

《产业政策同质竞争如何影响财政股权资本的空间配置？》附录

附录 1 产业政策判定与关键信息抽取：大语言模型的识别提示词

下述提示词用于政策文本“是否为产业政策”的识别与后续内容相似度计算。

任务说明

请严格按以下步骤分析政策文本：

1. ****产业政策判定****：根据以下定义判断文本是否为产业政策（输出：是/否/不知道 + 置信分数 0-100）
2. ****关键信息提取****：仅当判定为"是"时执行：
 - 主要行业：提取 1-5 个行业原文片段（分号分隔）
 - 政策目标：提取 1-5 个目标原文片段（分号分隔）
 - 政策工具：从 20 类工具中提取 1-5 个原文片段（分号分隔）

产业政策定义

1. 主体必须是政府机构（各级政府及部门），非政府主体（如企业）发动的政策不算。
2. 需包含具体政策措施（如资金支持、准入规则），仅描述愿景或政府日常活动（如搬迁/招聘）不算。
3. 必须直接针对特定行业（如"支持新能源汽车发展"），不针对行业的一般性政策（如改革开放）不算。
4. 旨在改变长期经济结构，应对短期冲击（如疫情/经济周期）的政策不算。

政策目标提取要求

- 仅提取明确提到特定产业发展目标或方向的原文片段
- 不包括一般性应用场景描述（如"推动数字技术在数字乡村的应用"）
- 如无符合条件的目标，输出"不知道"

政策工具类别

- 1.税收激励 2.财政补贴 3.信贷与融资支持 4.产业引导基金 5.股权投资支持 6.贸易保护 7.投资政策 8.技术研发与应用 9.政府采购 10.劳动力政策 11.基础设施投资 12.促进产业集群 13.环境政策 14.市场准入与监管 15.消费补贴 16.土地优先供应 17.本地化政策 18.促进创业 19.产业推广 20.其他制度性便利（简政放权、优化服务、法治环境等）

待分析文本

{policy_text}

输出格式

判定结果: [是/否/不知道]

置信分数: [0-100]

主要行业: [符合条件时的行业片段/否则留空]

政策目标: [符合条件时的目标片段/否则留空]

政策工具: [符合条件时的工具片段/否则留空]

""

附录 2 从政策原文到 *Similarity* 变量的提取过程：以淮安、泰州、南通为例

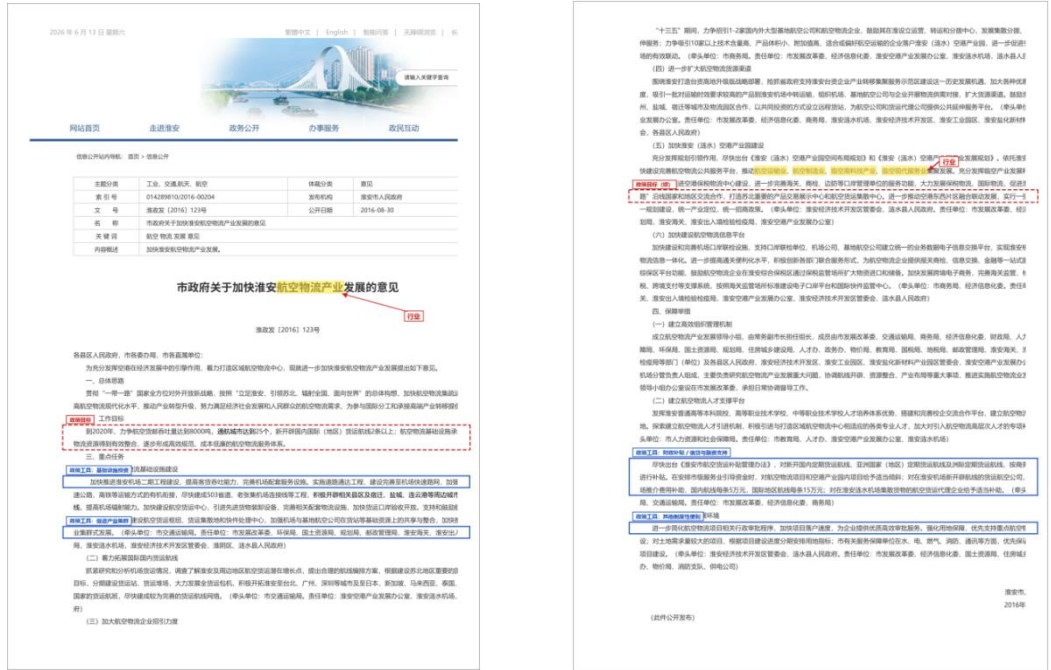
（一）单条政策的提取过程：以淮安市为例

以淮安市人民政府《关于加快淮安航空物流产业发展的意见》（淮政发〔2016〕123 号）为例，政策文本转化为结构化信息分四步进行：第一，识别产业政策。将政策文本及产业政策定义输入模型，要求给出判断、置信分和理由。本例置信分为 95，因其主体为市政府，包含财政补贴、用地保障等具体措施，直接指向航空物流产业，且目标具有五年期的长期结构性特征，经反向校验后确认为产业政策。第二，提取行业信息。模型识别出“航空物流产业”“航空运输业”“航空制造业”“临空高科技产业”“临空现代服务业”五个行业表述（见附图 1）。这些并非《国民经济行业分类》标准条目，而是政策语境中的自然语言描述，往往跨越多个行业门类。本文保留此类自然语言表述，而不强行映射到单一标准代码，以更准确反映政策主题及城市间表述差异。第三，提取政策工具。模型识别出财政补贴、信贷与融资支持、基础设施投资、促进产业集群和其他制度性便利五类工具，并直接摘录相应原文，如航线补贴、引导资金倾斜、机场二期建设、集群式发展和简化审批、强化用地保障等。第四，提取政策目标。模型直接摘录目标表述，如“到 2020 年，力争航空货邮吞吐量达到 8000 吨，通航城市达到 25 个，新开辟国内国际（地区）货运航线 2 条以上”，以及形成航空物流服务体系、打造区域货运集散中心等目标。经过上述处理，一份政策文本被压缩为行业、政策工具及政策目标三类信息。

（二）城市间相似度计算：以泰州和南通为例

为说明城市间相似度的计算逻辑，选取 2016 年泰州和南通两条代表性产业政策。泰州样本为《三大战略性主导产业奖励项目评估认定办法》（泰政办发〔2016〕69 号），识别出“生物医药与高性能医疗器械”“高技术船舶与海工装备”“节能与新能源”三类产业；南通样本为《关于促进市区工业发展的若干政策意见》（通政发〔2016〕27 号），识别出“船舶”“装备”“软件”“工业设计中心”“智能车间”五类产业。

两地行业表述既有重合，如船舶、装备，也有差异，如泰州侧重生物医药和新能源，南通侧重软件和智能制造载体。本文保留这种原始表述差异，不作强制统一。随后，将各城市当年所有产业政策识别出的行业信息汇总为“城市一年份”文档，经文本清洗和向量化后形成政策向量，再按正文式（1）计算城市间余弦相似度。该指标不仅反映是否涉及相同产业，也反映不同产业在政策文本中的强调结构。最后，将某城市与同省其他城市的两两相似度取均值，即得到城市一年份层面的核心解释变量 *Similarity*。泰州与南通 2016 年的相似度，只是两市各自与省内其他城市平均相似度中的一个组成部分。



附图 1 地方政府产业政策文件中关键信息的标识

附录 3 描述性统计

附表 1 报告主要变量的描述性统计。广义口径的异地政府基金投资 $\ln A_{away}$ 均值为 0.441、标准差为 0.759，取值范围[0,3.497]，分布右偏；狭义口径的异地政府基金投资 $\ln A_{away2}$ 均值为 0.308、标准差为 0.635，取值范围[0,3.091]，分布与 $\ln A_{away}$ 相近，表明城市-年份层面的异地政府股权投资离散度较高。产业政策相似度 $Similarity$ 的理论取值位于[0,1]区间，实际均值 0.255、标准差 0.102，范围[0.026,0.478]，显示各城市的产业政策相似度差异较大。

附表 1 描述性统计

变量类别	变量名称	样本量	均值	标准差	最小值	最大值
被解释变量	$\ln A_{away}$	3350	0.4407	0.7593	0	3.4965
	$\ln A_{away2}$	3350	0.3076	0.6354	0	3.0910
核心解释变量	$Similarity$	3350	0.2553	0.1018	0.0256	0.4782
控制变量	$GDPPC$	3350	10.7672	0.5654	9.4309	12.0655
	$Stru2$	3350	0.4500	0.1096	0.1560	0.7223
	$Stru3$	3350	0.4253	0.0966	0.2080	0.6915
	$Fixed$	3350	1.0136	0.6040	0.0823	3.6217
	Pop	3350	5.8333	0.7251	3.3865	7.0901
	$Fisr$	3350	0.4429	0.2171	0.0531	0.9886
	Fin	3350	1.0539	0.5944	0.3081	3.4141
	$AdminChange$	3350	0.2528	0.4347	0	1
	$Distance$	3350	2.0418	2.9172	0	8.3058
	HSR	3350	0.6042	0.4891	0	1

附录 4 稳健性检验

（一）替换被解释变量和估计模型

本文首先检验基准结论是否依赖特定的变量口径或估计方法。就被解释变量而言，剔除经股权穿透识别的政府基金投资，仅保留狭义口径下异地政府投资基金的直接和间接投资，并构造 $\ln Aways$ 2（加一后取自然对数）和以本地政府基金本地投资次数标准化的 $Aways2R$ 。就估计方法而言，鉴于并非所有城市一年份观测值均发生政府基金投资，被解释变量呈“零值堆积+连续正值”的混合分布，OLS 可能存在偏误，本文采用面板固定效应 Tobit 模型、Honoré（1992）的修整最小绝对离差法（TLAD）和伪泊松最大似然法（PPML）进行估计。附表 2 显示，在更换变量口径和估计方法后， $Similarity$ 的系数仍显著为负，基准结论稳健。

附表 2 稳健性检验：基于变量构造和模型替换

被解释变量：	去除股权穿透识别的政府基金投资				替换估计方法，被解释变量为 $\ln Aways_{i,t}$	
	$\ln Aways2_{i,t}$		$Aways2R_{i,t}$		TLAD	PPML
	(1)	(2)	(3)	(4)	(5)	(6)
$Similarity_{i,t}$	-0.3595** (0.1472)	-0.2800** (0.1170)	-0.3971** (0.1540)	-0.3363** (0.1454)	-0.6259** (0.2425)	-0.4422* (0.2591)
控制变量		Yes		Yes	Yes	Yes
城市 FE、年份 FE	Yes	Yes	Yes	Yes	Yes	Yes
观测值	3350	3350	3350	3350	3350	3051
R^2	0.5064	0.6634	0.1722	0.2114	-	-

注：（1）*、**和***分别为 10%、5%和 1%的显著性水平；（2）第（5）列括号为自助抽样调整的稳健标准误，其余各列括号为城市层面聚类的稳健标准误。

（二）产业政策相似度的衡量

本文进一步从影响时滞和测算口径两方面检验产业政策相似度指标的稳健性。第一，考虑到地方产业政策从出台、实施到被异地政府和基金识别并影响投资决策存在时滞，本文将核心解释变量构造成区间均值，分别为 $Similarity_{i,t-1-t}$ 、 $Similarity_{i,t-2-t}$ 和 $Similarity_{i,t-3-t}$ ，并对控制变量作相同处理。附表 3 第（1）至（3）列显示， $Similarity$ 对 $\ln Aways$ 的影响仍显著为负，且系数较正文表 1 第（2）列略有增大，说明产业政策同质竞争存在一定滞后效应。第二，调整产业政策文本范围：一是纳入区县政策文件，在大语言模型识别基础上重新计算相似度指标 $Similarity2$ ；二是仅保留使用财政补贴工具的政策样本。之所以聚焦补贴，是因为其是地方产业政策中使用最广、也最受关注的工具（Fang 等，2025）。具体做法是，以“补贴”为关键词提取相关政策，并据此重新汇总城市一年份层面的政策目标片段，计算包含政策工具和目标信息的相似度指标 $Similarity3$ 。附表 3 第（4）至（5）列显示，在调整文本范围后，产业政策相似度对被解释变量的影响仍显著为负。由此可见，城市与省内其他城市产业政策越趋同，政府间横向竞争越强，本地企业获得异地政府投资基金投资的机会越少。

附表 3 稳健性检验：产业政策相似度的度量方式

被解释变量： $\ln Aways_{i,t}$	解释变量分别为 $Similarity$ 滞后一至三期的平均值			考虑市以下 产业政策	“补贴”为工具 的产业政策
	(1)	(2)	(3)	(4)	(5)
$Similarity_{i,t-1-t}$	-0.4516***				

	(0.1451)				
$Similarity_{i,t-2-t}$		-0.5401***			
		(0.2015)			
$Similarity_{i,t-3-t}$			-0.7297***		
			(0.2572)		
$Similarity_{2,t}$				-0.2491***	
				(0.0932)	
$Similarity_{3,t}$					-0.2342**
					(0.0998)
控制变量	Yes	Yes	Yes	Yes	Yes
城市 FE、年份 FE	Yes	Yes	Yes	Yes	Yes
观测值	2987	2640	2305	3349	3212
R^2	0.8268	0.8355	0.8501	0.8155	0.8162

注：（1）*、**和***分别为 10%、5%和 1%的显著性水平；（2）括号内为城市层面聚类的稳健标准误。

（三）细化数据颗粒度：省内城市对面板及城市-行业面板

基准模型以城市与省内其他城市的平均政策相似度识别总量效应，本节进一步在“城市对一年份”和“城市—行业一年份”两个层面检验。

在“城市对 (i,j) —年份 t ”层面，样本包含 3856 个城市对、42308 个观测值¹。同一年每对城市生成两条方向性观测，即 $i \leftarrow j$ 和 $j \leftarrow i$ 。被解释变量 $\ln A_{way_{i,j,t}}$ 为注册地在 $j(i)$ 的政府基金对 $i(j)$ 地企业的投资次数，核心解释变量 $Similarity_{i,j,t}$ 为两城当年产业政策相似度，控制变量在 $i \leftarrow j$ 时使用 $\mathbf{X}_{i,t}$ ，在 $j \leftarrow i$ 时使用 $\mathbf{X}_{j,t}$ ，模型控制方向性城市对固定效应和年份固定效应，并采用 OLS、TLAD 和 PPML 估计。附表 4 第（1）至（3）列显示， $Similarity_{i,j,t}$ 的系数均显著为负，说明两城政策越相似，一地政府基金对另一地企业的异地投资越少。

其次，在“城市—行业一年份”层面，本文利用大语言模型提取政策文本中的“主要行业”信息，并映射至国民经济行业分类（GB/T 4754）：制造业细化至三位代码，其他行业采用一位代码。以城市 i 、行业 k 、年份 t 为单位，计算其与省内其他城市同一行业政策文本的平均相似度 $Similarity_{i,k,t}$ ；被解释变量 $\ln A_{way_{i,k,t}}$ 为该城市该行业获得的异地政府基金投资，模型控制城市 $\times$ 行业和行业 $\times$ 年份固定效应。本文还识别战略性新兴产业及高技术制造业进行分样本分析²。附表 4 第（4）至（8）列显示，行业层面的政策相似度仍显著抑制异地政府基金投资，且该效应主要集中于战略性新兴产业和高技术制造业，在其他行业中不显著。这表明，基准效应主要发生在地方布局最密集、最易重复竞争的重点赛道。

附表 4 稳健性检验：基于“城市-城市”配对及“城市-行业”面板的回归结果

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
被解释变量：	同省“城市对”匹配 $\ln A_{way_{i,j,t}}$			“城市-行业”面板 $\ln A_{way_{i,k,t}}$				
	OLS	TLAD	PPML	全样本	战略新兴	非战略新兴	高技术制造	非高技术制造
$Sim_{i,j,t}$	-0.0110**	-0.2689***	-0.7536**					

1 由于城市对层面的政府基金投资事件高度稀疏，被解释变量零值有 41209 条，该设定不宜作为基准回归模型，但能最大限度保留两两城市的信息。

2 战略性新兴产业包括：化学原料及制品（C26）、医药制造（C27）、化学纤维（C28）、非金属矿物制品（C30）、专用设备（C35）、汽车（C36）、铁路船舶航空航天（C37）、电气机械（C38）、计算机通信电子（C39）、电力热力（D44）、燃气（D45）、信息传输软件和信息技术服务业（I）、科学研究和技术服务业（M）、生态保护和环境治理业（N）和卫生和社会工作（Q）。高技术制造业：C26、C27、C34、C35、C36、C37、C38、C39。主要依据：国家统计局《战略性新兴产业分类（2018）》、国务院《中国制造 2025》。

	(0.0056)	(0.1026)	(0.3219)					
$Similarity_{i,k,t}$				-0.0042*	-0.0133**	0.0003	-0.0277**	0.0008
				(0.0025)	(0.0061)	(0.0026)	(0.0127)	(0.0022)
控制变量	Yes	Yes	Yes					
城市对 FE、年份 FE	Yes	Yes	Yes					
城市×行业 FE、行业×年份 FE				Yes	Yes	Yes	Yes	Yes
观测值	42308	42318	5712	92087	30030	62057	17723	74364
R^2	0.2582	-	-	0.1536	0.1938	0.0634	0.1787	0.0460

注：（1）*、**和***分别为 10%、5%和 1%的显著性水平；（2）第（1）和（3）列括号内为城市层面聚类的稳健标准误，第（2）列为自助抽样调整的稳健标准误。

（四）剔除潜在干扰样本

最后排除潜在干扰样本。一方面，剔除样本期间未获政府基金投资的城市，以缓解“零值堆积”偏误——这类城市多因项目供给不足、壁垒较高或信息不对称，其“零投资”与产业政策关联较弱。剔除后，附表 5 第（1）列结果显示， $Similarity$ 的系数为-0.343 并在 1%水平上显著，与正文表 1 相近，佐证结论稳健。另一方面，剔除投资标的丰富的城市——它们项目供给充沛、资金信息集聚、跨区域网络密集，或经济实力强而对省内产业政策影响较大，可能干扰识别。我们逐步剔除计划单列市（附表 5 第（2）列）、合肥和苏州这两个实务中的政府投资基金“模范”城市（附表 5 第（3）列）和全部副省级城市（附表 5 第（4）列），可发现 $Similarity$ 的系数仍显著，说明基准结论并非由少数“异常活跃”的城市驱动。

附表 5 稳健性检验：剔除零风投及风投活跃城市样本

	(1)	(2)	(3)	(4)
被解释变量： $\ln Away_{i,t}$	剔除样本期间未 受过投资的城市	计划单列市	剔除投资标的丰富的城市 合肥和苏州	副省级城市
$Similarity_{i,t}$	-0.3433*** (0.1069)	-0.3254*** (0.0980)	-0.3185*** (0.0967)	-0.2836*** (0.0927)
控制变量	Yes	Yes	Yes	Yes
城市 FE、年份 FE	Yes	Yes	Yes	Yes
观测值	3051	3290	3266	3170
R^2	0.8123	0.8073	0.8059	0.7953

注：（1）*、**和***分别为 10%、5%和 1%的显著性水平；（2）括号内为城市层面聚类的稳健标准误。

附录 5 市场化主体投资事件的变量构造过程

纯市场化主体异地投资变量的数据来源于 CNRDS，清洗和变量构造过程如下：

（1）将每笔投资事件分别与被投资企业注册信息和投资方机构信息匹配，获取双方的行政区划代码和行业分类，县级市统一归并至所属地级市，历史地名更新为现行区划，被投资企业归属为省直辖县级的予以剔除。投资方地址缺失的样本通过多轮辅助数据源补充（如 EPS、爱企查、企查查的批量导入搜索）；

（2）剔除债权融资、捐赠、政府补贴和股权转让等非股权投资事件；

（3）采用“数据库标识+名称关键词+人工校验”标注国资背景。首先利用清科数据库、投中数据库、原点数据库中的政府引导基金、国资背景基金名单进行初步标注；其次使用机构名称关键词扩展，当投资方名称含“引导基金”“政府”“财政”“国资委”“国有资本”等标注、含“城投”“开发区”“高新区”等地方平台特征词、含“中金”“中信”“招商局”“深创投”“华润”等已知国有控股机构名称、符合“省/市+投资/发展/建设+

集团/公司”地方投资平台命名模式的标注为国资。同时排除含“个人”“自然人”“未披露”以及“红杉”“高瓴”“IDG”“软银”等已知纯民营或外资机构的样本。当投资事件的出资方不含有国资背景标注时，记为纯市场化主体的投资事件，用以生成 $\ln Market_{i,t}$ 变量；

（4）机构类型识别。将名称含“基金”“股权投资”“创业投资”“创投”“资本”“投资管理”“合伙企业”及英文“Capital”“Venture”“Fund”等关键词的标注为 PE/VC 类机构，排除公募基金、保险、银行、证券、信托及个人投资者。当投资事件的出资方不含有国资背景标注且属 PE/VC 类的投资方时，记为纯市场化 PE/VC 基金的投资事件，用以生成 $\ln PEVC_{i,t}$ 变量；

（5）本异地判定。提取投资方与被投资企业区划代码前四位（地级市），相同为本地、不同为异地；直辖市以省级代码判定，同省即同城。投资方地址仅精确到省级而无法判定市级归属的样本予以排除（占比不足 0.2%）。

最终，本文以城市-年份为单位汇总，计算每个城市每年获得的纯市场化基金异地投资次数和纯市场化 PE/VC 基金异地投资次数，加一取自然对数后得到回归使用的变量。

参考文献

- [1] Fang, H., Li, M., and Lu, G., 2025, "Decoding China's industrial policies," NBER Working Paper, No. w33814.
- [2] Honoré, B. E., 1992, "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica: Journal of the Econometric Society*, pp.533~565.