

# 数字技能溢价、结构性极化与技能互补性

## 来自中国上市企业招聘数据（2016–2025）的证据

向稀

南京大学商学院

2026.4.3



- 1 研究背景与动机
- 2 数据与计量策略
- 3 时变工资溢价
- 4 技能互补性
- 5 结构性极化与工资不平等
- 6 稳健性检验
- 7 结论



- 1 研究背景与动机
- 2 数据与计量策略
- 3 时变工资溢价
- 4 技能互补性
- 5 结构性极化与工资不平等
- 6 稳健性检验
- 7 结论



- **核心问题**：数字技术是否提高了工资？对哪些人群有影响？
- 经典 SBTC 框架 (Katz & Murphy 1992; Acemoglu 2002)：技术提升对互补性抽象技能的相对需求，扩大技能溢价
- 任务导向模型 (Autor et al. 2003; Acemoglu & Autor 2011)：技术同时替代常规任务，产生就业极化而非简单需求上升

## 为什么选择中国？

- 1 2016–2025 年技术采纳速度极快（云计算 → 大数据 → 大语言模型）
- 2 劳动力市场结构张力突出：庞大低技能供给 + 快速扩张的大学毕业生群体
- 3 上市企业招聘数据提供高频雇主侧工资信号



# 三项主要贡献

## 1 时变溢价动态

逐年回归揭示 AI/机器学习溢价呈**倒 U 形**轨迹 (35% → 55% → 25%)，解释为需求冲击叠加滞后的供给响应

## 2 技能互补性

所有两两技能交互项均为**负值** (次可加性)：同时持有两项数字技能所获溢价低于各项溢价之和

## 3 结构性极化与工资不平等

行业数字技能需求扩散 (Gini: 0.20 → 0.14)，但上尾工资不平等加剧；分位数回归揭示背后机制



- 1 研究背景与动机
- 2 数据与计量策略
- 3 时变工资溢价
- 4 技能互补性
- 5 结构性极化与工资不平等
- 6 稳健性检验
- 7 结论



## 样本

- 693 万条来自中国上市企业的招聘职位，2016–2025 年
- 工资分析样本：含有效薪资信息的 820,807 条观测
- 57,460 家企业；3,824 个城市 × 年份单元
- 薪资标准化为月度中点值，取对数

## 技能分类

- 基于 NLP 的 13 类结构化词典
- 核心数字技能：
  - ▶ AI/机器学习 (1.3%)
  - ▶ 编程 (10.6%)
  - ▶ 数据工程 (4.8%)
  - ▶ 云计算/DevOps (3.5%)
  - ▶ 数据分析 (21.6%)
- *HasDigital* 综合指标：79.8%



## OLS 工资回归

$$\ln w_{icft} = \alpha + \sum_k \beta_k S_{icft}^k + \gamma_1 Edu_{it} + \gamma_2 Exp_{it} + \lambda_{ct} + \mu_j + \varepsilon_{icft}$$

- $\lambda_{ct}$ : **城市 × 年份固定效应** (3,824 组), 吸收地方劳动力市场冲击
- $\mu_j$ : **行业固定效应** (21 类)
- 标准误在**企业层面聚类** ( $G = 57,460$ )

**估计方法:** 城市 × 年份固定效应通过迭代组内变换消除 (Mundlak–Frisch–Waugh, 4 轮迭代), 在  $O(N)$  计算量内等价于引入全部虚拟变量



## 逐年回归

逐年估计  $\hat{\beta}_k(t)$ ，刻画溢价随时间的动态演变

## 交互项模型

$$+ \sum_{k < k'} \theta_{kk'} S^k S^{k'}$$

$\theta_{kk'} < 0$ : 次可加性

$\theta_{kk'} > 0$ : 超可加性

## 分位数回归

在  $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$   
估计，刻画分布效应

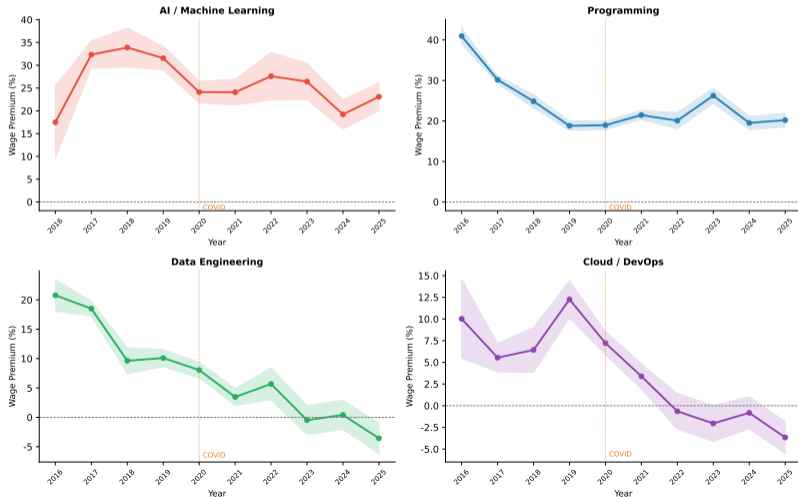


- 1 研究背景与动机
- 2 数据与计量策略
- 3 时变工资溢价**
- 4 技能互补性
- 5 结构性极化与工资不平等
- 6 稳健性检验
- 7 结论

# 年度工资溢价，2016-2025



Figure 5: Time-Varying Digital Skill Wage Premia (2016-2025)  
Annual OLS with 95% CI; Controls: Education, Experience, City Tier, Industry FE





# 核心发现：溢价动态

## ■ AI/机器学习——倒 U 形：

- ▶ 2016 年约 35% → 2018–2019 年峰值约 55% → 2022–2023 年约 25%
- ▶ 滞后供给响应：2017 年后入学 AI 专业的毕业生约 2020 年起大规模进入市场
- ▶ 2023 年后小幅回升：生成式 AI（ChatGPT 等）带动新一轮需求

## ■ 编程——渐进收窄：30% → 20%（供给持续稳定增长）

## ■ 数据工程——U 形：Hadoop 时代（2016–17）→ 低谷 → 流式处理时代（Flink/Kafka，2021 年后）回升

## ■ 2020 年新冠冲击：短暂下滑，2 年内恢复，表明数字技能回报未受永久性冲击

### 政策启示

横截面研究的结论高度依赖于采样时点（峰值还是谷底），时间维度动态至关重要



- 1 研究背景与动机
- 2 数据与计量策略
- 3 时变工资溢价
- 4 技能互补性**
- 5 结构性极化与工资不平等
- 6 稳健性检验
- 7 结论



# 交互项模型：普遍次可加性

两两交互项	$\hat{\theta}_{kk'}$	标准误
AI × 编程	-0.048***	(0.014)
AI × 数据工程	-0.106***	(0.016)
编程 × 数据工程	-0.229***	(0.009)
编程 × 云计算	-0.095***	(0.009)
AI × 云计算	-0.097***	(0.021)

## 隐含联合溢价：

- AI + 编程：可加性 61.1% → 实际**57.1%**
- AI + 数据工程：可加性 62.3% → 实际**46.1%**
- 编程 + 数据工程：可加性 55.5% → 实际**32.7%**

## 机制解释：

- 1 任务可替代性：ETL/查询优化任务重叠
- 2 雇主信号怀疑：多技能声称受到折价对待



# 启示：专而精优于博而浅

## ○ 型圈理论预测 (Kremer 1993)

超可加性：在多领域同时具备能力时，产出应呈乘数效应

## 本文发现：次可加性

所有五组两两交互项均**显著为负**

最大缺口：编程 × 数据工程 (-22.9 个百分点)

## 对课程设计的启示：

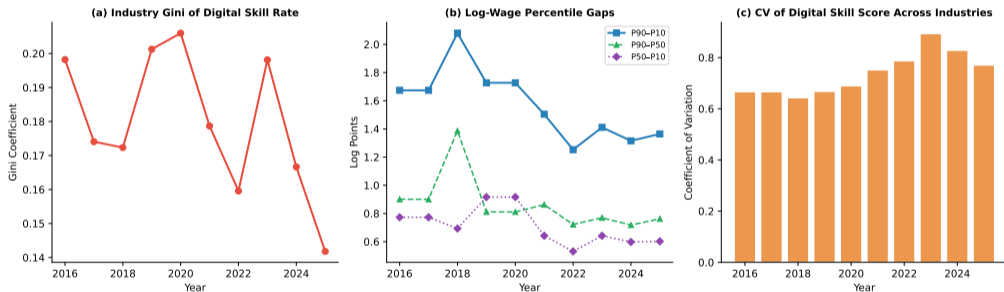
- 深度掌握一项高溢价技能（如 AI/机器学习）优于分散精力布局多个数字领域
- 最优技能组合策略：集中而非分散



- 1 研究背景与动机
- 2 数据与计量策略
- 3 时变工资溢价
- 4 技能互补性
- 5 结构性极化与工资不平等**
- 6 稳健性检验
- 7 结论



Figure 6: Polarization of Digital Skill Demand and Wage Inequality (2016-2025)



(a) 行业数字技能需求率 Gini 系数。(b) P90-P10/P90-P50/P50-P10 对数工资差距。(c) 行业间数字技能得分变异系数。



## 需求端：扩散

- 行业 Gini 从**0.20** (2016) 降至**0.14** (2025)，降幅 30%
- 数字技能需求从科技密集型行业向更广泛经济部门扩散

## 预期推断：

需求扩散  $\Rightarrow$  工资不平等收窄

## 工资端：上尾扩张

- P90-P10 对数工资差距：1.67 (2016)  $\rightarrow$  峰值 2.08 (2018)  $\rightarrow$  持续  $>1.25$  (2025)
- P90-P50 差距主导；P50-P10 差距相对稳定
- **上尾极化**，而非对称扩散

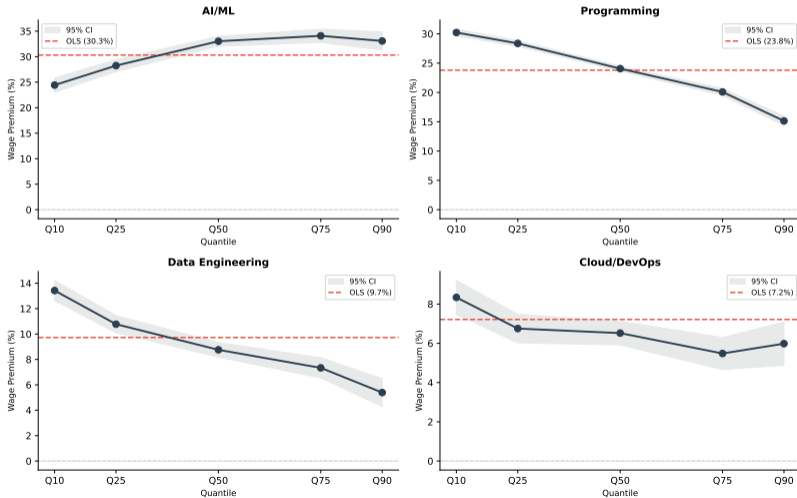
**悖论：扩散  $\neq$  平等**

当技能回报在工资分布上存在异质性时，需求侧扩散并不必然压缩工资不平等



# 分位数回归：破解悖论

Figure 9: Quantile Regression — Skill Wage Premia Across the Wage Distribution  
(Shaded area = 95% CI; dashed line = OLS benchmark)





# 异质的分位数溢价

## AI / 机器学习

- Q10: 25.3% → Q75: 34.8% → Q90: 33.1%
- 倒 U 形：溢价集中于高端技术岗位
- OLS 估计值  $\approx$  中位数分位数估计

## 编程

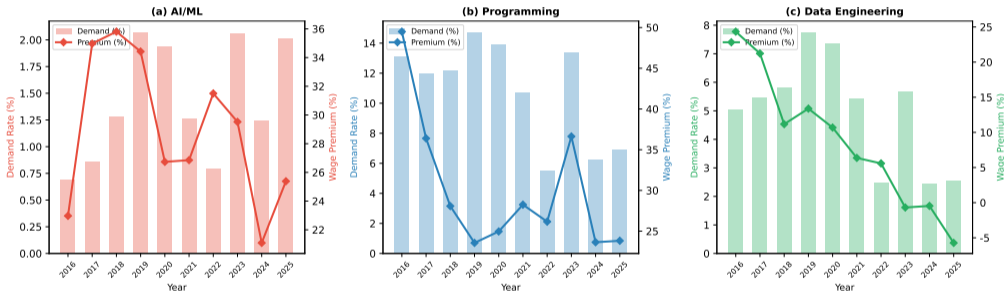
- Q10: 31.2% → Q90: 16.1% (单调递减)
- 编程技能已扩散至中低工资岗位
- 低端稀缺租金：低技能岗位中的罕见技能溢价更高

## 解释

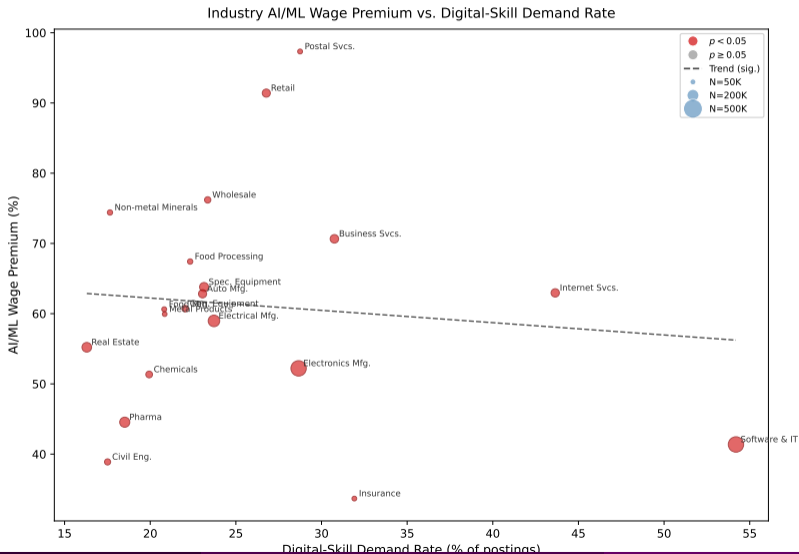
技能扩散的广度边际效应与特定工资分位数上的强度边际放大效应并存——两种机制同时运作



Figure 8: Decomposition of Skill Demand Rate vs. Wage Premium (2016-2025)



柱形图（左轴）：年度需求率。折线图（右轴）：条件工资溢价。AI/机器学习：需求单调上升，溢价先涨后跌——劳动力市场中经典的滞后供给响应。





- 1 研究背景与动机
- 2 数据与计量策略
- 3 时变工资溢价
- 4 技能互补性
- 5 结构性极化与工资不平等
- 6 稳健性检验**
- 7 结论



# 稳健性：子样本稳定性

子样本	N	AI/机器学习	编程	数据工程
全样本基准	820,807	26.0%***	24.8%***	9.9%***
2020 年前 (2016-19)	433,018	29.7%***	27.6%***	13.6%***
2020 年后 (2020-25)	387,789	23.7%***	20.4%***	4.2%***
制造业	378,413	26.7%***	27.1%***	8.9%***
服务业	285,803	24.9%***	21.8%***	8.9%***
一线城市	285,151	26.3%***	25.7%***	10.5%***
二三线城市	535,656	25.6%***	23.2%***	8.6%***
高学历 (≥ 本科)	397,319	19.2%***	18.0%***	6.8%***
低学历 (< 本科)	423,488	25.7%***	<b>33.3%***</b>	16.6%***

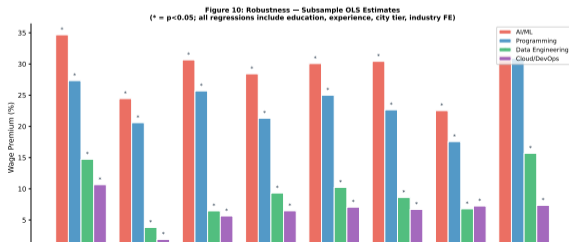
- 各子样本溢价均显著为正，结论稳健
- 数据工程和云计算溢价 2020 年后大幅收缩；AI 和编程溢价更具持续性
- 低学历群体的编程溢价反而更高：稀缺租金效应



# 稳健性：薪资截尾敏感性

薪资截尾范围	N	AI/机器学习	编程	数据工程
不截尾 [0%, 100%]	820,807	26.0%***	24.8%***	9.9%***
[1%, 99%]	804,975	25.8%***	23.8%***	9.9%***
[5%, 95%]	739,391	20.2%***	23.4%***	9.7%***
[10%, 90%]	671,137	14.9%***	21.9%***	9.0%***

- AI 溢价在截尾下降幅度更大：部分由高薪异常职位驱动
- 编程和数据工程溢价在截尾中更为稳定
- OLS 估计值接近 AI/机器学习的中位数分位数估计





- 1 研究背景与动机
- 2 数据与计量策略
- 3 时变工资溢价
- 4 技能互补性
- 5 结构性极化与工资不平等
- 6 稳健性检验
- 7 结论**



## 1 时变溢价

AI/机器学习溢价呈倒 U 形 (2016–2025)，与需求冲击叠加滞后供给响应一致。横截面研究对采样时点高度敏感。

## 2 技能组合次可加性

所有两两数字技能交互项均为负；最大缺口出现在编程 × 数据工程 (−22.9 个百分点)。应深度专精单项高价值技能。

## 3 扩散 ≠ 平等

行业数字需求扩散 (Gini 降 30%)，但上尾工资不平等加剧。AI/机器学习溢价在 Q75 处达峰；编程溢价从 Q10 到 Q90 单调递减。



## （一）课程设置的时机把握

高等教育系统存在 3-5 年的调整滞后，存在过度培养溢价已压缩领域人才的风险

## （二）深度优先于广度

次可加性结论支持技能培育政策中专注精通而非泛化数字素养

## （三）扩散与再分配并重

数字技能需求扩散不会自动收窄工资差距，需辅以互补性再分配政策工具

# 谢谢!

向稀  
南京大学商学院

欢迎批评指正