

《预期的博弈：央行沟通与国债收益率曲线》附录

附录 1 关于央行沟通文本处理的说明

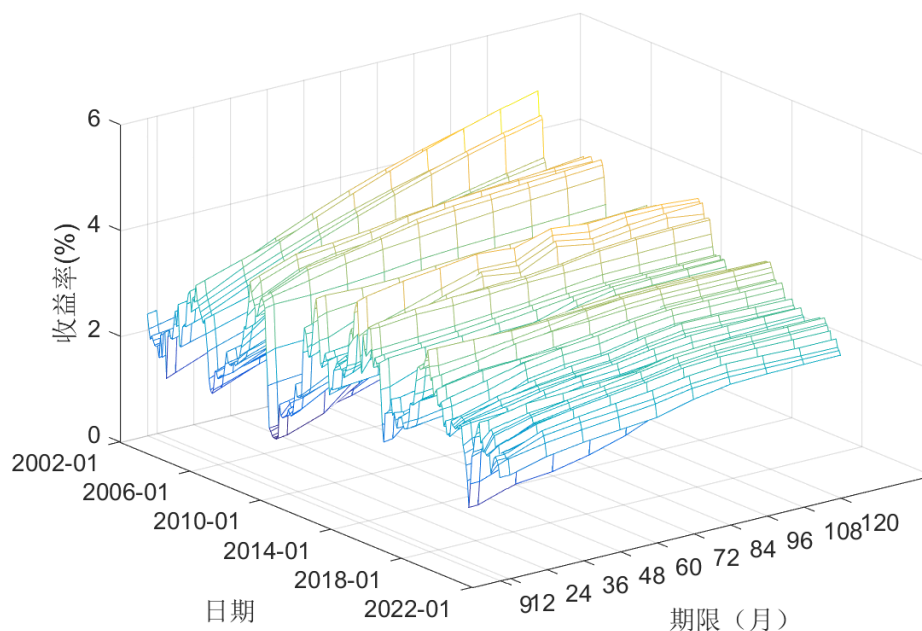
央行主要通过书面沟通和口头沟通两种形式向公众传递经济形势和货币政策信息。

本文从中国人民银行官网获取了自 2001 年第一季度至 2023 年第三季度公布的所有《中国货币政策执行报告》，共计 91 篇报告作为书面沟通文本语料。按照标准的文本数据预处理方法，将 PDF 文件格式转化成 TXT 纯文本格式以便于文本数据处理，删除文本中的数字、标点符号等特殊字符，并去除停用词，得到规范的书面沟通文本数据。

本文利用百度新闻搜索引擎，以“行长姓名+经济形势”等为关键词搜索并获取了含会议致辞、论坛演讲、答记者问等各类央行口头沟通信息。同时，从央行官网信息公开沟通交流专栏中下载行长重要讲话、重要活动等相关新闻报道予以补充和校对。收集整理了 2001 年 1 月至 2023 年 9 月所有的口头沟通文本数据，作为本研究的口头沟通文本语料。由于口头沟通文本信息大多来源于新闻报道的网页抓取数据，具有非结构化、非标准化的数据形式，网页中通常包含大量的导航、广告等无用信息以及重复信息，在使用前须对这些数据进行必要的清洗和预处理。首先，针对从网页中获取的原始口头沟通文本，过滤其 html 标签、广告文本、导航栏内容等噪声和冗余信息，只保留与本文研究有关的央行沟通主体内容。其次，网页文本存在不同媒体渠道对同一新闻事件进行重复报道或转载的情况。因此，需要对原始口头沟通文本语料进行规范化去重处理。最终得到去除噪声、冗余信息和重复内容后的 400 余篇有效央行口头沟通文本，将其转化成 TXT 纯文本格式，删除文本中的数字、标点符号等特殊字符，并去除停用词，得到规范的口头沟通文本数据。

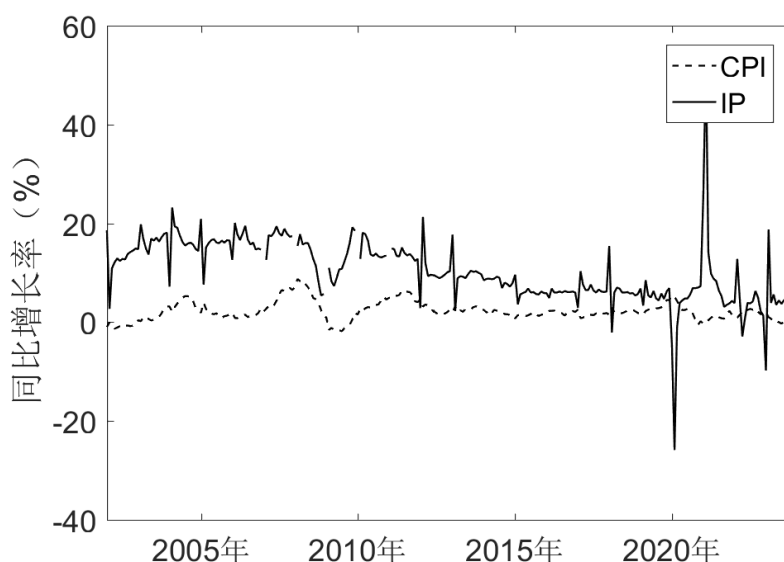
附录 2 国债收益率与宏观经济信息指标时序图

本文使用中债信息网公布的到期期限为 9 个月及 1~10 年期银行间市场国债收益率，样本区间为 2002 年 1 月至 2023 年 11 月，如附图 1 所示。



附图 1 各期国债收益率曲线

参照尚玉皇和张皓越（2023），本文选取了居民消费价格指数（CPI）和工业增加值的月度同比变化率作为通货膨胀率和实体经济活动的代理变量，数据如附图 2 所示。



附图 2 宏观经济指标走势图

附录 3 稳健性检验

中国人民银行每季度发布的《中国货币政策执行报告》一般会滞后两个月，按照沟通情绪的实际发生时间引入期限结构模型中，估计结果如附表 1 所示。

附表 1 按照实际发布时间的转移方程系数矩阵估计结果

$\mu_{f,L}^0 \times (10^{-5})$	8.0932	λ	0.0630	$\sigma^2 \times (10^{-10})$	1.4568		
	L_{t-1}	S_{t-1}	C_{t-1}	CPI_{t-1}	IP_{t-1}	$Senti_{t-1}$	μ
L_t	0.8967 [0.84, 0.96]	0.0245 [-0.02, 0.07]	0.012 [-0.01, 0.04]	0.0163 [-0.00, 0.03]	0.0042 [-0.00, 0.01]	-0.0775 [-0.11, -0.04]	0.0043 [0.00, 0.01]
S_t	0.0101 [-0.08, 0.10]	0.7927 [0.72, 0.86]	0.0309 [-0.01, 0.07]	0.0024 [-0.02, 0.03]	0.0003 [-0.01, 0.01]	0.1241 [0.07, 0.18]	-0.0028 [-0.01, 0.00]
C_t	0.147 [-0.01, 0.30]	0.1272 [0.01, 0.25]	0.7778 [0.71, 0.85]	-0.0507 [-0.09, -0.01]	-0.0028 [-0.01, 0.01]	0.3032 [0.20, 0.40]	-0.0094 [-0.01, -0.00]
CPI_t	0.0077 [-0.12, 0.14]	-0.0717 [-0.17, 0.03]	0.1042 [0.05, 0.16]	0.927 [0.89, 0.96]	0.0397 [0.03, 0.05]	0.1057 [0.03, 0.18]	-0.0009 [-0.00, 0.00]
IP_t	0.9305 [-0.26, 2.15]	-0.4499 [-1.42, 0.48]	-0.444 [-1.00, 0.11]	0.1458 [-0.18, 0.48]	0.5191 [0.43, 0.61]	-0.773 [-1.51, -0.04]	-0.0123 [-0.05, 0.03]
$Sent_t$	-0.0378 [-0.15, 0.07]	0.0027 [-0.09, 0.09]	0.1451 [0.08, 0.22]	-0.0179 [-0.05, 0.01]	0.0085 [-0.00, 0.02]	0.728 [0.57, 0.84]	0.0056 [0.00, 0.01]

为了降低央行沟通的文本情绪测量误差的影响，本文对姜富伟等（2021）的中文金融情感词典进行了一定的补充与完善以更好地适用于央行沟通文本。通过人工仔细阅读货币政策执行报告原文，同时参考 Heineman and Ulrich(2007)、卞志村和张义(2013)、林建浩和赵文庆（2015）等中外权威文献关于央行沟通措辞的选取，根据货币报告所传达的政策情绪信息，提取了 200 个针对央行沟通的专属情感措辞，添加至姜富伟的情感词语列表中，形成新的情感计算词典。最后，以该词典为基础，构造新的央行沟通情绪代理变量 $Sent_2$ ，并将该变量代入混频仿射利率期限结构模型中，估计结果如附表 2 所示。

附表 2 替换情绪变量后转移方程系数矩阵的估计结果

$\mu_{f,L}^Q \times (10^{-5})$	8.9330	λ	0.0675	$\sigma^2 \times (10^{-10})$	4.2508	
	L_{t-1}	S_{t-1}	C_{t-1}	CPI_{t-1}	IP_{t-1}	$Senti_{t-1}$
L_t	0.9003	0.028	0.0098	0.0059	0.0057	-0.0551
						μ
						0.0054

	[0.84, 0.96]	[-0.02, 0.08]	[-0.02, 0.04]	[-0.01, 0.02]	[0.00, 0.01]	[-0.09, -0.02]	[0.00, 0.01]
S_t	-0.0005	0.7812	0.0321	0.0196	-0.0021	0.0874	-0.003
	[-0.09, 0.09]	[0.71, 0.85]	[-0.01, 0.07]	[-0.01, 0.04]	[-0.01, 0.00]	[0.03, 0.14]	[-0.01, 0.00]
C_t	0.1443	0.1266	0.786	-0.0086	-0.0087	0.2535	-0.0147
	[-0.02, 0.31]	[-0.01, 0.26]	[0.72, 0.86]	[-0.06, 0.04]	[-0.02, 0.00]	[0.15, 0.36]	[-0.02, 0.01]
CPI_t	0.0041	-0.0746	0.0877	0.9459	0.0375	0.1352	0
	[-0.12, 0.13]	[-0.18, 0.03]	[0.03, 0.14]	[0.91, 0.98]	[0.03, 0.05]	[0.06, 0.21]	[-0.00, 0.01]
IP_t	0.9577	-0.5587	-0.1939	-0.0695	0.5405	-1.7043	-0.0197
	[-0.22, 2.11]	[-1.50, 0.39]	[-0.73, 0.33]	[-0.39, 0.27]	[0.46, 0.63]	[-2.44, -0.95]	[-0.07, 0.03]
$Senti_t$	-0.0126	-0.1421	0.1444	-0.0283	-0.0302	0.543	0.0093
	[-0.15, 0.13]	[-0.27, -0.03]	[0.08, 0.21]	[-0.07, 0.01]	[-0.05, -0.02]	[0.43, 0.65]	[0.00, 0.02]

央行沟通文本包括书面和口头两种形式。此前，我们基于最规范和系统的书面沟通文本——季度《中国货币政策执行报告》构造了央行沟通情绪指数。为了研究的完整性和准确性，此处我们将口头沟通文本纳入指标构造样本中。值得注意的是，纳入口头沟通的央行沟通指数理论上可以处理为月度指标，减少使用季度指标造成的信息损失。但在实际操作中，将口头沟通和书面沟通按照月度频率合并信息，会导致书面文本发布的那四个月的文本沟通信息量激增，使得沟通情绪数据具有明显的季节性。因此，本文按照书面和口头沟通所在季度进行了季度信息合并，构造了书面+口头的季度央行沟通情绪指数，如正文图 3 所示，新指数与原书面沟通指数的波动趋势大体一致。将新指数代入混频仿射利率期限结构模型，估计结果如附录表 3。

附表 3 基于书面+口头文本信息的央行沟通情绪指的转移方程系数矩阵估计结果

$\mu_{f,L}^Q \times (10^{-5})$	9.4469	λ	0.0680	$\sigma^2 \times (10^{-10})$	1.7792		
	L_{t-1}	S_{t-1}	C_{t-1}	CPI_{t-1}	IP_{t-1}	$Senti_{t-1}$	μ
L_t	0.9026	0.0286	0.0107	0.0157	0.0041	-0.0809	0.004
	[0.84, 0.96]	[-0.02, 0.08]	[-0.01, 0.04]	[-0.00, 0.03]	[-0.00, 0.01]	[-0.12, -0.04]	[0.00, 0.01]
S_t	0.0056	0.7839	0.0329	0.0036	0.0006	0.1138	-0.0024
	[-0.09, 0.10]	[0.71, 0.86]	[-0.01, 0.07]	[-0.02, 0.03]	[-0.01, 0.01]	[0.06, 0.17]	[-0.01, 0.00]
C_t	0.1419	0.1289	0.7869	-0.0583	-0.0012	0.359	-0.0092
	[-0.02, 0.31]	[-0.00, 0.26]	[0.72, 0.86]	[-0.10, -0.01]	[-0.01, 0.01]	[0.25, 0.47]	[-0.01, -0.00]
CPI_t	-0.0141	-0.0864	0.1048	0.9222	0.04	0.1207	0.0002
	[-0.15, 0.12]	[-0.19, 0.02]	[0.05, 0.16]	[0.89, 0.96]	[0.03, 0.05]	[0.04, 0.20]	[-0.00, 0.00]
IP_t	0.7201	-0.3763	-0.4647	0.1345	0.5629	-0.6171	-0.0057
	[-0.41, 1.93]	[-1.32, 0.58]	[-0.97, 0.04]	[-0.19, 0.45]	[0.47, 0.65]	[-1.35, 0.10]	[-0.04, 0.03]
$Senti_t$	-0.0279	-0.02	0.1831	-0.0133	0.0114	0.6266	0.0058
	[-0.17, 0.10]	[-0.14, 0.09]	[0.11, 0.26]	[-0.05, 0.02]	[0.00, 0.02]	[0.44, 0.77]	[0.00, 0.01]

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] 姜富伟，胡逸驰和黄楠，2021，《央行货币政策报告文本信息、宏观经济与股票市场》，《金融研究》第 6 期，第 95~113 页。
- [3] 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。
- [4] 尚玉皇和张皓越，2023，《国债收益率与高频宏观因子——基于非规则混频利率期限结构模型》，《统计研究》第 10 期，第 109~123。
- [5] Heinemann F. and K. Ullrich, 2007, “Does it Pay to Watch Central Bankers’ Lips? The Information Content of ECB Wording”, *Swiss Journal of Economics and Statistics*, 143(2), pp. 155~185.