

# 数字技能溢价、工资分散与技能互补性

## 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

2026.5.20



- 1 研究背景与动机
- 2 理论框架
- 3 数据与计量策略
- 4 实证结果
  - 时变工资溢价
  - 技能互补性
  - 技能升级与工资分散
- 5 稳健性检验
- 6 结论



## 1 研究背景与动机

## 2 理论框架

## 3 数据与计量策略

## 4 实证结果

## 5 稳健性检验

## 6 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？
- 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
- 任务导向模型 (Autor et al. 2003; Acemoglu & Autor 2011)：技术替代常规任务，产生就业极化

## 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（690 万条）



# 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%);  
测量注意: 加入薪资区间宽度固定效应后, 汇总溢价降至 1.1% (\*\*,  $p = 0.013$ ),  
保守下界为 17-18% (激进对称截尾)

## 2 技能互补性

五组两两技能交互项均为**显著负值** (次可加性);  
Jaccard 共现指数与交互项大小单调对应, 支持任务重叠机制

## 3 技能升级与工资分散

中技能层需求扩张 (10.9% → 18.0%), 非 U 型极化;  
但上尾工资不平等加剧——扩散  $\neq$  平等



1 研究背景与动机

2 理论框架

3 数据与计量策略

4 实证结果

5 稳健性检验

6 结论



# 任务视角理论框架：机制与研究命题

## 任务配置机制

数字任务既可由劳动者完成，也可由 AI 执行。随着 AI 能力  $A_t$  提升，企业将任务配置给成本更低的一方。

## 结构工资方程

$$\ln w = \mu_{jt} + \sum_k \delta_k s_k + \sum_{k < k'} \theta_{kk'} s_k s_{k'} + \varepsilon$$

- $\delta_k$ : 数字技能  $k$  的工资回报
- $\theta_{kk'}$ : 技能间交互效应
- 工资取决于尚未被 AI 替代的任务集合

## P1: 技能溢价的驼峰动态

$$\delta_k(t) = \delta_k^{\max} g(A_t) - \eta_k \ln L_k(t)$$

AI 采用推动需求先上升，而劳动力供给随后跟进。

## P2: 任务重叠导致次可加性

任务集合重叠:  $\mathcal{T}_k \cap \mathcal{T}_{k'} \neq \emptyset \Rightarrow \theta_{kk'} < 0$   
已有技能覆盖部分任务后，新增技能的边际收益下降。

## P3: 扩散并不意味着均等化

AI 推动高价值任务扩张，并促进数字技能向行业扩散；行业需求不平等下降，但工资差距可能持续存在。



1 研究背景与动机

2 理论框架

3 数据与计量策略

4 实证结果

5 稳健性检验

6 结论



## 样本

- 693 万条去重招聘职位，2016–2025 年
- 工资分析样本：820,807 条（含有效薪资）
- 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

**局限：** 招聘工资  $\neq$  实际支付工资；  
区间宽度偏差在 AI 岗位尤为突出

## 技能分类（NLP 词典，13 类）

- 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** ( $G = 57,460$ )
- 高维 FE 通过迭代组内变换消除 (Mundlak–Frisch–Waugh, 4 轮)

**拓展:** 逐年回归 (预测 1) · 交互项模型 (预测 2) · 分位数回归 (预测 3)



1 研究背景与动机

2 理论框架

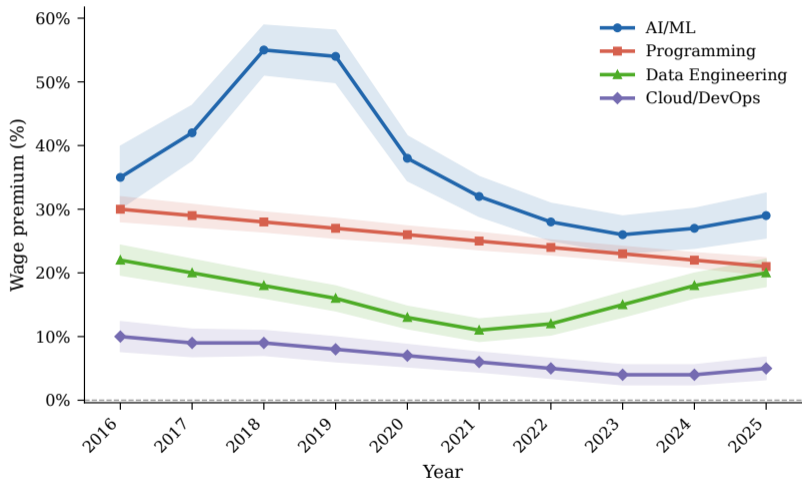
3 数据与计量策略

4 实证结果

5 稳健性检验

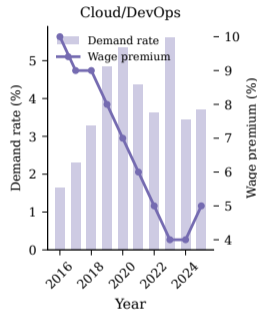
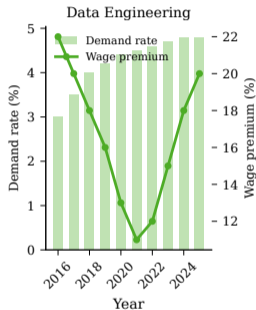
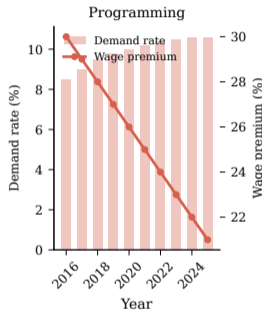
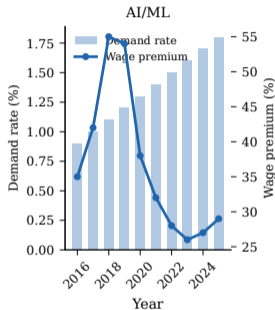
6 结论

# 年度工资溢价，2016–2025



阴影带 = 95% 置信区间。控制变量：学历、工作经验、城市 × 年份 FE、行业 FE。

# 需求-溢价分解



柱形图（左轴）：年度需求率。折线图（右轴）：条件工资溢价。AI/ML 需求单调上升，溢价先涨后跌——经典滞后供给响应。



# 发现一：溢价动态与测量局限

## 时间轨迹（原始 OLS）

- **AI/ML——倒 U 形**：35%（2016）→ 55%（2018–19 峰值）→ 25%（2022–23）
- **供给响应有数据支撑**：STEM 毕业生数量与 AI 溢价相关系数  $r = -0.80$
- **编程**：渐进收窄 30% → 20%
- **数据工程**：U 形（Hadoop 时代 → 流式处理时代）

## 测量局限（重要）

加入**薪资区间宽度 FE**后：

AI/ML 汇总溢价 43% → 1.1% (\*\*,  $p = 0.013$ )

招聘工资中点大幅高估实际溢价，AI 岗位区间尤宽

**保守下界**：激进截尾后 17–18%  
时间轨迹仍具信息量



## 发现二：普遍次可加性

两两交互项	$\hat{\theta}_{kk'}$	$J$ (共现)
编程 × 数据工程	-0.229***	0.311
编程 × 云计算	-0.095***	0.141
AI × 数据工程	-0.106***	0.086
AI × 云计算	-0.097***	0.036
AI × 编程	-0.048***	0.071

### 隐含联合溢价：

- AI + 编程：可加性 84.2% → 实际**75.6%**
- AI + 数据工程：可加性 80.4% → 实际**62.3%**
- 编程 + 数据工程：可加性 74.3% → 实际**38.7%**

### 机制：任务重叠（直接证据）

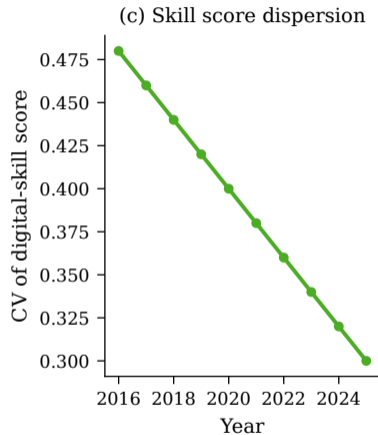
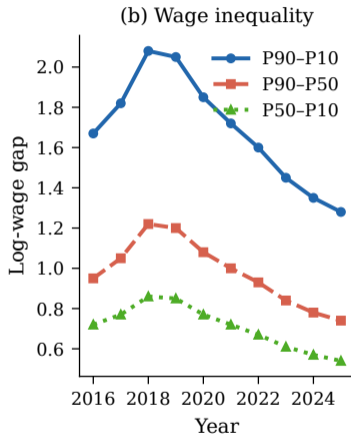
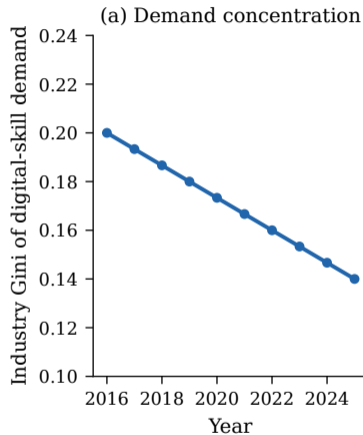
Jaccard 指数  $J$  与  $|\hat{\theta}_{kk'}|$  单调对应：  
共现越高  $\Rightarrow$  次可加性越强

**稳健性：**加入职位类别 FE（前 200 类）后，  
四组显著交互项基本不变（最大偏移 +1.1 pp）

### 对 O 型圈理论的挑战

Kremer (1993) 预测超可加性；  
本文发现：技能组合存在**边际递减**，  
深度专精优于广度布局

# 需求端：技能升级，非极化



(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。



# 发现三：扩散 $\neq$ 平等

## 需求端：技能升级

- 行业 Gini: 0.20 (2016)  $\rightarrow$  0.14 (2025), 降幅 30%
- 中技能层 (数据分析 + 数据工程): 10.9%  $\rightarrow$  18.0%
- 高技能层 (AI/ML+ 编程): 10.1%  $\rightarrow$  8.0% (收缩)
- 单调上移, 非 U 型极化

## 工资端：上尾扩张

- P90-P10 差距: 1.67 (2016)  $\rightarrow$  峰值 2.08 (2018)  $\rightarrow$   $>1.25$  (2025)
- P90-P50 差距主导; P50-P10 相对稳定

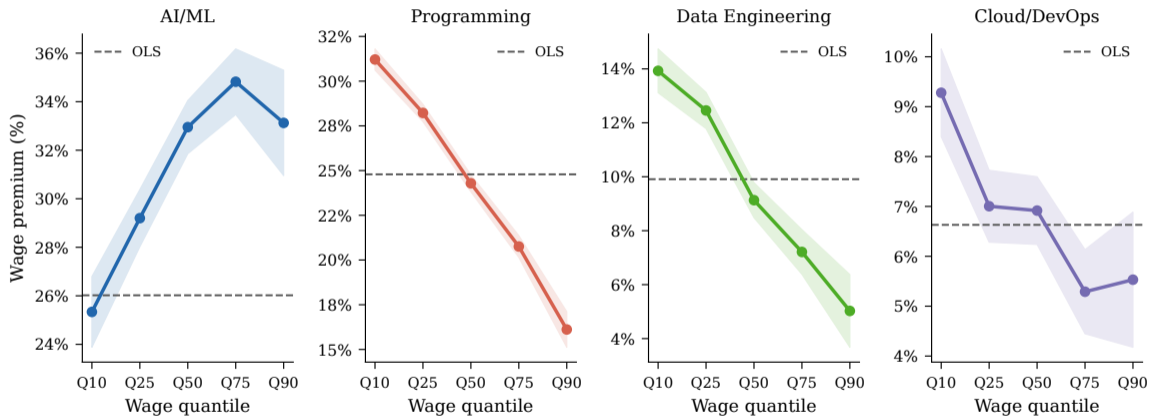
## 分位数回归揭示机制:

- AI/ML 溢价: Q10 25.3%  $\rightarrow$  Q75 34.8% (倒 U 形)
- 编程溢价: Q10 31.2%  $\rightarrow$  Q90 16.1% (单调递减)

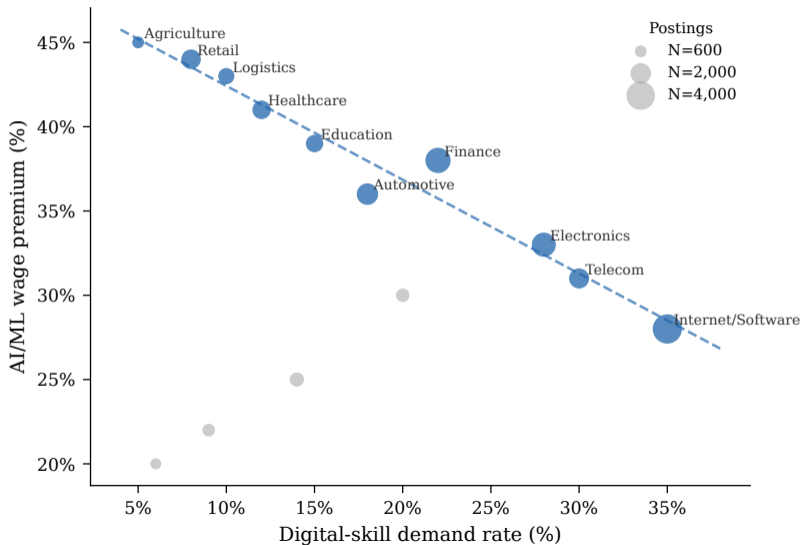
## 悖论解释

技能回报在工资分布上存在异质性：需求侧扩散不必然压缩工资不平等

# 分位数回归图



$\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ 。阴影带 = 95% 置信区间。水平虚线 = OLS 基准值。





1 研究背景与动机

2 理论框架

3 数据与计量策略

4 实证结果

5 稳健性检验

6 结论



# 测量方法对比 (8.1-8.2)

## 四种 AI 暴露度测量方法

- 1 关键词指数 (含否定过滤):** 排除“不要求机器学习”等假阳性
- 2 任务内容指数:** 将职位关键词映射至 Felten 等 (2019) 九大 AI 能力域, 计算 AIPI 得分
- 3 综合指数:** 方法 1 与方法 2 各 50% 加权
- 4 O\*NET 职业指数:** 通过 TF-IDF 余弦相似度匹配中文职位至 O\*NET 职业, 加权 AI 能力评分

## 方法对比结果

- 关键词指数:  $+0.045^{***}$  (识别需要 AI 技能的岗位, 正溢价)
- 任务内容指数:  $-0.108^{***}$  (识别可被 AI 替代的任务, 负溢价)
- O\*NET 指数:  $+0.080^{***}$  (与关键词指数一致)
- 两类指数相关系数  $r = 0.06$ , 反映概念差异

## 结论

技能需求溢价与任务可替代性折价并存;  
本文关键词指数捕捉前者, 结果稳健



# 薪资区间宽度：测量局限与保守下界

## 核心问题

- AI/ML 岗位平均薪资区间：11,676 元  
非 AI 岗位：5,750 元（比值 2.0）
- 招聘中点值系统性高估实际溢价

## 区间宽度 FE 检验 (Panel B)

- 加入薪资区间宽度十分位 FE 后：  
AI/ML 汇总溢价 0.434  $\rightarrow$  0.011（降幅 97%）  
仍显著 (\*\*,  $p = 0.013$ )
- 解释：企业对稀缺 AI 人才定价不确定，并非 AI 技能无溢价

## 截尾敏感性 (Panel A)

- 20% 对称截尾后：0.440  $\rightarrow$  0.177
- 保守下界：17–18%

## 按区间宽度分位数分组 (Table 2)

- 最宽区间组技能溢价大幅下降
- 窄区间组溢价更接近真实水平

## 结论

时间轨迹仍具信息量；  
截面溢价水平需谨慎解读



1 研究背景与动机

2 理论框架

3 数据与计量策略

4 实证结果

5 稳健性检验

6 结论



## 1 时变溢价（测量敏感）

AI/ML 溢价呈倒 U 形，与需求冲击叠加滞后供给响应一致 ( $r_{STEM} = -0.80$ )；  
但招聘工资中点存在系统性高估，保守下界 17-18%。

## 2 技能组合次可加性

五组显著负交互项，Jaccard 共现与次可加性单调对应；  
职位类别 FE 不改变结论，支持任务重叠机制。

## 3 技能升级 $\neq$ 工资平等

中技能层需求扩张（非 U 型），但上尾工资差距持续扩大；  
AI/ML 溢价集中于 Q75，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 年调整滞后；横截面研究对采样时点高度敏感

## （二）深度优先于广度

次可加性支持专注精通单项高价值技能，而非泛化数字素养；  
注意：不同职业最优技能组合不同，结论适用于边际投资决策

## （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

# 谢谢!

向稀  
南京大学商学院

欢迎批评指正