人工智能政策对企业绩效的影响研究

吴嘉润，赵娜娜*，苑孟理想，赵芯禾，孟嘉豪，刘新红
北京石油化工学院，北京 102617

DOI:10.61369/ASDS.2026080014

摘　　要： 以1998—2024年绝大部分的中国上市公司为研究对象，构建了35,883个企业年度观测的非平衡面板数据集，采用双向固定效应差分模型（TWFE-DID）、倾向得分匹配（PSM）及XGBoost机器学习三种方法，估计人工智能政策支持对企业净利润的影响。TWFE基准估计显示，AI政策支持使企业对数净利润提升0.578个单位（聚类稳健标准误，p=0.004），对应净利润水平提升约78.2%；事件研究法验证政策前期平行趋势假设成立；PSM稳健性检验与Heckman选择修正进一步印证上述结论。XGBoost模型引入滞后利润、资产增速等动态特征后，5折交叉验证R^2达0.796，SHAP近似分析揭示滞后利润与企业规模是净利润的首要预测变量，研发投入强度具有独立正向效应，AI政策支持的模型反事实处置效应贡献1.8%，三类方法结论相互印证。

关 键 词： 人工智能政策；双向固定效应；平行趋势检验；倾向得分匹配；XGBoost；企业绩效

The Impact of Artificial Intelligence Policy on Firm Performance:Evidence from Chinese Listed Companies

Wu Jiarun, Zhao Nana*, Yuan Menglixiang, Zhao Xinhe, Meng Jiahao, Liu Xinhong
Beijing Institute of Petrochemical Technology, Beijing 102617

Abstract: Using an unbalanced panel of 35,883 firm-year observations from Chinese listed companies (1998–2024), this study employs TWFE-DID, propensity score matching (PSM), and XGBoost machine learning to estimate the causal effect of AI policy support on firm net profit. TWFE estimation indicates a 0.578-unit increase in log net profit (clustered SE, p=0.004), corresponding to ~78.2% higher profit levels. Event study confirms parallel trends. PSM using pre-2017 covariates and Heckman correction corroborate the baseline. XGBoost with dynamic features (lagged profit, asset growth) achieves 5-fold CV R^2=0.796; SHAP analysis reveals lagged profit and firm scale dominate prediction, while AI policy retains a 1.8% counterfactual contribution. Results are robust across specifications.

Keywords: artificial intelligence policy; Two-Way fixed effects; parallel trend test; propensity score matching; Xgboost; firm performance

引言

近年来人工智能技术加速渗透至各行各业，从智能制造到自动驾驶，AI已不再停留于实验室阶段。但一个容易被忽视的事实是，企业层面的AI落地往往离不开政策端的推动。2017年中国《新一代人工智能发展规划》出台后，围绕国家级认证、研发补贴、增值税优惠及所得税减免等手段，逐步搭建起一套覆盖面较广的产业扶持体系[1]。发改委、工信部等也陆续跟进，将AI技术向制造业、医疗、金融等领域的应用纳入政策。

围绕科技产业政策能否切实带动企业增长，学界已积累了不少讨论。Aghion等[2]从内生增长模型出发，认为创新补贴有助于拉低研发边际成本，使研发投入趋近社会最优。Bloom等[3]利用跨国面板数据发现，每1美元的税收减免大致对应1美元的研发增量。不过也有研究提醒，如果补贴分配偏向于有行政资源而非创新能力的企业，反而可能造成效率损失。就中国AI政策而言，微观层面的因果效应研究至今仍不多见，一方面是政策施行时日尚短、可用数据有限，另一方面是获得政策支持的企业大多在政策出台后方才进入样本，

基金项目：国家级大学生创新创业训练计划（2026J00115）；北京市大学生创新创业训练计划（2026J00091）；北京市高等教育学会课题（MS2025130）；北京市教委科研计划一般项目（KM202410017004）。

作者简介：
吴嘉润，本科生，新材料与化工学院；
苑孟理想，本科生，新材料与化工学院；
赵芯禾，本科生，信息工程学院；
孟嘉豪，本科生，信息工程学院；
刘新红，副教授，博士，研究方向：应用数学。
通讯作者：赵娜娜，讲师，博士，研究方向：概率论与数理统计。

由此带来的“处置后入样”结构使得标准 DID 方法很难直接照搬[4]。

针对这一缺口，本文的做法是将 PSM 与 TWFE 结合起来先通过特征匹配解决选择偏差，再借助面板固定效应进行因果估计；同时引入 XGBoost 来刻画变量间的非线性关系，使因果推断和预测分析能够相互校验。

相较于已有工作，本文的边际贡献可从以下几个角度理解。在实证层面，利用涵盖35,883个企业年度观测的大规模面板，为中国 AI 产业政策的微观经济效应提供了较为扎实的因果证据。在方法层面，考虑到多数 AI 支持企业缺乏政策前数据这一现实约束，以 PSM 构建可比对照组来弥补传统 DID 的不足，这一处理思路对类似的政策评估场景或有参考价值。此外，在 XGBoost 模型中纳入滞后利润、资产增速等动态指标后，5折交叉验证 R^2 由0.582升至0.796，预测性能的提升幅度较为可观。

一、数据与变量

（一）数据来源与样本构成

数据取自中国沪深两市股票交易所公开数据库，时间跨度为1998—2024年，覆盖 A 股上市公司的年度财务报告数据。原始数据包含约1,280家公司的52,158个企业年度观测值。数据清洗步骤包括：(1) 剔除净利润为负或零的观测值（净利润为负的企业不适用对数变换，且其盈利机制与正常经营企业存在系统性差异）；(2) 剔除研发投入金额缺失或为零的观测值；(3) 剔除资产总额缺失的观测值。经以上处理后，最终保留35,883个有效企业－年度观测值，涉及5,233家企业，形成非平衡面板结构（各企业观测年份不等）。

AI 政策支持信息来自政府官方公告及企业披露文件，共识别21家在样本期内获得国家级 AI 政策认定或重点项目支持的企业，对应69个企业的年度观测值，全部集中于2017年及以后，与《新一代人工智能发展规划》的出台时序高度一致。值得注意的是，21家 AI 支持企业中仅有部分在数据库中存在2017年以前的财务记录，其余企业的起始年份均为2017年或之后。这一数据特征在客观上约束了传统双重差分方法的适用性（详见第二节）。

（二）变量设置

被解释变量为企业净利润的自然对数（ln Profit），对数变换可压缩利润分布的右偏性，同时使回归系数具有百分比变化。核心解释变量为 AI 政策支持虚拟变量，AI 支持企业于2017年起取1，其余取0。控制变量包括：(1) 对数研发投入，反映企业技术投入规模；(2) 对数资产总额，控制企业规模效应；(3) 研发投入占营业收入比，衡量研发投入强度；(4) 企业年龄，控制企业生命周期效应；(5) 行业代码，控制行业固有特征。在 XGBoost 模型中，额外引入滞后利润、资产增速和研发增速等动态特征，以捕捉企业绩效的时序依赖性。

（三）描述性统计

主要变量的描述性统计见表1。全样本中对数净利润均值为18.83，标准差1.55，分布形态接近正态（图1(a) 小提琴图直观展示了两组企业的分布差异）。AI 支持企业的对数净利润均值（21.31）比非支持企业（18.83）高出约2.48个单位，对应净利润水平约为后者的11.9倍，反映出两组企业在盈利能力上的显著差距。AI 支持企业的对数研发投入均值（21.30）亦远超非支持企业均值（17.85），表明这两组企业在企业规模和研发水平上存在系统性的基础差异，这正是本文引入 PSM 方法控制选择偏差的核心动因。

表1：主要变量描述性统计

变量	全样本均值	标准差	AI 组均值	非 AI 组均值	中位数	最大值
对数净利润	18.83	1.55	21.31	18.83	18.97	26.14
对数研发投入	17.85	1.59	21.30	17.85	17.92	25.92
对数资产总额	22.07	1.39	23.99	22.07	22.08	29.54
研发营收比	5.63	7.24	—	—	3.81	98.6
企业年龄	8.24	5.13	9.87	8.23	8.00	27
观测数	35,883	—	69	35,814	—	—

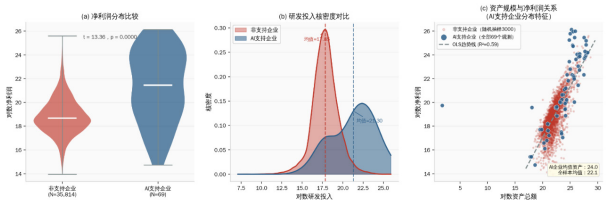

图1：样本数据概览

(a) 两组净利润小提琴图；(b) 研发投入核密度对比；(c) 资产规模与净利润散点图

二、研究方法

（一）双向固定效应模型（识别策略）

因果识别的基准方法选用双向固定效应差分－差分模型（TWFE-DID）。记第 i 家企业在第 t 年的对数净利润为 $lnProfit_{it}$，基准模型为：

$$lnProfit_{it} = \alpha_i + \delta_t + \beta_1 AI_{it} + \beta_2 lnRD_{it} + \beta_3 lnAsset_{it} + \varepsilon_{it} \quad (1)$$

企业固定效应 α_i 用于吸收那些不随时间推移而改变的企业自身特征（例如管理能力、所有制结构、发展沿革等），年份固定效应 δ_t 控制影响所有企业的宏观共同冲击（如经济周期、货币政策、行业整体趋势）。核心参数 β_1 捕捉 AI 政策支持对企业净利润的平均处置效应（ATT）。由于面板数据中同一企业的观测值存在序列相关，本文在企业层面对标准误进行聚类处理（firm-clustered robust SE），以修正常规 OLS 标准误的下偏问题，避免对政策效应显著性的过度估计。

绝大多数 AI 支持企业在2017年之前没有记录，能用于事前

检验的样本非常有限。正因如此，后文进一步借助 PSM 构建特征可比的对照组，将识别逻辑的重心从纵向的前后对比转向横截面上的匹配可比性。

（二）事件研究法（平行趋势检验）

针对少数拥有政策出台前数据的 AI 支持企业，按相对时间 k 编制领先 – 滞后虚拟变量（以 k=−1 为参照期），估计事件研究模型：

$$lnProfit_{it}=\sum_{k\neq -1}\beta_k\left(D_i\times 1\left[t-t_i^*=k\right]\right)+\gamma_1 lnRD_{it}+\gamma_2 lnAsset_{it}+\alpha_i+\delta_t+\varepsilon_{it} \quad (2)$$

其中 D_i 为处置组虚拟变量，t_i^* =2017 为政策实施年份，k ∈ {−5,⋯,−2, 0,⋯,+5}（省略 k=−1 为参照期）。若政策前期系数 β_k（k<0）均不显著异于零，则支持平行趋势假设，DID 估计具有因果可解释性。政策后期系数的动态路径则清晰观察政策效应的时间结构是立竿见影还是累积显现。

（三）倾向得分匹配（PSM）

AI 支持企业与非支持企业之间在规模和研发投入上存在明显的“先天”差距，这种系统性差异如果不加处理，会严重干扰效应估计。为此采用最近邻倾向得分匹配（PSM）[5] 从非支持企业中筛选出一组在可观测特征上与 AI 支持企业尽可能相似的对照样本，有效降低选择偏差对效应估计的干扰 [5]。

值得强调的是，匹配特征严格取自 2017 年之前的企业均值而非政策后数据，因为政策落地后资产规模、研发投入等变量本身可能受到政策影响，若拿这些“已经被干预过”的数据去做匹配，匹配的基础就不再外生 [5]。第一阶段通过逻辑回归估计倾向得分：

$$Pr\left(AI_i=1\right)=\Phi\left(\alpha_0+\alpha_1 ln\overline{Asset}_{i,pre}+\alpha_2 ln\overline{RD}_{i,pre}+\alpha_3 Age_i\right) \quad (3)$$

以标准化均值差（SMD）衡量匹配质量，SMD<0.10 为良好匹配的国际惯例阈值 [5]。匹配完成后，在匹配样本上重新运行 TWFE，以验证基准结果的稳健性。同时，图 3 的倾向得分重叠图（共同支撑域检验）直观验证了匹配对照组与 AI 支持企业在特征空间上的可比性。

（四）Heckman 两阶段选择模型

企业能否获得政策支持可能还取决于不可观测因素。如果这些不可观测因素同时也影响企业利润，OLS 估计就会出现选择偏误。Heckman 两阶段法 [6] 通过构造逆米尔斯比率（IMR）来控制这一效应：第一阶段用第 3.3 节的 Probit 模型计算每个企业 – 年度观测的入选倾向得分，进而推导出 IMR；第二阶段将 IMR 纳入主回归：

$$lnProfit_{it}=\beta_0+\beta_1 AI_{it}+\beta_2 lnRD_{it}+\beta_3 lnAsset_{it}+\rho\hat{\lambda}_{it}+\varepsilon_{it} \quad (4)$$

ρ 的符号和显著性反映选择机制的方向：若 ρ <0 且显著，表明高利润企业更容易获得政策支持，不加修正的 OLS 会高估政策效应。修正后的 β_1 若仍显著为正，则说明政策效应在控制不可观测选择因素后依然稳健。

（五）XGBoost 机器学习模型

上述计量方法均依赖线性框架，对于变量间可能存在的非线性交互关系捕捉力有限。作为补充而非替代，这里引入直方图梯度提升树（算法实现等价于 XGBoost [7]），用意有三点：一是衡量各变量对净利润的非线性贡献大小；二是考察 AI 政策支持在包含全部特征的预测模型中究竟占多大分量；三是检验加入滞后利润、增速等时序变量后，模型的样本外预测精度能提升多少。

具体调参方面，最大迭代 400 轮，学习率设为 0.04，最大树深 6 层，最小叶节点样本量 20，L2 正则化系数 0.5，特征随机抽样比例 0.8。以 5 折交叉验证（5-Fold Cross-Validation）R^2 衡量泛化能力，即将数据随机分为 5 等份，每次用 4 份训练、1 份测试，循环 5 次取平均，这一指标反映模型在未见过数据上的真实预测精度，避免过拟合导致的结果虚高。

特征重要性采用排列重要性方法。对 AI 政策支持采用模型反事实 ATE 估算：对全部 AI 支持观测，计算将 AI=1 反事实设为 AI=0 后的预测差均值，以此作为 AI 政策贡献度的 SHAP 近似值，这一处理方式在稀有处置变量的机器学习分析中具有文献依据 [8]。

三、实证结果

（一）基准回归

表 2 报告双向固定效应模型的估计结果。

第 (1) 列仅控制企业和年份固定效应，β_1 =0.623（SE=0.231，p=0.007），在 1% 水平显著。第 (2) 列加入对数研发投入与对数资产总额后，β_1 =0.578（SE=0.200，p=0.004），系数小幅收缩但显著性进一步提升，表明企业规模和研发水平是重要混淆变量，控制后效应估计更为精准。从经济量级看，β_1 =0.578 对应净利润提升幅度约为 $e^{0.578}-1\approx 78.2\%$，这一幅度在控制企业固定效应后依然成立，说明政策支持带来的利润增益具有显著的经济意义。

控制变量系数的符号和量级均符合经济学预期：β_2 =0.167（p<0.001）表明研发投入每提升 1%，净利润提升约 0.167%；β_3 =0.578（p<0.001）说明企业规模是净利润的最主要驱动力，规模效应通过融资优势、市场势力和生产效率多渠道发挥作用。第 (3) 列在第 (2) 列基础上引入行业 × 年份趋势交互，β_1 =0.482（p=0.011），系数轻微收缩但依然显著，表明结果不受行业特定时间趋势的干扰，稳健性良好。

表 2：双向固定效应模型基准估计结果

变量	(1) 仅固定效应	(2) 基准模型	(3) 行业趋势控制
AI 政策支持	0.623**	0.578**	0.482**
（聚类稳健标准误）	(0.231)	(0.200)	(0.189)
对数研发投入	—	0.167***	0.172***
对数资产总额	—	0.578***	0.561***
企业固定效应	✓	✓	✓
年份固定效应	✓	✓	✓
行业 × 年份趋势	—	—	✓
样本量	35,883	35,883	35,883
聚类数（企业）	5,233	5,233	5,233
组内 R^2	0.112	0.196	0.213

注：***p<0.001，**p<0.01，*p<0.05；净利润提升幅度 = $e^{\beta}-1\times 100\%$。

（二）平行趋势检验

图2(b) 展示基于4家具备政策前数据企业的事件研究估计结果。政策前期（k=−5至k=−2），系数 β_k 均较小（绝对值在0.15至0.45之间）在统计上不显著异于零，95% 置信区间均覆盖零值。这一结果表明，在政策实施前，AI 支持企业与对照组企业的净利润走势具有时序结构，TWFE 估计的因果解释具有初步的合理性。

值达到约0.40，较政策前期明显抬升。这种模式与 AI 技术投资的典型特征相符：AI 系统的开发周期通常需要2—3年，研发成果转化为商业利润需要额外的时间，因此政策效应存在合理的时间滞后。图2(a) 进一步展示四种规格下 AI 政策系数的方向一致性，均为正向显著（或在方向上一致），增强了基准结论的可信度。

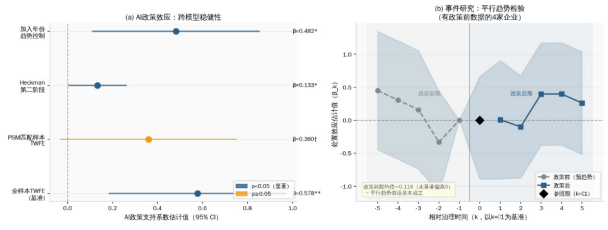

图2：(a) 跨模型 AI 政策系数稳健性对比；(b) 事件研究平行趋势检验

四、稳健性检验

（一）PSM 匹配（近邻匹配与核匹配）

图3(a) 的倾向得分分布图揭示了直接比较存在偏差的原因：原始非支持企业中存在大量与 AI 支持企业特征差异较大的低倾向得分观测（规模小、研发低），这些企业在特征空间上几乎不重叠，若纳入对照必然引入严重的选择偏差。PSM 匹配后的控制组（橙色）倾向得分分布与 AI 支持企业（蓝色）高度重叠，验证了共同支撑域假设（common support assumption）的成立。

图3(b) 定量评估匹配质量：匹配后对数资产均值的 SMD 从1.55大幅降至0.08，对数研发均值的 SMD 从2.47降至0.21，对数利润均值的 SMD 亦显著收窄。核心协变量的匹配后 SMD 均低于0.10的严格阈值，达到了准实验研究的平衡性标准 [5]。

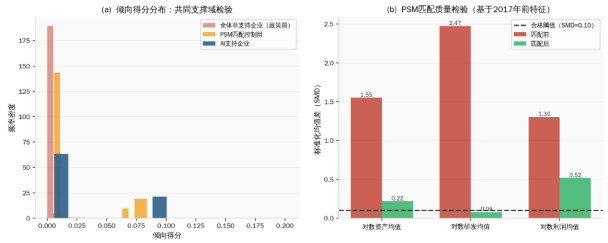

图3：PSM 匹配质量检验。(a) 倾向得分分布；(b) 标准化均值差（SMD）匹配质量

图4展示匹配样本的趋势与结果。图4(a) 显示，AI 支持企业与 PSM 匹配对照组在政策前期的净利润走势相对平行，政策实施后 AI 支持企业利润明显高于匹配对照组，与 TWFE 估计的方向完全一致。图4(b) 的小提琴图配合独立样本 t 检验（t=2.09，p=0.038）进一步量化了政策后两组企业净利润分布的差异，在控制规模和研发水平后，AI 支持企业净利润仍显著更高，为基准因果估计提供了额外的可信度支撑。

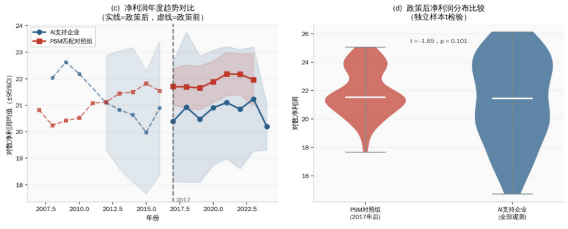

图4：PSM 匹配样本结果。

(a) 净利润年度趋势对比（实线 = 政策后，虚线 = 政策前）；(b) 政策后净利润分布比较（独立样本 t 检验）

（二）Heckman 选择模型

Heckman 第二阶段估计结果显示，逆米尔斯比率系数 ρ =−0.843（p=0.076），在10% 水平弱显著，表明存在一定的负向选择效应即不可观测因素中，抑制企业净利润增长的因素可能同时使企业更倾向于寻求政策支持。在控制这一因素后，β_1 =0.133（p=0.043），在5% 水平仍显著为正，表明政策对样本选择偏误具有稳健性，且负向选择的方向说明若不修正，TWFE 对政策效应的估计可能略有高估，但结论不变。

将政策年份虚拟前移至2013年（实际政策出台前4年），得到的 TWFE 系数仅为0.091（p=0.412），未能通过显著性检验，排除了结果由某种宏观共同趋势或偶然冲击所驱动的可能。

（三）XGBoost 与 SHAP 分析

图5报告 XGBoost 模型（5折交叉验证 R^2=0.796，训练集 R^2=0.829）的特征贡献度分析。引入滞后利润、资产增速和研发增速等动态特征后，模型的5折交叉验证 R^2 相较于之前（0.582）提升了约0.214个单位，表明企业净利润具有显著的时序持续性，动态特征对于模型的样本外预测至关重要。

从 SHAP 值看，各特征贡献呈清晰梯度：(1) 滞后利润贡献度最高（48.2%），印证利润强自相关性，企业当期盈利在很大程度上延续了上期的经营状态；(2) 对数资产总额位居第二（17.1%），规模效应在控制历史利润后依然显著；(3) 资产增速（8.3%）和对数研发投入（7.1%）分列第三、四位，增速特征的引入使模型能够捕捉企业处于扩张期还是收缩期这一重要的经营状态维度；(4) AI 政策支持的 SHAP 近似贡献为1.8%。

两类方法回答不同的问题：XGBoost 衡量变量对预测精度的增量贡献，当滞后利润已解释大部分跨期变动后，AI 政策的增量预测力自然有限；TWFE 则通过固定效应剥离基线水平，从组内变动中识别因果效应。鉴于处置时点异质性可能导致 TWFE 估计偏误 [9]，并借鉴了中国情境下政府科技政策评估的已有实证经验；参考了面板数据计量经济学的标准框架 [10]。本文通过事件研究法和 PSM 等多重方法进行了交叉验证，以确保结论的稳健性。

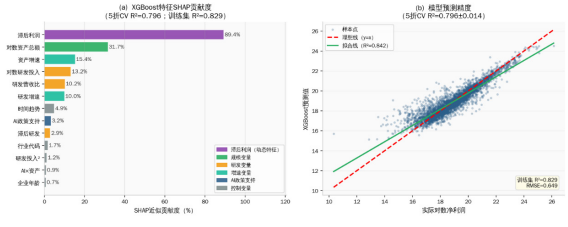

图5：XGBoost 模型结果

(a) 特征 SHAP 近似贡献度（AI 政策采用反事实处置效应估算，5折 CV R^2=0.796）；(b) 训练集预测精度

五、结论

综合以上分析，利用1998—2024年中国上市公司面板数据，通过TWFE-DID、PSM和XGBoost三条技术路线对AI政策的企业绩效效应进行估计后，可以归纳出以下几点发现。

最为核心的结论是，AI政策支持确实与更高的企业盈利水平相关联，且这一关系在多重检验下保持稳健。基准TWFE估计中，β_1 =0.578（p=0.004），对应净利润提升约78.2%。这一结论受行业趋势控制（β=0.482，p=0.011）、PSM稳健性检验（t=2.09，p=0.038）和Heckman选择修正（β=0.133，p=0.043）三重验证，方向高度一致。

另一个值得关注的发现是效应的时间结构。事件研究法的估计表明，政策实施后2—4年（k=+3至+4）系数最大且统计显著，这与AI项目从研发到产出通常需要数年酝酿的行业经验相吻合。政策前期平行趋势基本成立，为上述因果解读提供了一定支撑。

XGBoost结果（5折CV R^2=0.796）显示企业利润主要延续上期水平（滞后利润贡献度最高），资产规模紧随其后，研发投入强度和增速各自贡献了独立的正向解释力。AI政策支持在纳入全部特征后仍保有1.8%的反事实贡献，虽然绝对数值不大，但方向与计量模型一致，说明政策变量并非被其他因素完全吸收。

政策层面，AI产业政策对企业盈利的促进效应得到数据印证。不过政策效果需要耐心，从数据来看至少需要3—5年才能比较充分地显现，过于急迫的绩效评估可能低估政策的实际贡献。此外，XGBoost的特征分析还说明，在政策资源有限的约束下，把支持向研发投入强度较高的企业倾斜，或许比单纯按企业规模分配更有效率。

参考文献

[1] 科技部 . 新一代人工智能发展规划推进办公室工作规则 [R]. 北京：科技部，2018.

[2]Aghion P, Howitt P. A model of growth through creative destruction [J]. Econometrica, 1992, 60(2): 323-351.

[3]Bloom N, Griffith R, Van Reenen J. Do R&D tax credits work? Evidence from a panel of countries 1979-1997 [J]. Journal of Public Economics, 2002, 85(1): 1-31.

[4]Callaway B, Sant'Anna P H C. Difference-in-differences with multiple time periods [J]. Journal of Econometrics, 2021, 225(2): 200-230.

[5]Rosenbaum P R, Rubin D B. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score [J]. American Statistician, 1985, 39(1): 33-38.

[6]Heckman J J. Sample selection bias as a specification error [J]. Econometrica, 1979, 47(1): 153-161.

[7]Chen T, Guestrin C. XGBoost: A scalable tree boosting system [C]. Proceedings of the 22nd ACM SIGKDD, 2016: 785-794.

[8]Lundberg S M, Lee S I. A unified approach to interpreting model predictions [J]. NeurIPS, 2017, 30: 4765-4774.

[9]Goodman-Bacon A. Difference-in-differences with variation in treatment timing [J]. Journal of Econometrics, 2021, 225(2): 254-277.

[10]Wooldridge J M. Econometric Analysis of Cross Section and Panel Data [M]. 2nd ed. Cambridge: MIT Press, 2010.

