前瞻性信息披露与公司绩效

——基于文本分析和机器学习[1]

胡楠，薛付婧*

（西安交通大学管理学院，西安，710049）

【摘要】：以上市公司 2011-2016 年度财务报告为对象，本文分析了年报中所披露的前瞻性信息的有用性。通过 Word2Vec 机器学习技术获取在年报环境下代表前瞻性信息的词集，通过自然语言处理和文本分析获取所有词集的词频，本文构建出全新的前瞻性信息披露指标。研究结果发现，前瞻性信息披露水平与公司未来绩效呈正相关关系。进一步研究表明，公司年度报告可读性越强，前瞻性信息与公司未来绩效两者之间的关系越强。公司信息不对称程度越高，前瞻性信息与公司未来绩效两者之间的关系越强。后续证据发现，前瞻性信息能够有效抑制公司财务信息不透明程度，尤其是公司的正向盈余操纵，从而进一步证实了前瞻性信息的有用性。

【关键词】：前瞻性；未来业绩；财务信息不透明程度；文本分析；Word2Vec 机器学习

[1] 此论文在“第十七届中国实证会计国际研讨会”会议论文版本的基础上增加了附录内容

Forward-Looking Statement and Corporate Performance

--Based on Text Mining and Machine Learning Technology

Abstract: This paper analyzes the usefulness of the forward-looking information disclosed in the annual report based on the 2011-2016 financial report of listed companies. We construct a new forward-looking information disclosure index based on the forward-looking information in the annual report environment obtained through the Word2Vec machine learning technology and the word frequency of all the words calculated by the natural language processing and text analysis. Our results show that the level of forward-looking information disclosure is positively correlated with the company's future performance. Further research indicates that the stronger the readability of the company's annual report is, the stronger the relationship between forward-looking information and the company's future performance will be. And the relationship between forward-looking information and the company's future performance is more concentrated on firms with high information asymmetry. Besides, subsequent evidence finds that forward-looking information can effectively inhibit the opacity of the company's financial information, especially the company's positive earnings manipulation, thus further confirming the usefulness of forward-looking information.

Key Words: Forward-Looking; Future Performance; Financial Information Opacity; Text Mining; Machine Learning (Word2Vec)
JEL Codes: M41, G14,D82,D21

一、引言

随着人们对企业信息需求的不断增长，以会计报告为代表的财务信息已经不能满足广大投资者的决策需求，而非财务信息日益受到人们的关注（Plumlee et al.，2008；Dhaliwal et al.，2011；程新生等, 2012）。作为以文字性的非财务信息为主的前瞻性信息，指管理层根据公司过去各方面的财务表现得出的关于公司未来的预测性信息（何卫东，2003），包括公司下一年度的行业发展趋势和竞争对手、经营计划、发展战略及机遇挑战、资金的供需情况及使用情况、公司面临的风险和不利因素等。

目前，大部分前瞻性信息存在于年报中的管理层讨论与分析章节（MD&A），其以未来发展为视角，提供了基于公认准则而产生的表内信息及报表附注无法提供的信息，是上市公司对外披露的信息当中最具有价值的部分。上市公司自愿披露的公司未来发展前景的非财务信息在相关性和及时性方面表现比财务信息更加突出，对于投资者了解公司前景、预测发展趋势以及评估未来公司价值具有极大帮助。

AICPA、FASB、CICA 等机构发布了一系列研究报告鼓励上市公司自愿披露更多的非财务信息，尤其是有关企业未来发展前景的信息。中国证监会也制定了相应法规和准则要求和鼓励上市公司披露有关未来发展前景的非财务信息。比如在《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〉（2002）》中，要求在管理层讨论与分析中不能只重复财务报告的内容，应着重于其已知的、可能导致财务报告难以显示公司未来经营成果与财务状况的重大事项和不确定性因素。而在 2007 年修订版中更是规定上市公司应在董事会报告中专门披露对公司未来发展的展望，并在 2012 年修订版中进一步细化。从总体趋势来看，有关公司未来发展前景的非财务信息在资本市场上扮演了越来越重要的角色。

但到目前为止，我国前瞻性信息相关的实证研究还是不成熟的，这虽然与前瞻性信息的发展历史较短有关，但更与前瞻性信息研究所面临的特有挑战和困难是分不开的。由于前瞻性信息是一种文本信息，对其进行量化实证研究时缺乏相关数据库的支持，面临较大的技术壁垒。大部分学者们对前瞻性信息的度量还停留在调查评级（Barron et al., 1999）或人工分析（程新生等，2015；薛爽，2010；李常青，2008）等方法，这不但严重限制了研究样本的大小，也使研究使用的原始数据受个人主观因素的影响。部分学者采用管理层分析与讨论段落或字数多少来衡量前瞻性信息，但由于 MD&A 由回顾和展望两部分信息组成，此方法会受到经营状况总结的干扰。极少数学者采用特定关键词的字词数或语句的长度来度量年报中的前瞻性信息（汪炜和袁东任，2014；Schroeder and Gibson, 1990; Li,2008; 蒋艳辉和冯楚建，2014）。然而这些特定的关键词均是基于人为定义的词表确定而成。要么通过人工阅读方式来确定词表，缺乏客观性和充分的理论支持；要么通过现有战略计划词典或英文文献词集翻译来确定词表，忽略了在中文财务报告语料环境下能代表前瞻性的大量词汇。例如：计划、预计、未来、目标、以后、前景、希望、预测、将来等通识词汇，通常会被包含在前瞻性词典或英文文献词集中。而后续、新一轮、下一步、尚需、拟向、有望等在财务报告语料环境下频频出现的前瞻性词汇，通过目前的词典和翻译是得不到的，这样便会忽略大量有用信息。

此外，上述计算关键词词频需要用到文本分析中的分词技术。中文词语博大精深，年度财务报告又是财经领域的专业化信息披露数据，存在许多专业术语。而如何对财经专业类文

本进行准确分词，一直以来都是文本挖掘的难点，如果分词系统没有足够强大的训练词库或领域优化，分词准确率会有较大程度降低，所计算出的词频也会存在误差。目前通用的中科院（北理工）分词系统、哈工大 LTP 分词系统、结巴分词系统、THULAC（清华大学）分词系统，已被广泛应用于通用词汇的识别，但他们均无法较好识别财经金融专业术语、新行业通用词汇、法律文件名称、公司名与人名等实体名等。分词结果普遍存在不统一性拆分（例如：金融资产 和 金融 负债），不当拆分（例如：按摊 余 成本 计量）和过度拆分（例如：持有 至 到期 投资）等问题。因此，对于年度财务报告这类专业化的文本分析，需要改进目前的分词系统技术。

本文首先通过 Word2Vec 机器学习技术获取在年报环境下代表前瞻性信息的词集，然后基于财经专业类文本的分词系统技术，通过自然语言处理和文本分析计算出所有词集的词频，从而构建出全新的前瞻性信息披露指标。所谓“前瞻性”指标，指“前瞻性”词汇总词频在年报文本总词频中的占比，包含计划、预计、未来、目标、可能、下半年、预期、以后、挑战、后续等 120 个前瞻性词汇。具体来讲，前瞻性指标的构建过程如下：(1) 由专业的财经文本研究团队通过阅读大量前瞻性信息披露的政策法规、研究文献以及上市公司披露的文本信息，制定与前瞻性相关的种子词集；(2) 采用机器学习技术，通过 Word2Vec 神经网络语言模型基于上下文语义信息将词汇表示成多维向量，计算向量相似度从而获得词汇的相似词，对种子词集进行词汇扩充。该模型基于海量财经文本训练而成，所推荐的相似词更加适合财经文本语境，可有效避免人为定义词表的主观性和通用同近义词工具的弱相关性；(3) 将最终得到的前瞻性指标和目前文献已有的前瞻性指标进行交叉验证，并且邀请业界和学术界专家对指标词集进行审验。

与传统衡量前瞻性信息的手工打分方式相比，词频衡量方式更具备客观性、拓展性和海量性；与笼统的计算管理层分析与讨论段落或字数多少的方式相比，词频衡量方式能更好的捕获与公司未来相关的信息，剔除过去经营情况总结的干扰；与简单的计算关键词词汇词频占比的方式相比，本文的“前瞻性”指标采用机器学习方法，获取了在财务报告语料环境下所有表示“前瞻性”的词汇，具备全面性、客观性和财经特异性。

随后本文便展开对前瞻性信息的研究，第一步必然是基于该全新的前瞻性信息披露指标，通过量化实证的方法来研究前瞻性信息的有用性。具体来讲，是探讨前瞻信息是否有助于预测企业未来一年的经营业绩。如果在预测模型中，这些信息与企业未来一年的经营业绩相关，我们称这些信息是有用的。此外本文研究了前瞻性信息披露水平与公司财务信息透明度的关系，以期探讨前瞻性信息披露是否能发挥抑制管理层盈余操纵的作用，从而期望更进一步认识前瞻性信息的有用性。

研究结果发现，前瞻性信息披露水平与公司未来绩效呈正相关关系。进一步研究表明，公司年度报告可读性越强，前瞻性信息与公司未来绩效两者之间的关系越强。公司信息不对称程度越高，前瞻性信息与公司未来绩效两者之间的关系越强。后续证据发现，前瞻性信息能够有效抑制公司财务信息不透明程度，尤其是公司的正向盈余操纵，从而进一步证实了前瞻性信息的有用性。

本文的贡献体现在以下方面：（1）以往针对前瞻性信息的研究所采用的方法存在主观性和先验性的固有缺陷，不仅不适用于海量文本数据的处理，更无法精确测量每个年度财务报告文本中真正具有的信息含量的内容。本文将文本分析和机器学习的方法引入前瞻性信息

的研究，提出全新的前瞻性信息指标，为未来该领域的研究提供了参考和借鉴。（2）本文是首次用多年海量样本对前瞻性信息有用性进行实证研究，研究结果更具备客观性、拓展性和海量性。（3）不同于以往信息有用性的研究，主要通过探讨非财务信息对公司未来绩效的作用来确认，本文还考察了其对公司财务信息透明度的影响，拓宽了前瞻性信息的有用性研究。

后续文章安排如下：第二部分介绍本文的理论基础与研究假设，第三部分介绍本文的研究设计，第四部分列示本文的实证结果，最后一部分总结研究结论。

二、 理论基础与研究假设

（一）前瞻性信息的度量方法

除了传统的人工评级方法外，国外有越来越多的学者使用字典法等文本方法对前瞻性信息进行识别和度量（Li, 2010b; Muslu et al., 2015； Bozanic et al., 2013）。而由于中文的博大精深以及文本分析等技术壁垒，国内大部分学者们对前瞻性信息的度量还停留在调查评级（Barron et al., 1999）或人工分析（程新生等，2015；薛爽等，2010；李常青等 2008）等方法，这不但严重限制了研究样本的大小，也使研究使用的原始数据受个人主观因素的影响。少数学者采用特定关键词的字词数或语句的长度来度量年报中的前瞻性信息（汪炜和袁东任，2014；Schroeder and Gibson,1990; Li,2008; 蒋艳辉和冯楚建，2014）。然而这些特定的关键词均是基于人为定义的词表确定而成，要么通过人工阅读方式来确定词表，缺乏客观性和充分的理论支持；要么通过现有战略计划词典或英文文献词集翻译来确定词表，忽略了在中文财务报告语料环境下能代表前瞻性的大量词汇。例如：计划、预计、未来、目标、以后、前景、希望、预测、将来等通识词汇，通常会被包含在前瞻性词典或英文文献词集中。而后续、新一轮、下一步、尚需、拟向、有望等在财务报告语料环境下频频出现的前瞻性词汇，通过目前的词典和翻译是得不到的，这样便会忽略大量有用信息。

（二）前瞻性信息的有用性

根据信息不对称理论，前瞻性信息以未来发展为视角，提供了基于公认准则而产生的表内信息及报表附注无法提供的信息，因此它应该能在某种程度上缓解公司的信息不对称程度，有助于投资者了解公司现状并预测公司未来业绩。继 Copeland（1978）强调了叙述性信息对投资者决策的重要性之后，很多学者针对包含大量的前瞻性信息的管理层讨论与分析的内容进行了研究。国外文献对 MD&A 有用性进行实证研究，一般通过两条途径：一是研究 MD&A 是否有助于预测公司长期业绩；二是研究市场对 MD&A 披露的反应。如果 MD&A 确实有助于对公司未来业绩进行预测，那我们就可以认为它是一种有用的信息披露。

国外研究发现，前瞻性信息为外部投资者提供了解企业现状和预测未来的有用信息，在一定程度上可以缓解投资者、分析师等外部信息使用者与上市公司之间的信息不对称。（Bryan，1997； Muslu et al.，2014）。具体来讲前瞻性信息有助于信息使用者获得预测未来的增量信息，不仅能够帮助投资者更好地预测公司未来业绩（Cole and Jones, 2004）和股

票价格（Eli and Baruch，1996；Cole and Jones, 2004；Francis et al., 2003；Behn et al., 1996），而且有助于提高分析师预测的准确性，降低预测的分歧度（Clarkson etal., 1999；Barron et al., 1999），提高分析师预测的准确度（Bozzolan et al., 2009）。更进一步地，MD&A 中的前瞻性信息具有较强的预测性作用，有助于预测公司下期存货变动情况（Sun，2010）和破产风险（Mayew et al., 2014）。Li (2010b) 研究发现，MD&A中前瞻性描述的语调与公司未来盈利以及流动性均正相关。

国内针对前瞻性信息的实证研究仍较少，但有关此类信息的重要性已逐渐收到国内学者的重视。但由于我国引进MD&A制度的时间尚短，披露制度尚不完善，加上相关数据的难以获取限制了实证研究的开展。而且研究样本、方法的限制，现有文献在前瞻性信息的有用性这方面的研究结果上分歧比较明显。如，李峰森、李常青（2008）探讨了MD&A 信息的有用性。他们的研究表明，我国MD&A 信息总体上对预测公司未来收入、每股盈余和经营现金流量的变化有显著的辅助作用，而且股票市场也对此做出了及时迅速的反应。这是中国目前考察MD&A 信息有用性的第一篇文章，然而，作者的样本数据仅为部分2004 年的半年报，而且是采取随机抽取的办法，因此样本具有较大的异质性。薛爽、肖泽忠和潘妙丽（2010）通过对前瞻性信息采用手动评分法，专门研究了亏损上市公司的MD&A 信息披露的有用性，研究发现MD&A 中关于亏损原因的分析和下一年度的战略部署可以为投资者提供关于企业未来经营业绩的增量信息。蒋艳辉和冯楚建 (2014) 采用特定关键词的字词数或语句的长度来度量年报中的前瞻性信息，通过对多个指标进行研究也得到前瞻性深度与未来财务业绩正相关。但李常青等研究也发现，我国MD&A也存在明显的“报喜不报忧”倾向，好消息的数量大致为坏消息的2至3倍，有不少公司明明经营状况不佳，但如果只看MD&A，却会觉得公司形势还不错。

国外虽然有较多关于前瞻性信息的研究，其结论能否适用于中国这样一个发展中的资本市场，显然不能先知先觉地妄下结论。而国内有关前瞻性信息有用性的实证研究，多是基于人工调查评级或人工分析等方法，样本具有较大的异质性，且基于中国当时MD&A 信息披露质量不高的状况，针对所有项目的考察可能难以获得统一的结论。而随着我国披露制度的不断完善，资本市场的不断健全，近年来前瞻性信息披露的质量理论上应当远高于之前。尤其是在2012年初，上交所发布《编制要求》对2011 年上市公司MD&A 信息披露提出了更具体的要求，如突出公司个性、简洁易懂等，并提出要加强审核，因而2011 年以来的MD&A 披露理论上应当比往年披露的质量有所提高。这为我们的研究提供了基础。

根据自愿披露动机中的资本市场交易假说，公司和外部投资者之间存在信息不对称，投资者会进行逆向选择，即将隐藏信息视为坏消息，并对公司资产进行折价。管理层为了避免“次品车”市场条件下的“价值折价”，缓解信息不对称，有动力去披露更多信息来突出自己的竞争优势和向投资者表达未来良好的发展前景从而达到与其他类型公司相分离的目的（乔旭东，2003；张宗新等2005）。当管理层预计未来业绩好时，他们对公司真实的经营情况充满信心，也更担心投资者进行逆向选择，于是有强烈的动机向信息使用者提供更多准确的公司未来经营和管理状况，客观公正的对公司经营情况进行分析与预计，使信息使用者能更好的预计公司未来业绩发展趋势；反之，当预计未来业绩差时，管理层对未来发展持保守态度，没有强烈的动机向信息使用者传递未来经营状况（蒋艳辉和冯楚建，2014）。因此前瞻性信息披露水平越高，越能代表管理层对未来业绩的看好，那么前瞻性信息披露水平与公

司未来一期的财务绩效应呈正相关关系。故本文提出以下假设1A。然而，价值较低的公司管理者也可能会出于迷惑投资者的目的，披露更多的信息（何卫东，2003；唐跃军等，2008)；拥有更多关系网的价值越高的公司也可能为了防止专有成本，披露较少的信息（程新生等，2012）。当这两类公司占比较多时，公司管理者会披露更多数量但是质量较差的信息。那么所披露的前瞻性信息则不能够有效预测上市公司未来一期的财务绩效，前瞻性信息的有用性不足。故本文提出竞争假设1B。

H1A: 前瞻性信息能够预测上市公司未来一期的财务绩效。前瞻性信息披露水平越高，公司未来一期的财务绩效越高。

H1B: 前瞻性信息不能够预测上市公司未来一期的财务绩效。前瞻性信息披露水平与公司未来一期的财务绩效不相关。

在假设 1A 成立的条件下：预计未来业绩较好的管理层为防止逆向选择会披露较多的前瞻性信息。而当信息不对称程度越高时，投资者的信息需求越强，逆向选择的可能性越强，对公司未来发展有信心的管理层越有动机去披露更多的高质量的信息，包括前瞻性信息。而这些较高质量的前瞻性信息，理应对上市公司未来一年财务绩效的预测能力越强。

H2: 信息不对称程度越高时，前瞻性信息对上市公司未来一年财务绩效的预测能力越强。

在假设 1A 成立的条件下：作为文本信息，前瞻性信息本质上属于自愿披露的信息，管理层拥有较高的自由度，因此其信息揭示作用更多地受到可读性等因素的影响。当企业存在困难和问题时，管理层可能存在刻意使用晦涩难懂的语言增大阅读难度，从而掩盖企业所面临问题的主观动因（程新生等，2015；Lo et al.，2017）。反而年报可读性越高时，越能体现管理者欲提供更多有用的信息而非为隐藏坏消息模糊绩效，越能体现出管理层可信度。故在此情况下，前瞻性信息质量越高，理应对上市公司未来一年财务绩效的预测能力越强。

H3: 文本可读性越高时，前瞻性信息对上市公司未来一年财务绩效的预测能力越强。

基于信息不对称理论，前瞻性信息以未来发展为视角，提供了基于公认准则而产生的表内信息及报表附注无法提供的信息，增加了管理层隐藏负面信息以及盈余操控的难度，从而会降低公司财务信息的不透明度。借鉴 Hutton et al.(2009)和周晓苏等（2016）的研究，我们采用修正 Jones 模型计算的可操纵应计利润的绝对值（AB_DA）衡量公司财务信息的不透明程度，AB_DA 越高，公司财务信息越不透明。

H4: 前瞻性信息披露水平可有效降低公司财务信息的不透明度。

三、 研究设计

（一）样本选择和数据来源

鉴于 2012 年初，上交所发布《编制要求》对 2011 年上市公司包含大量前瞻性内容的 MD&A 信息披露提出了更具体的要求，如突出公司个性、简洁易懂等，并提出要加强审核，因而 2011 年以来的前瞻性信息披露理论上应当比往年披露的质量有所提高。故本文选择了深、沪两市 2011-2016 年间所有上市公司年度财务报告作为初始研究样本。本文所使用的上市公司年度报告来自于上交所、深交所以及巨潮资讯网，其他研究数据均来自来源于 CSMAR 数据库。

数据处理步骤如下：

（1）从上交所、深交所及巨潮资讯网上下载 2011-2016 年所有 A 股上市公司年度报告。

（2）借助 WinGo 财经文本数据平台，将 PDF 文档转化成 txt 文档，并对数据进行以下清理：1）表格问题的处理。年报中的表格在转换后变为文本框，从而无法直接对其内容进行查询或分析，且由于表格中内容大多为数字，少数文字内容模板化严重，所含增量前瞻性文本信息较少。故文本识别并剔除了所有表格，然后进行分析。2）页眉页脚的清理。年报中页眉的内容基本为“XX 公司 XX 年度报告”，页脚的内容大多为数字、“第 XX 页”或“XX 页”等，因此文本识别并剔除了这类页眉页脚。3）剔除扫描文件和缺失文件。

（3）采用 WinGo 财经文本分词系统，对文档内容进行分词等文本处理，从而将文本的非结构化数据结构化成词向量进行存储。

（4）通过 R 语言计算前瞻性指标词集的词频。

（5）从 CSMAR 中下载 2011-2016 年所有 A 股上市公司的其他研究数据。

（6）剔除金融保险类企业、ST 和*ST 类企业、以及不超过两年的观测值，得到 14890 个样本。

（7）剔除所有缺失观测，为避免极端值影响，本文对回归模型连续变量进行上下 1%缩尾处理，模型 1 最终得到 10599 个样本，模型 2 最终得到 12555 个样本。

（8）为了控制误差项的异方差和时间序列相关问题对估计系数标准误的影响，我们的回归分析采用稳健的标准误（Robust），并将误差在公司层面聚类（Cluster）。

（二）模型设计与变量衡量

1. 前瞻性信息披露指标的衡量

所谓“前瞻性”指标，指“前瞻性”词汇总词频在年报文本总词频中的占比，包含计划、预计、未来、目标、可能、下半年、预期、以后、挑战、后续等 120 个前瞻性词汇（见附录）。本文首先通过 Word2Vec 机器学习技术获取在年报环境下代表前瞻性信息的词集，然后基于财经专业类文本的分词系统技术，通过自然语言处理和文本分析计算出所有词集的词频，从而构建出全新的前瞻性信息披露指标。具体来讲，前瞻性指标的构建过程如下：

(1) 借鉴 Muslu et al.(2014)和 Li(2010）中衡量前瞻性信息的词集，并结合中国前瞻性信

息披露的政策法规以及上市公司披露的中文文本信息特点，本文制定出衡量中文年度报告中前瞻性信息的种子词集，包括：计划、预计、未来、目标、可能、如果、机遇、预期、挑战、预测、今后、目的、契机、前景、希望、展望、相信、愿景、期待、明年、期望、打算、来年。

(2) 采用 Word2Vec 机器学习技术，先后采用 Skip-gram 模型（Continuous Skip-gram model）和 CBOW 模型（Continuous Bag-of-Words Model）基于上下文语义信息将年度报告的词汇表示成多维向量，计算向量相似度，从而获得上述种子词汇在财务报告环境下的相似词，对种子词集进行词汇扩充。根据结果及效率评估，最终选取基于 CBOW 模型的相似词结果。该技术本质是基于神经网络来完成的 Word Embedding 方法，根据海量财经文本训练而成，所推荐的相似词更加适合财经文本语境，可有效避免人为定义词表的主观性和通用同近义词工具的弱相关性。

Skip-gram 模型：

$$\max \prod_{w \in C} \prod_{u \in Content(w)} p(u|w)$$

其中，上式等价于

$$\max \sum_{w \in C} \sum_{u \in Content(w)} \log p(u|w)$$

CBOW 模型：

$$\max \sum_{w \in C} \log p(w|Context(w))$$

其中，C 表示语料；w 表示中心词；Content(w)表示中心词的上下文。CBOW 模型是在已知当期词上下文的前提下预测当前词；Skip-gram 模型是在已知当前词的前提下预测其上下文。

(3) 将最终得到的前瞻性指标和目前文献已有的定量前瞻性指标（管理层分析与讨论段落多少）进行交叉验证，并且邀请业界和学术界专家对指标词集进行审验。确定最终前瞻性词集为 120 个。

(4) 计算前瞻性词汇总词频占年度报告全文总词频的比例，即得到前瞻性信息披露指标。该指标越大，表明公司的前瞻性信息披露水平越高。

$$Forward\_Index_{(i,t)} = \frac{120\text{个前瞻性词汇总词频}_{(i,t)}}{\text{年度报告全文总词频}_{(i,t)}}$$

2. 财务信息的不透明程度的衡量

借鉴 Hutton et al.(2009) 和周晓苏等（2016）的研究，我们采用修正 Jones 模型计算的可操纵应计利润的绝对值（AB_DA1）衡量公司财务信息的不透明程度，AB_DA1 越高，公司财务信息越不透明。同时，我们还采用 Jones 模型计算的可操纵应计利润的绝对值（AB_DA2）和 Kothari (2005) 提出的业绩修正的琼斯模型计算的可操纵应急利润的绝对值（AB_DA3）作为稳健性检验。

3. 模型设计

借鉴 Davis et al. (2012)、Tetlock et al. (2008) 、Davis and Tran (2012)、陈小悦和徐晓东（2001）等文献，建立了模型（1）

$$\begin{aligned}\mathrm{ROE}_{i,t+1} = {} & \alpha_0 + \alpha_1 \mathrm{Forward}_{\mathrm{Index}_{it}} + \alpha_2 \mathrm{ROE}_{it} + \alpha_3 \mathrm{SIZE}_{it} + \alpha_4 \mathrm{AGE}_{it} + \alpha_5 \mathrm{LEV}_{it} + \alpha_6 \mathrm{BM}_{it} \\ & + \alpha_7 \mathrm{Growth}_{it} + \alpha_8 \mathrm{Loss}_{it} + \alpha_9 \mathrm{FSHR}_{it} + \alpha_{10} \mathrm{Manager}_{it} + \alpha_{11} \mathrm{Nation}_{it} \\ & + \alpha_{12} \mathrm{CEO}_{it} + \alpha_{13} \mathrm{YRET}_{it} + \alpha_{14} \mathrm{Volatility}_{it} + \sum \mathrm{Year} + \sum \mathrm{Industry} + \varepsilon_{it}\end{aligned} \tag{1}$$

模型（1）用于检验研究假设 1A 和 1B。因变量是 T+1 公司业绩（ROE 或 ROA），解释变量为 T 年年报中前瞻性披露水平（Forward_Index）。若 Forward_Index 前面的系数显著为正，则研究假设 1A 得以通过检验；控制变量选取 T 年的：业绩（ROE 或 ROA）、公司规模（Size）、上市年限（Age）、财务杠杆（LEV）、市账比（MTB）、总资产增长率（Growth）、亏损状况（Loss）、股权集中度（FSHR）、管理者持股（Manager）、国有股股东持股（Nation）、二职合一（CEO）、股票收益率（YRET）、股票波动率（Volatility）。

$$\begin{aligned}\mathrm{AB\_DA}_{i,t} = {} & \alpha_0 + \alpha_1 \mathrm{Forward}_{\mathrm{Index}_{it}} + \alpha_2 \mathrm{ROE}_{it} + \alpha_3 \mathrm{SIZE}_{it} + \alpha_4 \mathrm{AGE}_{it} + \alpha_5 \mathrm{LEV}_{it} + \alpha_6 \mathrm{BM}_{it} \\ & + \alpha_7 \mathrm{Growth}_{it} + \alpha_8 \mathrm{Loss}_{it} + \alpha_9 \mathrm{FSHR}_{it} + \alpha_{10} \mathrm{Manager}_{it} + \alpha_{11} \mathrm{Nation}_{it} \\ & + \alpha_{12} \mathrm{CEO}_{it} + \alpha_{13} \mathrm{YRET}_{it} + \alpha_{14} \mathrm{Volatility}_{it} + \sum \mathrm{Year} + \sum \mathrm{Industry} + \varepsilon_{it}\end{aligned} \tag{2}$$

模型（2）用于检验研究假设 4。因变量是公司当年财务信息的不透明程度，借鉴 Hutton et al.(2009)和周晓苏等（2016）的研究，我们采用修正 Jones 模型计算的可操纵应计利润的绝对值（AB_DA1）衡量公司财务信息的不透明程度，AB_DA1 越高，公司财务信息越不透明。同时，我们还采用 Jones 模型计算的可操纵应计利润的绝对值（AB_DA2）和 Kothari (2005) 提出的业绩修正的琼斯模型计算的可操纵应急利润的绝对值（AB_DA3）作为稳健性检验。解释变量为公司当年年报中前瞻性披露水平（Forward_Index）。若 Forward_Index 前面的系数显著为负，表明公司的前瞻性信息披露水平越高，公司进行盈余操纵的程度越低，从而财务信息的不透明程度越高，则研究假说 H4 得以通过检验。

表 1：变量定义表

变量符号	变量名	变量注释
Panel A. 公司绩效与财务信息的不透明程度（因变量）		
ROE	净资产收益率	公司净利润与该年年末股东权益余额的比值
ROA	总资产收益率	公司净利润与该年年末总资产余额的比值

AB_DA1	可操纵应计利润的绝对值（基于修正 Jones 模型）	根据修正 Jones 模型计算的可操纵应计利润的绝对值来衡量公司财务信息的不透明程度（基于资产负债表计算总应计利润，模型不包含截距，取残差作为操纵性应计利润）
AB_DA2	可操纵应计利润的绝对值（基于 Jones 模型）	根据 Jones 模型计算的可操纵应计利润的绝对值来衡量公司财务信息的不透明程度（基于资产负债表计算总应计利润，模型不包含截距）
AB_DA3	可操纵应计利润的绝对值（基于业绩修正 Jones 模型）	根据业绩修正的 Jones 模型计算的可操纵应计利润的绝对值来衡量公司财务信息的不透明程度（基于资产负债表计算总应计利润，修正的琼斯模型中包含 ROA 与截距）
Panel B. 前瞻性信息披露水平（自变量）		
Forward_Index	前瞻性披露水平	财报中 120 个前瞻性词汇总词频占年度报告全文总词频的比例
#Word	年报总词频	财务报告中分词后的总词数
#Sentence	年报总句频	财务报告全文包含的句子总数
#Rep_size	年报大小	财务报告文件大小
#Forward_count	前瞻性信息词频	财务报告中分词后 120 个前瞻性词汇总词频
Panel C. 控制变量		
Size	公司规模	公司年末总资产的自然对数
Age	上市年限	Ln（公司上市的年数+1）
LEV	财务杠杆	公司年末总负债与总资产比值
BM	账面市值比	公司年末账面价值与市场价值的比值
Growth	总资产增长率	公司本期总资产增长额与期初总资产的比值
Loss	亏损状况	虚拟变量，当公司上一年度净利润为负时，Loss 为 1，否则为 0
FSHR	股权集中度	第一大股东持股份数占总股数的百分比
Manager	管理者持股	公司年末管理层持股份数占总股数的百分比
Nation	国有股股东持股比例	公司年末国有股股东持股份数占总股数的百分比
CEO	二职合一	虚拟变量，当 CEO 与董事长同一人时，则取值为 1；否则为 0
YRET	股票收益率	不考虑现金红利再投资的年个股回报率
Volatility	股票波动率	最近 250 个交易日对数收益率估计出来的波动率
Panel E. 其他变量		
IO	机构持股比例	公司机构持股份数占总股数的百分比
MTB	市值账面比	公司市场价值与年末账面价值的比值
RD	研发费用	公司年末研发费用
Readability	可读性	从文本构成上，基于神经概率语言模型提出的顺序简易型大小 $\text{order simplicity} = \frac{1}{\mathrm{N}}\sum_{s=1}^{N} logp_s$

（三）样本描述

各变量的描述性统计结果见表 2。为了避免异常值的影响，文本在进行描述性统计和回归分析前，进行了 1%和 99%水平上的缩尾处理。从表 2 中可以看出，公司绩效（ROA 和 ROE）的均值分别是 0.063 和 0.037，标准差分别为 0.112 和 0.052。在前瞻性信息披露水平中，无论是前瞻性指标（Forward_Index）、 财务报告中分词后 120 个前瞻性词汇总词频（#Forward_count）、财务报告中分词后的总词数（#Word）都存在比较大的差异，保证了充分的变异性。此外，图 1 展示了所有样本前瞻性指标（Forward_Index）的分布，其服从正态分布，这为本文的计量分析提供了基础，可以继续进行下一步研究。图 2 展示了财报和前瞻性信息的描述性统计图，可以看到前瞻性词频和前瞻性词频占比（Forward_Index）呈现逐年上升状态，这反应了前瞻性信息确实受到大家的越来越多的重视。

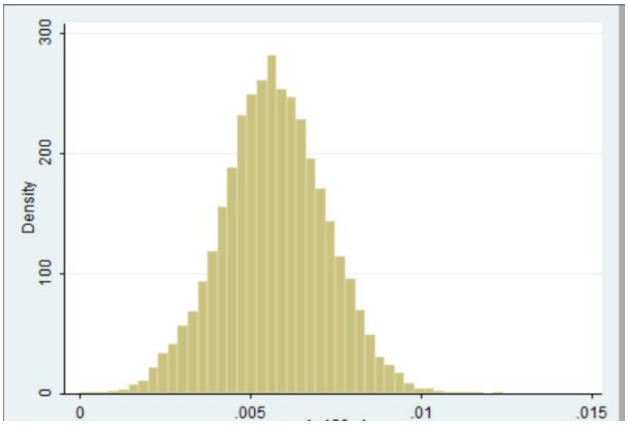

图 1 前瞻性指标的分布状态

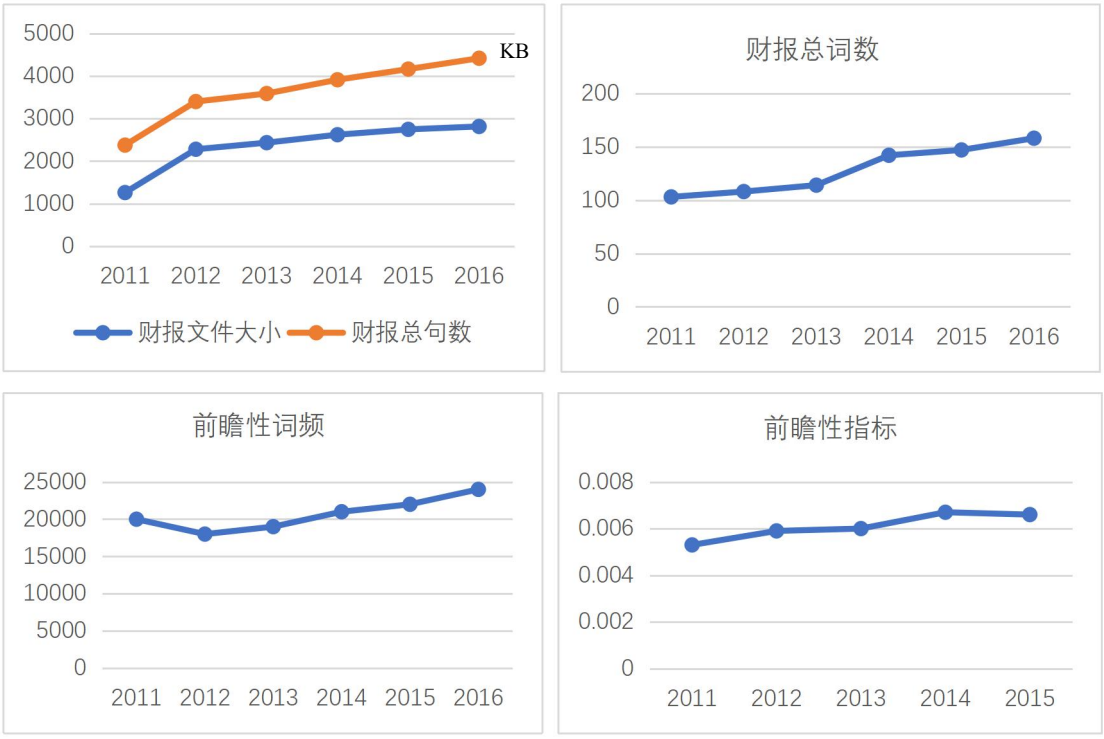

图 2 财报与前瞻性信息描述统计图

表 2：描述性统计表

变量	样本量	均值	标准差	最小值	中位数	最大值
ROE	10599	0.063	0.112	-0.558	0.066	0.354
ROA	10599	0.037	0.052	-0.175	0.034	0.195
Forward_Index	10599	0.006	0.001	0.003	0.006	0.009
#Word	10599	46412	11330	176	45032	115193
#Sentence	10599	20763	5065	11243	20122	36555
#Rep_size	10599	1244	257	732	1223	1944
#Forward_count	10599	2475	1311	586	2452	7883
Size	10599	130	43	55	123	261
Age	10599	22.000	1.263	19.370	21.830	25.830
LEV	10599	15.520	5.331	4.011	15.350	28.530
BM	10599	0.441	0.220	0.046	0.437	0.940
Growth	10599	0.959	0.942	0.077	0.644	5.275
RAR	10599	0.154	0.263	-0.270	0.095	1.574
Loss	10599	0.097	0.296	0.000	0.000	1.000
FSHR	10599	0.357	0.152	0.088	0.338	0.754
Manager	10599	0.041	0.124	0.000	0.000	0.643
Nation	10599	0.091	0.169	0.000	0.000	0.656
CEO	10599	0.243	0.429	0.000	0.000	1.000
YRET	10599	0.261	0.575	-0.549	0.147	2.499
Volatility	10599	0.470	0.152	0.241	0.425	0.886
IO	10599	0.058	0.083	0.000	0.033	0.529
MTB	10599	2.174	2.115	0.190	1.553	13.020
RD	10599	0.014	0.017	0.000	0.009	0.085
Readability	10599	-26.660	2.907	-34.900	-26.550	-20.540
AB_DA1	12555	0.132	0.13	0.001	0.091	0.619
AB_DA2	12555	0.132	0.13	0.002	0.092	0.621
AB_DA3	12555	0.134	0.134	0.001	0.091	0.661

表 3：主要变量相关系数

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
（1）ROE(t+1)	1												
（2）ROA(t+1)	0.825	1											
（3）Forward_Index	0.023	0.032	1										
（4）Size	0.066	-0.030	0.093	1									
（5）Age	-0.004	-0.070	0.097	0.134	1								
（6）LEV	-0.076	-0.305	-0.009	0.486	0.260	1							
（7）BM	-0.116	-0.248	-0.038	0.642	0.122	0.582	1						
（8）Growth	0.131	0.137	-0.019	0.086	-0.056	0.01	-0.050	1					
（9）Loss	-0.158	-0.213	0.01	-0.068	0.059	0.204	0.058	-0.186	1				
（10）YRET	0.079	0.105	0.198	-0.059	0.113	-0.059	-0.274	0.205	-0.029	1			
（11）Volatility	-0.009	-0.001	0.139	-0.110	0.079	-0.073	-0.307	0.144	0.051	0.534	1		
（12）AB_DA1	0.028	0.066	-0.089	-0.249	-0.121	-0.221	-0.188	0.153	-0.021	-0.037	0.038	1	
（13）AB_DA2	0.028	0.066	-0.089	-0.250	-0.119	-0.220	-0.190	0.154	-0.016	-0.036	0.038	0.991	1
（14）AB_DA3	0.018	0.017	-0.024	-0.025	0.004	-0.011	-0.021	0.031	-0.011	0.017	0.026	0.142	0.148

四、 实证结果

（一）前瞻性信息披露水平和公司未来绩效

利用模型（1），本文对假设 1A 和 1B 进行了检验，考察前瞻性信息披露对公司避税行为的影响，检验结果如表 4 所示：1）前瞻性指标对公司未来绩效 ROA 的回归系数是 0.889，在 5%以下的置信度下回归结果显著（t 值=2.58）；2）前瞻性指标对公司未来绩效 ROE 的回归系数是 1.888，在 5%以下的置信度下回归结果显著（t 值=2.19）。因此可以接受假设 1A：前瞻性信息披露水平与上市公司未来绩效具有正相关关系。从而证实了前瞻性信息能在某种程度上缓解公司的信息不对称程度，有助于投资者了解公司现状并预测公司未来业绩。

在控制变量中，公司本期资产报酬率（ROA）、总资产增长率（Growth）、个股回报率（YRET）、股权集中度（FSHR）、国有股股东持股比例（Nation）等于公司未来绩效显著正相关。而公司本期资产负债率（LEV）、账面市值比（BM）、股票波动性（Volatility）等与公司未来绩效显著负相关。这些结果与前人的研究结论保持一致。

表 4：前瞻性信息披露指标和公司绩效

	(1)	(2)
Dependent	**ROA(t+1)**	**ROE(t+1)**
Forward_Index	0.889**	1.888**
	(2.58)	(2.19)
ROA	0.544***	0.785***
	(23.69)	(17.30)
Size	0.005***	0.013***
	(7.38)	(7.51)
Age	0.000	0.000**
	(1.51)	(2.19)
LEV	-0.020***	0.036***
	(-5.52)	(3.93)
BM	-0.008***	-0.022***
	(-9.96)	(-8.56)
Growth	0.007***	0.020***
	(4.09)	(4.52)
Loss	0.036***	0.048***
	(14.98)	(7.50)
FSHR	0.015***	0.032***
	(5.00)	(4.47)
Manager	0.002	-0.003
	(0.54)	(-0.31)
Nation	0.014***	0.026***
	(4.75)	(4.54)
CEO	0.000	0.000
	(0.22)	(0.19)
YRET	0.008***	0.016***
	(7.88)	(6.79)
Volatility	-0.035***	-0.073***
	(-5.96)	(-4.99)
Constant	-0.090***	-0.279***
	(-5.57)	(-6.89)
Industry Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Observations	10599	10599
Adjusted R^2	0.373	0.197

注：括号内为 t 值，估计采用稳健标准误，误差项在公司层面聚类。*、**、***分别表示显著性水平(双尾)为 10%、5%、1%

（二）前瞻性信息披露与公司绩效：信息不对称的影响

为验证假设 2，本文借鉴（Chen et al., 2017; Ajinkya et al., 2005; Smith and Watts 1992; Aboody and Lev 2000; 杨海燕等，2012），选取机构持股比例、公司市值账面比、公司研发费用作为信息不对称的代理变量。

1. 机构持股比例的影响

机构持股比例很大程度能影响公司的信息不对称程度：一方面，机构投资者比例较高时，迫于机构投资者的压力，上市公司不得不提高信息披露质量（Chen et al., 2017）。另一方面，作为一类重要的外部投资者，机构投资者具有普通投资者所不具备的信息收集和处理能力，而由于其他投资者对机构投资者持股的跟踪和模仿，使得公司信息产生了溢出效应，有助于公司信息透明度的提高（Ajinkya et al., 2005; 杨海燕等，2012）。

本文以每年度公司的机构持股比例（IO）的中位数为界限，将样本分为低机构持股比例和高机构持股比例样本组，并采用模型（1）进行分组回归。回归结果见表 5。表 5 中第（1）、（3）列为高机构持股比例样本组的回归结果，第（2）、（4）列为低机构持股比例样本组的回归结果。通过比较第（1）、（2）列中前瞻性信息（Forward_Index）的回归系数的显著性可以看到，对于机构持股比例较少的公司，前瞻性信息能够显著预测未来一年公司的业绩；而对于机构持股比例较多的公司，前瞻性信息对公司未来业绩并没有显著影响。该结果与本文的分析一致，说明机构持股比例较低的公司，信息不对称程度越高，故投资者的信息需求越强，披露信息的估值作用越显著，管理层有动机去披露更多的质量较高的信息，包括前瞻性信息。而这些质量较高的前瞻性信息，对上市公司未来一年财务绩效的预测能力越强。

表 5：前瞻性信息披露与公司绩效：机构持股比例的影响

	(1)	**(2)**	**(3)**	**(4)**
Dependent	**ROA(t+1)**	**ROA(t+1)**	**ROE(t+1)**	**ROE(t+1)**
IO	**High**	**Low**	**High**	**Low**
Forward_Index	0.263	1.171**	0.570	2.432*
	(0.68)	(2.06)	(0.64)	(1.69)
Constant	-0.069***	-0.068***	-0.262***	-0.181***
	(-3.19)	(-3.01)	(-4.79)	(-3.23)
Control	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	5299	5300	5299	5300
Adjusted R^2	0.483	0.257	0.299	0.118
Test: p-value	0.000		0.000	

注：括号内为 t 值，估计采用稳健标准误，误差项在公司层面聚类。*、**、***分别表示显著性水平(双尾)为 10%、5%、1%

2. 公司市值账面比

公司市值账面比也可作为信息不对称的代理变量(Smith and Watts 1992; Aboody and Lev 2000)。本文以每年度公司的市值账面比（MTB）的中位数为界限，将样本分为低市值账面比和高市值账面比样本组，并采用模型（1）进行分组回归。回归结果见表 6。表 6 中第（1）、（3）列为高市值账面比样本组的回归结果，第（2）、（4）列为低市值账面比样本组的回归结果。通过比较第（1）、（2）列中前瞻性信息（Forward_Index）的回归系数的显著性可以看到，对于市值账面比较高的公司，前瞻性信息能够显著预测未来一年公司的业绩；而对于市值账面比较低的公司，前瞻性信息对公司未来业绩并没有显著影响。该结果与本文的分析一致，说明市值账面比较高的公司，信息不对称程度越高，为了防止逆向选择，管理层有动机去披露更多的质量较高的信息，质量越高的前瞻性信息，对上市公司未来一年财务绩效的预测能力越强。

表 6：前瞻性信息披露与公司绩效：公司账面市值比的影响

	(1)	**(2)**	**(3)**	**(4)**
Dependent	**ROA(t+1)**	**ROA(t+1)**	**ROE(t+1)**	**ROE(t+1)**
MTB	**High**	**Low**	**High**	**Low**
Forward_Index	1.329**	0.605	2.843***	1.164
	(2.47)	(1.42)	(2.69)	(0.89)
Constant	-0.190***	-0.079***	-0.190***	-0.079***
	(-6.81)	(-4.51)	(-6.81)	(-4.51)
Control	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	5299	5300	5299	5300
Adjusted R^2	0.369	0.328	0.369	0.328
Test: p-value	0.000		0.000	

注：括号内为 t 值，估计采用稳健标准误，误差项在公司层面聚类。*、**、***分别表示显著性水平(双尾)为 10%、5%、1%

3. 研发费用

研发费用亦可代表信息不对称程度（Smith and Watts 1992; Aboody and Lev 2000）。本文以每年度公司的研发费用（RD）的中位数为界限，将样本分为低研发费用和高研发费用样本组，并采用模型（1）进行分组回归。回归结果见表 7。表 7 中第（1）、（3）列为高研发费用样本组的回归结果，第（2）、（4）列为低研发费用样本组的回归结果。通过比较第（1）、（2）列中前瞻性信息（Forward_Index）的回归系数的显著性可以看到，对于研发费用比较高的公司，前瞻性信息能够显著预测未来一年公司的业绩；而对于研发费用比较低的公司，前瞻性信息对公司未来业绩并没有显著影响。该结果与本文的分析一致，说明研

发费用比较高的公司，信息不对称程度越高，同上文所述，此时前瞻性信息对上市公司未来一年财务绩效的预测能力越强。

表 7：前瞻性信息披露与公司绩效：公司研发费用的影响

	(1)	(2)	(3)	(4)
Dependent	**ROA(t+1)**	**ROA(t+1)**	**ROE(t+1)**	**ROE(t+1)**
RD	**High**	**Low**	**High**	**Low**
Forward_Index	1.053***	0.371	2.556***	0.560
	(2.69)	(0.70)	(2.93)	(0.41)
Constant	-0.053***	-0.105***	-0.211***	-0.309***
	(-2.94)	(-4.84)	(-4.98)	(-5.57)
Control	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	4915	5684	4915	5684
Adjusted R^2	0.523	0.262	0.341	0.143
Test: p-value	0.000		0.000	

注：括号内为 t 值，估计采用稳健标准误，误差项在公司层面聚类。*、**、***分别表示显著性水平(双尾)为 10%、5%、1%

（三）前瞻性信息披露指标与公司绩效：年报可读性的影响

可读性是用来衡量人们理解文本内容的程度。文本的可读性越高，表示文本越容易被理解；文本的可读性越低，表示文本越不容易被理解。目前，比较常用的衡量文本可读性的指标如 Flesch reading-ease score，Flesch-Kincaid grade level，SMOG index，Gunning fog index 等都是针对英文语料的，由于语言的差异以及指标构建时针对具体语言考虑的特征等因素的影响，使得这些指标不能够直接的被用于衡量中文文本的可读性。一些学者通过研究中文语料的特点，考虑将文本的总字数，平均句长以及专业术语密度作为主要考虑因素来衡量中文文本的可读性，但由于没有统一的术语词典以及词典中含有许多并不会影响可读性的词等因素的影响，使得这种衡量中文可读性的方式不具有普适性。

为验证假设 3，本文从文本构成上，基于神经概率语言模型提出的顺序简易型大小定义可读性，计算公式为：

$$\text{Readability} = \frac{1}{\text{N}}\sum_{s=1}^{N} logp_s$$

其中，Ps 表示句子 s 生成的概率，N 表示构成文本的句子数。即将语料分词后用向量表示，计算单词序列（句子）出现的概率，概率越高，表明句子是由一些常见的词对组合而成，表示词对越被人们所熟知，可读性越好。而概率越低表示句子是由一些不常见的词对组合而成，表示这样的词对搭配不常用，代表越不被人们所熟知，可读性越差。

本文以每年度公司财务报告可读性的中位数为界限，将样本分为低可读性和高可读性样本组，并采用模型（1）进行分组回归。回归结果见表 8。表 8 中第（1）、（3）列为高财报可读性用样本组的回归结果，第（2）、（4）列为研发费用样本组的回归结果。通过比较

第（1）、（2）列中前瞻性信息（Forward_Index）的回归系数的显著性可以看到，对于可读性比较高的公司，前瞻性信息能够显著预测未来一年公司的业绩；而对于可读性比较低的公司，前瞻性信息对公司未来业绩并没有显著影响。该结果与本文的分析一致，说明可读性比较高时，越能体现管理者愿意提供更多有用的信息而非为隐藏坏消息模糊绩效，管理层可信度越高。在此情况下，前瞻性信息质量较高，对上市公司未来一年财务绩效的预测能力越强。

表 8：前瞻性信息披露与公司绩效：年报可读性的影响

	(1)	(2)	(3)	(4)
Dependent	**ROA(t+1)**	**ROA(t+1)**	**ROE(t+1)**	**ROE(t+1)**
Readability	**High**	**Low**	**High**	**Low**
Forward_Index	1.739***	0.104	3.029**	1.127
	(3.38)	(0.22)	(2.54)	(0.98)
Constant	-0.091***	-0.100***	-0.271***	-0.292***
	(-3.33)	(-5.09)	(-4.19)	(-5.77)
Control	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	5299	5300	5299	5300
Adjusted R^2	0.361	0.392	0.189	0.205
Test: p-value	0.078		0.383	

注：括号内为 t 值，估计采用稳健标准误，误差项在公司层面聚类。*、**、***分别表示显著性水平(双尾)为 10%、5%、1%

（四）前瞻性信息披露与公司不透明程度

利用模型（2），本文对假设 4 进行了检验，考察前瞻性信息披露对公司财务信息不透明程度的影响，检验结果如表 9 所示：1）前瞻性指标对可操纵应计利润的绝对值（AB_DA1）的回归系数是-1.940，在 5%以下的置信度下回归结果显著（t 值=-2.01）；2）前瞻性指标对可操纵应计利润的绝对值（AB_DA2）的回归系数是-1.924，在 5%以下的置信度下回归结果显著（t 值=-1.99）；3）前瞻性指标对可操纵应计利润的绝对值（AB_DA3）的回归系数是-2.286，在 5%以下的置信度下回归结果显著（t 值=-2.31）。因此可以接受假设 4：前瞻性信息披露水平与上市公司信息不透明程度具有负相关关系。从而证实了前瞻性信息披露增加了管理层隐藏负面信息以及盈余操控的难度，从而会降低公司财务信息的不透明度。

表 9：前瞻性信息与公司不透明程度

	(1)	(2)	(3)
Dependent	**AB_DA1**	**AB_DA2**	**AB_DA3**
Forward_Index	-1.940**	-1.924**	-2.286**
	(-2.01)	(-1.99)	(-2.31)
ROA	0.003***	0.003***	0.003***
	(7.59)	(7.23)	(7.01)
Size	-0.016***	-0.015***	-0.014***
	(-10.38)	(-10.25)	(-9.07)
Age	0.000	0.000	0.001*
	(1.42)	(1.22)	(1.91)
LEV	-0.070***	-0.070***	-0.070***
	(-7.62)	(-7.72)	(-7.35)
BM	-0.002	-0.002	-0.001
	(-1.00)	(-0.92)	(-0.64)
Growth	0.082***	0.082***	0.088***
	(14.15)	(14.38)	(14.23)
Loss	0.021***	0.018***	0.018***
	(5.37)	(4.85)	(4.72)
FSHR	0.029***	0.028***	0.032***
	(3.39)	(3.28)	(3.62)
Manager	0.065***	0.063***	0.067***
	(6.12)	(5.92)	(5.84)
Nation	0.086***	0.086***	0.086***
	(9.36)	(9.39)	(9.15)
CEO	0.016***	0.016***	0.016***
	(5.09)	(5.14)	(5.03)
YRET	-0.010***	-0.009***	-0.009***
	(-3.43)	(-3.32)	(-3.04)
Volatility	0.025	0.027*	0.058***
	(1.58)	(1.75)	(3.29)
Constant	0.519***	0.508***	0.481***
	(13.36)	(13.33)	(11.99)
Industry Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Observations	12555	12555	12555
Adjusted R^2	0.149	0.148	0.147

注：括号内为 t 值，估计采用稳健标准误，误差项在公司层面聚类。*、**、***分别表示显著性水平(双尾)为 10%、5%、1%

为进一步探究，本文将可操纵应计利润分为正负两组，分别代表正向的盈余操纵和负向的盈余操纵，重新带入模型（2）进行回归分析。如表 10 中第（1）、（2）列所示，（1）列是基于修正的 Jones Model 计算得到的可操纵应计利润为正的样本，（2）列式可操纵应计利润为负的样本。通过比较第（1）、（2）列中前瞻性指标（Forward_Index）的回归系数的显著性可以看到，前瞻性信息对正向盈余操纵的抑制作用更强。同理，本文基于 Jones Model 和业绩修正的 Jones Model 计算得到的可操纵应计利润进行分组回归，通过比较第（3）和（4）列，第（5）和（6）列，均得到同样的实证结果。

表 10：前瞻性信息与公司不透明程度

	(1)	**(2)**	**(3)**	**(4)**	**（5）**	**（6）**
Dependent	**AB_DA1**	**AB_DA1**	**AB_DA2**	**AB_DA2**	**AB_DA3**	**AB_DA3**
AB_DA	**Positive**	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**
Forward_Index	-3.703**	0.446	-3.738***	0.587	-3.716**	-0.153
	(-2.55)	(0.42)	(-2.61)	(0.55)	(-2.57)	(-0.14)
Constant	0.485***	0.645***	0.485***	0.612***	0.522***	0.523***
	(8.85)	(15.83)	(9.06)	(15.03)	(9.32)	(11.89)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6115	6351	6127	6338	6215	6251
Adjusted R^2	0.201	0.154	0.204	0.153	0.199	0.142
Test: p-value	0.000		0.000		0.000	

注：括号内为 t 值，估计采用稳健标准误，误差项在公司层面聚类。*、**、***分别表示显著性水平(双尾)为 10%、5%、1%

五、 结论

作为以文字性的非财务信息为主的前瞻性信息，是以未来发展为视角，提供了基于公认准则而产生的表内信息及报表附注无法提供的信息，是上市公司对外披露的信息当中最具有价值的部分。然而以往国内针对前瞻性信息的研究所采用的方法存在主观性和先验性的固有缺陷，不仅不适用于海量文本数据的处理，更无法精确测量每个年度财务报告文本中真正具有的信息含量的内容。

本文尝试将文本分析和机器学习的方法引入前瞻性信息的研究，首先通过 Word2Vec 机器学习技术获取在年报环境下代表前瞻性信息的词集，然后基于财经专业类文本的分词系统技术，通过自然语言处理和文本分析计算出所有词集的词频，从而构建出全新的前瞻性信息披露指标，以期为未来该领域的研究提供了参考和借鉴。与传统衡量前瞻性信息的手工打分方式相比，词频衡量方式更具备客观性、拓展性和海量性；与笼统的计算管理层分析与讨论段落多少的方式相比，词频衡量方式能更好的捕获与公司未来相关的信息，剔除过去经营情

况总结的干扰；与简单的计算关键词词汇词频占比的方式相比，本文的“前瞻性”指标采用机器学习方法，获取了在财务报告语料环境下所有表示“前瞻性”的词汇，具备全面性、客观性和财经特异性。

研究前瞻性信息的第一步，必然是是研究其有用性。因为国内有关前瞻性信息有用性的实证研究，多是基于人工调查评级或人工分析等方法，样本具有较大的异质性，而且中国当时 MD&A 信息披露质量不高，学者们针对所有项目的考察可能难以获得统一的结论。所以本文在提出全新的前瞻性指标后，希望基于该前瞻性信息披露指标，通过量化实证的方法来研究前瞻性信息的有用性。

研究结果发现，前瞻性信息披露水平与公司未来绩效呈正相关关系。进一步研究表明，公司年度报告可读性越强，前瞻性信息与公司未来绩效两者之间的关系越强。公司信息不对称程度越高，前瞻性信息与公司未来绩效两者之间的关系越强。后续证据发现，前瞻性信息能够有效抑制公司财务信息不透明程度，尤其是公司的正向盈余操纵，从而进一步证实了前瞻性信息的有用性。

参考文献

[1] 程新生、刘建梅、程悦，2015，相得益彰抑或掩人耳目：盈余操纵与 MD&A 中非财务信息披露，《会计研究》，第 8 期，第 11-18 页。

[2] 程新生、谭有超、刘建梅，2012，非财务信息、外部融资与投资效率——基于外部制度约束的研究，《管理世界》,第 7 期，第 137-150 页。

[3] 陈小悦、徐晓东，2001，股权结构、企业绩效与投资者利益保护，《经济研究》, 第 11 期，第 11 页。

[4] 贺建刚、孙铮、周友梅，2013，金字塔结构、审计质量和管理层讨论与分析——基于会计重述视角，《审计研究》，第 6 期，第 68-75 页。

[5] 何卫东，2003，上市公司自愿性信息披露的动机、策略和监管，《深交所》，第 1 期，第 31-32 页。

[6] 李慧云、张林、张玥，2015，MD&A 信息披露、财务绩效与市场反应——来自中国沪市的经验证据，《北京理工大学学报：社会科学版》, 第 1 期，第 89-96 页。

[7] 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，《中国工业经济》，第 12 期，第 132-150 页。

[8] 蒋艳辉、冯楚建，2014，MD&A 语言特征, 管理层预期与未来财务业绩——来自中国创业板上市公司的经验证据，《中国软科学》，第 11 期，第 115-130 页。

[9] 李锋森、李常青，2008，上市公司 “管理层讨论与分析” 的有用性研究，《证券市场导报》，第 12 期，第 67-73 页。

[10] 乔旭东，2003，上市公司年度报告自愿披露行为的实证研究，《当代经济科学》，第 2 期。

[11] 唐跃军, 吕斐适, 程新生，2008，大股东制衡、治理战略与信息披露一一来自 2003 年中国上市公司的证据，《经济学(季刊)》，第 7 期，第 647-664 页。

[12] 汪炜、袁东任，2014，盈余质量与前瞻性披露: 正向补充还是负向替代?，《审计与经济研究》，第 1 期，第 48-57 页。

[13] 谢德仁、林乐，管理层语调能预示公司未来业绩吗?——基于我国上市公司年度业绩说明会的文本分析，《会计研究》，第 2 期，第 20-27 页。

[14] 薛爽、肖泽忠、潘妙丽，管理层讨论与分析是否提供了有用信息?——基于亏损上市公司的实证探索，《管理世界》,第 5 期，第 130-140 页。

[15] 杨海燕、韦德洪、孙健，2012，机构投资者持股能提高上市公司会计信息质量吗———兼论不同类型机构投资者的差异，《会计研究》，第 9 期，第 16-23 页。

[16] 周晓苏、王磊、陈沉，环境不确定性, 财务报告透明度和股价暴跌风险，《审计与经济研究》，第 31 期，第 57-66 页。

[17] 张宗新, 张晓荣, 廖士光，2006，上市公司自愿性信息披露行为有效吗?——基于 1998—2003 年中国证券市场的检验，《经济学(季刊)》, 第 4 期，第 369-386 页。

[18] Aboody D, Lev B,2000, Information asymmetry, R&D, and insider gains， The Journal of Finance, 55(6): 2747-2766.

[19] Ajinkya B, Bhojraj S, Sengupta P,2005, The association between outside directors, institutional investors and the properties of management earnings forecasts[J]. Journal of Accounting Research, 43(3): 343-376.

[20] Brown S V, Tucker J W, 2011 Large-sample evidence on firms’ year-over-year MD&A modifications, Journal of Accounting Research, 49(2): 309-346.

[21] Barron O E, Kile C O, O'KEEFE T B, 1999, MD&A quality as measured by the SEC and analysts' earnings forecasts, Contemporary Accounting Research, 16(1): 75-109.

[22] Bryan S H, 1997 Incremental information content of required disclosures contained in management discussion and analysis, Accounting Review, 285-301.

[23] Bradshaw M, Ertimur Y, O'Brien P, 2017, Financial Analysts and Their Contribution to Well-Functioning

Capital Markets, Foundations and Trends® in Accounting, 11(3): 119-191.

[24] Chen X, Harford J, Li K, 2007, Monitoring: Which institutions matter? Journal of Financial Economics, 86(2): 279-305.

[25] Cole C J, Jones C L, 2004, The usefulness of MD&A disclosures in the retail industry, Journal of Accounting, Auditing & Finance, 19(4): 361-388.

[26] Copeland T, 1978, Efficient capital markets: Evidence and implications for financial reporting, Journal of Accounting, Auditing and Finance, 2(1): 33-48.

[27] Clarkson P M, Kao J L, Richardson G D, 1999, Evidence that management discussion and analysis (MD&A) is a part of a firm's overall disclosure package, Contemporary Accounting Research, 16(1): 111-134.

[28] Davis A K, Piger J M, Sedor L M, 2012, Beyond the numbers: Measuring the information content of earnings press release language, Contemporary Accounting Research, 29(3): 845-868.

[29] Davis A K, Tama-Sweet I, 2012, Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A, Contemporary Accounting Research, 29(3): 804-837.

[30] Dhaliwal D S, Li O Z, Tsang A, et al, 2011, Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting, The Accounting Review, 86(1): 59-100.

[31] Francis J, Schipper K, Vincent L, 2003, The relative and incremental explanatory power of earnings and alternative (to earnings) performance measures for returns, Contemporary Accounting Research, 20(1): 121-164.

[32] Frazier K B, Ingram R W, Tennyson B M, 1984, A methodology for the analysis of narrative accounting disclosures, Journal of Accounting Research, 318-331.

[33] Hutton A P, Marcus A J, Tehranian H, 2009, Opaque financial reports, R2, and crash risk, Journal of Financial Economics, 94(1): 67-86.

[34] Kothari S P, Leone A J, Wasley C E, 2005, Performance matched discretionary accrual measures, Journal of Accounting and Economics, 39(1): 163-197.

[35] Li F, 2008, Annual report readability, current earnings, and earnings persistence, Journal of Accounting and Economics, 45(2-3): 221-247.

[36] Li F, 2010, The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach, Journal of Accounting Research, 48(5): 1049-1102.

[37] Lo K, Ramos F, Rogo R, 2017, Earnings management and annual report readability, Journal of Accounting and Economics, 63(1): 1-25.

[38] Mayew W J, Sethuraman M, Venkatachalam M, 2014, MD&A Disclosure and the Firm's Ability to Continue as a Going Concern, The Accounting Review, 90(4): 1621-1651.

[39] Muslu V, Radhakrishnan S, Subramanyam K R, et al, 2014, Forward-looking MD&A disclosures and the information environment, Management Science, 61(5): 931-948.

[40] Plumlee M, Brown D, Marshall S, 2008, The impact of voluntary environmental disclosure quality on firm value.

[41] Schroeder N, Gibson C, 1990, Readability of management's discussion and analysis, Accounting Horizons, 4(4): 78-87.

[42] Smith Jr C W, Watts R L, 1992, The investment opportunity set and corporate financing, dividend, and compensation policies, Journal of Financial Economics, 32(3): 263-292.

[43] Sun Y, 2010, Do MD&A disclosures help users interpret disproportionate inventory increases? The Accounting Review, 85(4): 1411-1440.

[44] Tetlock P C, Saar-Tsechansky M, Macskassy S, 2008, More than words: Quantifying language to measure firms' fundamentals, The Journal of Finance, 63(3): 1437-1467.

附录

附录（一）Word2Vec 相似词示例

Word2Vec 神经网络模型由 Mikolov et al. (2013) 提出，是近年来深度学习领域的里程碑式成果 (LeCun et al., 2015)。Word2Vec 模型根据上下文内容将词语表征为实数值向量，并通过向量的相似度计算得到词语之间的语义相似性 (Bengio et al., 2003)。在实际应用中，Woed2Vec 又分为 CBOW (Continuous Bag of Word) 和 Skip-gram（Continuous Skip-gram model）两种模型，我们分别采用两种模型进行训练，根据结果及效率评估，最终采用 CBOW 模型。

通过 Word2Vec 神经网络语言模型基于上下文语义信息将词汇表示成多维向量，计算向量相似度从而获得词汇的相似词。该模型基于海量财经文本训练而成，所推荐的相似词更加适合财经文本语境，可有效避免人为定义词表的主观性和通用同近义词工具的弱相关性。下表为“今后”词汇的 Word2Vec 相似词结果示例（取相似度排序为前 15 个示例）：

附表一 Word2Vec 相似词结果示例（前 15 个）

原始词	**相似词**	**相似度（基于年报语境）**	**词频**
今后	未来	0.6892	119072
今后	下一步	0.6282	4888
今后	下一阶段	0.5399	546
今后	下步	0.524	159
今后	后续	0.4816	26107
今后	未来五年	0.4779	1699
今后	后期	0.4379	4664
今后	新年度	0.4317	1806
今后	展望未来	0.4273	468
今后	未来市场	0.4259	1950
今后	未来发展	0.414	22154
今后	下半年	0.4039	35817
今后	将来	0.3992	5285
今后	明年	0.3937	1352
今后	近期	0.3923	7774

附录（二）前瞻性指标词集

1. 前瞻性指标构建词集

Step 1 种子词提取

来源	Muslu et al. (2014) 和 Li (2010）、证监会、年报
种子词	计划、预计、未来、目标、可能、如果、机遇、预期、挑战、预测、今后、目的、契机、前景、希望、展望、相信、愿景、期待、明年、期望、打算、来年（共 23 个）

Step 2 扩充年报环境下“前瞻性”词集

来源	机器学习-Word2Vec 相似词扩充、剔除并不表示将来或词汇、剔除在年报中出现频率过小的词汇
扩充词集	下半年、以后、机会、不确定性、后续、未来发展、发展机遇、尚需、拟向、还将、长远发展、发展空间、近期、有望、追求、新一轮、拟将、日趋、短期内、将来、预见、下一步、后期、趋于、尚待、必将、将向、先机、仍需、预估、未来市场、新年度、未来五年、拟于、新形势下、下一阶段、下一年、大好时机、长远规划、发展良机等（共 97 个）

2. 年报中前瞻性词集的语句示例

词汇	**年报语句示例**
Panel A	**有关行业的未来发展趋势，市场竞争格局，发展战略**
展望	**展望** 2013 年，尽管中国经济面临的内外部环境依然复杂多变，但在前期「稳增长」调控政策和新一届政府着力推进新一轮经济改革措施的拉动下，国内经济有望延续 2012 年四季度以来的回升走势，并继续保持平稳较快增长，铁路运输行业亦有望在经济回升、高铁成网运营创造新的客货运输需求的背景下步入新的较快增长期。
未来	政府部门针对高速公路行业出台政策的频率越来越高，政策的约束性越来越强，社会舆论压力也较为集中，多方面因素共同决定了国家**未来**对收费公路的管理必然趋于更加严格。
可能	公司**可能**面临宏观经济下行，电力需求减弱的问题。
机遇	由于雾霾天气越来越严重，政府将进一步扩大排放法规执行的范围和力度，这虽然对新车的销售产生制约，但同时也蕴含着市场**机遇**，比如强制淘汰黄标车，给车辆更新提供了机会。
前景	存在宏观经济复苏**前景**不明朗的风险。
Panel B	**有关公司战略，经营计划，资金需求与使用计划**
目标	本公司将以转型引领发展，以“打造大资管经营体系，争当交易银行排头兵”为**目标**，全力构建大同业、大资管、大交易体系。
愿景	公司将以“打造全球综合竞争力最强的特种船公司，成为国际领先的工程物流服务商”为**愿**

	景。
今后	**今后**，除本次配套募集资金投资项目外，公司计划在以下几点着力突击：一是充分发挥公司的资源优势和市场优势，不断增强资本实力，提升品牌影响力，促进业务拓展，实现快速增长，二是加大研发投入，努力提高单井产量。
未来	**未来**公司将进一步加大体制创新力度，提升资源配置效率，把南航的存量优势转化为增量优势。
计划	本集团**计划**使用自有资金和银行贷款等方式来满足上述项目需求。
预计	2012 年公司**预计**固定资产投资 48.00 亿元，主要用于一硅钢技术改造、轧板厂技术改造等项目。

Panel C 有关公司经营业绩预测和经营目标	
预计	本集团设定 2013 年的总体通行费收入目标约为人民币 20.95 亿元，财务费用**预计**比 2012 年略有增长。
预计	**预计**本集团收费公路项目的车流量和收入增长将存在不确定性。
期待	我们对 2012 年集团的发展充满信心与**期待**。
可能	公司存在经营业绩**可能**下降甚至亏损。

Panel D 有关公司所面临的风险，挑战以及应对措施	
预期	2014 年上半年市场流动性较为宽松，尽管 5 月份存在财政存款季节性上缴、年中效应临近等不利因素，但在央行公开市场操作的引导下，机构对于后市资金面**预期**稳定，银行间资金面维持宽松态势，本公司流动性风险处于中低偏下水平。
未来	公司旗下郑州日产汽车有限公司产品销售在**未来**存在受中日关系影响的风险。
挑战	行业竞争加剧风险随着国内民航运输业市场的逐步开放，三大航空公司、外国航空公司以及中小航空公司在规模、航班班次、价格、服务等方面竞争日趋激烈，对公司的经营模式和管理水平提出了较大**挑战**。
期望	本集团在监管框架及市场环境允许的情况下，主要通过匹配投资资产的期限与对应保险责任的到期日来管理流动性风险，以**期望**本集团能及时偿还债务并为投资活动提供资金。
希望	在主营业务持续疲弱的情况下，公司寄**希望**于通过资产重组，收购海外优质矿山，从根本上解决公司可持续发展问题。