗长文本与重复叙述造成噪声放大；在治理层面，应通过积极所有权与投票政策将披露质量改进纳入参与议题，以技术优势和人才优势反向推动企业提升可验证的社会责任实践。

本文对于上市企业有着重要的启示作用。企业社会责任报告不应只满足于符合规定和套用模板，而应真实反映企业在社会责任方面的实际行动与资源投入。企业应在内部治理中明确责任分工，建立跨部门协同的数据管理机制，通过第三方审核提升可信度。报告撰写应层次清晰，正文突出年度进展与目标完成情况，详细数据应以可读取形式提供，并保持历年数据可比性。不同类型企业可突出自身重点，例如民营企业可强调创新与治理成效，国企可侧重政策对接与公共绩效，高环境敏感行业可增加实时披露。通过这些方式，企业能够增强信息透明度，改善资本市场评价并缓解融资压力。

本文对于投资者有着重要的启示作用。应将 CSR 文本信号纳入资产配置与选股的系统证据，而非作为形象化的附属信息。披露内容异质性高、跨期一致性强企业，往往更可能带来超额收益。投资者可借助文本分析工具筛选投资对象，并结合事件分析评估市场反应。同时，应警惕文本冗长、重复或频繁调整口径的报告，避免信息过载与误判风险。长期来看，披露的一致性可反映企业战略和治理的稳定性，投资者可借此推动企业提升社会责任表现，从而实现财务回报与社会价值的共同提升。

参考文献

[1] 朱红军，何贤杰，陶林. 中国的证券分析师能够提高资本市场的效率吗——基于股价同步性和股价信息含量的经验证据[J]. 金融研究，2007(2)：110-121.

[2] CHAN K，HAMEED A. Stock price synchronicity and analyst coverage in emerging markets[J]. Journal of Financial Economics，2006，80：115-147.

[3] 王茂斌，叶涛，孔东民. 绿色制造与企业环境信息披露——基于中国绿色工厂创建的政策实验[J]. 经济研究，2024，59(2)：116-134.

[4] 李新丽，万寿义，张雪梅. 企业社会责任能降低股价同步性吗？——基于强制披露企业社会责任报告的准自然实验[J]. 科学决策，2018(9)：1-27.

[5] XU D，HUANG J，REN X，et al. ESG report textual similarity and stock price synchronicity：Evidence from China[J]. Pacific-Basin Finance Journal，2024，85：102343.

[6] 翁舟杰，赖政. 公司 ESG 表现与防范股票市场风险——基于股价同步性的路径分析[J]. 统计研究，2024，41(8)：56-68.

[7] 王明涛，李茜. 分析师跟踪网络与公司特质信息研究——基于股价同步性视角的实证分析[J]. 金融经济学研究，2023，38(3)：98-113.

[8] 吕英，方文艳. 党员 CEO 促进了 ESG 责任履行吗？——内部治理和外部治理的协同作用[J]. 技术经济，2024，43(6)：125-140.

[9] CHRISTENSEN D M，SERAFEIM G，SIKOCHI A. Why is corporate virtue in the eye of the beholder? The case of ESG ratings[J]. The Accounting Review，2022，97(1)：147-175.

[10] 陈宏韬，殷海锋，张天舒，等. ESG 评级分歧影响资本市场定价效率吗？——基于上市公司股价同步性的研究[J]. 财经研究，2025，51(5)：65-80.

[11] HU J，ZOU Q，YIN Q. Research on the effect of ESG performance on stock price synchronicity：Empirical evidence from China's capital markets[J]. Finance Research Letters，2023，55：103847.

[12] MORCK R，YEUNG B，YU W. The information content of stock markets：Why do emerging markets have synchronous stock price movements?[J]. Journal of Financial Economics，2000，58(1/2)：215-260.

[13] 巫岑，饶品贵，岳衡. 注册制的溢出效应：基于股价同步性的研究[J]. 管理世界，2022，38(12)：177-202.

[14] 孙凡，孙泽宇. 公共数据开放何以改善资本市场定价效率[J]. 经济管理，2025，47(1)：166-189.

[15] 王雪, 丁亚楠. 中小投资者服务中心行权与股价同步性[J]. 金融监管研究, 2024(1): 96-113.
[16] 郑祖玄, 张森, 张军, 等. 公司数字化可以影响股价同步性吗? ——基于 ChatGPT 对分析师深度报告的解读[J]. 世界经济文汇, 2025(4): 87-103.
[17] 于连超, 董晋亭, 毕茜. 环境法治强化与股价崩盘风险——基于环境法庭设立的准自然实验[J]. 当代财经, 2025(1): 141-153.
[18] 付强, 张呈, 廖益兴. 直面挑战: 审计师数字化专长是否有助于提高审计质量?[J]. 审计与经济研究, 2024, 39(2): 42-51.
[19] 刘智宇, 郭范勇, 张云. 投资者实地调研能降低股价同步性吗[J]. 财经科学, 2022(12): 34-49.
[20] 王建琼, 廖晓芮, 曹世蛟. 高管连锁网络与股价同步性[J]. 技术经济, 2022, 41(9): 157-169.
[21] 化兵, 孙武军. 多个大股东股权结构与股价同步性[J]. 北京工商大学学报(社会科学版), 2025, 40(1): 81-91.
[22] 赵彦锋. 国有股东参股如何降低民营企业股价同步性[J]. 财经问题研究, 2025(9): 46-61.
[23] GROSSMAN S J, STIGLITZ J E. On the Impossibility of informationally efficient markets[J]. The American Economic Review, 1980, 70(3): 393-408.
[24] WU C. Mandatory CSR disclosure regulation and stock price synchronicity: Evidence from a quasi-natural experiment[J]. Humanities and Social Sciences Communications, 2025, 12(1): 1-14.
[25] 李子健, 李春涛, 冯旭南. 非财务信息披露与资本市场定价效率[J]. 财贸经济, 2022, 43(9): 38-52.
[26] 毕晓方, 邢晓辉, 刘晟勇. 稳定型机构投资者、盈余平滑与股价同步性[J]. 北京工商大学学报(社会科学版), 2023, 38(3): 93-106.
[27] 李玉敏, 吴晓晖, 秦利宾, 等. 共享的力量: 共同机构投资者与股价同步性[J]. 管理科学学报, 2024, 27(11): 78-101.
[28] 杨科, 胡楠, 田凤平. 有限关注、投资者情绪与绿色溢酬——基于融合深度学习技术的动态集成网络定价模型[J]. 统计研究, 2024, 41(11): 129-141.
[29] SPENCE M. Job market signaling[J]. The Quarterly Journal of Economics, 1973, 87(3): 355.
[30] BARNEY J. Firm resources and sustained competitive advantage[J]. Journal of Management, 1991, 17(1): 99-120.
[31] 李四海, 李震. 企业社会责任报告特质信息含量的信号效应研究——基于自然语言处理技术的分析[J]. 中国工业经济, 2023(1): 114-131.
[32] 何玉润, 刘曈, 王鹏程, 等. 监管压力下分析师策略性信息披露行为研究[J]. 会计研究, 2025(4): 31-46.
[33] GEORGE A. AKERLOF. The market for "Lemons": Quality uncertainty and the market mechanism[J]. The Quarterly Journal of Economics, 1970, 84: 488-500.
[34] DIAMOND D W. Reputation acquisition in debt markets[J]. Journal of Political Economy, 1989, 97(4): 828-862.
[35] 张光利, 薛慧丽, 高皓. 企业 IPO 价值审核与股票市场表现[J]. 经济研究, 2021, 56(10): 155-171.
[36] 张慧. 上市公司 ESG 责任履行有助于提升股票收益吗?[J]. 技术经济, 2024, 43(8): 85-100.
[37] 郑海埃, 倪古强, 戚赞栎. 审计师重复披露关键审计事项: 基于事实还是迫于压力[J/OL]. 南开管理评论, 1-26[2025-09-27]. https://link.cnki.net/urlid/12.1288.f.20250716.1901.006.
[38] LANG M, LUNDHOLM R. The relation between security returns, firm earnings, and industry earnings[J]. Contemporary Accounting Research, 1996, 13(2): 607-629.
[39] 张忠良, 费秦君, 陈愉予, 等. 基于 BERT 模型和动态集成选择的多分类文本情感识别研究[J]. 中国管理科学, 2024, 32(6): 140-150.
[40] HUANG A H, WANG H, YANG Y. FinBERT: A large language model for extracting information from financial text[J]. Contemp Account Res, 2023, 40: 806-841.
[41] KELLY B, PAPANIKOLAOU D, SERU A, et al. Measuring technological innovation over the long run[J]. American Economic Review: Insights, 2021, 3(3): 303-320.
[42] 吕璐成, 赵亚娟. 基于客观数据的专利价值评估指标进展与评述[J]. 图书情报工作, 2023, 67(18): 129-139.
[43] LI Y, LI Z, LIU T. Does university-industry collaboration improve firm productivity? Evidence from China[J]. PLOS ONE, 2024, 19(7): e0305695.
[44] 王亚平, 刘慧龙, 吴联生. 信息透明度、机构投资者与股价同步性[J]. 金融研究, 2009(12): 162-174.
[45] 伊志宏, 杨圣之, 陈钦源. 分析师能降低股价同步性吗——基于研究报告文本分析的实证研究[J]. 中国工业经济, 2019(1): 156-173.
[46] 付金存, 曹觊乐偲, 李皓月, 等. 企业"漂绿"研究的发展脉络、研究述评与未来展望[J]. 技术经济, 2024, 43(4): 100-117.
[47] 师展, 樊重俊. 企业社会责任对会计文本语调操纵的治理效应[J]. 企业经济, 2023, 42(12): 57-67.
[48] HOLLWECK T, ROBERT K, YI N. Case study research design and methods[J]. Canadian Journal of Program Evaluation, 2015, 30(1): 108-110.

Corporate Social Responsibility Disclosure and Stock Price Synchronicity: An Analysis Based on NER and TF-BIDF Textual Techniques

Yang Suchang, Cao Yining

(School of Economics, Lanzhou University, Lanzhou 730000, China)

Abstract: Based on corporate social responsibility (CSR) reports publicly released by listed companies from 2008 to 2022, the textual content was analyzed using Named Entity Recognition (NER) and Term Frequency-Back Inverse Document Frequency (TF-BIDF) techniques, and the mechanism explored through which the signals transmitted by CSR disclosure affect stock price synchronicity. The analysis showed that two types of signals play important roles. Heterogeneous information content transmits firm-specific information, and the intertemporal consistency index reflects consistent communication. Both signals reduce information asymmetry between firms and the market, thereby decreasing stock price synchronicity. Further analysis indicated that signaling heterogeneity is more significant in non-state-owned enterprises. Higher levels of social responsibility fulfillment strengthen the effects of heterogeneous content and consistency index, while higher disclosure levels may weaken the signaling effect. Robustness tests were conducted by replacing samples and applying the PSM-DID and Heckman models. From the perspective of signal characteristics, internal content analysis and external longitudinal comparison were employed. The findings provide a new perspective on the impact of CSR reports on market performance. The results also offer theoretical support for improving information efficiency in China's capital market and policy suggestions for enhancing the CSR disclosure system.

Keywords: CSR; stock price synchronicity; machine learning; NLP; information disclosure