

央行货币政策报告文本信息、 宏观经济与股票市场

姜富伟 胡逸驰 黄楠

(中央财经大学金融学院,北京 100089;北京大学经济学院,北京 100871;
嘉实基金管理有限公司,北京 100005)

摘要: 本文利用金融情感词典和文本分析技术,分析中国人民银行货币政策执行报告的文本情绪、文本相似度和文本可读性等多维文本信息,刻画央行货币政策执行报告的文本特征,探究货币政策报告的文本信息与宏观经济和股票市场的关系。实证研究发现,货币政策报告的文本情绪的改善会引起显著为正的股票市场价格反应,报告文本相似度的增加会引起股票市场波动性的显著降低,报告可读性对公布后股票市场的波动性影响不显著。货币政策报告文本情绪还与诸多宏观经济指标显著相关。进一步研究发现,引起股票市场显著反应的是报告文本情绪中反映货币政策指引的部分,而反映宏观经济历史状态的部分对股票市场的影响不显著。本文从文本大数据分析角度证明了我国央行沟通的有效性,对国内央行沟通相关研究形成了有益补充。

关键词: 文本情绪分析; 中央银行沟通; 股票市场; 宏观经济

JEL 分类号: C51, G12, G18 **文献标识码:** A **文章编号:** 1002-7246(2021)06-0095-19

一、引言

中央银行沟通(central bank communication)是指中央银行向公众披露货币政策目标、货币政策策略及未来货币政策走向等信息的过程。20世纪90年代以来,中央银行沟通日益引起学术界和业界的重视。迄今为止,学术界已经就中央银行沟通的度量、中央银行沟通与通货膨胀预期和金融市场稳定的关系等问题进行了许多有意义的研究。近年来,中国人民银行也日益重视对外界的货币政策沟通。在人民银行的政策沟通工具

收稿日期:2018-12-03

作者简介:姜富伟(通讯作者),金融学博士,教授,中央财经大学金融学院,E-mail:jfuwei@gmail.com.

胡逸驰,硕士研究生,北京大学经济学院,E-mail:15600583022@163.com.

黄楠,经济学博士,嘉实基金管理有限公司,E-mail:huangnan@jsfund.cn.

* 本文感谢国家自然科学基金项目(72072193,71872195,71602198)的资助。感谢《金融研究》编辑部和匿名审稿人的宝贵意见,文责自负。

中,《货币政策执行报告》(以下简称“报告”)历史最为悠久,形式最为成熟,受到金融市场的广泛关注。自 2001 年第一季度起,人民银行开始每季度在其官方网站公布报告,回顾过往的经济形势,解释货币政策效果,展望未来的国内外局势,并阐明下一个阶段的货币政策思路。

本文利用金融情感词典和文本分析技术手段,对我国央行货币政策报告进行文本分析,并探究货币政策报告的文本信息与宏观经济和股票市场的关系。本文实证分析主要包括四部分:第一,我们计算出了报告的文本情绪、文本相似度、文本可读性等指标,对报告的文本特征进行了细致的统计描述;第二,我们围绕计算出的文本指标,研究报告文本情绪和宏观经济状况的关系,发现总体上报告的文本情绪与诸多宏观经济金融指标显著相关;第三,利用 EGARCH 模型的实证分析,发现报告文本情绪的上升对报告发布后一个交易日的股票市场走势有显著的正向影响,报告文本相似度的增加能够显著降低报告发布后一个交易日的股票市场波动性,报告文本可读性对发布后一个交易日股票市场的波动性并无显著影响;第四,将报告文本情绪分解为反映宏观经济状况的部分和反映央行货币政策指引的部分,并探究两者是否都能对股票市场走势有显著影响,结果显示报告文本情绪中对市场有显著影响的主要是反映央行货币政策指引的部分。

与已有的关于中央银行沟通和金融经济文本情绪分析的文献相比,本文开展了更深入的实证和理论探索,主要有四个创新和贡献。第一,本文使用前沿的文本分析技术对货币政策报告文本进行全方面多维度分析,在内容和技术手段上与以往文献都有明显差异。从内容上看,以往文献大都从央行沟通的货币政策倾向的角度来对央行沟通进行量化。而本文从从央行书面沟通的语言文字出发,进一步探究报告文字层面的特征信息含量,及其与宏观经济和股票市场的关系。从技术手段上看,已有的文献多通过人工阅读或统计关键词的方法来构建央行沟通的量化指标,技术手段较单一,结果主观性较强;而本文基于中文文本分词方法和情绪词典,用计算机编程自动计算出文本情绪、相似度和可读性等指标,保证了研究结果不受主观人为因素干扰。

第二,本文对中文金融经济文本情绪分析中关于央行沟通情绪分析的研究形成了有益补充。目前国内金融文本情绪的相关文献主要针对财经媒体报道和公司财务报告的文本情绪,较少涉及中央银行沟通文本情绪。本文在我国央行沟通文本的基础上,做出了文本情绪分析尝试,为将来围绕央行沟通开展更加全面和深入的研究打下了坚实的基础。

第三,本文对央行沟通文本情绪影响股票市场的传导机制开展了深入分析。中央银行沟通包含两方面信息:一方面是经济金融基本面信息,另一方面是央行的政策指引。现有文献的一个普遍不足是:没有探究影响资产价格的因素是沟通中的宏观经济基本面历史信息,还是沟通中的央行未来货币政策主张和指引。本文的实证结果表明能够显著影响股票市场的是文本情绪中反映央行货币政策指引的部分,而非反映宏观经济基本面的部分。这一结论意味着报告的文本情绪能够影响股票市场,是由于央行主张和政策指引改变了投资者的市场预期。

第四,本文的研究对于我国进行货币政策调整和投资者投资管理也有十分重要的意义。从货币政策当局角度而言,本文发现货币政策报告文本(尤其是体现政策指引部分的文本)可以影响我国资产价格,这一结果证明了我国央行沟通的有效性,有助于提高政策调控效率,还有助于实现经济稳定和金融稳定的双重目标。从投资者角度而言,本文证明货币政策报告文本会影响我国资产价格回报和资产波动率,投资者要注意货币政策报告文本内容的分析,据此动态管理投资组合,提高投资回报(姜富伟等,2011)。

二、理论与文献综述

(一) 中央银行沟通与金融市场

货币政策报告是央行沟通的一种形式,因而本文属于对央行沟通的研究。国外学者关于金融市场对央行沟通反应的研究有三类研究对象:股票、利率和汇率。在针对股票的研究中,Lucca and Moench(2015)发现美国联邦公开市场委员会(FOMC)发布会议公告前的几个小时,股票市场会出现较大的超额收益。而Gieslak et al.(2019)对FOMC公告效应的进一步研究发现,股票的超额收益具有明显的周期性。针对利率的研究中,Guthrie and Wright(2000)提出了著名的“Open Mouth Operation”,并发现新西兰央行的公开市场操作公告对不同期限的利率都有显著影响。Demiralp and Jorda(2002)则发现在美联储开始公布联邦基金目标利率后,达到目标利率所需的公开市场操作规模便显著减小。在针对汇率的研究中,部分学者的实证研究发现汇率政策公告可以驱动汇率走势变化(Beine et al. 2009),但有些学者的研究则表明关于汇率的告示并不能有效驱动汇率走势,反而增加了汇率波动性(Jansen et al. 2007)。

类似地,国内关于中央银行沟通对金融市场和资产价格影响的研究也可分为利率、股票和汇率三个方向。首先,诸多关于利率对中央银行沟通反应的研究证实,央行沟通对短期利率走势有显著影响(冀志斌和宋清华,2012;张强和胡荣尚,2014;吴国培和潘再见,2014)。另外,冀志斌和宋清华(2012)、张强和胡荣尚(2014)还发现沟通对利率的影响具有不对称性,即紧缩取向的沟通提升利率的作用大于宽松取向的沟通降低利率的作用;吴国培和潘再见(2014)发现以央行行长冠名的沟通效力大于以他人冠名的沟通。其次,央行沟通对股票市场的影响并不显著,且在统计意义和经济意义上,口头沟通的效果比书面沟通更好(冀志斌和宋清华,2012;张强和胡荣尚,2013;吴国培和潘再见,2014)。最后,朱宁等(2016)关于央行沟通对汇率影响的研究表明,沟通对汇率影响较为显著;而与利率和股票不同的是,书面沟通对汇率的作用要优于口头沟通。

(二) 经济金融文本大数据分析

经济金融文本分析的核心在于用适当的方法提取文本样本的特征信息,主要集中于探究文本情绪(tone)、文本相似度(similarity)和文本可读性(readability)等几类文本指标对经济走势和金融市场的反映或影响。

经济金融文本情绪相关的国外文献可以分为三类:第一类文献研究公司公告的文本情绪对公司股票价格的影响。Loughran and McDonald(2011)与 Jegadeesh and Wu(2013)指出公司年度报表(10-K filings)文本情绪对公司股票异常收益率有显著影响。Jiang et al.(2019)发现管理层情绪对未来股票市场回报有显著的负向作用,加入管理层情绪会十分显著地提高股票回报预测模型的准确性。第二类文献研究媒体报道文本情绪对股票市场的影响。Tetlock(2007)统计了每日金融新闻中 77 种不同类别词语的数量,用主成分分析的方法提取新闻文本情绪的代表变量,并用 VAR 模型来检验新闻文本情绪代理变量的变动是否对股票市场有显著影响;Garcia(2013)发现,在经济衰退时期,纽约时报的文本情绪对股票市场收益率的预测能力较强;Huang et al.(2014)提取了分析师报告的文本情绪,发现该指标能显著解释报告发布后相关公司股票的价格变动。第三类文献研究中央银行沟通的文本情绪与宏观经济和金融市场的关系,Amaya and Filbien(2015)研究了欧洲央行沟通文本情绪对股票市场走势的影响;Ehrmann and Talmi(2017)用 EGARCH 模型研究了加拿大央行货币政策公告文本情绪对债券市场和股票市场收益率的影响。Ricardo et al.(2020)利用文本情绪分析法构建了 30 多个国家的金融稳定情绪指数(FSS),发现其对信贷、资产价格、系统性风险和货币政策利率的变动都有着较强的解释力。

国内经济金融文本情绪分析主要集中于两类文本对象:媒体报道和公司财务报告。游家兴和吴静(2012)发现,媒体报道传递出的文本情绪过于高涨和低落时,股票价格更有可能偏离内在价值,且公司信息不确定性越强时,媒体情绪对定价错误的影响越显著。谢德仁和林乐(2015)发现公司管理层在年度业绩说明会上的正面语气与公司下一年度的业绩显著正相关,说明管理层情绪可信度较高;罗炜等(2017)发现,财经媒体报道的正面情绪可以给创业企业带来更多的风险投资数额。相比于媒体报道和公司财务报告,国内也有许多研究对中国人民银行政策沟通使用人工识别法进行文本分析。譬如,卞志村和张义(2012)分析了央行信息披露引导居民通胀预期的作用效果,发现相比于传统货币政策工具,信息披露有着时滞更短的优点。张强和胡荣尚(2013)发现央行沟通对金融市场的影响存在但不明显,且书面沟通的效力弱于口头沟通。林建浩和赵文庆(2015)基于货币政策执行报告构建了中央银行沟通指数,并对数量型、价格型政策工具和央行沟通进行谱密度分析,发现央行沟通既能及时反映经济金融形势变化,在长期也有连续性。

文本相似度(similarity)是指两个文本在遣词造句或表达含义上的相似程度,货币政策报告的文本相似度即报告文本用词的一致性,刻画着不同期报告所含信息差异的大小。文本相似度的研究集中于对中央银行沟通的研究。譬如,Jansen et al.(2007)研究了欧洲央行沟通的文本相似度,发现欧央行的信息公开文本总体存在一致性,但也会通过灵活的遣词造句来应对变化;Acosta and Meade(2015)研究了美联储公开市场委员会例会后公告(statement)的文本相似度,发现公告的文本相似度越来越高,表明美联储在公告措辞的变化上变得谨小慎微;Ehrmann and Talmi(2017)发现加拿大央行沟通文本相似度的增加能

显著降低股票市场收益率的波动性。

文本可读性(readability)指阅读和理解文本的难易程度,阅读和理解文本的难度越大,其可读性越低;阅读和理解文本越容易,可读性越高。文本可读性在经济金融文本分析领域的运用多局限于英文文本。譬如,Jansen(2011)用可读性指标衡量美联储沟通的清晰度(clarity)并发现沟通清晰度的上升可减少股票市场和债券市场的波动性,Loughran and McDonald(2014)发现公司财务报告的可读性强会减少报告发布后股票价格的波动性。

三、实证分析方法

本文研究的主题是央行货币政策执行报告的文本信息与宏观经济、股票市场走势的关系。本部分第一节介绍本研究如何对报告进行分词,第二节介绍如何计算报告的文本情绪,第三节介绍计算报告文本相似度和文本可读性的方法,第四节介绍选用的计量模型,即EGARCH模型。

(一) 报告分词方法

文本分析的基础是对文本进行分词,分词的基础是分词工具自带的词库。中文文本中较为常用的分词方法为jieba分词,然而jieba分词自带的通用词库无法很好地识别经济金融文本中的专有名词。Python中jieba模块的词库为开源词库,用户可以自行添加词库中不包含的词语,我们将经济金融相关的一些专业术语添加到jieba模块的词库中,以提高分词的准确性。目前国内中文专业词汇的词库较少,其中搜狗词库作为覆盖面相对较广、知名度相对较高的专业词库,在语义学和文本分析研究中得以广泛应用。搜狗词库容纳了许多经济金融专业词库,我们挑选了三个下载量较大的经济、金融、财会词库¹。在jieba开源词库中加入搜狗词库后,绝大多数经济金融专业名词都被正确分词,分词效果提升显著。

(二) 文本情绪计算与中文金融情感词典

Loughran and McDonald(2011)指出,经济金融文本用词的情感倾向与通用文本有较大区别,为此两位学者开发出了目前在英文经济金融文本分析中占有重要地位的LM词典。实证研究表明LM词典应用于经济金融文本分析的结果要明显优于通用情感词典。姜富伟等(2020)开发出了一个用于经济金融文本分析的中文情感词典(以下简称“新词典”),实证结果表明新词典在经济金融文本上的适用性要显著高于通用的知网情感词典。本文先采用姜富伟等(2020)的情感词典计算货币政策报告的文本情绪,在稳健性检验中再用知网情感词典计算出的文本情绪检验我们的结果。但是,由于姜富伟等(2020)的新情感词典是基于财经媒体报道开发的,其对央行沟通文本情感词判断的准确性还需

¹ 本文添加的词库包括“财会词汇大全”、“财经金融词汇大全”和“金融词库”,下载地址为 <https://pinyin.sogou.com/dictcate/index/80>。

要改进。针对央行沟通文本的特点,我们仔细阅读了央行货币政策报告的文本,并参照冀志斌和宋清华(2012)、林建浩和赵文庆(2015)等以往文献关于选词的设定,在新词典中添加了适用于央行沟通文本的一些常用情感词。

本文采用情感单元法计算货币政策报告的文本情绪。一个情感单元由否定词(可无)和情感词组成,一篇文章包含许多情感单元。我们识别第 i 个情感单元的方法如下:从第 $i-1$ 个情感词结束后开始逐个词检索,直到找到第 i 个情感词,之后在第 $i-1$ 和第 i 个情感词之间统计否定词的数量,记为 m_i ,从而识别出如图 1 所示的完整情感单元。对于一个情感单元,其情绪如(1)式所示。(1)式中 $sig(m_i)$ 为否定指数, $weight(word_i)$ 表示情绪词的情绪指向,两者计算公式如(2)式和(3)式所示。将一篇货币政策报告中所有情感单元的情绪加总再除以报告的总字数,即可得到该篇货币政策报告的文本情绪。

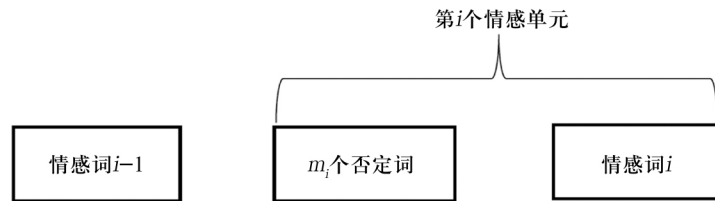


图 1 情感单元图示

$$tone_i = sig(m_i) \times weight(word_i) \quad (1)$$

$$sig(m_i) = (-1)^{m_i} \quad (2)$$

$$weight(word_i) = \begin{cases} 1 & \text{若 } word_i \text{ 是积极情感词} \\ -1 & \text{若 } word_i \text{ 是消极情感词} \end{cases} \quad (3)$$

参考国内外关于央行沟通测度的研究(冀志斌和宋清华 2012; Baker et al. 2016),可以发现过往文献中央行沟通指数在各期的值有正有负,能够更为直接地体现央行对经济的总体看法和货币政策倾向。由于人民行政策沟通中积极情感词明显多于消极情感词,本文基于央行文本情绪构造的央行沟通指数都为正值。为了与主流文献相符,也便于讨论实证分析的结果,我们对计算出的原始文本情绪指标进行 Z 值标准化处理,并将标准化后的文本情绪指标用于后文的实证分析。

(三) 文本相似度与可读性

本文采用基于 TF-IDF 加权的余弦相似度法计算报告的文本相似度,计算过程分为如下步骤:首先,去除报告文本中只有语气停顿、转折等作用的词语,即停用词(stop word)。其次,将每个文本看成一个词汇向量,向量的某个维度表示对应词语在这个文本中出现的次数,得到所有文本的向量表示。再次,计算所有词语的 idf 系数,对于文本样本中出现过的词语 j ,计算 idf 系数如(4)式所示。其中, N 为文本集中文本的数量, n_j 为文本集中出现词语 j 的文本数量。最后,在每个文本向量中表示词语 j 频率的维度上乘

以词语 j 的 idf 系数,再用经 idf 系数调整后的文本向量计算余弦相似度,可得到结果,即 TF-IDF 加权的文本相似度。

$$idf_j = \ln\left(\frac{N}{n_j}\right) \quad (4)$$

关于文本可读性,本文将平均句子长度作为可读性的衡量指标。具体而言,我们将句号、感叹号和问号看作一个句子的结尾对文本句子总数进行统计,进而以文本总字数除以文本句子总数,得到平均句子长度。在英文文献中,Loughran and McDonald(2014)指出“迷雾指数”(Fog Index)等可读性指标在预测股票价格波动性方面效果比不上平均句子长度(平均每个句子包含的英文单词数量)。在中文文本分析的研究中,陈银娥和江媛(2017)在实证研究中也采用平均句子长度(平均每个句子包含的字符数量)来表征中文文本的可读性。

(四) 计量方法: EGARCH 模型

一般而言,金融资产收益率呈现出有偏和尖峰肥尾的特点,下文对本文所用的股票市场收益率数据的描述性统计也可以看出这些特点。ARCH 类模型的应用能够有效地刻画金融资产收益率的这些特点造成的异方差性。本文使用 EGARCH(1,1) 模型来对股票市场收益率进行计量分析。

四、数据介绍

(一) 文本数据

本文计算文本情绪的对象是人民银行每季度公布一次的《货币政策执行报告》。报告是人民银行用于向社会各界介绍货币信贷情况、分析货币政策效果、回顾金融市场情况、阐述宏观经济走势、点明货币政策方向的重要书面沟通手段。第一期报告(2001年第一季度报告)于2001年5月公布于人民银行官方网站上,此后每隔3个月发布新的报告。一般而言,每个季度的报告会在该季度过后的第二个月发布。本文使用的报告从2001年第一季度起,至2018年第三季度止,共71篇文本。在对报告进行文本分析前,我们对报告进行了预处理(preprocessing),删除了所有标点符号,去掉了所有标题和非正文的文本内容。

(二) 宏观经济与金融数据

表1展示了本文实证分析所用的经济金融数据的起止时间、频率和来源。我们所选取的数据涵盖宏观经济景气状况、价格水平、就业情况、利率走势、汇率走势和资产价格等国民经济和宏观金融运行的诸多方面。因文本情绪数据频率为季度,需要把所有经济金融数据调整为季度频率,变频方式如表1所示。

表 1 宏观经济与金融数据简介

宏观经济金融数据	起止时间	频率	数据来源	变频方式
制造业采购经理指数	2005—2018 年	月度	国家统计局	取季度均值
消费者价格指数	2001—2018 年	月度	国家统计局	取季度同比增长率
城镇登记失业率	2002—2018 年	季度	国家统计局	—
1 年期国债到期收益率	2002—2018 年	日度	中债估值中心	取季度均值
美元兑人民币中间价	2001—2018 年	日度	中国人民银行	取季度均值
Wind 全 A 指数静态市盈率	2002—2018 年	日度	Wind 资讯	取季度均值

(三) 股票市场数据

本文选取的股票指数包括上证综指、深证成指、沪深 300 和中证 500，涵盖大市、大盘股和中小盘股等不同股票类型的收益状况，涵盖的时间范围为 2005 年 10 月 1 日至 2018 年 12 月 31 日。所有指数的日度涨跌幅数据均来源于 Wind 资讯。

五、实证分析

(一) 描述性统计

本节对基于报告计算出的文本情绪、相似度、可读性等文本指标进行介绍。文本长度是文本指标中较为基础的，图 2 展示了报告文本长度的三季度移动平均线。可以看出，报告的总字数从最初几个季度的一万余字稳定增长到 2018 年的三万字左右。报告文本长度的增长，也伴随着报告内容的日渐丰富，涵盖到了货币金融、货币政策、宏观经济、金融市场等方面的状况。

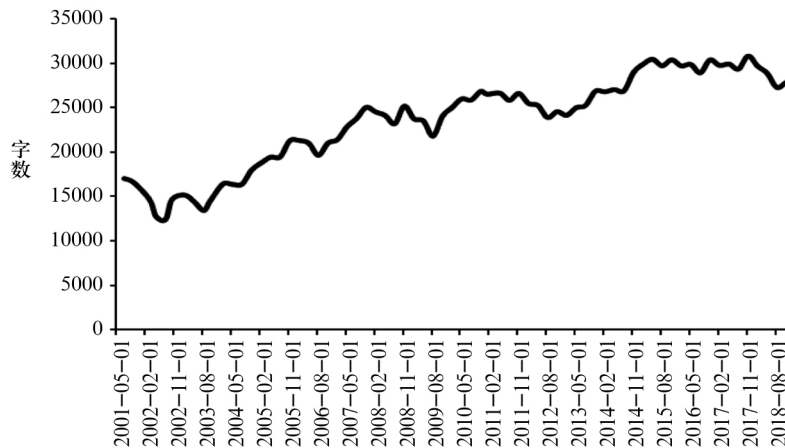


图 2 报告文本长度变化

观察报告的长度后,还需进一步探究报告关注的重点。在将文本进行分词后,我们统计了不同词语在整个文本中出现的频率,可以看出,“发展”、“商业银行”、“中国人民银行”、“房地产开发”、“基础货币”、“国债收益率”、“银行体系”、“流动性”、“宏观审慎”等词出现频率较高,表明央行对宏观金融运行、金融市场与机构、经济发展和金融监管相关的内容最为关注,还有一些词汇在不同期的报告中频繁出现,这就促使我们计算报告的文本相似度,以进一步探究不同期报告用词的相似性。图3展示了2001年第一季度至2018年第三季度报告文本相似度的三季度移动平均线。从趋势上看,报告的文本相似度在逐渐上升,报告用词趋于相近。虽然报告存在许多重复用词,但在撰写每一期报告时央行都会根据当时的宏观经济金融状况进行一定程度的修改。

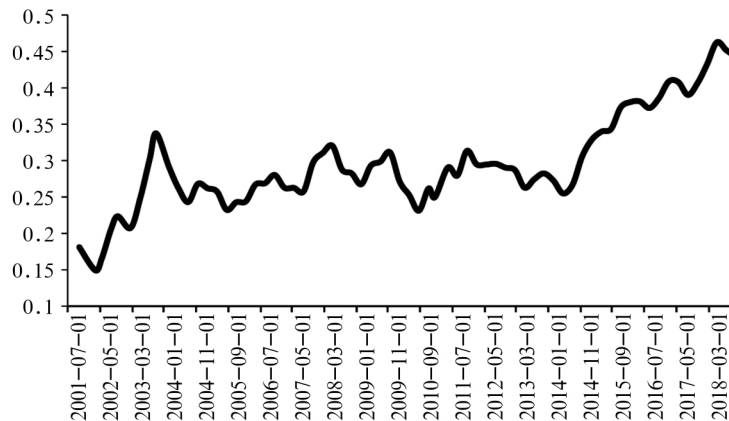


图3 报告文本相似度

之后,以每句所包含字数为衡量指标,计算了报告的文本可读性。图4展示了2001年第一季度至2018年第三季度报告可读性的变化。可以看出,2001—2008年,报告句子明显变短,可读性上升。2009年至今,报告的可读性维持在较稳定的水平。

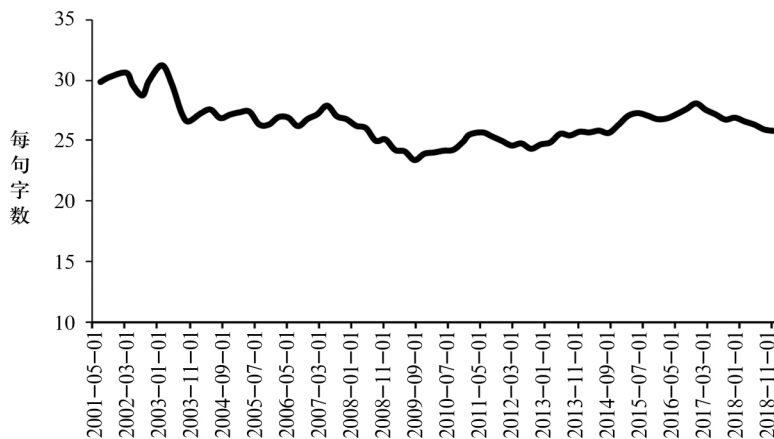


图4 报告文本可读性

最后,我们计算出报告的文本情绪。图 5 展示了从 2001 年第一季度至 2018 年第三季度报告文本情绪的三季度移动平均线。可以看出报告的文本情绪在不同时期表现出不同的趋势,趋势的转变通常伴随着重大的宏观经济金融事件,如中国加入世贸组织到 2007 年美国爆发次贷危机前,报告文本情绪呈现出上升状态;2007 年次贷危机爆发至 2009 年“四万亿”计划出台前,报告文本情绪呈下滑趋势;而“四万亿”计划出台后,报告文本情绪迅速上升。2015 年夏季股票市场大幅波动期间,报告文本情绪也出现了显著下滑,但下滑幅度相对较小。

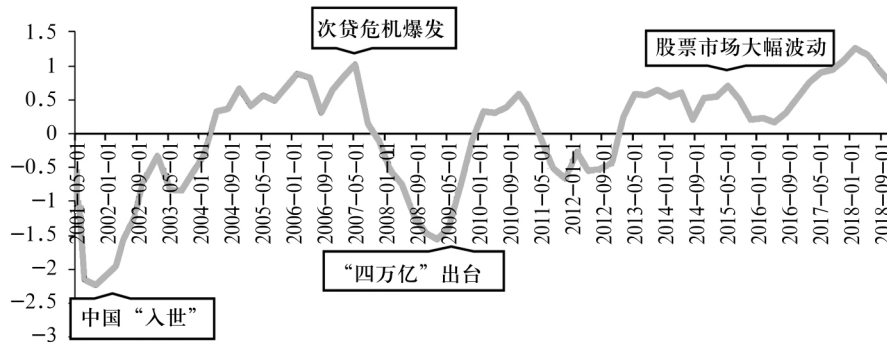


图 5 用情绪单元法得出的报告情绪走势

接下来,我们对估计下文计量模型所使用的股票市场数据、文本情绪和宏观经济金融数据进行描述性统计。表 2 展示了描述性统计的结果。从描述性统计指标中可以看出,本文所使用的四个股票指数收益率数据均存在有偏、尖峰厚尾的特点,这也说明 EGARCH 模型适用于本文的计量分析。

表 2 文本指标、宏观经济与金融数据与股票市场数据描述性统计

变量	最小值	最大值	平均值	标准差	偏度	峰度	样本数
shcomposite(%)	-8.84	9.46	0.0376	1.65	-0.48	7.23	3220
szcomposite(%)	-9.29	9.59	0.0510	1.89	-0.40	5.75	3220
hs300(%)	-9.24	9.34	0.0526	1.76	-0.42	6.52	3220
zz500(%)	-8.95	9.87	0.0692	2.01	-0.74	6.02	3220
tone	-2.42	2.41	0.00	1.00	-0.35	2.90	71
GDP(%)	6.40	15.00	9.20	2.52	0.70	2.21	56
NX(%)	-0.14	8.88	4.12	1.99	0.28	3.01	56
CPI growth(%)	-1.70	8.37	2.46	2.19	0.61	-0.02	72
unemployment(%)	3.80	4.30	4.10	0.12	-0.24	0.14	65

续表

变量	最小值	最大值	平均值	标准差	偏度	峰度	样本数
bond yield(%)	0.98	3.94	2.58	0.70	-0.16	-0.71	67
forex	6.12	8.28	7.15	0.83	0.32	-1.57	72
PMI	41.53	56.27	51.92	2.46	-1.02	4.31	56
P/E ratio	12.05	67.45	29.68	15.92	0.99	-0.25	72

注:表 2 展示了本文所用的文本指标、宏观经济与金融数据和股票市场数据的描述性统计结果。shcomposite 代表上证综指的日度对数收益率, szccomposite 代表深证成指的日度对数收益率, hs300 代表沪深 300 的日度对数收益率, zz500 代表中证 500 的日度对数收益率, tone 代表情绪单元法计算出的报告文本情绪, GDP 代表中国实际 GDP 季度同比增长率, NX 代表净出口占同期 GDP 的比例, CPI growth 代表消费者价格指数的季度同比增长率, unemploy 代表城镇登记失业率数据, bond yield 代表 1 年期国债到期收益率的季度均值, forex 代表人民币兑美元汇率的季度均值, PMI 代表制造业采购经理指数的季度均值, P/E ratio 代表 Wind 全 A 指数静态市盈率的季度均值。

(二) 文本情绪与宏观经济金融指标相关性分析

在实证研究中,首先对报告文本情绪与宏观经济金融指标是否存在较强的相关性进行分析。首先就相关性分析所涉及的变量展开 ADF 检验。除 GDP 增长率、净出口指标、城镇登记失业率和美元兑人民币中间价一阶差分后平稳外,其他指标均通过检验。

接下来就报告文本情绪和宏观金融经济指标的相关性进行 Pearson 相关系数检验,我们关注报告文本情绪的预测效果,故这里检验滞后一期的报告文本情绪序列与经济金融指标的相关性。表 3 展示了 Pearson 相关系数检验的结果,可以看出报告文本情绪和 GDP 增长、CPI 指数增长、PMI 和失业率之间都存在显著的相关性,与净出口、利率水平、股票市场估值和汇率变动之间无显著的相关性。从相关系数的符号上看,较高的文本情绪指数预示着经济增长较快、通货膨胀水平较高、产业景气度较高、失业率较低,即文本情绪与宏观经济的相关性是正向的。

表 3 文本情绪与宏观经济与金融指标相关性分析结果

	tone(-1)	D_GDP	D_NX	CPI growth	bond yield	PMI	P/E ratio	D_unemploy	D_forex
tone(-1)	1								
D_GDP	0.42 *** (3.276)	1							
D_NX	-0.05 (-0.373)	-0.14 (-1.043)	1						
CPI growth	0.30 ** (2.125)	-0.29 ** (-2.179)	0.07 (0.524)	1					
bond yield	0.02 (0.150)	-0.30 ** (-2.263)	0.10 (0.736)	0.38 *** (3.035)	1				
PMI	0.29 ** (2.119)	0.46 *** (3.719)	-0.08 (-0.620)	0.39 *** (3.116)	-0.17 (-1.251)	1			

续表

	tone(-1)	D_GDP	D_NX	CPI growth	bond yield	PMI	P/E ratio	D_unemploy	D_forex
P/E ratio	-0.07 (-0.520)	0.08 (0.571)	-0.17 (-1.231)	0.53*** (4.583)	-0.10 (-0.761)	0.60*** (5.525)	1		
D_unemploy- ment	-0.30** (-2.225)	-0.14 (-0.995)	0.02 (0.139)	-0.19 (-1.399)	-0.12 (-0.889)	-0.50*** (-4.233)	-0.23* (-1.692)	1	
D_forex	0.14 (1.004)	0.14 (1.050)	0.07 (0.544)	-0.46*** (-3.754)	-0.22 (-1.638)	-0.31*** (-2.404)	-0.36*** (-2.798)	0.02 (0.144)	1

注:表3展示了相关性分析的结果。tone(-1)代表情绪单元法计算出的报告文本情绪的一阶滞后序列,D_GDP代表GDP同比增长率的一阶差分序列,D_NX代表净出口占同期GDP比重的一阶差分序列,CPI growth代表消费者价格指数的季度同比增长率,bond yield代表1年期国债到期收益率的季度均值,PMI代表制造业采购经理指数的季度均值,P/E ratio代表Wind全A指数静态市盈率的季度均值,D_unemployment代表城镇登记失业率数据的一阶差分序列,D_forex代表人民币兑美元汇率季度均值的一阶差分序列。括号内为Pearson相关性检验的t值。*/**/**分别代表系数在10%/15%/1%的水平上显著。

(三) 文本情绪与股票市场

1. 模型设定与变量选择

为了探究报告的文本情绪等文本指标是否会对股票市场有明显影响,我们参考 Ehrmann and Fratzscher(2009)、张强和胡荣尚(2014)等对央行沟通效力的研究的设定,采用 EGARCH 模型进行计量,并将上文计算出的文本指标加入均值方程和方差方程作为解释变量。我们设定均值方程如(5)式所示。其中 r 代表股票指数的涨跌幅, S_{gdp} 代表未预期到的GDP增长, S_{policy} 代表未预期到的货币政策操作,tone 代表报告的文本情绪。我们设定方差方程如(6)式所示,其中 D_{gdp} 、 D_{policy} 和 D_{report} 都为虚拟变量,分别表示是否公布GDP增长率数据、是否有货币政策操作和是否有报告发布。若在某一期有发布,则变量取1;反之,变量取0。参考 Ehrmann and Talmi(2017)的模型设定,我们将 TF-IDF 加权的文本相似度指标加入 EGARCH 模型的方差方程,检验货币政策报告的文本相似度是否会影响股票市场的波动性;参考 Loughran and McDonald(2014)将文本可读性指标加入 EGARCH 模型的方差方程,检验货币政策报告的可读性是否会影响股票市场的波动性。

$$r_t = \alpha + \beta r_{t-1} + \gamma S_{gdp_t} + \delta S_{policy_t} + \theta tone_t \quad (5)$$

$$\ln(\sigma_t^2) = a + b \left(\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + c \left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right) + d \ln(\sigma_{t-1}^2) + e D_{gdp_t} + f D_{policy_t} + g D_{report_t} + h Read_t + i Sim_t \quad (6)$$

在估计模型时,我们采用的股票收益数据即为上文提及的几个指数的日度收益率数据。衡量未预期到的GDP增长的方法,是以GDP同比增长率的实际数减去之前的预测数。预测数来源于北京大学中国经济研究中心公布的朗润经济预测数据。由于朗润经济预测数据的公开发布于2015年截止,我们使用第一财经首席经济学家调研的GDP同比增长率数据作为2016—2018年的预测数据。参考 Reid(2009)与熊海芳和王志强(2012)的设定,采用货币政策公布前后银行间市场7天同业拆借加权利率的差值作为未预期到

的货币政策操作的代理变量,如(7)式所示。需要说明的是,本文计入变量的货币政策操作包括法定存款准备金率变动公告、再贴现利率变动公告、1年期存贷款基准利率变动和常备借贷便利每月操作金额和利率的公告。

$$S_policy_t = interbank_rate_{t+1} - interbank_rate_{t-1} \quad (7)$$

2. 估计结果

表4给出了上述模型估计的结果。对均值方程而言,在控制GDP增速和货币政策操作两个变量后,报告的文本情绪对四个股票市场指数的走势都存在显著的正向影响。对方差方程而言,报告的发布会增加股票市场的波动性,但只在两个股票指数的模型估计中表现出统计显著性,影响比较微弱;报告的可读性对四个股票指数的影响都不具有较高的统计显著性,表明报告信息的理解难度大小并不会显著影响股票市场的波动性,说明央行报告的阅读和理解难度在市场可接受的范围内;报告文本相似度的增加会降低股票市场的波动性,且在三个股票指数的模型估计中表现出统计显著性,表明当央行报告包含较多新信息时,市场容易受到冲击,波动性增大;反之则波动性减小。总体来看,虽然报告可读性的增加并不会显著降低股票市场的波动性,但报告发布本身会对市场波动性造成微弱影响,报告文本情绪的提高会对股票市场收益率造成非常显著的正向影响;报告文本相似度的提高会较为显著地降低股票市场的波动性,表明央行货币政策报告的发布在整体上对股票市场有明显作用,侧面说明我国央行沟通具备较强的有效性。

表4 用情绪单元法计算出的文本情绪估计的EGARCH模型结果

变量	上证综指	深证成指	沪深300	中证500
均值方程:				
constant	0.0333* (1.688)	0.0224 (0.824)	0.0463** (2.130)	0.0293 (0.971)
r_{t-1}	0.00906 (0.501)	0.0277 (1.501)	0.0223 (1.199)	0.0527*** (2.960)
S_gdp_t	0.347** (2.540)	0.497** (2.576)	0.426*** (2.837)	0.258 (1.394)
S_policy_t	-0.413* (-1.832)	-0.625** (-2.133)	-0.534** (-2.072)	-0.652** (-2.268)
$tone_t$	0.541*** (3.648)	0.726*** (3.126)	0.744*** (4.186)	0.683*** (2.932)
方差方程:				
D_gdp_t	0.0217 (0.304)	0.0980 (1.403)	0.0377 (0.515)	0.0674 (0.918)
D_policy_t	0.0335 (1.013)	0.0770** (2.114)	0.0160 (0.472)	0.149*** (3.862)

续表

变量	上证综指	深证成指	沪深 300	中证 500
D_{report}_t	0.282 (1.140)	0.450 (1.584)	0.567* (1.918)	0.277 (1.047)
$Read_t$	0.00163 (0.092)	0.00648 (0.332)	0.0128 (0.678)	0.0294 (0.786)
Sim_t	-1.76* (-1.849)	-1.90** (-2.148)	-1.72* (-1.764)	-1.34 (-1.401)
R squared	0.008	0.011	0.011	0.013
N	3220	3220	3220	3220

注:表 4 展示了用情绪单元法计算出的文本情绪估计上文 EGARCH 模型的结果。其中 r 是股票价格指数的涨跌幅(单位为%) S_{gdp} 代表未预期到的 GDP 增长 S_{policy} 代表未预期到的货币政策操作 $tone$ 代表报告的文本情绪。 D_{gdp} 、 D_{policy} 和 D_{report} 都为虚拟变量,分别表示是否公布 GDP 增长率数据、是否有货币政策操作和是否有报告发布;若某一期有发布,则变量取 1;反之,则变量取 0。 $Read$ 代表发布的报告的可读性 Sim 代表本期报告与上一期报告的文本相似度。括号内为估计系数的 t 值。* / ** / *** 分别代表系数在 10% / 5% / 1% 的水平上显著。

3. 机制分析

上节的实证结果已经表明央行货币政策报告的文本情绪会显著影响股票市场走势。然而有另一个问题尚待解决:通常认为,央行货币政策报告既反映客观经济金融基本面,也反映央行的政策主张和对宏观经济金融的观点。文本情绪作为针对报告总体计算出的指标,自然也应包括这两部分,那么股票市场反应的对象是客观经济金融形势,还是央行的指引和政策主张?本节中,我们重新设定 EGARCH 模型来探究这个问题。

首先需要分离出报告中反映基本面和反映央行指引的部分。通过仔细阅读和比较各期报告,我们将报告分为八个部分:货币信贷概况、金融市场分析、国际经济回顾、国内经济回顾、国际经济展望、国内经济展望、货币政策操作和货币政策思路。前四部分为已有的经济金融信息,可以归类为反映基本面的部分;后四部分为央行对未来国内外局势的预估和货币政策相关分析,可以归类为反映央行指引的部分。我们分别基于金融情绪词典、用情绪单元法计算两部分的文本情绪,并将两个原始的文本情绪序列作 Z 值标准化,作为反映经济金融基本面和央行指引的指标。表 5 给出了新模型的估计结果。方程(1)和(4)仅加入基本面指标作为解释变量,且两个方程中基本面指标前的系数为正但统计上不显著,表明基本面指标对股票市场不具有显著影响。方程(2)和(5)仅加入央行指引指标作为解释变量,两个方程中央行指引指标前的系数为正且在统计上显著,表明央行指引指标对股票市场具有显著的正向影响。方程(3)和(6)同时加入了基本面和央行指引指标作为均值方程中的解释变量,两个方程中央行指引指标前系数均为正且显著,而基本面指标前系数为正但并不显著,证明报告中主要影响市场的是反映央行指引的部分,而反映基本面的部分对市场影响并不显著。综合以上分析,可以看出投资者更关注央行在报告中的政策指引,对报告中回顾的宏观经济金融基本面信息反应不明显。考虑到报告中的宏观经济金融数据在报告发布前都已发布,市场已经对基本面信息作出反应,以上实证分

析的结果符合现实经济信息发布和市场运行的规律。

表 5 将文本情绪分离估计 EGARCH 模型的结果

变量	上证综指			深证成指		
	(1)	(2)	(3)	(4)	(5)	(6)
均值方程:						
constant	0.034* (1.745)	0.034* (1.723)	0.035* (1.763)	0.0235 (0.865)	0.0250 (0.923)	0.0249 (0.918)
r_{t-1}	0.00836 (0.461)	0.00819 (0.451)	0.00859 (0.473)	0.0265 (1.431)	0.0272 (1.475)	0.0271 (1.465)
S_gdp_t	0.346** (2.514)	0.346** (2.525)	0.347** (2.539)	0.494** (2.548)	0.496*** (2.572)	0.497*** (2.574)
S_policy_t	-0.417* (-1.841)	-0.416* (-1.840)	-0.416* (-1.844)	-0.624** (-2.114)	-0.625** (-2.115)	-0.625** (-2.118)
$real_t$	0.143 (0.744)		0.250 (1.367)	0.192 (0.954)		0.102 (0.510)
$guid_t$		0.291** (1.988)	0.324** (2.131)		0.421** (2.063)	0.423** (2.039)
方差方程:						
D_gdp_t	0.0153 (0.214)	0.0169 (0.237)	0.0172 (0.241)	0.106 (1.512)	0.103 (1.475)	0.103 (1.475)
D_policy_t	0.0343 (1.033)	0.0333 (1.006)	0.0345 (1.039)	0.0756** (2.077)	0.0762** (2.086)	0.0761** (2.084)
D_report_t	0.312 (1.255)	0.321 (1.300)	0.291 (1.169)	0.460 (1.596)	0.478* (1.661)	0.469 (1.631)
$Read_t$	0.000415 (0.023)	0.00217 (0.122)	0.00168 (0.094)	-0.00628 (-0.316)	-0.00939 (-0.475)	-0.00815 (-0.412)
Sim_t	-1.81** (-1.973)	-1.71** (-2.001)	-1.75** (-2.017)	-2.04** (-2.314)	-2.05** (-2.437)	-2.03** (-2.311)
R squared	0.005	0.006	0.007	0.007	0.009	0.009
N	3220	3220	3220	3220	3220	3220

注:表 5 展示了将文本情绪分离估计 EGARCH 模型的结果。其中 r 是股票价格指数的涨跌幅(单位为%) S_gdp 代表未预期到的 GDP 增长 S_policy 代表未预期到的货币政策操作 $guid$ 代表报告中反映央行指引部分的文本情绪, $real$ 代表报告中反映基本面部分的文本情绪。 D_gdp 、 D_policy 和 D_report 都为虚拟变量,分别表示是否公布 GDP 增长率数据、是否有货币政策操作和是否有报告发布;若某一期有发布,则变量取 1;反之,则变量取 0。 $Read$ 代表报告的可读性, Sim 代表本期报告与上一期报告的文本相似度。括号内为估计系数的 t 值。* / ** / *** 分别代表系数在 10% / 5% / 1% 的水平上显著。

4. 稳健性检验

前文估计 EGARCH 模型时,我们使用了基于姜富伟等(2020)开发的金融情感词典计算出的文本情绪。本节换用权威的通用情感词典,即知网(HowNet)情感词典来计算文本

情绪,并在不改变其他变量的条件下重新估计上述两个 EGARCH 模型。模型估计结果与前述实证分析结果一致。

六、结论与建议

本文利用文本大数据分析技术和中文财经情感字典方法,分析中国人民银行货币政策执行报告的文本情绪、文本相似度和文本可读性等语义文本指标信息,探究了货币政策报告文本情绪和宏观经济状况的相关性,用 EGARCH 模型实证分析政策报告的文本指标对股票市场的影响。我们发现,货币政策报告的文本情绪与诸多宏观经济指标有着显著的相关性。在控制经济增长和货币政策后,报告文本情绪对报告发布后股票市场的收益率有显著的正向影响。报告文本相似程度越高,股票市场波动性越小;反之,报告文本相似程度越低,股票市场波动性越大。报告文本可读性高低对股票市场波动性的影响并不显著。进一步研究发现,报告文本情绪中对股票市场有显著影响的是其反映央行政策指引的部分,而反映宏观经济基本面状况的部分对股票市场影响并不显著。

本文的研究结论有一定政策含义。一方面,本文的实证结果表明央行沟通可以显著影响股票市场的收益率和波动性,从而证明我国央行沟通具备有效性,监管当局可通过沟通方式更好地传达政策意图,引导市场预期从而防范金融风险。另一方面,本文进一步指出,央行报告的文本情绪中影响市场的部分是反映央行对未来经济形势的判断和政策前瞻的部分,从数据层面证实了央行指引对股票市场的显著影响,表明央行有效预期管理的重要性。

参考文献

- [1] 卞志村和张义 2012,《央行信息披露、实际干预与通胀预期管理》,《经济研究》第 12 期,第 15~28 页。
- [2] 陈银娥和江媛 2017,《管理层权力、制度环境与董事会报告可读性——来自我国上市公司的经验证据》,《珞珈管理评论》第 3 期,第 84~99 页。
- [3] 冀志斌和宋清华 2012,《中央银行沟通的金融市场效应——基于中国数据的实证分析》,《宏观经济研究》第 9 期,第 45~53 页。
- [4] 姜富伟、孟令超和唐国豪 2020,《媒体文本情绪与股票回报预测》,《经济学(季刊)》已接收。
- [5] 姜富伟、涂俊、David E. Rapach、Jack K. Strauss 和周国富 2011,《中国股票市场可预测性的实证研究》,《金融研究》第 9 期,第 107~121 页。
- [6] 林建浩和赵文庆 2015,《中国央行沟通指数的测度与谱分析》,《统计研究》第 1 期,第 52~58 页。
- [7] 罗炜、何顶、洪莉莎和常国珍 2017,《媒体报道可以预测创业企业的发展前景吗》,《金融研究》第 8 期,第 177~191 页。
- [8] 吴国培和潘再见 2014,《中央银行沟通对金融资产价格的影响——基于中国的实证研究》,《金融研究》第 5 期,第 34~47 页。
- [9] 谢德仁和林乐 2015,《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+第 93 页。
- [10] 熊海芳和王志强 2012,《货币政策意外、利率期限结构与通货膨胀预期管理》,《世界经济》第 6 期,第 30~55 页。

- [11] 游家兴和吴静 2012,《沉默的螺旋: 媒体情绪与资产误定价》,《经济研究》第 7 期,第 141 ~ 152 页。
- [12] 张强和胡荣尚 2013,《中央银行沟通对金融资产价格的影响——以股票市场为例》,《财贸经济》第 8 期,第 67 ~ 77 页。
- [13] 张强和胡荣尚 2014,《中央银行沟通对利率期限结构的影响研究》,《国际金融研究》第 6 期,第 10 ~ 20 页。
- [14] 朱宁、许艺焯和邱光辉 2016,《中央银行沟通对人民币汇率波动的影响》,《金融研究》第 11 期,第 32 ~ 46 页。
- [15] Acosta M. , and E. Meade 2015. “Hanging on Every Word: Semantic Analysis of the FOMC’s Postmeeting statement” , *FEDS Notes*.
- [16] Amaya D. , and J. -Y. Filbien. 2015. “The Similarity of ECB’s Communication” , *Finance Research Letters* , 13: 234-242.
- [17] Baker , S. , N. Bloom , and S. Davis , 2016. “Measuring Economic Policy Uncertainty” , *The Quarterly Journal of Economics* , 131(4) : 1593 ~ 1636.
- [18] Beine M. , G. Janssen , and Lecourt C. 2009. “Should Central Bankers Talk to the Foreign Exchange Markets?” , *Journal of International Money and Finance* , 28(5) : 776 ~ 803.
- [19] Cieslak A. , A. Morse and Vissing Jorgensen A. 2019. “Stock Returns over the FOMC Cycle” , *The Journal of Finance* , 74: 2201 ~ 2248.
- [20] Demiralp , S. and J. Oscar. 2002. “The Announcement Effect Evidence from Open Market Desk Data” , *FRBNY Economic Policy Review* , 8(1) : 1 ~ 20.
- [21] Ehrmann , M. , and M. Fratzscher 2009. “Purdah – on The Rationale for Central Bank Silence around Policy Meetings” , *Journal of Money , Credit and Banking* , 41(2 - 3) : 517 ~ 528.
- [22] Ehrmann , M. , and J. Talmi 2017. “Starting from a Blank Page? Semantic Similarity in Central Bank Communication and Market Volatility” Staff Working Papers.
- [23] Garcia , D. 2013. “Sentiment During Recessions” , *Journal of Finance* , 68(3) : 1267 ~ 1300.
- [24] Guthrie , G. , and J. Wright. 2000. “Open Mouth Operations” , *Journal of Monetary Economics* , 46: 489 ~ 516.
- [25] Huang , A. , A. Zang , and R. Zheng 2014. “Evidence on the Information Content of Text in Analyst Reports” , *Accounting Review* , 89(6) : 2151 ~ 2180.
- [26] Jansen , David - Jan , and Jakob De Haan. 2007. “Were Verbal Efforts to Support the Euro Effective: A High - frequency Analysis of ECB Statements” , *European Journal of Political Economy* , 23: 245 ~ 259.
- [27] Jansen , D. J. 2011. “Does the Clarity of Central Bank Communication Affect Volatility in Financial Markets? evidence from humphrey hawkins Testimonies” , *Contemporary Economic Policy* , 29(4) : 494 ~ 509.
- [28] Jegadeesh , N. , and D. Wu 2013. “Word Power: A New Approach for Content Analysis” , *Journal of Financial Economics* , 110(3) : 712 ~ 729.
- [29] Jiang , F. , Lee J. , X. Martin , and G. Zhou 2019. “Manager Sentiment and Stock Returns” , *Journal of Financial Economics* , 132(1) : 126 ~ 149.
- [30] Loughran , T. , and B. Mcdonald 2011. “When is a Liability not a Liability? Textual Analysis , Dictionaries , and 10 - Ks” , *Journal of Finance* , 66(1) : 35 ~ 65.
- [31] Loughran , T. , and B. Mcdonald 2014. “Measuring Readability in Financial Disclosures” , *The Journal of Finance* , 69(4) : 1643 ~ 1671.
- [32] Lucca , D. O. , and E. Moench 2015. “The pre - fomic announcement drift” , *Journal of Finance* , 70(1) : 329 ~ 371.
- [33] Reid , M. 2009. “The Sensitivity of South African Inflation Expectations to Surprises” , *South African Journal of Economics* , 77(3) : 414 ~ 429.
- [34] Ricardo , C. , G. Keshav , Londono J. M. , and Nathan M. 2021. “Sentiment in Central Banks’ Financial Stability Reports” , *Review of Finance* 25(1) 85 ~ 120.

[35] Tetlock, P. C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", *Journal of Finance*, 62(3): 1139 ~ 1168.

Textual Information of Central Bank Monetary Policy Report , Macroeconomy and Stock Market Performance

JIANG Fuwei HU Yichi HUANG Nan

(School of Finance , Central University of Finance and Economics;
School of Economics , Peking University; Harvest Fund Management Co. , Ltd)

Summary: Since the 1990s , central bank communication has become a hot issue in macroeconomics and finance. Many scholars have conducted meaningful research on the issues such as the measurement of central bank communication , central bank communication and inflation expectations , and central bank communication and financial markets. Among them , the influence of central bank communication on financial markets and asset prices has received wide attention. A large number of empirical studies have shown that central bank communication has a significant impact on the stock market , the bond market and the foreign exchange market. As the People's Bank of China (PBOC) has paid increasing attention to policy communication in recent years , many Chinese scholars have conducted research on the impact of PBOC's communication on China's financial markets. However , there are two major shortcomings in the existing studies. First , they only focus on monetary policy tendency of PBOC's communication and ignore other information contained in the communication. Second , most scholars construct quantitative indicators by manual reading and scoring , making the results highly subjective.

This paper uses text analysis techniques to analyze 71 Monetary Policy Implementation Reports (hereinafter referred to as "the reports") of PBOC , calculates the text sentiment (tone) , the similarity and readability and other text indicators of the reports , and explores the relationship between these text indicators and the macro economy and the stock market. Based on the Chinese financial sentiment dictionary developed by Jiang et al. (2020) , this paper uses the sentiment unit method to calculate the tone of the reports. In addition , this paper uses TF - IDF weighted cosine similarity to characterize the similarity of the reports , and uses average sentence length to characterize the readability of the reports. The paper then uses correlation analysis to examine the relationship between the tone of the reports and macroeconomic indicators such as economic growth , inflation , and interest rates. With reference to Ehrmann and Fratzscher (2009) , Zhang and Hu (2014) , this paper adds tone , similarity and readability to the EGARCH model to explore whether textual indicators of the reports affect stock market returns and the volatility on the trading day after the release. Furthermore , this paper decomposes the content of the reports into two parts: economic and financial fundamentals and central bank policy guidelines , calculates the tone of the two parts and examines their impacts on the stock market respectively.

The empirical results show that the tone of the reports is significantly correlated with macroeconomic indicators such as economic growth , inflation , and employment levels , and higher value of tone indicates better economic situation. After controlling for variables such as economic growth and monetary policy , the tone of the

report has a significant positive impact on stock market return after the report is released. The similarity of the report has a significantly negative impact on stock market volatility, whereas the readability of the report does not have a significant impact on stock market volatility. Further research shows that it is the part reflecting the central bank's policy guidelines rather than the part reflecting macroeconomic and financial fundamentals that has a significant impact on stock market returns.

This paper fills a number of gaps in the field of central bank communication and text analysis. First, this paper is the first to use cutting-edge text analysis techniques to conduct a comprehensive analysis of the monetary policy reports of PBOC. Second, this paper fills the gap in quantitative analysis of the sentiment of central bank communication in China's academia. Third, this paper conducts an empirical study on the mechanism of PBOC's communication affecting the stock market, and proves that the report's tone affects the stock market only through the policy guidance channel.

The findings of this paper are of great significance to strengthening financial supervision and promoting macro-prudential management in China. The results show that PBOC communication can significantly affect the stock market, which fully affirms the effectiveness of PBOC's communication. Through adequate communication with the market, PBOC can influence asset prices, thereby achieving the purpose of monetary policy regulation and maintaining financial stability. In addition, this paper points out that the part of PBOC's reports that affects the market is the part that reflects central bank's predictions of the future economic situation and its policy guidance. Therefore, PBOC should use its authority and influence to manage market expectations more effectively through timely announcements and clear explanations of economic and financial situation predictions and monetary policy guidance.

Keywords: Textual Sentiment Analysis, Central Bank Communication, Stock Market, Economy

JEL Classification: C51, G12, G18

(责任编辑: 林梦瑶) (校对: ZL)