

《监管一体化对上市公司绩效的影响》附录

附录 1 监管一体化指标测算流程

第一步是基于法律法规文本数据归纳出监管议题。LDA（潜在狄利克雷分配）算法能够在给定主题数量的情况下，基于文本进行主题分类，每个主题是由在文本中倾向于共同出现的词语聚类而成；除此之外，LDA 还提供了每个文档中各类主题出现的比例。本文基于北大法宝中央法规数据库给定的 99 个一级法规类别作为基准，设定 99 个监管议题。本研究采用 TF-IDF 加权的 LDA 方法，以更好地揭示中央法规涵盖的议题结构与政策关注点。

第二步是在使用 LDA 归纳出监管议题的基础上计算每类议题的一体化程度。首先，使用训练后的 LDA 模型对所有法律法规文件进行拟合，得到各个法律法规文件在各个议题上的词频。其次，根据各个法律法规文件的起草机构，按“机构-年份”汇总文件，并计算每个“机构-议题-年份”的词数。然后，对于每个“议题-年份”，确定每个机构对该议题的监管比例。以下为计算公式（1），其中， $\omega_{i,a,t}$ 为议题 i 第 t 年在机构 a 的占比， $Integ_{i,t}$ 表示议题 i 在第 t 年的监管一体化程度。

$$Integ_{i,t} = \sum_a \omega_{i,a,t}^2 \quad (A1)$$

最后，计算企业层面的监管一体化程度。本文整理了 2007 年至 2023 年间的公司年报文本数据，首先将之前在中央法规文件上训练过的 LDA 模型拟合每个上市公司的年报，得到公司 f 在年份 t 的年报中议题 i 的词频比例 $P_{f,i,t}$ ，该词频表示了公司对特定监管议题的监管一体化水平的暴露程度；然后以该词频为权重，将各议题的监管一体化程度加权求和，得到上市公司 f 在年份 t 时的监管一体化程度 $Integ_{f,t}$ ，如公式（2）所示。

$$Integ_{f,t} = \sum_i P_{f,i,t} \times Integ_{i,t} \quad (A2)$$

附录 2 议题分类结果示例

本文将北大法宝的 99 个法规类别以及每个议题下出现频率最高的 30 个词提供给大语言模型¹，并要求它基于议题下的词汇以及北大法宝的法规类别对各议题进行命名。本文允许大语言模型在北大法宝的法规类别不能有效命名时给出新的命名。附图 1 展示了部分分类结果，图中词汇的大小表示其在该议题下出现的频率。附图 1 左侧为议题 2：证券，其中最广泛出现的前五个词汇为“基金、发行、证券、投资、公司”，附图 1 右侧为议题 9：银行，其中最广泛出现的前五个词汇为“银行、分行、金融、商业银行、履职”。



附图 1 议题分类示例

¹ 平均来看，前 30 个词在各主题下出现的频率占 58.9%，能够充分代表该主题的主要内容；本文使用的大语言模型为 DeepSeek-V3。

附录 3 内生性处理与稳健性检验结果

(一) 内生性分析

附表 1 估计结果表明,工具变量回归与基准回归结果仍大致保持一致,同时 K-P rk Wald F 统计量拒绝了该工具变量存在识别不足或弱工具变量的原假设。上述分析表明,该工具变量较好地满足了外生性条件,有效缓解了基准结果中可能存在的内生性问题。

附表 1 稳健性检验:工具变量回归

面板A: 二阶段回归结果				
变量名	(1) <i>ROA</i>	(2) <i>ROA</i>	(3) <i>Gsales</i>	(4) <i>Gsales</i>
<i>Integ</i>	1.7116*** (2.9148)	1.3345** (2.5452)	10.1723*** (2.6495)	8.4530** (2.2051)
观测值	42,747	42,747	42,740	42,740
R^2	-0.0194	0.0659	-0.0076	0.0194
控制变量	No	Yes	No	Yes
企业固定效应	Yes	Yes	Yes	Yes
行业-年份固定效应	Yes	Yes	Yes	Yes
K-P rk LM	219.6956	216.8831	219.8659	216.9603
K-P rk Wald F	113.5490	110.6586	113.5969	110.6980
面板B: 一阶段回归结果				
变量名	(1) <i>Integ</i>	(2) <i>Integ</i>	(3) <i>Integ</i>	(4) <i>Integ</i>
<i>Integ_{IV}</i>	0.1761*** (10.6559)	0.1756*** (10.5194)	0.1762*** (10.6582)	0.1757*** (10.5213)
观测值	42,747	42,747	42,740	42,740
控制变量	No	Yes	No	Yes
企业固定效应	Yes	Yes	Yes	Yes
行业-年份固定效应	Yes	Yes	Yes	Yes

本文还进一步利用 2018 年国务院发布的《关于加快推进全国一体化在线政务服务平台建设的指导意见》这一准自然实验进行政策冲击分析。本文计算企业在 2018 年之前的平均监管碎片化程度,记作 *Frag_pre* ($1-Integ_pre$);同时定义政策时间虚拟变量 *Post*,若年份大于等于 2018 记为 1,否则为 0。最后将交互项 *Frag_pre*Post* 并代入基准模型进行估计。回归结果如附表 2 所示。《意见》实施后,原先监管一体化程度较低的企业经营绩效显著提升。

附表 2 稳健性检验:政策冲击检验

变量名	(1) <i>ROA</i>	(2) <i>Gsales</i>
<i>Fragmentation_pre*Post</i>	0.3197*** (5.4905)	2.3860*** (8.0769)
常数项	-0.0741** (-2.2211)	1.3982*** (7.3401)
观测值	37,170	37,165
R^2	0.5186	0.2158
控制变量	Yes	Yes
企业固定效应	Yes	Yes
行业-年份固定效应	Yes	Yes

(二) 剔除疫情期间样本

新冠疫情很可能在降低了企业的监管一体化程度的同时，放大了市场运行的波动性进而导致企业经营绩效下滑。本文使用 2007-2019 年的样本对基准结果重新进行估计，结果如附表 3 列（1）列（2）所示，监管一体化的影响与基准结果仍基本保持一致。

（三）替换核心变量

本文使用净资产回报率 *ROE* 和资产增长率 *Gasset* 来替换基准回归中使用的 *ROA* 和 *Gsales*，结果如附表 3 列（3）列（4）所示。本文还使用企业的全要素生产率 *TFP* 作为衡量企业经营绩效的综合指标，代入基准模型进行回归。结果如附表 3 列（5）所示。

本文改变预设议题数为 75 个重新训练 LDA 模型，并基于新的分类结果重新测算企业一体化程度，记为 *Integ_robust*，代入基准模型重新估计。结果见附表 3 列（6）列（7）。结果表明在更改预设议题数后，监管一体化程度对各变量的影响仍与基准结果保持一致，确保了基准结果的稳健性。

附表 3 稳健性检验：剔除新冠疫情影响及替换变量

变量名	(1) <i>ROA</i>	(2) <i>Gsales</i>	(3) <i>ROE</i>	(4) <i>Gasset</i>	(5) <i>TFP</i>	(6) <i>ROA</i>	(7) <i>Gsales</i>
<i>Integ</i>	0.2371*** (3.4002)	2.0945*** (3.9526)	0.6307*** (4.3557)	0.8229*** (2.6987)	2.4650*** (4.6024)		
<i>Integ_robust</i>						0.1046** (2.3418)	0.7274** (2.2886)
常数项	0.0420 (1.3031)	2.4659*** (10.1747)	-0.0154 (-0.2509)	3.0467*** (21.8740)	0.0718 (0.2612)	0.0591** (2.4509)	2.4618*** (15.4475)
观测值	26,675	26,670	42,686	42,747	42,500	42,747	42,740
<i>R</i> ²	0.5563	0.2433	0.4382	0.3014	0.8172	0.5447	0.2345
控制变量	Yes	Yes	Yes	Yes	Yes	Yes	Yes
企业固定效应	Yes	Yes	Yes	Yes	Yes	Yes	Yes
行业-年份固定效应	Yes	Yes	Yes	Yes	Yes	Yes	Yes

（四）增加控制变量

本文进一步控制年报文本的信息质量相关的控制变量，包括年报文本语调（Jiang et al., 2019; 姜富伟等, 2021）、文本信息的确定程度（阮睿等, 2021）、风险披露情况（王雄元等, 2017）、前瞻性信息含量（许帅等, 2025）以及年报文本的可读性和简洁性。年报文本语调使用年报中积极词汇数减去消极词汇数，并使用年报总词汇数进行标准化。文本信息的确定程度使用（程度确定词语数量－程度模糊词语数量）/程度词语总数量衡量；风险披露情况用年报中风险相关词汇占总词汇数的词频衡量；前瞻性信息使用年报中前瞻词汇占总词频的比重衡量。年报可读性由年报句子的平均词汇数量衡量，简洁性则由年报词汇数量的自然对数衡量。其次，我们加入了董事长个人固定效应。最后，我们还控制了深交所和上交所对上市企业信息披露质量考评结果。回归结果如附表 4 所示。

附表 4 稳健性检验：更换模型设定

变量名	(1) <i>ROA</i>	(2) <i>Gsales</i>	(3) <i>ROA</i>	(4) <i>Gsales</i>	(5) <i>ROA</i>	(6) <i>Gsales</i>
<i>Integ</i>	0.3786*** (6.1726)	2.8955*** (7.1193)	0.3257*** (4.7401)	2.4619*** (4.9971)	0.3302*** (4.1994)	2.5972*** (4.6413)
常数项	0.0707* (1.7464)	-1.3647*** (-5.3968)	0.3195*** (5.9047)	-0.0265 (-0.0667)	0.3717*** (5.8059)	-0.3642 (-0.7939)
观测值	42,747	42,740	41,070	41,067	33,361	33,359
<i>R</i> ²	0.5473	0.2528	0.6465	0.3589	0.6635	0.3754
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
年报文本控制变量	Yes	Yes	Yes	Yes	Yes	Yes

披露质量控制变量	No	No	No	No	Yes	Yes
董事长固定效应	No	No	Yes	Yes	Yes	Yes
企业固定效应	Yes	Yes	Yes	Yes	Yes	Yes
行业-年份固定效应	Yes	Yes	Yes	Yes	Yes	Yes

附录 4 监管一致性测算流程

首先，基于本文已训练的潜在狄利克雷分布（LDA）主题模型，对于每一篇法规文档 d ，将其分配至概率最大的主导监管议题 z_k ($k = 1, \dots, 99$)。在此基础上，提取出每个监管议题 z_k 下由不同监管机构 i 发布的法规文本集合 $D_{k,i}$ 。

其次，本文构建监管议题语义向量。对于每一个监管议题 k ，提取 LDA 模型生成的该监管议题下概率最高的前 30 个关键词，构成全局监管议题词表 $W_k = \{w_1, w_2, \dots, w_{30}\}$ 。针对特定机构 i 于 t 年在监管议题 k 下的所有法规文本，将其分词后的去重关键词集合记为 $S_{k,i,t}$ 。随后，构建一个长度为 N 的二进制向量 $V_{k,i,t}$ ，其第 j 维元素定义为：

$$V_{k,i,t}^{(j)} = \begin{cases} 1, & \text{if } w_j \in S_{k,i,t} \\ 0, & \text{if } w_j \notin S_{k,i,t} \end{cases} \quad (\text{A3})$$

其中， $V_{k,i,t}^{(j)} = 1$ 表示机构 i 在议题 k 的监管实践中涵盖了全局核心词汇 w_j ，反之则为 0。该向量保留了监管议题的核心语义特征。

再次，在获得各机构的语义向量后，本文采用余弦相似度来衡量同一监管议题下任意两个不同机构 i 和 m 之间的监管一致性。对于监管议题 k ，机构对 (i, m, t) 的一致性得分 $Sim_{k,i,m,t}$ 计算如下：

$$Sim_{k,i,m,t} = \cos(V_{k,i,t}, V_{k,m,t}) = \frac{V_{k,i,t} \cdot V_{k,m,t}}{\|V_{k,i,t}\| \times \|V_{k,m,t}\|} = \frac{\sum_{j=1}^N V_{k,i,t}^{(j)} V_{k,m,t}^{(j)}}{\sqrt{\sum_{j=1}^N (V_{k,i,t}^{(j)})^2} \sqrt{\sum_{j=1}^N (V_{k,m,t}^{(j)})^2}} \quad (\text{A4})$$

该指标取值范围为 $[0, 1]$ 。值越接近 1，表明两机构在该议题下的监管要求高度趋同。

最后，为得到监管议题 k 层面的整体监管一致性水平 $C_{k,t}$ ，本文计算该监管议题下所有可能的跨机构配对相似度的算术平均值：

$$Consistency_{k,t} = \frac{1}{M_{k,t}} \sum_{i \neq m} Sim_{k,i,m,t} \quad (\text{A5})$$

其中， $M_{k,t}$ 为 t 年时监管议题 k 下涉及的两个及以上不同机构的配对总数。最终得到的 $Consistency_{k,t}$ 即为各监管议题的监管一致性水平。我们使用与前文一致的方法将监管议题的监管一致性水平映射至企业层面，即得企业层面的监管一致性测度。

$$Consistency_{f,t} = \sum_k P_{f,k,t} \times Consistency_{k,t} \quad (\text{A6})$$

附录 5 共同起草结果

由于监管一致性测度难度大，本文构建的监管一致性度量 $Consistency_{f,t}$ 可能存在一定的局限性。因此我们进行补充检验：在我国，部分中央法律法规文件采取多机构共同起草与联合发布的模式，在此类法规框架下，监管一体化能够更有效地提升监管一致性，进而对企业业绩产生更为显著的积极作用。我们参考 Kalmenovitz et al. (2025)，构建共同起草的代理变量 $Coauthor_{f,t}$ 。第一步，在“议题-年份”层面上计算每一监管议题中来自多机构联合发布文件的词频占比 ($Coauthor_{i,t}$)；第二步，将该议题层面指标按照企业年报中的词频占比 $P_{f,i,t}$ 进行加权平均，得到企业在特定年份面对的监管要求中来自共同起草的法律法规的比例 ($Coauthor_{f,t}$)。我们将 $Coauthor_{f,t}$ 与 $Integ_{f,t}$ 的交互项代入基准模型进行回归分析。附表 5 展示了加入交互项后的实证结果，表明 $Coauthor$ 增强了 $Integ$ 的正向作用，在共同起草的情况下，不同部门之间的协调成本进一步降低，放大了监管一体化的积极效应。

附表 5 共同起草结果

	(1) <i>ROA</i>	(2) <i>Gsales</i>
<i>Integ</i>	0.3083*** (4.8931)	2.6267*** (6.1999)
<i>Integ</i> × <i>Coauthor</i>	1.3480** (2.2130)	15.2989*** (4.1811)
<i>Coauthor</i>	0.0546** (2.2073)	0.4936*** (3.0120)
常数项	0.0653*** (2.7560)	2.4930*** (16.3623)
观测值	42,747	42,740
R^2	0.5451	0.2357
控制变量	Yes	Yes
企业固定效应	Yes	Yes
行业-年份固定效应	Yes	Yes

参考文献

- [1] 姜富伟、孟令超和唐国豪, 2021,《媒体文本情绪与股票回报预测》,《经济学(季刊)》第4期,第1323~1344页。
- [2] 阮睿、孙宇辰、唐悦等, 2021,《资本市场开放能否提高企业信息披露质量?——基于“沪港通”和年报文本挖掘的分析》,《金融研究》第2期,第188~206页。
- [3] 王雄元、李岩琼和肖恣, 2017,《年报风险信息披露有助于提高分析师预测准确度吗?》,《会计研究》第10期,第37~43,96页。
- [4] 许帅、邵帅和何贤杰, 2025,《业绩说明会前瞻性信息对分析师盈余预测准确性的影响——信口雌黄还是言而有征》,《中国管理科学》第3期,第34~44页。
- [5] Jiang, F., J. Lee, X. Martin and G. Zhou, 2019, “Manager Sentiment and Stock Returns”, *Journal of Financial Economics*, 132(1), pp.126~149.
- [6] Kalmenovitz, J., M. Lowry and E. Volkova, 2025, “Regulatory Fragmentation”, *The Journal of Finance*, 80(2), pp.1081~1126.