

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

□王克敏 王华杰 李栋栋 戴杏云

摘要: 本文研究管理者出于自利动机的年报文本信息复杂性披露策略。研究发现,相比于业绩较好公司,业绩较差公司年报文本信息的复杂性更高,且上述关系在盈余管理空间小、管理层持股比例大、两职合一和法律风险低的公司中更为显著;进一步地,年报文本信息复杂性越高,管理者获得的超额薪酬越高;相比业绩较好公司,业绩较差公司年报文本信息复杂性的短期、长期市场反应更积极。研究表明,出于自利动机,管理者会操纵年报文本信息复杂性;文本信息复杂性操纵对数字信息操纵有替代作用,且管理层持股及公司内、外部治理均会影响文本信息复杂性操纵;通过操纵年报文本信息复杂性,管理者能够获取更高超额薪酬,并提高公司市场估值。本文首次基于中文年报文本信息大样本分析,为已有文本信息复杂性披露策略研究提供了实证证据,验证了数字信息和文本信息操纵手段间的替代效应,拓展了机会主义披露研究。本文的研究结论对于完善公司年报文本信息披露监管法规具有重要启示。

关键词: 文本信息复杂性 披露策略 管理者自利 年报

一、引言

现代企业制度所有权与经营权分离,与外部股东相比,公司管理者存在信息优势,拥有更多关于公司财务状况、经营业绩、发展前景的信息(Berle et al., 1935)。作为足智多谋的“经济人”(Jensen and Meckling, 1994),上市公司管理者在信息披露过程中存在机会主义行为。已有研究发现,管理者会通过操纵数字信息获利(DeAngelo, 1988; Defond and Park, 1997)。如在盈余披露时,管理者会机会主义地选择盈余基准以夸大盈余增长程度,导致投资者看不清其业绩衡量指标(Schrand and Walther, 2000)。除数字信息外,文本信息也能成为管理者自利的工具(Schipper, 1989)。基于美国上市公司年报(10-K)的研究发现,管理者会通过操纵文本信息可读性,隐藏公司的坏消息(Courtis, 1998; Li, 2008)。与英语文化相比,汉语文化更强调圆式思维,在语言使用上更加含蓄、委婉(顾曰国, 1992; 孟庆涛, 2009),其同样存在中文年报的解读成本问题,抑或为管理者基于自利而操纵文本信息披露提供了更大可能。然而,截至目前,中文年报文本分析尚属探索性研究领域,中文年报文本信息披露的大样本研究还十分鲜见。因此,本文借助文本分析技术,结合汉语语言学特性,构建中文年报文本信息复杂性指标,并基于信息不对称理论和代理理论,研究管理者出于自利动机的中国上市公司年报文本信息复杂性披露策略。

本文认为,已有关于年报披露策略的研究主要将数字信息操纵和文本信息操纵作为两类独立的信息披露操纵手段。正如Huang等(2014)尝试性地考察了盈余管理空间和业绩快

^{*} 本研究得到国家自然科学基金项目(71272072、71572042)的资助,特此致谢。文责自负。

报文本语气操纵之间的相关性,发现当公司盈余管理空间较小时,管理者更可能操纵文本语气。然而,截至目前,尚无关于盈余管理空间与年报文本信息复杂性操纵关系的研究。本文强调,同为管理者操纵市场预期的手段,数字信息操纵和文本信息操纵可能并非独立存在,盈余操纵受限很可能促使管理者寻求其他手段管理市场预期;不仅文本语气操纵,文本复杂性操纵也可作为盈余管理的替代,即当盈余可操纵空间有限时,管理者会增加年报文本信息复杂性操纵。

已有英文财报文本信息复杂性变量主要为文本长度和Fog指数(Li, 2008; Loughran and McDonald, 2014),其中,Fog指数基于平均句子长度和句子中复杂单词比例构建,其主要从单词角度出发,忽略了文本逻辑复杂性。除对字词的理解外,文本间的逻辑关系是导致读者理解难度上升的重要因素(Pretorius, 2006)。基于此,在借鉴英文文本信息披露研究注重字词复杂性的基础上,结合汉语更委婉含蓄的语言逻辑特点,本文从文本逻辑和字词的复杂性两个角度刻画中文年报文本信息复杂性,既有助于深入挖掘中文文本信息复杂性,也为其他语言文本信息披露特征的度量提供了重要思路。

本文将从以下4个方面展开。第一,考虑到年报文本信息复杂性操纵策略,主要是管理者基于薪酬等私利而隐藏公司的负面信息,本文检验公司当期业绩与年报文本信息复杂性的关系。第二,考虑到数字信息操纵和文本信息操纵均为管理者操纵信息披露的手段,两者可能存在替代关系,本文检验盈余管理空间对年报文本信息复杂性操纵的影响;与此同时,考虑到管理者所持股份价值激励及公司内、外部治理会影响年报文本信息复杂性操纵的收益和成本,本文分别检验管理层持股比例、两职合一及法律风险对当期业绩与年报文本信息复杂性关系的影响。第三,考虑到薪酬是管理者操纵年报文本信息复杂性,掩盖公司负面信息的直接动因,且管理者操纵复杂性更可能是为了获取应得薪酬以外的超额薪酬,本文检验年报文本信息复杂性与管理者超额薪酬的关系。第四,考虑到管理者操纵年报文本信息复杂性,旨在增加普通投资者的信息解读成本,影响市场估值;具体而言,业绩较好公司年报文本信息复杂性的客观性更强,业绩较差公司年报文本信息复杂性的机会主义动机更强,投资者更加难以识别业绩较差公司的文本信息复杂性操纵;鉴于此,本文在区分业绩好、坏的基础上,检验年报文本信息复杂性与短期、长期市场反应的关系。

本文研究发现:第一,公司当期业绩表现越差,年报文本信息复杂性越高;第二,上述关系在盈余管理空间小、管理层持股比例大、两职合一和法律风险低的公司中更显著;第三,年报文本信息复杂性越高,管理者获得的超额薪酬越高;第四,与业绩较好公司相比,业绩较差公司年报文本信息复杂性的短期、长期市场反应更积极。研究表明:出于自利动机,管理者会操纵年报文本信息复杂性;文本信息复杂性操纵对数字信息操纵有替代作用,且管理层持股及公司内、外部治理均会影响文本信息复杂性操纵;管理者通过年报文本信息复杂性操纵能够获取更高的超额薪酬,并提高公司的市场估值。

本文的研究贡献为:第一,本文首次通过中文年报文本信息大样本分析,为已有文本信息复杂性披露策略研究提供了基于中文年报的实证证据。本文基于汉语语言学研究成果,从文本逻辑和字词复杂性两个层面构建中文文本信息复杂性变量,填补了英文文本研究只关注字词复杂性的空白;与此同时,本文引入中文文本分析与分词技术,文本信息提取与变量生成全部由计算机程序自动完成,与传统手工搜集和阅读打分方法相比(孙蔓莉等, 2013),研究样本大幅增加,研究结果更加客观。第二,本文发展了机会主义披露策略的研究。已有关于年报披露策略的研究将数字信息操纵和文本信息操纵作为两类独立的信息披露操纵手段(Schrand and Walther, 2000; Li, 2008)。Huang等(2014)以公司发布的业绩快报为研究对象,发现文本语气操纵更易发生在盈余管理空间小的公司。本文则深入研究了盈余管理空间对年报文本信息复杂性披露策略的影响,为数字信息操纵和文本信息操纵的替代机制提供了证据。第三,本文的研究发现具有一定实践价值。目前,中国财务报告披露相关法规重点关注信息披露的全面性和透明度,对文本信息复杂性关注较少,仅在《公开发行证券的公司信息披露内容与格式准则第2号——年度报告的内容与格式(2014年修订)》中提到“鼓励公司以简明易懂的方式对投资者特别是中小投资者披露信息”。本文的研究发现提示监管者应加强对财务报告文本信息复杂性的关注,指导上市公司发布更加简明易懂的财务报告,同时,本文也可帮助市场参与者了解公司文本信息披露策略,提示投资者

当公司年报文本信息复杂性较高时,应加强对公司的全面分析。

本文剩余内容安排如下:第二部分为已有相关研究与本文逻辑论证,第三部分为实证分析,第四部分为研究结论与政策启示。

二、已有相关研究与本文逻辑论证

公司信息披露策略一直是资本市场研究的焦点之一。根据不完全反应假说(Incomplete Revelation Hypothesis),资产价格反映信息的程度取决于信息的解读成本(Grossman and Stiglitz, 1980; Bloomfield, 2002)。因此,出于薪酬、帝国构建等自利动机(Kothari et al., 2009),管理者会通过操纵信息披露提高信息解读成本,引导市场做出对其自身能力和公司价值有利的判断(Bloomfield, 2002)。已有研究发现,管理者操纵数字信息披露的机会主义行为普遍存在(Healy and Wahlen, 1999; Schrand and Walther, 2000; Healy and Palepu, 2001)。Defond和Park(1998)发现,出于职位安全考虑,管理者会进行更多的盈余管理。与此同时,由于数字信息反映的主要是公司过去的经营业绩,无法满足利益相关者分析公司未来发展情况的需求,文本信息越来越受到重视(Li, 2010)。近年来,随着计算机算法以及信息处理技术的发展,文本分析逐渐成为财务会计领域的研究热点。基于英语环境的文本分析研究,通过检验年报可读性与公司业绩的关系,考察管理者的机会主义披露行为,但所得结论并不一致。以Li(2008)为代表的相关文献发现,公司业绩越差,年报可读性也越差;但Merkley(2014)通过对年报中研发信息的分析发现,文本信息可读性与公司业绩不存在显著相关性。值得注意的是,相关研究中采用的可读性度量方法和样本区间等均会对结果产生很大影响。特别地,中文文本分析技术发展相对缓慢,目前尚未形成公认、统一的文本信息可读性度量方法,致使中文年报的文本信息披露策略研究严重滞后。已有关于中文年报文本的研究不仅数量较少,且均是通过手工收集数据与人工阅读的方法分析文本信息披露特征(孙蔓莉等, 2013),存在研究样本量较小、变量取值较为主观等问题。基于此,本文深入分析中文文本信息披露特征,构建中文年报文本信息复杂性指标,并基于信息不对称理论和代理理论,研究管理者出于自利动机的年报文本信息复杂性披露策略。

第一,汉语语言特征为中国管理者操纵年报文本信息复杂性提供了更大可能。已有关于英文年报文本信息披露质量的研究发现,上市公司年报文本信息可读性差,超出普通人的理解范围(Pashalian and Crissy, 1950, 1952),给投资者正确解读公司信息带来很大挑战。相比英语文化的线性思维,汉语文化更强调圆式思维,反映在语言上是更加委婉含蓄(顾曰国, 1992; 孟庆涛, 2009),这导致中文语义更为曲折隐晦,增加了上市公司年报文本信息的解读难度。例如,星湖科技2015年面临大幅亏损,在其当年年报中,使用了较多转折句,如:“虽然随着国家整体经济进入发展的新常态,生物发酵产业也进入了发展的调整期,但生物发酵又归属于玉米深加工产业范畴”、“公司进一步深化改革,优化管理流程,……,但受限于盈利产品的规模、市场容量,以及目前的用途等,公司整体经营未能全部抵御市场的风险”,这些转折词语大大增加了解读信息的难度。相比而言,同行业的海天味业2015年业绩稳步增长,在其当年年报中,使用了较多简洁易懂的句子,如:“产能释放与市场发展的良性互动,使投入能最大化产出,也保证了酱油、蚝油、酱等核心品类能大胆发展”、“报告期内,公司加快产品结构调整,高端酱油保持了较大发展,各产品产销量继续保持同步增长,各产品产销率保持在较高水平,期末产成品库存量良性下降”。可见,管理者在分析公司业绩表现时在语言表达方面确有选择,中文语义的委婉曲折导致普通投资者很难看清公司年报文本信息的真实面目,中国上市公司年报文本信息披露的管理者机会主义土壤更肥沃。基于此,针对中文年报,本文认为,当公司当期业绩较差时,为了隐藏坏消息,管理者会增加年报文本信息复杂性。

第二,数字信息操纵空间及管理者激励和公司内、外部治理机制是影响年报文本信息复杂性披露策略的重要因素。一方面,作为操纵市场预期不同手段,管理者可能会基于数字信息操纵选择文本信息披露策略作为替代。基于美国公司业绩快报,Huang等(2014)考察了盈余管理空间对文本语气披露策略的影响,发现当公司盈余管理空间较小时,管理者更可能操纵文本语气。本文认为,文本信息复杂性操纵也可作为数字信

息操纵的替代,即当公司盈余操纵空间有限时,管理者会增加年报文本信息复杂性操纵。另一方面,管理者财富激励及公司内、外部治理机制会影响年报文本信息复杂性操纵的收益和成本。管理者持股比例越大,公司市值对其自身财富的影响也越大,管理者越有动机操纵年报文本信息复杂性以提高市场估值;董事会是解决代理问题的公司内部治理机制,当公司总经理兼任董事长即两职合一时,董事会对高管的监管力度被削弱,代理问题会更严重(Jensen and Meckling, 1976),管理者更有可能操纵年报文本信息复杂性以满足自利动机;与此同时,公司面临的法律风险越高,管理者越倾向于提前披露坏消息(Skinner, 1994),并发布可读性更高的年报(Nelson and Pritchard, 2007),即法律风险会约束管理者对年报文本信息复杂性的操纵行为。由此,本文认为,当公司盈余管理空间较小、管理层持股比例较大、两职合一及法律风险较低时,管理者对年报文本信息复杂性的操纵更强。

第三,管理者出于自利动机操纵年报文本信息复杂性,旨在隐藏不利于个人能力评价和公司价值评估的坏消息。那么,此操纵行为是否能够实现管理者的自利动机便是一个值得检验的问题。首先,公司治理理论强调对管理者的激励约束,通常将管理者薪酬与其经营业绩挂钩(Healy and Palepu, 2001)。Murphy(1985)利用公司股东回报率、销售增长率和股票市场表现衡量公司业绩,发现管理者薪酬与其显著正相关;Healy(1985)则发现,管理者会通过盈余管理虚增盈余,从而获取更高薪酬。其次,本文认为,由于文本信息复杂性操纵无法改变盈余数字信息,管理者操纵复杂性更可能是为了获取应得薪酬以外的超额薪酬。基于此,本文强调,年报文本信息越复杂,投资者解读年报信息的成本越高,越无法识别不利于评价管理者能力的坏消息,管理者将获得更高的超额薪酬。

第四,管理者操纵年报文本信息复杂性,旨在提高市场普通投资者的信息解读成本,使其无法看清公司真实业绩表现,从而影响市场估值。首先,已有研究发现,投资者无法识别管理者对数字信息的操纵行为(Sloan, 1996)。Xie(2001)发现,市场会高估可操纵应计的持续性,最终给予公司更高估值。本文认为,相较于数字信息操纵,年报文本信息复杂性操纵的隐蔽性更强,更不易被察觉,投资者更容易受到迷惑而无法看清公司当期业绩不佳的事实。其次,本文认为,业绩较好和较差公司年报文本信息复杂性的特征不同;具体地,业绩较好公司年报文本信息复杂性的客观性更强,业绩较差公司年报文本信息复杂性的机会主义动机更强,投资者更加无法识别业绩较差公司的文本信息复杂性操纵。基于此,本文强调,相比业绩较好公司,业绩较差公司年报文本信息复杂性的短期、长期市场反应更积极。

三、实证分析

(一)样本选择与数据来源

本文以2001~2015年A股非金融行业上市公司年报为研究样本。年报来源于上海证券交易所、深圳证券交易所网站及巨潮资讯网。本文运用Perl文本分析及汉语分词技术对年报文本信息进行提取、分析,处理步骤详见附录。公司财务数据、市场交易数据、违规惩处数据及管理者薪酬数据均来自国泰安数据库,两职合一数据来自于色诺芬(CCER)上市公司治理结构数据库。

参考Li(2008),样本筛选过程如下:(1)删除因格式问题无法提取文本信息的样本;(2)删除所有者权益为负或零的样本;(3)删除ROE数据缺失或绝对值大于1的样本;(4)考虑到复杂性变量需用文本长度进行标准化,若文本长度过短,则复杂性变量易出现极端值,即使缩尾处理也无法解决,因此,删除年报全文长度小于1%分位数的样本;(5)删除总资产、市值账面比等控制变量缺失的样本。本文最终样本包括2812家上市公司的23925份年报,样本筛选过程详见表1。

(二)变量定义与描述性统计

1. 变量定义

(1)年报文本信息复杂性。

考虑到复杂性会反映在文本信息的多个层面,本文从文本逻辑和字词复杂性两个角度设计3个变量考察

中文年报文本信息复杂性,分别为:逆接成分密度($Adverse_i$)、会计术语密度($ATDensity_i$)、次常用字密度($UCDensity_i$)。

逆接成分密度($Adverse_i$)。考虑到文本的逻辑关系越复杂,越难被读者理解(Pretorius, 2006),本文利用逻辑关系复杂性衡量年报文本信息复杂性。根据廖秋忠(1986),文本中逆接成分(“然而”、“但是”、“即使”等)越多,句子、段落间描述的事件越不协调或在本质上差异更大,这与人们的思维习惯相悖,会给阅读理解带来很大困难。基于此,本文设计年报文本信息复杂性第一个变量 $Adverse_i$,表示逆接成分密度,具体为每百字中包含逆接关系连接成分的个数。 $Adverse_i$ 值越大,表示年报文本信息复杂性越高。

会计术语密度($ATDensity_i$)。已有关于可读性的研究发现,年报难度接近高度专业化的科技或学术论文,主要原因是其中含有大量财务会计专业术语。考虑到专业术语在年报中出现的频率在一定程度上能够反映年报的理解难度,本文基于灵格斯词霸(2008),设计年报文本信息复杂性第二个变量 $ATDensity_i$,表示会计术语密度,具体为每百字中包含会计术语个数(“减值”、“资本化”、“损益”等)。 $ATDensity_i$ 值越大,表示年报文本信息复杂性越高。

次常用字密度($UCDensity_i$)。考虑到文本中的生僻字会降低读者阅读的顺畅性,提高阅读难度,本文基于《现代汉语次常用字表》(1988),设计年报文本信息复杂性第三个变量 $UCDensity_i$,表示次常用字密度,具体为每百字中包含的次常用字个数(“擎”、“棘”、“淤”等)。 $UCDensity_i$ 值越大,表示年报文本信息复杂性越高。

(2)公司业绩。

公司业绩(ROE_i)。参考Li(2008),本文用当年净资产收益率 ROE_i 衡量公司业绩。 ROE_i 越大,公司当期业绩表现越好。

(3)年报文本信息复杂性披露策略影响因素。

盈余管理空间(R_NOA_{i-1})。参考Barton和Simko(2002)、Das等(2011),本文用t-1期净经营资产(Net Operating Assets)占t-2期总资产的比例衡量当期盈余管理空间。具体地,净经营资产=经营资产-经营负债;经营资产=总资产-货币资金-短期金融投资^①,经营负债=总负债-短期借款-长期借款。净经营资产比例衡量盈余管理的限制程度,即净经营资产比例越大,公司盈余管理空间越小。如果公司净经营资产比例大于当年研究样本的中位数则 $R_NOA_{i-1}=1$,否则为0。 $R_NOA_{i-1}=1$ 表示公司当期盈余管理空间较小, $R_NOA_{i-1}=0$ 表示公司当期盈余管理空间较大。由于计算净经营资产比例需t-1期及t-2期数据,故盈余管理空间(R_NOA_{i-1})变量观测值为20776。

管理层持股($R_MANAHL D_i$)。参考Ang等(2000),本文用当年年末管理层持股数量占公司总股本的比例衡量管理层持股比例,若管理层持股数据缺失,则认为持股比例为零。如果公司管理层持股比例大于当年研究样本的中位数,则 $R_MANAHL D_i=1$,否则为0。 $R_MANAHL D_i=1$ 表示公司当期管理层持股较多, $R_MANAHL D_i=0$ 表示公司当期管理层持股较少。管理层持股($R_MANAHL D_i$)变量观测值为23925。

两职合一($Dual_i$)。本文用公司总经理与公司董事长是否为同一人衡量公司两职合一,如果总经理兼任公司董事长,则 $Dual_i=1$,否则为0。两职合一($Dual_i$)变量观测值为23913。

法律风险(R_LR_i)。本文用Kim和Skinner(2012)的Litigation Probit模型计算公司预期的法律风险。如果公司法律风险大于当年研究样本的中位数则 $R_LR_i=1$,否则为0。 $R_LR_i=1$ 表示公司当期法律风险较高, $R_LR_i=0$ 表示公司当期法律风险较低。

表1 样本筛选过程

	样本观测 (公司一年度)	样本观测 (公司)
2001~2015年间所有A股非金融行业上市公司年报	26781	2871
减:因格式问题无法提取文本信息的样本	748	8
减:所有者权益为负或零的样本	514	9
减:ROE数据缺失或绝对值大于1的样本	336	17
减:年报全文长度小于1%分位数的样本	251	3
减:总资产、市值账面比等控制变量缺失的样本	1007	22
最终样本	23925	2812

具体估计模型如下:

$$\begin{aligned}
 SUE D_{i,t} = & \beta_0 + \beta_1 FPS_i + \beta_2 LN ASSET_{i,t-1} \\
 & + \beta_3 SALES_GROWTH_{i,t-1} + \beta_4 RETURN_{i,t} \\
 & + \beta_5 RETURN_SKEWNESS_{i,t} \\
 & + \beta_6 RETURN_STD_DEV_{i,t} + \beta_7 TURNOVER_{i,t} + \varepsilon_{i,t}
 \end{aligned}$$

其中, $SUE D_i$ 为虚拟变量,如果公司当年受到违规

惩处则为1,否则为0; FPS_i 为虚拟变量,如果公司处于高法律风险行业则为1,否则为0,如果2001~2015年,某行业中受违规惩处的企业数量在所有行业中排前1/3,则该行业为高法律风险行业; $LNASSET_{t-1}$ 为t-1期总资产取对数; $SALES\ GROWTH_{t-1}$ 为t-1期营业收入减去t-2期营业收入后,除以t-2期营业收入; $RETURN_i$ 为12个月的累计异常回报,用总市值加权的市场收益进行调整,如果公司当年受到违规惩处,则累计期为当年最后一次受惩处日前12个月,如果公司当年未受惩处,则累计期为当年年末前12个月; $RETURN\ SKEWNESS_i$ 为上述12个月回报的偏度; $RETURN\ STD\ DEV_i$ 为上述12个月回报的标准差; $TURNOVER_i$ 为上述12个月的换手率,即上述12个月的交易总量除以年初总股份数。由于计算法律风险需t-1期及t-2期数据,且模型变量无缺失值,故法律风险(R_LR_i)变量观测值为20786。

(4)管理者自利。

首先,考虑到经营复杂性可能会同时影响文本信息复杂性与管理者薪酬,即文本信息复杂性与管理者薪酬之间存在内生性问题。其次,基于本文逻辑,管理者操纵复杂性更可能是为了获取应得薪酬以外的薪酬,而超额薪酬是管理者薪酬减去应得薪酬的部分,可用来控制公司业绩、经营复杂性及其他基本面因素对管理者薪酬的影响。鉴于此,本文用管理者超额薪酬衡量管理者自利。

管理者超额薪酬($ABMANACOMP_{i,t}$)。参考Core等(1999)、吴联生等(2010),本文用管理者薪酬减去应得薪酬衡量管理者超额薪酬。具体地,管理者薪酬为高管前三名薪酬总额平均数,并取对数;应得薪酬为通过分年、分行业回归如下管理者薪酬决定模型估计所得(每年每行业至少10个观测值)。

$$MANACOMP_{i,t} = \beta_0 + \beta_1 ROA_{i,t} + \beta_2 ROA_{i,t-1} + \beta_3 SOE_{i,t} + \beta_4 DUAL_{i,t} + \beta_5 MANAHL_{i,t} + \beta_6 LEV_i + \beta_7 BDS_{i,t} + \beta_8 SIZE_{i,t} + \varepsilon_{i,t}$$

其中, $MANACOMP_i$ 为管理者薪酬,定义同上; ROA_i 和 ROA_{i-1} 为t期和t-1期总资产收益率; SOE_i 为国有企业虚拟变量,如果公司为国有企业则为1,否则为0; $DUAL_i$ 为两职合一虚拟变量,如果公司总经理与董事长两职合一则为1,否则为0; $MANAHL_i$ 为管理层持股比例,即当年年末管理层持股数量除以公司总股本; LEV_i 为资产负债率,即总负债除以总资产; BDS_i 为董事会规模,即董事会人数; $SIZE_i$ 为公司规模,即总资产对数。因从2005年起才强制要求公司在年报中披露管理者薪酬数据,故超额薪酬数据始于2005年;且计算超额薪酬需模型变量无缺失值和年度一行业观测值大于10,故超额薪酬($ABMANACOMP_{i,t}$)变量观测值为19163。

(5)市场识别。

本文分别从短期、长期市场反应两方面考察市场能否识别管理者对年报文本信息复杂性披露的操纵行为。

短期市场反应($CAR[-10,+1]$)。考虑到中国存在业绩预告、业绩快报等年报信息提前透露机制,且年报信息提前泄露情况比较严重,故本文用年报公告日前10天至公告日后1天的累计异常回报衡量短期市场反应,并用总市值加权的市场收益进行调整。短期市场反应($CAR[-10,+1]$)变量观测值为23007。

长期市场反应($BHAR_{i,t}$)。考虑到中国上市公司年报需在每个会计年度结束之日起4个月内编制完成,本文用公司会计年度结束后第5个月起12个月的持有到期收益衡量长期市场反应,并用总市值加权的市场收益进行调整。长期市场反应($BHAR_{i,t}$)变量观测值为23814。

(6)控制变量。

为控制其他因素对研究变量间关系的影响,参考Li(2008),本文设计以下控制变量:公司规模($Size_i$)、市值账面比(MTB_i)、资产负债率($Leverage_i$)、所有权性质(SOE_i)、公司年龄(Age_i)、再融资事件(SEO_i)、并购事件(MA_i)、非经常性损益(SI_i),定义详见表2。

2. 描述性统计

为防止极值效应,本文对所有连续变量均在上下1%的水平进行缩尾处理(Winsorize),极值处理后的描述性统计详见表3。

表3显示,逆接成分密度($Adverse_i$)、会计术语密度($ATDensity_i$)和次常用字密度($UCDensity_i$)的均值分别为0.068、12.038和0.784。根据本文文本提取结果,中国上市公司年报平均长度为5.2万字。假设一份年报为5

万字,则该年报包含约34个逆接词、6000个会计专业术语、400个次常用字。结果表明,年报文本中各句子、段落间的衔接较为复杂,存在较多语义转折的句子与段落,且包含较多会计专业术语和次常用字,对缺乏专业知识的普通投资者而言,年报文本信息的理解难度较大。

图1为年报文本信息复杂性时间趋势图。从时间序列上看,中国上市公司年报复杂性呈上升之势。2001年,中国上市公司年报每百字中仅包含0.058个逆接词、11.348个会计术语和0.702个次常用字,该数值在2015年分别上升到了0.076、12.395和0.852。可见,上市公司年报文本信息越来越复杂,增加了投资者的解读难度。

表2 变量定义

年报文本信息复杂性变量	
$Adverse_i$	逆接成分密度,即年报每百字包含逆接成分个数,逆接成分表来自廖秋忠(1986)。
$ATDensity_i$	会计术语密度,即年报每百字包含会计术语个数,会计术语表来自灵格斯词霸(2008)。
$UCDensity_i$	次常用字密度,即年报每百字包含次常用字个数,次常用字表来自《现代汉语次常用字表》(1988)。
公司业绩变量	
ROE_i	净资产收益率,即公司净利润/所有者权益。
影响因素变量	
R_NOA_{i-1}	盈余管理空间虚拟变量,如果公司净经营资产比例大于当年研究样本的中位数则为1,否则为0。净经营资产比例为t-1期净经营资产占t-2期总资产的比例。
R_MANAHL	管理层持股虚拟变量,如果公司管理层持股比例大于当年样本的中位数则为1,否则为0。管理层持股比例为年末管理层持股数量占公司总股本的比例。
$Dual_i$	两职合一虚拟变量,如果公司总经理和董事长为同一人则为1,否则为0。
R_LR	法律风险虚拟变量,如果公司法律风险大于当年研究样本的中位数则为1,否则为0。法律风险为基于Kim和Skinner(2012)中Litigation Probit模型计算得到的公司预期法律风险。
管理者自利变量	
$ABMANACOMP_{i+1}$	t+1期管理者超额薪酬,即管理者薪酬减去应得薪酬,其中,管理者薪酬为高管前三名薪酬总额平均数,并取对数;应得薪酬为通过分年、分行业回归管理者薪酬决定模型估计得到。
市场识别变量	
$CAR[-10,+1]$	短期市场反应,即年报公告日前10天至公告日后1天的累计异常回报,用总市值加权的市场收益调整。
$BHAR_{i+1}$	长期市场反应,即公司会计年度结束后第五个月起12个月的持有到期收益,用总市值加权的市场收益调整。
控制变量	
$Size_i$	公司规模,即年末总资产对数。
MTB_i	公司市值账面比,即公司市值/所有者权益。
$Leverage_i$	公司资产负债率,即公司总负债/总资产。
SOE_i	所有权性质虚拟变量,若公司为国有企业则为1,否则为0。
Age_i	公司年龄,即公司上市至今年数。
SEO_i	再融资事件虚拟变量,若公司当年有再融资行为(增发、配股、可转债)则为1,否则为0。
MA_i	并购事件虚拟变量,若公司当年有兼并收购行为则为1,否则为0。
SL_i	非经常性损益,即公司非经常性损益占总资产比例,其中非经常性损益=营业外收入-营业外支出。
$Year$	年度虚拟变量。
$Industry$	行业虚拟变量,基于证监会2012年行业分类标准,制造业取前两位字符,其余取第一位字符。

表3 样本描述性统计

变量	观测值	均值	中位数	标准差	分位数			
					1%	25%	75%	99%
Panel A: 年报文本信息复杂性								
$Adverse_i$	23925	0.068	0.067	0.023	0.020	0.052	0.083	0.129
$ATDensity_i$	23925	12.038	12.071	0.886	9.984	11.497	12.625	13.783
$UCDensity_i$	23925	0.784	0.762	0.178	0.437	0.669	0.871	1.339
Panel B: 公司业绩								
ROE_i	23925	0.063	0.068	0.107	-0.466	0.028	0.112	0.319
Panel C: 管理者自利								
$ABMANACOMP_{i+1}$	19163	0.003	0.017	0.597	-1.598	-0.368	0.399	1.436
Panel D: 市场识别								
$CAR[-10,+1]$	23007	0.008	-0.001	0.084	-0.193	-0.043	0.050	0.291
$BHAR_{i+1}$	23814	0.076	-0.018	0.470	-0.829	-0.175	0.203	2.333
Panel E: 控制变量								
$Size_i$	23925	21.676	21.517	1.194	19.311	20.824	22.328	25.384
MTB_i	23925	2.276	1.526	2.298	0.228	0.846	2.810	13.936
$Leverage_i$	23925	0.451	0.458	0.204	0.047	0.295	0.610	0.870
SOE_i	23925	0.539	1.000	0.498	0.000	0.000	1.000	1.000
Age_i	23925	9.024	8.000	5.749	1.000	4.000	13.000	22.000
SEO_i	23925	0.102	0.000	0.303	0.000	0.000	0.000	1.000
MA_i	23925	0.681	1.000	0.466	0.000	0.000	1.000	1.000
SL_i	23925	0.006	0.002	0.012	-0.021	0.000	0.007	0.070

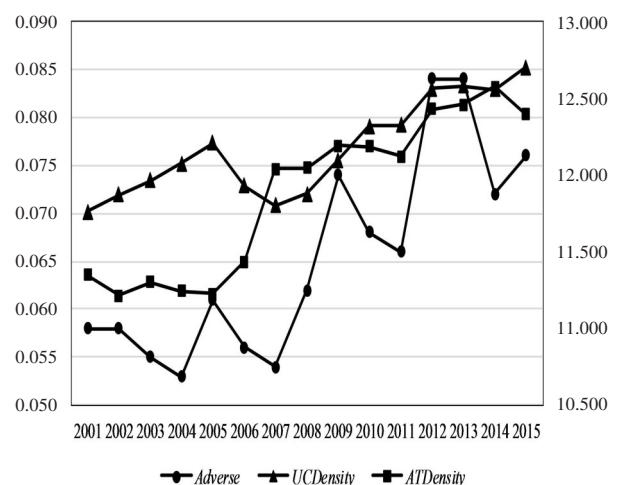


图1 年报文本信息复杂性时间趋势

注: $Adverse$ 和 $UCDensity$ 为左侧坐标轴, $ATDensity$ 为右侧坐标轴, 为量纲一致, $UCDensity$ 在原定义基础上除以 10 在图中呈现。

(三)年报文本信息复杂性披露策略的检验

1. 公司当期业绩与年报文本信息复杂性

考虑到管理者操纵年报文本信息复杂性主要是为了隐藏公司的负面信息。首先,基于公司业绩好坏分组,检验业绩好公司和差公司年报文本信息复杂性的差异。具体地,本文基于样本ROE_{*i*}的中位数区分业绩好公司、差公司,若当期业绩小于当年研究样本的中位数则为业绩差公司,否则为业绩好公司,结果详见表4和图2。表4和图2显示,业绩差公司逆接成分密度(*Adverse_i*)、会计术语密度(*ATDensity_i*)和次常用字密度(*UCDensity_i*)的均值分别为0.071、12.108和0.791,业绩好公司对应变量的均值分别为0.066、11.968和0.777,两者差异均在1%水平显著为正。表4和图2说明,公司当期业绩越差,年报文本信息复杂性越高。

其次,为控制其他因素对年报文本信息复杂性的影响,本文设计模型(1)进行回归分析,检验公司当期业绩与年报文本信息复杂性的关系。

$$Complexity_{i,t}=\beta_0+\beta_1ROE_{i,t}+\beta_2ControlVariable_{i,t}+Year+Industry+\varepsilon_{i,t} \tag{1}$$

其中,*Complexity_i*表示3个复杂性变量,分别为*Adverse_i*、*ATDensity_i*和*UCDensity_i*,*ControlVariable_i*表示控制变量,其他变量定义详见表2。回归结果详见表5。

表5的回归结果显示,逆接成分密度(*Adverse_i*)回归中ROE_{*i*}系数为-0.017,T值为-7.28;会计术语密度(*ATDensity_i*)回归中ROE_{*i*}系数为-0.153,T值为-2.92;次常用字密度(*UCDensity_i*)回归中ROE_{*i*}系数为-0.044,T值为-4.11,均在1%水平显著为负。结果表明,公司当期业绩越差,年报文本信息复杂性越高,其与已有基于英文年报文本信息复杂性披露策略(Li,2008)的研究结论一致。

2. 年报文本信息复杂性披露策略影响因素检验

(1)数字信息操纵的影响。

考虑到数字信息操纵和文本信息操纵可能并非独立存在,文本信息操纵会受到数字信息操纵的影响,本文检验盈余管理空间对当期业绩与年报文本信息复杂性关系的影响,结果详见表6的Panel A。

表6 Panel A结果显示,公司当期业绩与盈余管理空间的交叉项ROE_{*i*}×*R_NOA_{i-1}*对逆接成分密度(*Adverse_i*)、会计术语密度(*ATDensity_i*)和次常用字密度(*UCDensity_i*)的系数分别为-0.007、-0.243和-0.069,分别在5%、5%

表4 当期业绩与年报文本信息复杂性的分组比较

	业绩差		业绩好			
	均值	中位数	均值	中位数	均值差异	中位数差异
	(1)	(2)	(3)	(4)	(1)-(3)	(2)-(4)
<i>Adverse_i</i>	0.071	0.069	0.066	0.064	16.22***	15.99***
<i>ATDensity_i</i>	12.108	12.157	11.968	11.993	12.21***	13.60***
<i>UCDensity_i</i>	0.791	0.770	0.777	0.754	6.23***	7.42***
Obs.	11884		12041			

注:***、**、*分别表示1%、5%、10%的显著性水平,下表同。

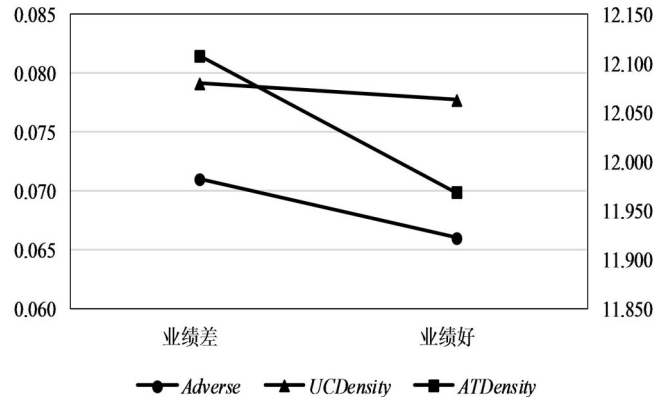


图2 当期业绩与年报文本信息复杂性分组比较

注:*Adverse*和*UCDensity*为左侧坐标轴,*ATDensity*为右侧坐标轴,为量纲一致,*UCDensity*在原定义基础上除以10在图中呈现。

表5 当期业绩与年报文本信息复杂性的回归分析

	(1)	(2)	(3)
	<i>Adverse_i</i>	<i>ATDensity_i</i>	<i>UCDensity_i</i>
<i>ROE_i</i>	-0.017*** [-7.28]	-0.153*** [-2.92]	-0.044*** [-4.11]
<i>Size_i</i>	-0.003*** [-6.13]	-0.176*** [-8.51]	0.010 [1.62]
<i>MTB_i</i>	0.000 [1.67]	-0.003 [-1.18]	-0.002** [-2.46]
<i>Leverage_i</i>	-0.001 [-0.30]	0.040 [0.50]	-0.022 [-1.17]
<i>SOE_i</i>	0.001 [1.06]	0.066*** [3.40]	-0.003 [-0.36]
<i>Age_i</i>	-0.004*** [-22.20]	0.202*** [17.06]	-0.002 [-0.85]
<i>SEO_i</i>	-0.001** [-2.22]	0.000 [0.01]	0.004 [1.28]
<i>MA_i</i>	-0.001*** [-5.72]	-0.030** [-2.64]	-0.006*** [-2.97]
<i>SI_i</i>	0.028** [2.37]	-0.355 [-0.60]	0.115** [2.48]
<i>Cons</i>	0.217*** [20.13]	12.346*** [19.89]	0.677*** [4.14]
<i>Year</i>	控制	控制	控制
<i>Industry</i>	控制	控制	控制
N	23925	23925	23925
Adj. R ²	0.522	0.663	0.707

注:***、**、*分别表示1%、5%、10%的显著性水平,[]中系数为T值,下表同。

和1%水平显著为负。Panel A结果表明,当公司盈余可操纵空间较小时,管理者会增加文本信息复杂性操纵作为替代。

(2)所持股份价值激励和公司内、外部治理的影响。

考虑到管理层持股价值激励和公司内、外部治理会影响年报文本信息复杂性操纵的收益和成本,本文以管理层持股比例衡量持股价值激励,两职合一衡量内部治理,法律风险衡量外部治理,分别检验其对当期业绩与年报文本信息复杂性关系的影响,结果详见表6的Panel B。

表6 Panel B的列(1)~(3)检验管理层持股的影响,列(4)~(6)检验两职合一的影响,列(7)~(9)检验法律风险的影响。列(1)~(3)结果显示,公司当期业绩与管理层持股的交叉项 $ROE_i \times R_MANAHL_{i-1}$ 对逆接成分密度($Adverse_i$)、会计术语密度($ATDensity_i$)和次常用字密度($UCDensity_i$)的影响系数分别为-0.010、-0.179和-0.046,分别在1%、5%和5%水平显著为负;列(4)~(6)结果显示,公司当期业绩与两职合一的交叉项 $ROE_i \times Dual_i$ 对逆接成分密度($Adverse_i$)、会计术语密度($ATDensity_i$)和次常用字密度($UCDensity_i$)的影响系数分别为-0.010、-0.237和-0.034,分别在1%、5%和10%水平显著为负;列(7)~(9)结果显示,公司当期业绩与法律风险的交叉项 $ROE_i \times R_LR_i$ 对逆接成分密度($Adverse_i$)、会计术语密度($ATDensity_i$)和次常用字密度($UCDensity_i$)的影响系数分别为0.007、0.121和0.042,均在5%水平显著为正。Panel B结果表明,管理层持股比例越大、两职合一、法律风险越低,管理者对年报文本信息复杂性的操纵越强。

表6 文本信息复杂性披露影响因素的回归分析

Panel A: 数字信息操纵

	(1)	(2)	(3)
	$Adverse_i$	$ATDensity_i$	$UCDensity_i$
$ROE_i \times R_NOA_{i-1}$	-0.007** [-2.16]	-0.243** [-2.05]	-0.069*** [-3.89]
ROE_i	-0.024*** [-9.91]	-0.269*** [-3.96]	-0.075*** [-4.24]
R_NOA_{i-1}	0.000 [0.15]	0.018 [0.93]	0.009*** [5.01]
Cons	0.129*** [22.09]	14.717*** [54.77]	0.792*** [10.87]
ControlVariable _i	控制	控制	控制
Year	控制	控制	控制
Industry	控制	控制	控制
N	20776	20776	20776
Adj. R ²	0.246	0.355	0.128

Panel B: 管理层持股、两职合一、法律风险

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$Adverse_i$	$ATDensity_i$	$UCDensity_i$	$Adverse_i$	$ATDensity_i$	$UCDensity_i$	$Adverse_i$	$ATDensity_i$	$UCDensity_i$
$ROE_i \times R_MANAHL_{i-1}$	-0.010*** [-4.16]	-0.179** [-1.98]	-0.046** [-2.22]						
$ROE_i \times Dual_i$				-0.010*** [-3.24]	-0.237** [-2.37]	-0.034* [-1.75]			
$ROE_i \times R_LR_i$							0.007** [2.27]	0.121** [2.43]	0.042** [2.00]
ROE_i	-0.017*** [-10.43]	-0.137** [-2.37]	-0.060*** [-4.49]	-0.019*** [-8.19]	-0.068 [-1.28]	-0.073*** [-5.32]	-0.024*** [-9.76]	-0.194** [-2.16]	-0.100*** [-5.20]
R_MANAHL_{i-1}	0.000 [0.23]	-0.099*** [-8.44]	-0.010*** [-3.56]						
$Dual_i$				0.001 [1.16]	0.003 [0.12]	0.001 [0.24]			
R_LR_i							0.000 [0.11]	0.025 [1.54]	0.005** [2.17]
Cons	0.124*** [37.25]	14.563*** [122.49]	0.832*** [30.14]	0.124*** [23.45]	14.480*** [52.98]	0.820*** [11.91]	0.127*** [18.40]	14.564*** [51.61]	0.753*** [10.33]
ControlVariable _i	控制	控制	控制	控制	控制	控制	控制	控制	控制
Year	控制	控制	控制	控制	控制	控制	控制	控制	控制
Industry	控制	控制	控制	控制	控制	控制	控制	控制	控制
N	23925	23925	23925	23913	23913	23913	20786	20786	20786
Adj. R ²	0.250	0.351	0.130	0.250	0.349	0.129	0.245	0.349	0.128

注:限于篇幅,本表未报告控制变量回归结果,控制变量包括Size_i、MTB_i、Leverage_i、SOE_i、Age_i、SEO_i、MA_i、SI_i。

(四) 年报文本信息复杂性披露策略的管理者自利行为检验

本文研究发现,管理者出于自利动机会通过增加年报文本信息复杂性隐藏公司当期的坏消息。那么该操纵行为是否能够实现管理者自利便是一个待检验的问题。考虑到薪酬是管理者操纵年报文本信息复杂性、掩盖坏消息的直接动因,且管理者操纵复杂性更可能是为了获取应得薪酬以外的薪酬,本文设计模型(2)检验年报文本信息复杂性对管理者超额薪酬的影响。

$$ABMANACOMP_{i,t+1} = \gamma_0 + \gamma_1 Complexity_{i,t} + \gamma_2 ROE_{i,t} + \gamma_3 ControlVariable_{i,t} + Year + Industry + \varepsilon_{i,t} \quad (2)$$

其中, $Complexity_{i,t}$ 表示3个复杂性变量,分别为 $Adverse_{i,t}$ 、 $ATDensity_{i,t}$ 和 $UCDensity_{i,t}$, $ControlVariable_{i,t}$ 表示控制变量,其他变量定义详见表2。回归结果详见表7。

表7的回归结果显示,年报文本信息复杂性变量 $Adverse_{i,t}$ 、 $ATDensity_{i,t}$ 和 $UCDensity_{i,t}$ 的系数分别为2.199、0.051和0.138, T值分别为8.25、7.59和1.99, 分别在1%、1%和5%水平显著为正。结果表明,年报文本信息复杂性与管理者超额薪酬显著正相关,即管理者通过操纵年报文本信息复杂性实现了其自利动机。

(五) 年报文本信息复杂性披露策略市场识别的检验

考虑到管理者操纵年报文本信息复杂性,旨在隐藏坏消息,影响市场对公司业绩的解读,从而影响市场估值。具体而言,考虑到业绩较好公司年报文本信息复杂性的客观性更强,业绩较差公司年报文本信息复杂性的机会主义动机更强,即业绩较好、较差公司年报文本信息复杂性的特征不同。基于此,本文设计模型(3)在区分业绩好、坏的基础上检验年报文本信息复杂性对投资者市场反应的影响。

$$CAR[-10, +1] = \gamma_0 + \gamma_1 Complexity_{i,t} \times BAD_{i,t} + \gamma_2 Complexity_{i,t} + \gamma_3 BAD_{i,t} + \gamma_4 ROE_{i,t} + \gamma_5 ControlVariable_{i,t} + Year + Industry + \varepsilon_{i,t} \quad (3)$$

$$BHAR_{i,t+1} = \gamma_0 + \gamma_1 Complexity_{i,t} \times BAD_{i,t} + \gamma_2 Complexity_{i,t} + \gamma_3 BAD_{i,t} + \gamma_4 ROE_{i,t} + \gamma_5 ControlVariable_{i,t} + Year + Industry + \varepsilon_{i,t} \quad (4)$$

其中, $Complexity_{i,t}$ 表示3个复杂性变量,分别为 $Adverse_{i,t}$ 、 $ATDensity_{i,t}$ 和 $UCDensity_{i,t}$, $BAD_{i,t}$ 为当期业绩好坏虚拟变

表7 年报文本信息复杂性与管理者超额薪酬的回归分析

	ABMANACOMP _{t+1}		
	(1)	(2)	(3)
<i>Adverse_t</i>	2.199*** [8.25]		
<i>ATDensity_t</i>		0.051*** [7.59]	
<i>UCDensity_t</i>			0.138** [1.99]
<i>ROE_t</i>	0.103** [2.00]	0.064 [1.21]	0.057 [1.00]
<i>Size_t</i>	0.023*** [4.51]	0.023*** [4.86]	0.021*** [3.99]
<i>MTB_t</i>	0.037*** [10.34]	0.036*** [10.79]	0.039*** [11.04]
<i>Leverage_t</i>	-0.166*** [-5.49]	-0.162*** [-5.18]	-0.185*** [-5.61]
<i>SOE_t</i>	-0.007 [-0.74]	-0.014 [-1.51]	-0.016* [-1.93]
<i>Age_t</i>	0.008*** [3.82]	0.007*** [3.55]	0.009*** [4.41]
<i>SEO_t</i>	-0.046*** [-7.16]	-0.051*** [-7.96]	-0.051*** [-7.55]
<i>MA_t</i>	0.045*** [5.61]	0.044*** [5.64]	0.042*** [5.59]
<i>SL_t</i>	3.044*** [4.71]	3.096*** [4.65]	3.192*** [4.84]
<i>Cons</i>	-0.805*** [-8.75]	-1.256*** [-14.61]	-0.714*** [-7.80]
<i>Year</i>	控制	控制	控制
<i>Industry</i>	控制	控制	控制
<i>N</i>	19163	19163	19163
<i>Adj. R²</i>	0.051	0.050	0.047

量,如果公司当期业绩低于行业的中位数,则 $BAD_{i,t}=1$, 否则为0。 $ControlVariable_{i,t}$ 表示控制变量,其他变量定义详见表2。回归结果详见表8。

表8的列(1)~(3)为短期市场反应的结果,列(4)~(6)为长期市场反应的结果。列(1)~(3)结果显示,年报文本信息复杂性变量 $Adverse_{i,t}$ 、 $ATDensity_{i,t}$ 和 $UCDensity_{i,t}$ 的系数分别为-0.088、-0.001和-0.007, 分别1%、10%和1%水平显著为负,业绩好坏与逆接成分密度的交叉项 $BAD_{i,t} \times Adverse_{i,t}$ 系数为0.117,与会计术语密度的交叉项 $BAD_{i,t} \times ATDensity_{i,t}$ 系数为0.004,与次常用字密度的交叉项 $BAD_{i,t} \times UCDensity_{i,t}$ 系数为0.007, 分别在5%、1%和10%水平显著为正。列(4)~(6)结果显示,年报文本信息复杂性变量 $Adverse_{i,t}$ 、 $ATDensity_{i,t}$ 和 $UCDensity_{i,t}$ 的系数分别为-0.402、-0.011和-0.071, 分别在1%、5%和5%水平显著为负,交叉项 $BAD_{i,t} \times Adverse_{i,t}$ 、 $BAD_{i,t} \times ATDensity_{i,t}$ 、 $BAD_{i,t} \times UCDensity_{i,t}$ 的系数分别为1.313、0.010和0.073, 分别在1%、1%和5%水平显著为正。表8结果表明,业绩较好公司年报文本信息复杂性与短期、长期市场反应显著负相关,业绩较差公司年报文本信息复杂性与短期、长期市场反应间的负相关关系显著减弱。也就是说,从市场估值角度表明,无论是短期还是长期,鉴于文本信息复

杂性增加了阅读难度,提高了信息解读成本,会影响投资者估值。针对业绩较好公司的文本信息复杂性,投资者更偏好向下估值;而针对业绩较差公司,因其操纵年报文本信息复杂性更强,对投资者造成的迷惑更多,导致投资者向下估值的程度被削弱,即投资者更无法看清业绩较差公司年报文本信息复杂性的机会主义操纵行为。

(六)稳健性检验

考虑到管理者操纵年报文本信息复杂性时,除基于公司自身业绩外,也会基于公司当年同行业其他公司的业绩情况。基于此,本文利用相对业绩变量 ROE_IA_i 度量公司业绩表现,具体为上市公司净资产收益率与当年同行业的中位数的差值,行业划分基于证监会2012年行业分类标准,制造业取前两位字符,其余取第一位字符。 ROE_IA_i 值越大,表示上市公司与当年同行业其他公司相比业绩越好。回归结果详见表9。

表9的回归结果显示,逆接成分密度($Adverse_i$)回归中 ROE_IA_i 的系数为-0.017,会计术语密度($ATDensity_i$)回归中 ROE_IA_i 的系数为-0.161,次常用字密度($UCDensity_i$)回归中 ROE_IA_i 的系数为-0.049,均在1%水平显著为负。结果表明,公司当期相对业绩越差,年报文本信息复杂性越高。该结果与表5结果一致。

四、研究结论与政策启示

基于信息不对称理论和代理理论,本文研究管理者出于自利动机的年报文本信息复杂性披露策略。研究

表8 年报文本信息复杂性与市场反应的回归分析

	CAR[-10,+1]			BHAR _{t-1}		
	(1)	(2)	(3)	(4)	(5)	(6)
$Adverse \times BAD_i$	0.117** [2.09]			1.313*** [8.54]		
$ATDensity_i \times BAD_i$		0.004*** [3.52]			0.010*** [2.87]	
$UCDensity_i \times BAD_i$			0.007* [1.75]			0.073** [2.53]
$Adverse_i$	-0.088*** [-2.86]			-0.402*** [-3.36]		
$ATDensity_i$		-0.001* [-1.87]			-0.011** [-2.78]	
$UCDensity_i$			-0.007*** [-3.16]			-0.071** [-2.49]
BAD_i	-0.010** [-2.35]	-0.052*** [-3.63]	-0.008** [-2.16]	-0.082*** [-8.24]	-0.117** [-2.76]	-0.050* [-1.99]
ROE_i	0.080*** [9.25]	0.080*** [10.39]	0.079*** [9.92]	-0.060 [-1.55]	-0.073* [-1.85]	-0.076* [-1.90]
$Size_i$	-0.004*** [-5.29]	-0.004*** [-4.94]	-0.004*** [-4.96]	-0.040*** [-8.05]	-0.038*** [-8.01]	-0.041*** [-8.61]
MTB_i	-0.001*** [-4.01]	-0.001*** [-4.19]	-0.001*** [-3.79]	-0.015*** [-5.81]	-0.014*** [-5.60]	-0.015*** [-6.01]
$Leverage_i$	0.019*** [4.61]	0.019*** [4.87]	0.018*** [4.65]	0.150*** [6.23]	0.149*** [6.18]	0.151*** [6.54]
SOE_i	0.000 [-0.22]	0.000 [-0.30]	-0.001 [-0.69]	-0.010* [-1.78]	-0.012** [-2.31]	-0.011* [-1.98]
Age_i	0.000 [0.71]	0.000 [0.54]	0.000 [0.86]	0.001 [0.71]	0.000 [0.29]	0.000 [0.63]
SEO_i	0.001 [0.51]	0.001 [0.69]	0.001 [0.79]	-0.002 [-0.51]	-0.002 [-0.44]	-0.002 [-0.34]
MA_i	-0.001 [-0.54]	-0.001 [-0.75]	0.000 [-0.56]	-0.021*** [-5.21]	-0.020*** [-5.20]	-0.022*** [-5.61]
SI_i	0.003 [0.05]	0.005 [0.13]	0.012 [0.29]	0.143 [0.58]	0.169 [0.70]	0.139 [0.58]
$Cons$	0.092*** [5.28]	0.099*** [4.25]	0.111*** [5.34]	0.715*** [6.57]	0.528*** [5.19]	0.760*** [7.51]
$Year$	控制	控制	控制	控制	控制	控制
$Industry$	控制	控制	控制	控制	控制	控制
N	23007	23007	23007	23814	23814	23814
Adj. R ²	0.037	0.037	0.040	0.098	0.098	0.098

发现:第一,公司当期业绩与年报文本信息复杂性显著负相关;第二,该负相关关系在盈余管理空间小、管理层持股比例大、两职合一和法律风险低的公司中更加显著;第三,年报文本信息复杂性能提高管理者超额薪酬;第四,相比业绩较好公司,业绩较差公司年报文本信息复杂性的短期、长期市场反应更积极。本文研究表明,出于自利动机,管理者的确会通过增加年报文本信息复杂性隐藏公司当期坏消息;管理者将文本信息复杂性操纵作

表9 当期相对业绩与年报文本信息复杂性的回归分析

	(1)	(2)	(3)
	$Adverse_i$	$ATDensity_i$	$UCDensity_i$
ROE_IA_i	-0.017*** [-7.14]	-0.161*** [-3.23]	-0.049*** [-4.86]
$Size_i$	-0.003*** [-6.17]	-0.176*** [-8.53]	0.010 [1.63]
MTB_i	0.000 [1.65]	-0.004 [-1.19]	-0.002** [-2.46]
$Leverage_i$	-0.001 [-0.28]	0.040 [0.50]	-0.023 [-1.20]
SOE_i	0.001 [1.05]	0.066*** [3.39]	-0.004 [-0.38]
Age_i	-0.004*** [-22.26]	0.202*** [17.15]	-0.002 [-0.74]
SEO_i	-0.001** [-2.22]	0.000 [0.01]	0.004 [1.29]
MA_i	-0.001*** [-5.66]	-0.030** [-2.64]	-0.006*** [-2.97]
SI_i	0.029** [2.57]	-0.333 [-0.56]	0.127** [2.79]
$Cons$	0.215*** [19.86]	12.323*** [19.89]	0.667*** [4.03]
$Year$	控制	控制	控制
$Industry$	控制	控制	控制
N	23925	23925	23925
Adj. R ²	0.523	0.663	0.707

为数字信息操纵的替代,且管理者财富激励及公司内、外部治理均会影响年报文本信息复杂性披露策略;年报文本信息复杂性操纵能使管理者获得更高超额薪酬,并增加市场普通投资者对年报文本信息的解读成本,影响公司市场估值。

本研究在理论和实践方面具有重要启示。在理论层面,一方面,已有英文文本复杂性研究主要从字词角度出发,忽略了文本逻辑复杂性,本文从逻辑和字词的复杂性两个角度构建中文年报文本信息复杂性指标,提供了更加契合中文文本披露的度量方法,克服了已有中文文本研究通过定性、小样本分析、主观打分等判定文本信息披露质量的不足,为文本信息复杂性披露策略研究(Li, 2008; Loughran and McDonald, 2014)提供了基于中文年报分析的实证证据。另一方面,本文通过研究盈余管理空间与年报文本信息复杂性操纵的替代机制,发展了机会主义披露策略的研究。在政策实践层面,本文提示监管者应加强对公司财务报告文本信息复杂性的关注,指导上市公司发布更加简明易懂的财务报告,降低管理者与投资者间的信息不对称;同时,本文还可帮助投资者识别公司年报文本信息复杂性披露的机会主义策略,提示投资者加强对公司基本面等信息的全面分析。

(作者单位:王克敏、李栋栋,复旦大学管理学院;王华杰,上海立信会计金融学院会计学院;戴杏云,中国金融期货交易所。责任编辑:张世国)

注释

①2007年之前,“短期金融投资”取值为资产负债表中的“短期投资净额”,2007年及之后取值为资产负债表中的“交易性金融资产”、“可供出售金融资产净额”与“持有至到期投资净额”三者之和。

参考文献

- (1)顾曰国:《礼貌、语用与文化》,《外语教学与研究》,1992年第4期。
- (2)国家语言工作委员会和国家教育委员会:《现代汉语次常用字表》,http://www.zdic.net/z/zb/cc2.htm,1988年。
- (3)廖秋忠:《现代汉语篇章中的连接成分》,《中国语文》,1986年。
- (4)灵格斯词霸:《灵格斯汉英会计词典》,http://www.lingoes.cn/zh/dictionary,2008年。
- (5)孟庆涛:《英汉语用礼貌原则跨文化交际运用探析》,《现代语文:语言研究》,2009年第2期。
- (6)孙蔓莉、蒋璐、孙健:《业绩归因的自利性披露及市场反应研究——汇率单边升值情境下的纺织业表现》,《会计研究》,2013年第4期。
- (7)吴联生、林景艺、王亚平:《薪酬外部公平性、股权性质与公司业绩》,《管理世界》,2010年第3期。
- (8)Ang, J. S., Cole, R. A., Lin, J. W., 2000, “Agency Costs and Ownership Structure”, *The Journal of Finance*, Vol.55, pp.81~106.
- (9)Barton, J. and Simko, P. J., 2002, “The Balance Sheet as an Earnings Management Constraint”, *The Accounting Review*, Vol.77, pp.1~27.
- (10)Berle, A. A., Means, G. C., Weidenbaum, M. L., Jensen, M., 1935, “The Modern Corporation and Private Property”, *Economic Journal*, Vol.20, pp.119~129.
- (11)Bloomfield, R. J., 2002, “The ‘Incomplete Revelation Hypothesis’ and Financial Reporting”, *Accounting Horizons*, Vol.16, pp.233~243.
- (12)Core, J. E., Holthausen, R. W., Larcker, D. F., 1999, “Corporate Governance, Chief Executive Officer Compensation and Firm Performance”, *Journal of Financial Economics*, Vol.51, pp.371~406.
- (13)Courtis, J. K., 1998, “Annual Report Readability Variability: Tests of the Obfuscation Hypothesis”, *Accounting, Auditing and Accountability Journal*, Vol.11, pp.459~472.
- (14)Das, S., Kim, K. and Patro, S., 2011, “An Analysis of Managerial Use and Market Consequences of Earnings Management and Expectation Management”, *The Accounting Review*, Vol.86, pp.1935~1967.
- (15)DeAngelo, L. E., 1988, “Managerial Competition, Information Costs and Corporate Governance: The Use of Accounting Performance Measures in Proxy Contests”, *Journal of Accounting and Economics*, Vol.10, pp.3~36.
- (16)Defond, M. L. and Park, C. W., 1997, “Smoothing Income in Anticipation of Future Earnings”, *Journal of Accounting and Economics*, Vol.23, pp.115~139.
- (17)Grossman, S. J. and Stiglitz, J. E., 1980, “On the Impossibility of Informationally Efficient Markets”, *The American Economic Review*, Vol.70, pp.393~408.
- (18)Healy, P. M., 1985, “The Effect of Bonus Schemes on Accounting Decisions”, *Journal of Accounting and Economics*, Vol.7, pp.85~107.
- (19)Healy, P. M. and Palepu, K. G., 2001, “Information Asymmetry, Corporate Disclosure and the Capital Markets: A Review of the Empirical Disclosure Literature”, *Journal of Accounting and Economics*, Vol.31, pp.405~440.
- (20)Hirshleifer, D., Hou, K., Teoh, S. H. and Zhang, Y., 2004, “Do Investors Overvalue Firms with Bloated Balance Sheets?”, *Journal*

of Accounting and Economics, Vol.38, pp.297~331.

(21) Huang, X., Teoh, S. H. and Zhang Y., 2014, "Tone Management", *The Accounting Review*, Vol.89, pp.1083~1113.

(22) Jensen, M. C. and Meckling, W. H., 1994, "The Nature of Man", *Journal of Applied Corporate Finance*, Vol.7, pp.4~19.

(23) Jensen, M. C. and Meckling, W. H., 1976, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", *Journal of Financial Economics*, Vol.4, pp.305~360.

(24) Kim, I. and Skinner, D. J., 2012, "Measuring Securities Litigation Risk", *Journal of Accounting and Economics*, Vol.53, pp.290~310.

(25) Kothari, S. P., Shu S. and Wysocki P. D., 2009, "Do Managers Withhold Bad News?", *Journal of Accounting Research*, Vol.47, pp.241~276.

(26) Li, F., 2008, "Annual Report Readability, Current Earnings and Earnings Persistence", *Journal of Accounting and Economics*, Vol.45, pp.221~247.

(27) Li, F., 2010, "Textual Analysis of Corporate Disclosures: A Survey of the Literature", *Journal of Accounting Literature*, 29, pp.143.

(28) Loughran, T. and McDonald, B., 2014, "Measuring Readability in Financial Disclosures", *The Journal of Finance*, Vol.69, pp.1643~1671.

(29) Merkley, K. J., 2014, "Narrative Disclosure and Earnings Performance: Evidence from R&D Disclosures", *Accounting Review*, Vol.89, pp.725~757.

(30) Murphy, K. J., 1985, "Corporate Performance and Managerial Remuneration: An Empirical Analysis", *Journal of Accounting and Economics*, Vol.7, pp.11~42.

(31) Nelson, K. K. and Pritchard, A. C., 2007, "Litigation Risk and Voluntary Disclosure: The Use of Meaningful Cautionary Language", SSRN Working Paper.

(32) Pashalian, S. and Crissy, W. J. E., 1950, "How Readable are Corporate Annual Reports?", *Journal of Applied Psychology*, Vol.34, pp.244~248.

(33) Pashalian, S. and Crissy, W. J. E., 1952, "Corporate Annual Reports Are Difficult, Dull Reading, Human Interest Value Low, Survey Shows", *Journal of Accountancy*, Vol.94, pp.215~219.

(34) Pretorius, E. J., 2006, "The Comprehension of Logical Relations in Expository Texts by Students Who Study Through the Medium of ESL", *System*, Vol.34, pp.432~450.

(35) Schipper, K., 1989, "Commentary on Earnings Management", *Accounting Horizons*, Vol.3, pp.91~102.

(36) Schrand, C. M. and Walther, B. R., 2000, "Strategic Benchmarks in Earnings Announcements: The Selective Disclosure of Prior-period Earnings Components", *The Accounting Review*, Vol.75, pp.151~177.

(37) Skinner, D. J., 1994, "Why Firms Voluntarily Disclose Bad News", *Journal of Accounting Research*, Vol.32, pp.38~60.

(38) Sloan, R., 1996, "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?", *The Accounting Review*, Vol.71, pp.289~315.

(39) Xie, H., 2001, "The Mispricing of Abnormal Accruals", *The Accounting Review*, Vol.76, pp.357~373.

附录

本附录介绍了年报文本处理的主要步骤。

1. 下载年报

从上海证券交易所、深圳证券交易所、巨潮资讯网网站获得2001~2015年间所有A股上市公司年报下载地址,利用批量下载软件下载所有PDF格式的年报文件。

2. 将PDF文件转换为HTML文件

利用格式转换软件(Solid Converter PDF)将PDF格式的年报文件批量转换为HTML格式。

3. 利用Perl的HTML模块,整理年报文本

(1)删除所有图片(HTML tag为)

(2)针对表格(HTML tag为

)，逐一单元格进行判断,如果单元格内包含句子符号(包括“,”、“。”、“;”等)或者包含15个及以上中文字符,则保留单元格内信息,否则删除单元格。

(3)段落处理

① 删除不包含任何中文字符的段落,多为未标记“td”的表格或者页码行。

② 删除数字及空格字符占段落总字符数大于30%的段落,多为未标记“td”的表格。

③ 删除年报固定格式行;如√ 适用 □ 不适用、(元)、(股)等。

(4)去除所有HTML tag,输出纯文本。

4. 分词、统计词频

(1)借助简易中文分词系统SCWS(Simple Chinese Word Segmentation)对年报文字信息进行分词,并统计词频。

本文使用到逆接成分、会计术语和次常用字3个词典。其中,逆接成分词典来自《现代汉语篇章中的连接成分》(廖秋忠,1986)的逆接连接成分;会计术语词典来自《灵格斯汉英会计词典》(灵格斯词霸,2008);次常用字词典来自《现代汉语次常用字表》(国家语言文字工作委员会和国家教育委员会,1988)。

(2)用Perl计算文本中总中文字符数。

(3)将词典词频除以总中文字符数,再乘以100,即得到本文所用复杂性变量。

vation preference. In the eastern and western regions, fiscal decentralization can improve regional innovation efficiency, but in the central region, fiscal decentralization has no significant impact on innovation efficiency. As far as the mechanism is concerned, in the eastern region, fiscal decentralization will also improve the efficiency of regional innovation, but also inhibit the local government's innovation preference and lead to the loss of innovation efficiency. In the western region, after controlling the endogenous problem of government's innovation preference, this paper finds that fiscal decentralization will reduce the government's excessive intervention in regional innovation activities and indirectly enhance the efficiency of regional innovation.

Does Discovery Opportunity and Creation Opportunity Transform into Each Other?:

Based on Multi-Actors Perspective

Cai Li, Lu Xifeng, Shan Biaoan and Yu Haijing

As central elements of entrepreneurship, discovery opportunity and creation opportunity have gained extensive attention. Recently, the transformation between these two different types of opportunity has become frontier issue. However, the current research on this issue is still in an exploratory stage. Existing research concerned more about the role of entrepreneurs in entrepreneurial process and ignored the effects of other stakeholders. Consequently, this study utilizes the RBV, stakeholder theory and organizational learning theory to construct the transformation model between discovery opportunity and creation opportunity and analyzes the effect of multi-actor in the transformation process. Based on theoretical analysis and the case study of Alipay and Yu'e Bao, we propose that multi-actors who participate in the opportunity development process contributes to creating opportunity set, and the creation and expansion of opportunity set realizes the transformation between the two types of opportunity. Therefore, we reveal the transformation mechanism between these two types of opportunity based on multi-actor perspective.

Complexity of Annual Report and Management Self-interest:

Empirical Evidence from Chinese Listed Firms

Wang Kemin and other authors

We investigate the manager's self-interest narrative disclosure strategy of annual reports. We find that compared to firms with good performance, firms with bad performance have more complicated annual reports. The negative relation is stronger among firms with high earnings management constraint, high management shareholding, CEO duality or low litigation risk. Moreover, the complexity of annual report is positively related with management abnormal compensation. Furthermore, compared to firms with good performance, the short-run and long-run market return to complexity is more positive for firms with bad performance. The above evidence is consistent with the conjecture that managers will manipulate the complexity of annual reports to hide adverse information to get more abnormal compensation and affect market valuation. The manipulation of annual report complexity has a substitution role to the manipulation of number information; and managers' self-interest, internal and external governance will also affect the annual report complexity. For the first time through large sample analysis of Chinese annual report, we provide empirical evidences for the study in complexity of narrative information validate the substitution effect between number and narrative information manipulation and expand the study of opportunistic disclosure. Besides, findings of this paper have important implications for the construction of the financial narrative disclosure system.