丘成桐中学科学奖 · 经济金融建模奖赛道候选课题 20 则

> 本方案面向 **2027 赛季**（2026 赛季提交截止为 2026 年 9 月 15 日，从零启动已来不及）。倒排时间表以 2026 年 9 月为第 0 周、2027 年 9 月 15 日提交为终点，共约 52 周。

学生画像与资源假设：普通重点高中生，有一定编程与数学基础；计算资源为个人笔记本、纯 CPU、无 GPU 无集群；实验资源不超过中学常规实验室；全部使用公开数据；由中学教师即可指导。凡超出这一假设的路线，都在标签里标明了资源门槛。

四条必须先知道的赛制事实（取自官方公告，2020 – 2026 全部公示文章）：指导老师不得来自培训机构或任何谋利机构，违者取消资格；总决赛全程英文（研究报告、PPT、答辩）；入围总决赛的研究报告会在 11 月全网公示；提交材料含论文查重报告与实验视频。规则每年会改，报名开放后须重新核对当年《参赛规则》与本学科《参赛指南》。

这个赛道的评委在奖励什么

统计 2020 – 2025 共 65 项经济金融建模奖获奖论文，这个赛道的口味比名字暗示的窄得多。

"经济金融建模"实际奖励的主要是应用微观计量与理论博弈，不是量化金融。 两大主族几乎平分了全部席位：

其一是**政策与事件的因果识别**：双重差分、事件研究、断点回归、合成控制。碳交易对企业绿色创新的影响、澳大利亚 JobSeeker 补贴的乘数效应、中美贸易战与专利布局、隔代抚养与子女学业、清洁空气行动与心理福祉、新农保与劳动供给的道德风险、AI 对劳动需求的企业层面影响（2025 金奖）。这一族是本赛道的绝对主流，也是公开数据最能支撑的一族。

其二是**理论建模与机制设计**：统计性歧视为何低效的代理理论（2020 金奖）、无限市场的双边匹配、非对称竞赛中的最优偏袒、燃油车规制的税收 vs 碳配额理论（2021 金奖）、搭便车问题的博弈论分析（2023 金奖）、信用体系如何打破重复囚徒困境、政府碳抵消机制设计。**这一族拿了六个金奖里的三个**，且只需要纸笔与少量数值求解——对资源受限学生极其友好，却常被误以为"太难"。

资产定价与量化策略是少数派，且天花板低。 六年里只有谱偏度与股票估值、LASSO-BiLSTM 汇率预测、动量投资与马科维茨对比等寥寥几篇，全部停在优胜奖或 finalist。这是一个重要信号：单纯的"回测出了超额收益"在这个赛道不吃香，而且它的技术陷阱（数据窥探、幸存者偏差、前视偏差、交易成本）恰好是评委最会追问的。本方案 E01 – E10 的资产定价路线一律定位为"已发表异象在新样本上的独立检验"——结论正负都成立，且免疫于数据窥探的指控。

数据可得性是这个赛道的第一杀手。 中学生拿不到 Wind、Bloomberg、CRSP、Compustat。历届获奖里用的多是 CFPS、CGSS、大众点评公开数据、上市公司年报、专利数据库、招聘网站公开职位——全部是免费或半免费来源。本方案每条路线都写明了免费数据源的时间覆盖、频率与缺失问题，凡不确定的一律标注"需核实"而非假装确定。

近年主题漂移：2023 年起碳政策与绿色金融、平台经济与区块链溯源、AI 对就业与企业绩效的影响明显上升。2025 年有四篇与 AI 经济学相关。这些题材的公开数据正在变多，是当前性价比最高的切入方向。

本赛道 20 条路线的分族逻辑

- **E01 – E10 金融市场与计量经济族**：以已发表结果的独立检验与政策因果识别为主。统计口径全部预先写死——Newey-West 或聚类稳健标准误、多重检验校正、按时间划分而非随机划分、结构性断点检验。凡涉及策略的路线都强制要求交易成本假设与样本外检验，并明确"策略不显著同样是有效结论"。
- **E11 – E20 社会经济系统与建模族**：博弈论、机制设计、基于主体的建模、网络扩散、运筹优化、行为实验。这一族的最大风险不是算力而是"自说自话"——模型参数拍脑袋、结论无法被数据检验。每条路线都要求模型被外部数据或已发表实证结果校准，纯模拟路线必须在有解析解的极限上验证并做全局敏感性扫描。涉及校内人类被试的路线写明了知情同意、匿名化与检验力估计要求（"找几十个同学填问卷"不是有效设计，方案里写明了为什么）。

两族独立排序：E01/E11 最稳，E10/E20 风险与上限最高。

E01 · 美股已发表因子溢价的"发表后衰减"：基于 Ken French 数据库 2013–2026 延展窗口的独立检验

`优先级:最高` · `族:资产定价实证` · `数据:纯免费(Ken French)` · `算力:极低(CPU 秒级)` · `技能:统计推断`

1 · 研究问题

> 规模、价值、动量、盈利、投资五个已发表因子（factor premium）在 McLean – Pontiff (2016) 样本截止（2013）之后的 2013 – 2026 窗口内，月均溢价相对发表前样本衰减了多少个百分点？衰减幅度是否与"套利资本消除错定价"假说预测的方向一致，且在 95% 置信区间内可与"纯数据窥探"（data snooping）情形区分？

2 · 研究背景与空白

因子溢价指按公司特征（市值、账面市值比、过去收益等）排序构造的多空组合的平均超额收益，是资产定价实证的核心对象。若溢价源于数据窥探，发表后应立即消失；若源于错定价，发表后随套利资本进入而逐步衰减；若源于风险补偿，则应持续存在。三种机制对"发表后样本"给出可区分的预测。

McLean & Pontiff (2016, Journal of Finance) 用 97 个已发表异象测得：样本外收益衰减约 26%，发表后五年衰减约 58%（样本截止约 2013 年）。Jacobs & Müller (2020, JFE) 在全球样本上发现发表后衰减主要是美国现象。Jensen, Kelly & Pedersen (2023, JF) 用贝叶斯框架认为多数因子可复现且样本外仍成立（代码公开于 GitHub bkelly-lab/ReplicationCrisis）。三者结论并不一致。

空白在时间覆盖：2013 – 2026 这 13 年（含 2018 – 2020 年著名的"价值因子失灵"与 2022 年后的反弹）尚未被上述框架以统一口径覆盖。方法完全公开、数据免费、结论正负都成立，是典型的"新样本独立检验"型学生课题。

3 · 可检验假设

- H1：五个因子（SMB、HML、UMD、RMW、CMA）在 2013-01 – 2026-06 窗口的月均溢价，相对各自"发表前样本"均值平均衰减 ⩾ 40%，且至少 3 个因子的发表后溢价在 t > 3.0 门槛下不再显著。
- H2（机制备择）：衰减幅度与因子换手率/套利成本代理（如小市值腿占比）正相关——若成立支持"套利资本"解释；若衰减均匀分布则更接近整体风险环境变化。H1 被否证（溢价未衰减）同样是有效结论，支持风险补偿解释。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用 Ken French 数据库提供的 6 个 Size – B/M 组合自建管线重构 SMB 与 HML，在 Fama – French (1993) 原始样本区间 1963-07 – 1991-12 上与官方因子序列比对：相关系数 ⩾ 0.995，月均值偏差 ⩽ 0.02 个百分点，t 值偏差 ⩽ 5%。UMD 用 6 个 Size – Momentum 组合同法校验。**此步不过关，后续全部结论无效。**
2. 统计口径预先写死：月度序列均值检验一律用 Newey – West 标准误，滞后阶按 floor(4(T/100)^(2/9)) 并另报滞后 6 与 12 的稳健性；报月均溢价点估计 + 95% CI + 年化夏普，而非只报 p 值。
3. 多重检验校正：仅预注册 5 个因子，按 Bonferroni 校正后判显著性，并同时报告 Harvey – Liu – Zhu (2016) 的 t > 3.0 门槛结论。

4. 结构性断点：对每个因子做 Bai – Perron 多断点检验与以发表年份为断点的 Chow 检验，报断点位置的 90% 置信区间。
5. "随机划分 vs 时间划分"对照：额外做一次随机打乱月份的伪"前后"对比，量化若不按时间划分会高估/低估多少衰减。
6. 可复现性：全部脚本（数据下载→因子重构→检验→图表）开源，第三方一键重跑 < 10 分钟。

5 · 数据与工具

用途	来源 / 工具
因子与组合月度收益（1926–07 至今，每月更新，CSV 免费）	Ken French Data Library（mba.tuck.dartmouth.edu），下载 F–F Research Data Factors、Momentum、6 Portfolios 系列；总量 < 20 MB
无风险利率交叉核对	FRED（TB3MS），免费 API
官方因子序列（仅用于校验与对比，**不计入本项目数据贡献**）	同 Ken French 库
分析环境	Python：pandas、statsmodels（Newey–West、Chow）、ruptures 或自写 Bai–Perron；纯 CPU，全流程秒—分钟级
对照文献结果（仅对比，不计入贡献）	McLean–Pontiff (2016) 表 3；Jensen et al. (2023) GitHub 复现代码

6 · 方法路径

1. 搭环境，下载 Ken French 全套 CSV，完成第 4.1 条因子重构硬门槛校验。
2. 整理五因子的"发表前 / 发表后 / 2013 后延展"三段样本定义，预注册全部检验与滞后阶选择。
3. 计算各段月均溢价、Newey – West t 值、CI 与夏普，复算 McLean – Pontiff 口径的衰减百分比。
4. 做 Bai – Perron / Chow 断点检验与滚动 10 年窗口溢价图，定位衰减发生时点。
5. 构造套利成本代理，检验 H2 的横截面相关（仅 5 个观测点，须用置换检验并明示功效有限）。
6. 做随机划分对照与滞后阶敏感性分析。
7. 交叉校验：用 Jensen et al. 公开代码的美国子集重算同口径衰减，比对差异并解释。

7 · 新颖性边界

- 本课题不提出新因子、不声称发现新异象；"发表后衰减"现象本身由 McLean – Pontiff (2016) 确立（97 个异象，样本至约 2013），Jacobs & Müller (2020) 与 Jensen – Kelly – Pedersen (2023) 已做大规模延伸。
- 本项目贡献（主结论）：以统一、完全可复现的免费数据管线，给出这 5 个核心因子在 2013 – 2026 延展窗口的衰减幅度点估计与置信区间，并区分三种机制的预测。溢价"未衰减"与"已衰减"都是有效结论，误差棒必须证明有能力分辨 40% 量级的衰减。
- 历届丘奖对照：2024 年提名奖 "Spectral Coskewness and the Valuation of Stocks"（深圳中学）做的是新定价量的横截面检验；本课题定位相反——不造新量，只做已发表量的独立延展检验，评审谱系上属于该赛道罕见但方法学最扎实的类型。

8 · 决策门槛（go / no-go）

- **第 2 周末**：第 4.1 条硬门槛必须通过。若 SMB/HML 重构相关系数达不到 0.995，先排查组合权重口径（价值加权、NYSE 断点）；第 4 周仍不达标则**降级**：放弃自建重构，直接采用官方因子序列做全部检验，方法学校验改为"复算 McLean – Pontiff 表 3 中可用 Ken French 数据覆盖的因子行，衰减百分比偏差 ⩽ 5 个百分点"。主结论框架（延展窗口衰减估计）完全保留。
- **第 8 周末**：五因子三段样本的全部点估计与 CI 完成。若 H2 横截面检验功效过低（置换检验 p 值区间过宽），降级为纯描述性机制讨论，H1 仍为主结论。
- 已知困难即研究内容：2018 – 2020 价值因子巨幅回撤会使"衰减 vs 断点"难以区分——这正是断点检验要回答的问题，不是障碍。
- 适配前提:适合统计功底较好、接受"全程无策略回测、无超额收益故事"定位的学生;本题的卖点是推断严谨性，不是收益率。

E02 · Liu–Stambaugh–Yuan 中国三因子模型（CH-3）在 2017–2026 年 A 股的样本外独立检验

`优先级:很高` · `族:资产定价实证` · `数据:akshare+官方因子序列` · `算力:低(CPU 分钟级)` · `技能:数据工程+统计`

1 · 研究问题

> CH-3 模型的规模因子与价值因子（EP 口径的 VMG）在其原始样本（2000 – 2016）之后的 2017-01 – 2026-06 窗口内，月均溢价与 t 值是否仍达到原文量级？注册制改革与退市常态化削弱"壳价值"后，CH-3 剔除市值最小 30% 股票的核心设计是否仍必要？

2 · 研究背景与空白

Liu, Stambaugh & Yuan (2019, JFE)《Size and Value in China》指出：A 股小盘股价格被"壳价值"（借壳上市预期）污染，剔除市值最小 30% 后，用盈利市值比（EP）构造的价值因子 VMG 与规模因子构成的 CH-3 模型显著优于照搬 Fama – French 三因子，样本区间 2000 – 2016。该文是中国因子研究的方法学源头，因子构造规则完全公开，作者主页持续发布更新后的 CH-3/CH-4 因子序列（需核实当前更新时点）。

2019 年科创板注册制、2020 年创业板注册制、2021 年起退市新规、2023 年全面注册制——这一系列制度变化直接打击借壳预期，理论上应使"壳价值污染"减弱，即 CH-3 的核心设计前提在新样本中可能弱化。检索未见以 2017 – 2026 完整十年为样本、按原文方法逐条重构并检验该前提的公开英文论文（少数会议文章做过局部再评估；"未找到"不等于"不存在"，第 4 周须再核）。

空白在时间覆盖与前提检验：方法公开、有官方因子序列可校验、结论正负都成立，且制度变迁提供了明确的经济学故事，适合一年期课题。

3 · 可检验假设

- H1：2017 – 2026 窗口内 VMG 月均溢价仍为正，但相对 2000 – 2016 原文样本衰减 ⩾ 30%，Newey – West t 值降至 3.0 以下。
- H2（前提检验）：市值最小 30% 组的"壳价值溢价"（用原文的壳概率代理或最简化的小市值组超额收益代理）在 2019 年注册制后显著收缩——若成立，说明 CH-3 的剔除规则在新制度下收益递减；若小盘异常依旧，则剔除规则仍必要。两个方向都是有效结论。

4 · 量化验收标准

1. 方法学校验（硬门槛）：用 akshare 全市场数据自建管线重构 CH-3 三因子，在与作者官网因子序列的重叠区间（至少 2005 – 2016）上比对：月度相关系数 ⩾ 0.90，年化均值偏差 ⩽ 1.5 个百分点。达不到须逐项归因（退市股缺失、股本口径、财报时点对齐）。**此步不过关，后续全部结论无效。**
2. 幸存者偏差量化：统计管线中缺失的退市股数量与市值占比，按年列表；给出"剔除最小 30% 后缺失退市股对因子收益的影响上界"估算。
3. 统计口径：月度溢价用 Newey – West（滞后阶同 E01 规则并报敏感性）；前后期差异报点估计 + 95% CI；断点用以 2019-07（科创板开市）为先验断点的 Chow 检验加 Bai – Perron 无先验检验。
4. 多重检验：预注册的检验不超过 6 个，Bonferroni 校正；引用性比较一律用 t > 3.0 门槛。

5. 财报数据使用披露日之后的信息（避免前视偏差）：所有会计变量滞后至公告日后次月末才进入排序，管线中写死并单测。
6. 可复现性：数据快照存档 + 脚本开源，重跑（除下载外）< 30 分钟。

5 · 数据与工具

用途	来源 / 工具
A 股日行情、总股本、市值（约 2000 至今;退市股覆盖不全,幸存者偏差须按第 4.2 条量化）	akshare（免费、无积分限制）;tushare 免费额度作交叉核对(积分制,历史财务接口可能受限,需核实)
财务数据（净利润、公告日期）	akshare 财报接口 + 巨潮资讯网公告日期核对;数据量约 5000 股 × 26 年,清洗后 < 2 GB
官方 CH-3/CH-4 因子序列（**仅用于校验与对比，不计入本项目数据贡献**）	Stambaugh 个人主页 finance.wharton.upenn.edu/~stambaug（更新时点需核实;若停更,重叠区间以已发布部分为准）
无风险利率	中国人民银行/中债一年期国债或原文口径的一年定存利率,免费
分析环境	Python: pandas、statsmodels;纯 CPU,因子重构全量约分钟—小时级

6 · 方法路径

1. 搭环境，用 akshare 拉取全市场月末市值与 EP 所需财务数据，建立含公告日对齐的特征表。
2. 按原文规则（剔除上市 < 6 个月、剔除最小 30% 市值、2×3 独立排序、价值加权）重构 CH-3，完成第 4.1 条硬门槛校验并写偏差归因报告。
3. 量化退市股缺失（第 4.2 条），给出偏差上界。
4. 计算 2017 – 2026 窗口的因子溢价、t 值、CI，与原文样本对比，检验 H1。
5. 构造壳价值代理，做 2019 断点前后的对比与 Chow/Bai – Perron 检验，检验 H2。
6. 稳健性：改用现金分红率/BM 口径的价值因子、改剔除比例（20%/40%）做敏感性。
7. 交叉校验：用官方更新因子序列直接重算 H1（若其覆盖 2017 后），比对自建管线结论。

7 · 新颖性边界

> - 本课题不声称提出新模型；CH-3 的构造、壳价值机制、2000 – 2016 的全部实证结论归 Liu – Stambaugh – Yuan (2019)。中国市场"动量缺失、反转显著"等横截面事实也已有大量文献（如 Anomalies in the China A-share market, PBFJ 2021）。
> - 本项目贡献（主结论）：2017 – 2026 十年样本外窗口上 CH-3 溢价的独立再估计，以及注册制冲击下"剔除最小 30%"设计前提的直接检验。溢价衰减与否、前提弱化与否，四种组合都是可发表的有效结论。
> - 历届丘奖对照：该赛道尚无 CH-3 类因子检验获奖论文;最接近的是 2024 提名奖 Spectral Coskewness（新定价量）与 2022 金奖 Herding Behavior around "Smart Money"（交易行为），本课题的"制度变迁 × 模型前提"角度与两者均不重叠。

8 · 决策门槛（go / no-go）

- **第 6 周末**：核实 akshare 全市场历史行情+财务的实际覆盖（特别是 2010 年前与退市股）。若 2005 – 2016 重叠校验区间可用股票不足官方口径的 85%，**降级路径 A**：把自建重构窗口缩至 2010 – 2026，2005 – 2009 用官

方序列衔接；**降级路径 B**：完全采用官方因子序列做 H1（前提是其更新至 2024 后），自建部分只做 H2 的壳价值代理。两条路径主结论框架不变。

- **第 10 周末**：硬门槛校验须通过（相关系数 $\geqslant 0.90$）。若卡在 0.80 – 0.90，允许以"偏差归因分析"替代精确复现作为方法学章节，但 H1 结论须同时用自建与官方两套序列报告，且以官方序列为准。
- **第 20 周末**：H2 壳价值代理若数据不可得（借壳事件表难以免费获取），降级为"最小 30% 组超额收益的断点检验"这一弱化版前提检验，明示其局限。
- 适配前提：需要较强的数据清洗耐心;财报对齐是本题最大工作量（估计占 40% 工时），选题前学生须明确接受。

E03 · 波动率预测模型赛马：GARCH(1,1) 对 HAR 与极差估计量在中美指数 2005–2026 上的 QLIKE/DM 检验

`优先级:高` · `族:波动率建模` · `数据:yfinance+akshare(免费)` · `算力:低(CPU 小时级)` · `技能:时间序列计量`

1 · 研究问题

> 在只用免费日频 OHLC 数据的约束下，HAR 类模型对 GARCH(1,1) 的样本外波动率预测优势（以 QLIKE 损失与 Diebold – Mariano 检验衡量）在沪深 300、标普 500、恒生指数三个市场、1/5/22 日三个预测期上是否稳健存在？该优势在 2015 股灾、2020 疫情、2024-09 政策行情等高波动区间是否放大？

2 · 研究背景与空白

波动率不可直接观测，评估预测模型需要代理：学术标准是 5 分钟已实现波动率（realized volatility），但免费渠道拿不到长历史高频数据；基于日内最高/最低价的极差估计量（Parkinson 1980、Garman – Klass 1980）效率约为日收益平方的 5 – 7 倍，且只需 OHLC——这是本课题绕开数据墙的关键。

Hansen & Lunde (2005) 比较 330 个模型后发现汇率上没有模型显著击败 GARCH(1,1)，个股上含杠杆效应的模型更优。Corsi (2009) 的 HAR 模型以日/周/月三尺度已实现量做线性回归，成为 RV 预测基准。中国市场已有大量用 5 分钟 RV 的比较研究（如 Realized GARCH 类，样本多止于 2020 年前后），结论普遍是含已实现量的模型更优。

空白在评价维度与时间覆盖：已有中国研究几乎全部以高频 RV 为信息集与代理；"仅用免费 OHLC 极差信息还能保留多少 HAR 优势"这一对学生和散户真实可得的信息集问题未见系统回答，且 2021 – 2026（含 2024 年极端行情）样本未被覆盖。预测优势不显著同样是有效结论。

3 · 可检验假设

- H1：以 Garman – Klass 极差方差为目标与回归元的 HAR-RG 模型，在 1 日预测期上对 GARCH(1,1) 的 QLIKE 损失改进 $\geqslant 10\%$，DM 检验 $p < 0.05$（三个市场中至少两个成立）。
- H2：预测期拉长到 22 日后，HAR-RG 对 GARCH(1,1) 的优势缩小到 QLIKE 改进 $< 5\%$ 或不显著——即极差信息的优势主要在短期。若 H1 本身不成立（GARCH(1,1) 未被击败），呼应 Hansen – Lunde 的"难以击败"结论，同为有效结果。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：(a) 自写 GARCH(1,1) 极大似然估计，在同一数据上与成熟库 arch 的参数估计偏差 $\leqslant 1\%$、对数似然偏差 $\leqslant 0.1\%$；(b) 在模拟 GBM 路径上验证 Parkinson/Garman – Klass 估计量的无偏性（10^4 条路径，偏差 $\leqslant 2\%$，自助法 CI 覆盖）。**此步不过关，后续全部结论无效。**
2. 统计口径：训练/测试严格按时间划分（滚动 1000 交易日窗口，逐日或每 5 日重估）;绝不随机划分，并附一次"随机划分对照"量化高估幅度。损失函数只用 QLIKE 与 MSE（Patton 2011 证明对噪声代理稳健的两类），**不用 MAE/MAPE 并说明原因**。
3. 模型比较：两两用 Diebold – Mariano（HAC 方差，滞后阶 = 预测期-1 起报并做敏感性）；全体模型用 Model Confidence Set (Hansen – Lunde – Nason 2011) 控制多重比较，$\alpha = 0.10$。

4. 报告效应量：QLIKE 改进百分比 + 自助法 95% CI；分高/低波动子样本（以 VIX/中国波指或滚动波动率中位数划分）报告。
5. 每市场样本外预测点 $\geqslant$ 2500 个交易日。
6. 可复现性：脚本开源，全流程 CPU 重跑 $\leqslant$ 2 小时（约 6 模型 $\times$ 3 市场 $\times$ 5000 次滚动重估，GARCH 单次拟合亚秒级）。

5 · 数据与工具

用途	来源 / 工具
标普 500（^GSPC，1950 至今）、恒生（^HSI，1987 至今）日频 OHLC	yfinance，免费；缺失率与拆分调整须自检
沪深 300 日频 OHLC（2005-04 指数发布至今）	akshare 指数接口（免费）；与 yfinance 000300.SS 交叉核对，二者不一致处以交易所公开数据为准
高波动区间划分参考	CBOE VIX（FRED 免费）;中国波指已停发,用滚动实现波动率替代并注明
5 分钟 RV 小样本旁证（可选,近 1–2 年）	akshare 分钟数据（历史深度有限,需核实;仅作附录旁证,不进主结论）
分析环境	Python: arch、statsmodels、numpy;MCS 自写或用现成实现(需核实维护状态);纯 CPU
对照基准（**仅校验对比,不计入贡献**）	Hansen–Lunde (2005)、Corsi (2009) 原文结果;中国市场 Realized GARCH 文献结论

6 · 方法路径

1. 搭环境，完成第 4.1 条两项硬门槛校验。
2. 下载三市场 OHLC，清洗（停牌、涨跌停日的极差退化处理规则须预先写死并可复现）。
3. 实现模型族：GARCH(1,1)、GJR-GARCH、EWMA(RiskMetrics)、HAR-RG（极差版 HAR）、AR(1)-RG、混合 GARCH-X（极差作外生量）。
4. 滚动样本外预测 1/5/22 日波动率，计算 QLIKE/MSE。
5. 做 DM 两两检验与 MCS，输出主结果表。
6. 分高/低波动子样本与三大危机窗口做异质性分析，检验 H2。
7. 交叉校验：用近 1 – 2 年分钟数据算 5 分钟 RV，检验极差代理与 RV 的相关性（预期 $\geqslant$ 0.9），为主结论的代理有效性提供旁证。

7 · 新颖性边界

- 本课题不提出新模型；GARCH/HAR/极差估计量与 QLIKE/DM/MCS 全部方法归原作者。中国市场"含已实现量模型更优"的结论已被高频 RV 文献反复得到，不得声称为本项目发现。
- 本项目贡献（主结论）：在"免费 OHLC 信息集"这一明确约束下的跨市场、跨预测期系统赛马，与 2021 – 2026 新样本（含 2024 极端行情）的覆盖。若 HAR 优势在此信息集下消失，是对高频文献结论适用边界的有效刻画。
- 历届丘奖对照：2022 铜奖 "LASSO-BiLSTM ensemble for exchange rate forecasting" 属预测建模但无严格损失函数比较与 DM/MCS 推断；本课题反其道而行——模型简单、推断从严，差异化清晰。

8 · 决策门槛（go / no-go）

- **第 4 周末**：三市场 OHLC 数据质量核查完成（缺失率 < 1%、极差为零的涨跌停日占比统计）。若沪深 300 的 OHLC 在早期年份质量不达标，**降级**：样本起点后移至 2010，或以上证综指（1990 至今）替换，主结论框架不变。
- **第 8 周末**：硬门槛与基线两模型（GARCH(1,1)、HAR-RG）滚动预测跑通。若滚动重估耗时超预算 10 倍，降级为每 20 日重估一次（文献常规做法之一），并报告重估频率敏感性。
- **第 16 周末**：主结果表若显示所有差异均不显著且 CI 极宽（无法分辨 10% 的 QLIKE 改进），把功效分析写成研究内容：用模拟标定"该样本量下可分辨的最小改进"，结论改述为适用边界刻画。
- 适配前提:适合喜欢时间序列与编程、能接受"没有交易策略、只有预测精度"叙事的学生;全程无须任何付费数据。

E04 · 尾部风险模型回测：GARCH–EVT、GARCH–t 与历史模拟在 A 股与美股指数上的 VaR/ES 违约检验

`优先级:高` · `族:风险度量与极值理论` · `数据:yfinance+akshare(免费)` · `算力:低(CPU 小时级)` · `技能:统计+极值理论`

1 · 研究问题

> McNeil – Frey (2000) 的 GARCH-EVT 两步法在沪深 300、创业板指与标普 500 的 2010 – 2026 滚动样本外回测中，其 99% VaR 与 97.5% ES（expected shortfall，预期损失）的违约表现是否仍系统优于 GARCH-t 与历史模拟？A 股涨跌停制度截断的收益分布是否改变 EVT 尾部拟合的相对优势？

2 · 研究背景与空白

VaR（value at risk）与 ES 是监管与风控的标准尾部风险度量；巴塞尔协议 IV 已把市场风险标准度量从 99% VaR 转向 97.5% ES。模型好坏由回测（backtesting）判定：VaR 看违约次数与独立性（Kupiec 1995 的 POF 检验、Christoffersen 1998 的独立性检验），ES 用 Acerbi – Szekely (2014) 类检验。

McNeil & Frey (2000, J. Empirical Finance) 提出两步法：先用 GARCH 过滤收益得到近似 i.i.d. 的标准化残差，再用广义帕累托分布（GPD）的超阈值方法（POT）拟合残差尾部；在 1990 年代美欧数据上显著优于正态与无条件方法。后续文献（如 Novales & Garcia-Jorcano 2019）在多市场确认条件 EVT 类模型的 ES 预测更准。中国市场亦有近作以 e-backtesting 框架在巴塞尔 IV 口径下评估 ES 方法（Risks, 2026），但其关注监管审慎性而非涨跌停制度对 EVT 拟合的影响。

空白在评价维度与样本：A 股 ±10%/±20% 涨跌停对收益分布形成硬截断，理论上压缩单日尾部并把极端风险转移为连续跌停——这对 GPD 尾部拟合是结构性挑战，未见以 2010 – 2026 样本、跨涨跌停宽度（主板 10% vs 创业板 2020 后 20%）系统检验。"EVT 在截断分布上无优势"同样是有效结论。

3 · 可检验假设

- H1：在标普 500 上，GARCH-EVT 的 99% VaR 违约率落入 Kupiec 95% 接受域且 ES 检验不拒绝，而历史模拟至少在一个口径被拒绝——复现经典结论。
- H2：在主板指数（沪深 300）上，GARCH-EVT 相对 GARCH-t 的优势小于其在标普 500 上的优势（用违约率偏离与 ES 检验统计量差值衡量）；在涨跌停放宽后的创业板（2020-08 后）该优势部分恢复。方向相反则说明截断主要由过滤步骤吸收，同为有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：(a) 自写 GPD 极大似然拟合，在形状参数 ξ 已知的模拟数据（学生 t、Fréchet 域，n=250 超阈值样本，10^3 次重复）上偏差 $\leqslant$ 5%，自助法 CI 覆盖率 90 – 98%；(b) 用教科书公开算例复算 Kupiec 与 Christoffersen 检验统计量，逐位吻合；(c) 在标普 500 上定性复现 McNeil – Frey 的结论排序（条件 EVT $\geqslant$ 条件 t > 无条件）。**此步不过关，后续全部结论无效。**
2. 统计口径：滚动窗口 1000 交易日，每 10 日重估，严格时间序;阈值选取写死为残差的 90% 分位并报 85%/95% 敏感性;每指数样本外 $\geqslant$ 2500 日、99% VaR 期望违约 $\geqslant$ 25 次。

3. 多模型 × 多指数 × 多口径的检验族做 FDR（Benjamini – Hochberg，q=0.10）校正，并同时报未校正结果。
4. 报效应量：违约率点估计 + Clopper – Pearson 95% CI，ES 检验统计量 + 模拟 p 值，不只报"拒绝/不拒绝"。
5. 连续跌停的处理规则（视为单日观测还是合并事件）预先写死，作为稳健性的两套口径都报。
6. 可复现性：脚本开源，全流程 CPU ⩽ 3 小时（约 4 模型 × 4 指数 × 900 次滚动重估）。

5 · 数据与工具

用途	来源 / 工具
标普 500（1950 至今）、沪深 300（2005 至今）、中证 500、创业板指（2010 至今）日收盘	yfinance + akshare 交叉核对，免费;指数无个股幸存者偏差问题
涨跌停制度时间表（2020–08–24 创业板 10%→20% 等）	交易所公告（深交所官网），一级来源
分析环境	Python: arch(GARCH)、scipy(GPD MLE 自写为主、genpareto 对照)、statsmodels;纯 CPU
对照基准（**仅校验对比,不计入贡献**）	McNeil–Frey (2000) 原文表;Novales & Garcia–Jorcano (2019) 结论

6 · 方法路径

1. 搭环境，完成第 4.1 条三项硬门槛校验（模拟校验 → 检验统计量复算 → 经典结论复现）。
2. 下载四指数收益，清洗并标注涨跌停制度区间与连续跌停事件。
3. 实现四模型：历史模拟、GARCH-正态、GARCH-t、GARCH-EVT(POT)。
4. 滚动样本外预测 99%/95% VaR 与 97.5% ES，记录违约序列。
5. 做 Kupiec/Christoffersen/ES 检验与 FDR 校正，输出主结果矩阵。
6. 分制度子样本（创业板 2020-08 前后）检验 H2，附残差尾部 QQ 图与 ξ 估计的时变图。
7. 交叉校验：用 Hill 估计量独立估计尾指数，与 GPD 的 ξ 比对（预期同数量级），并做阈值敏感性图。

7 · 新颖性边界

- 本课题**不**提出新风险模型；两步法归 McNeil – Frey (2000)，回测检验归 Kupiec/Christoffersen/Acerbi – Szekely。条件 EVT 更优的一般结论已被多市场文献确立。
- 本项目贡献（主结论）：涨跌停截断这一制度特征对 EVT 相对优势的影响估计——利用创业板 2020 年放宽这一制度变化做同一市场内的前后对照，这一评价维度未见系统处理;以及 2010 – 2026（含 2015 股灾、2024-02 微盘股流动性危机）的新样本回测记录。
- 历届丘奖对照：该赛道历届无 VaR/ES 回测类获奖论文，风险度量方向整体空缺;与 2020 提名奖 G20 股市关联网络论文（相关结构）无方法重叠。

8 · 决策门槛（go / no-go）

- **第 6 周末**：硬门槛 (a)(b) 必须通过;(c) 若标普 500 上无法复现经典排序，先查过滤残差的自相关（Ljung – Box），第 8 周仍失败则**降级**：放弃"复现经典"表述，改为"在现代样本上的全新回测记录"，方法学校验以 (a)(b) 模拟与算例校验为准，主结论框架（制度截断 × 模型比较）不变。

- **第 12 周末**：确认样本外违约次数达到第 4.2 条下限。若创业板 2020 后子样本违约数 < 15 次（功效不足），降级为 95% VaR 口径为主、99% 为辅，并报功效模拟。
- 已知困难即研究内容：连续跌停使"单日损失"定义模糊——两套口径的差异本身就是涨跌停如何扭曲风险度量的证据，写入正文而非附录。
- 适配前提:需要一定的概率论兴趣（GPD、尾指数）;数学强、编程中等的学生最适配。

E05 · A 股隔夜负收益之谜（T+1 制度）的 2020–2026 新样本独立检验与 A/H 双重上市对照

优先级:较高 · 族:市场微观结构 · 数据:akshare+yfinance(免费) · 算力:低 · 技能:实证金融+数据工程

1 · 研究问题

> A 股平均隔夜收益为负、日内收益为正的独特分解格局（与美股相反）在 2020-01 – 2026-06 的新样本中是否依然成立、量级几何？同一公司 A 股（受 T+1 限制）与 H 股（无 T+1）的隔夜收益差异是否与 T+1 解释的预测一致？

2 · 研究背景与空白

把日收益分解为隔夜（前收盘→开盘）与日内（开盘→收盘）两段，是微观结构文献识别投资者群体行为的标准手法。美股隔夜收益系统为正（Lou, Polk & Skouras 2019, JFE）;而 A 股相反：Qiao & Dam (2020, Journal of Financial Markets)《The overnight return puzzle and the "T+1" trading rule in Chinese stock markets》测得 A 股平均隔夜收益显著为负，归因于 T+1 制度（当日买入不得当日卖出）使短线买家在开盘价上要求约 14 个基点的折价补偿。后续有 2023 – 2024 年论文研究隔夜-日内反转与异象收益的隔夜/日内分布（Overnight versus intraday returns of anomalies in China, PBFJ 2023）。

空白在时间覆盖与识别设计：已发表样本大多止于 2010 年代后期;2020 年以来量化交易占比上升、2024 年程序化交易新规、融券收紧等都可能改变隔夜折价的量级。A/H 双重上市公司提供了"同一现金流、不同交易制度"的天然对照，现有公开英文文献中以 2020 后样本系统使用该对照检验 T+1 解释的工作未见（第 6 周须再核;"未找到"不等于"不存在"）。隔夜折价消失或收缩本身就是制度成本变化的有效证据。

3 · 可检验假设

- H1：2020 – 2026 样本中 A 股等权平均隔夜收益仍显著为负（Newey – West t ⩽ -3.0），但量级较 Qiao – Dam 原样本收缩 ⩾ 30%。
- H2：约 140 家 A/H 双重上市公司中，同公司 A 股隔夜收益系统低于 H 股隔夜收益，差值与该股换手率（短线交易需求代理）正相关;若 A/H 差值不显著，则 T+1 解释被削弱，指向卖方隔夜风险等替代解释——两个方向都是有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用自建管线在与 Qiao – Dam 原文重叠的样本区间上复算平均隔夜收益：符号一致，量级偏差 ⩽ 20%（免费数据与 CSMAR 的口径差须归因：复权方式、剔除规则、新股处理）。**此步不过关，后续全部结论无效。**
2. 统计口径：等权与市值加权都报;均值检验用 Newey – West（日频滞后 21 起报并做敏感性）;横截面回归用按股票与月份双向聚类的稳健标准误;报点估计（基点/日）+ 95% CI。
3. 前视与幸存者偏差控制：样本为逐月重构的"当月在市股票"名单;akshare 退市股缺失比例按年披露，并做"仅沪深 300 成分股"（成分历史可得）的稳健性对照。
4. 复权与除权除息处理规则写死：隔夜收益必须用前收盘价（含除权调整）计算，管线单测覆盖分红除权日。
5. 多重检验：主检验族 ⩽ 8 个，Bonferroni 校正。
6. 可复现性：脚本开源，CPU 全流程 ⩽ 1 小时。

5 · 数据与工具

用途	来源 / 工具
A 股全市场日频开/收盘与前收盘（约 2000 至今;退市覆盖不全,按第 4.3 条处理）	akshare，免费;抽样 50 只与交易所行情核对
H 股日频开/收盘	yfinance（.HK 代码），免费;港股无 T+1、无涨跌停
A/H 双重上市名单及对应关系	恒生 AH 股溢价指数成分（恒生官网/akshare AH 比价接口），免费
汇率（A/H 收益可比化）	FRED 或央行中间价，免费
对照基准（**仅校验对比,不计入贡献**）	Qiao—Dam (2020) 原文表;Lou—Polk—Skouras (2019) 美股分解
分析环境	Python: pandas、statsmodels(聚类稳健 SE);纯 CPU,数据 < 3 GB

6 · 方法路径

1. 搭环境，拉取全市场日频 OHLC 与前收盘，写复权单测。
2. 在重叠区间复算 Qiao – Dam 的平均隔夜收益，完成硬门槛校验与偏差归因。
3. 计算 2020 – 2026 隔夜/日内收益分解的时间序列与横截面分布，检验 H1。
4. 构建 A/H 配对面板，计算同公司隔夜收益差，做双向聚类回归检验 H2。
5. 横截面异质性：按换手率、市值、涨跌停触及频率分组，检验 T+1 解释的可检验推论。
6. 事件检验：2024 年程序化交易监管等制度节点前后的隔夜折价变化（探索性，明确标注）。
7. 交叉校验：用沪深 300 成分子样本（无幸存者偏差）重跑全部主检验，比对结论稳定性。

7 · 新颖性边界

- 本课题不声称发现隔夜负收益现象——该现象及 T+1 解释归 Qiao – Dam (2020)，美股隔夜溢价归 Lou – Polk – Skouras (2019)，异象的隔夜/日内分布已有 PBFJ 2023 论文。
- 本项目贡献（主结论）：2020 – 2026 新样本的独立再估计 + A/H 同公司对照这一识别设计在新样本上的系统执行。折价收缩/不变/消失三种结果都直接回答"T+1 的隐性成本是否随市场结构演化"这一政策相关问题（T+0 改革是反复出现的公共议题）。
- 历届丘奖对照：2022 金奖 Herding Behavior around "Smart Money" 属交易行为实证但不涉及隔夜分解;该赛道无隔夜收益类获奖论文。

8 · 决策门槛（go / no-go）

- **第 6 周末**：核实 akshare 前收盘/复权数据质量（50 只抽样与交易所数据全对齐）与 Qiao – Dam 复算结果。若量级偏差 > 20% 且无法归因，**降级**：主样本改为沪深 300 + 中证 500 成分股（约 800 只，数据质量可控），全市场结果降为附录，主结论框架不变。
- **第 10 周末**：A/H 配对面板建成（≥ 100 对，≥ 1000 交易日）。若 H 股开盘价数据缺失严重，降级：H2 改用"A 股内部横截面"版本——按融券可得性/换手率分组检验 T+1 推论，识别力下降但框架保留。

- 已知困难即研究内容：A/H 不同交易时区与假日错位使"同一隔夜"定义需要仔细对齐——对齐规则本身写入方法章节。
- 适配前提:适合对交易制度感兴趣、细心程度高的学生;无需高深数学，但数据对齐工作量大（约占 40% 工时）。

E06 · 上证 50ETF 期权定价的数值方法收敛性验证与隐含波动率微笑的免费数据测量

`优先级:中` · `族:衍生品定价数值方法` · `数据:akshare 期权接口(部分需核实)` · `算力:低(CPU)` · `技能:数值计算+金融工程`

1 · 研究问题

> 自建的二叉树（CRR）、蒙特卡洛与显式有限差分三种定价器在 Black – Scholes 解析解上的收敛阶各是多少、达到 0.1% 精度各需多大计算量？用经此校验的定价器反解 2023 – 2027 年上证 50ETF 期权的隐含波动率，其微笑/偏斜形态与平价关系（put-call parity）偏离是否被交易成本边界所约束？

2 · 研究背景与空白

期权定价的数值方法（树、蒙特卡洛、有限差分）是金融工程的基本功，三者在 BS 模型下有解析解可作精确标尺，收敛阶（CRR 为 O(1/N)，MC 为 O(N^-1/2)，显式差分受稳定性条件约束）是可测量、可复现的对象。隐含波动率（implied volatility, IV）微笑指同一到期日不同行权价的 IV 呈非平坦形态，直接否证 BS 常数波动率假设。

上证 50ETF 期权 2015-02 上市，是境内最活跃的场内期权。已有研究充分记录其 IV 微笑：Sui et al. (2024, Applied Economics)《Implied volatility curve: an empirical study on Chinese SSE 50 ETF options》刻画了微笑形态的决定因素;更早的 Finance Research Letters 2022 文章比较了中国期权市场的定价模型。这些工作依赖 Wind 等付费数据。

空白在可复现性与数据路径：用纯免费数据 + 自建并经收敛性验证的定价管线，对 2023 – 2027 最新样本（含 2024-09 波动率跳升）的微笑与平价偏离做出可一键重跑的测量，未见公开工作。本课题的方法学贡献（收敛性对照与精度-成本曲线）在计算类课题中是被承认的贡献形态，但定位必须讲清，否则会被误读为教科书练习——差异化全在验收标准的严格度与新样本测量上。

3 · 可检验假设

- H1：50ETF 认沽期权 IV 系统高于同 delta 认购（负偏斜），且偏斜在 2024-09 行情中加深 ⩾ 5 个 IV 百分点（vega 加权）。
- H2：平价关系偏离的分布以零为中心、95% 的观测落在双边交易成本界（佣金+买卖价差估计）之内;若系统性单侧偏离，则指向 ETF 融券约束——两个方向都是有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：三种定价器对欧式看涨/看跌在 ⩾ 20 组参数（覆盖实值/虚值、短/长到期）上与 BS 解析解偏差 ⩽ 0.1%;实测收敛阶与理论值偏差 ⩽ 15%（对数-对数回归斜率）;MC 报标准误并验证 CI 覆盖;显式差分给出稳定性条件的实测边界。**此步不过关，后续全部结论无效。**
2. IV 反解用 Brent 法，收敛容差 10^-8;对深度实值/临近到期合约的失败率单独报告，剔除规则写死（如剩余期限 < 5 天、成交量为零的合约剔除）。
3. 统计口径：微笑形态用每日横截面回归（IV 对 moneyness 的二次拟合）的系数时间序列刻画，均值检验用 Newey – West;2024-09 前后对比报效应量 + 95% CI;检验族 ⩽ 6 个，Bonferroni 校正。

4. 平价偏离以基点计，交易成本界的每个组成部分（佣金、价差、冲击）给出来源与保守/宽松两套假设。
5. 无风险利率与分红口径写死：用 SHIBOR/国债到期匹配利率，ETF 分红处理规则单列。
6. 可复现性：定价器 + 数据管线开源，全流程 CPU $\leqslant$ 2 小时。

5 · 数据与工具

用途	来源 / 工具
50ETF 期权日行情（结算价、收盘价、成交量、行权价、到期日）	akshare 期权接口（option_ 系列）：当日快照确定可得; 日度历史深度需核实*——若不足,自 2026-09 起每日自动抓取快照建库(52 周可积累约 250 交易日)
备选期权历史数据	中金所股指期权(IO,2019-12 上市)每日结算行情公开文件(cffex.com.cn 下载区,**需核实批量可得性**)
50ETF 现货日行情、分红	akshare/yfinance(510050.SS),免费
无风险利率	SHIBOR(全国银行间同业拆借中心官网)/中债国债收益率曲线,免费
分析环境	Python: numpy、scipy;三种定价器全部自写(这是课题核心);纯 CPU,单次全链重跑分钟级
对照基准(**仅校验对比,不计入贡献**)	BS 解析解;Sui et al. (2024) 的微笑形态结论

6 · 方法路径

1. 自写 BS 解析定价与三种数值定价器，完成第 4.1 条收敛性硬门槛，绘制精度-计算量曲线。
2. 第 1 周即启动期权数据每日抓取脚本（cron），同时核实历史接口深度——这是全课题最早的不可逆动作。
3. 建立合约主数据表（行权价、到期日、调整合约标记），实现分红调整与利率匹配。
4. 反解全样本 IV，生成微笑/期限结构的每日快照与时间序列。
5. 检验 H1：偏斜量级、2024-09 前后（若历史数据可得）或 2026-27 内高低波动区间对比。
6. 检验 H2：平价偏离分布与交易成本界比对。
7. 交叉校验：用 CRR 定价器对美式特征（50ETF 期权为欧式，此步用于说明方法边界）与用两条独立利率曲线重算 IV，量化口径敏感性。

7 · 新颖性边界

- 本课题不声称发现 IV 微笑或平价偏离——两者是教科书事实，中国 50ETF 期权的微笑已被 Sui et al. (2024) 等刻画;数值方法的收敛阶是已知理论结果。
- 本项目贡献（主结论）：(1) 三种方法在统一测试集上的收敛性对照与一条可操作的"精度-成本"选择判据（方法学可复现性贡献）;(2) 纯免费数据管线对 2023/2026 – 2027 最新样本微笑与平价偏离的测量及其交易成本归因。二者合并构成主结论，缺一则降为练习。
- 历届丘奖对照：该赛道历届无期权/衍生品定价获奖论文，属空缺方向;需在引言中主动与"教科书复现"划清界线（验收标准第 1、4 条即为界线）。

8 · 决策门槛（go / no-go）

- 第 3 周末：核实 akshare 期权历史深度与中金所文件批量下载可行性。若两者都拿不到 2023 – 2025 历史，降级：主样本改为"2026-09 起自建快照库"（约 250 交易日），H1 的 2024-09 对比改为样本内高/低波动区间对

比，主结论框架（收敛性 + 微笑测量）不变。

- **第 8 周末**：收敛性硬门槛必须全部通过——这部分无外部依赖，不通过说明实现有误，只能修复不能降级。
- **第 20 周末**：若 IV 反解失败率 > 10%，收紧合约剔除规则并把失败模式（深度实值、零成交）写成数据质量小节。
- 适配前提:适合数理与编程双强、对金融工程感兴趣的学生;是 10 题中唯一"理论校验占半壁江山"的课题，数据风险由自建快照库兜底，但要接受历史深度受限的可能。

E07 · 创业板涨跌停放宽（2020-08-24）对波动率与尾部厚度的长窗口 DID 再检验

`优先级:中` · `族:政策评估(DID)×微观结构` · `数据:akshare(免费)` · `算力:低` · `技能:因果推断`

1 · 研究问题

> 创业板日涨跌幅限制由 ±10% 放宽至 ±20%（2020-08-24）后，相对主板匹配对照组，个股波动率与收益分布尾部厚度（Hill 尾指数）在 1 年、3 年、5 年三个时间尺度上分别变化了多少？已发表文献报告的短期波动上升在长窗口内是收敛还是持续？

2 · 研究背景与空白

涨跌停板是价格稳定机制与价格发现障碍之争的经典对象。2020-08-24 创业板注册制改革同步把涨跌幅从 10% 放宽到 20%（存量股票一并适用），构成罕见的准自然实验：同为 A 股、同一监管环境，只有创业板被处理。

已有工作：Wan & Zhang (2022, Economics Letters) 用 DID 发现放宽后流动性改善、波动率显著上升，价格效率无显著变化;International Review of Financial Analysis (2023) 与 PLOS One (2023) 的类似设计确认波动与流动性上升、短期效应大于长期，且在散户关注度高的股票上更强;另有研究发现股价崩盘风险下降。这些论文的样本窗口普遍止于 2021 – 2022，即事件后 1 – 2 年。

空白在时间覆盖与评价维度：(1) 事件后 5 年（至 2026）的长窗口未被覆盖——期间创业板经历 2021 高点、2022 – 2024 回撤与 2024-02 微盘股流动性危机，短期结论能否外推是开放问题;(2) 已有文献用方差/振幅衡量波动，未系统检验**尾部厚度**（Hill 尾指数）——而放宽限价的理论争点恰恰在极端日收益的再分布（10% 截断把大跌转化为连续跌停）。效应收敛与持续都是有效结论。

3 · 可检验假设

- H1：DID 估计的波动率上升（日收益标准差口径）在事件后第 1 年显著为正（聚类 t ⩾ 3.0），但第 3 – 5 年衰减 ⩾ 50%——即市场逐步吸收更宽的限价。
- H2：创业板个股收益的 Hill 尾指数在放宽后显著下降（尾部变厚），且连续跌停事件频率下降 ⩾ 50%——即极端风险从"多日连续跌停"形态转为"单日大跌"形态。若尾指数不变，说明 10% 截断本就少有约束力，同为有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：在 Wan – Zhang (2022) 的原样本窗口（事件前后各约半年）与可比口径下复现其 DID 波动率系数：符号一致、量级偏差 ⩽ 30%（免费数据与其 CSMAR 源的差异须归因）。**此步不过关，后续全部结论无效。**
2. 识别假设检验写死：事件前 24 个月逐月动态系数图（event-study plot）检验平行趋势，事件前系数联合 F 检验 $p > 0.10$ 方可继续;两个安慰剂——伪事件日（2018-08）与伪处理组（中小板/主板内部随机分组，500 次置换）。
3. 统计口径：标准误按股票聚类，另报按行业×月双向聚类;报每个时间尺度的 DID 点估计 + 95% CI（效应量以事件前波动率的百分比表述）;PSM 匹配（市值、换手、行业、事件前波动）的平衡性表必须通过（标准化偏差 < 0.1）。

4. Hill 尾指数：每股票-年用超过 95% 分位的观测估计，阈值敏感性（90%/97.5%）必报;尾指数 DID 用自助法 CI。
5. 幸存者偏差：2021 起退市加速直接与结论相关——退市股缺失名单按年披露，并以"事件时点在市股票固定面板"为主口径避免构成漂移。
6. 可复现性：脚本开源，CPU 全流程 ⩽ 2 小时（约 1000 处理股 + 1000 对照股 × 8 年日频）。

5 · 数据与工具

用途	来源 / 工具
创业板与主板个股日频行情（2016–2026）	akshare，免费;退市股覆盖问题按第 4.5 条处理
事件与制度时间表	深交所官方公告（一级来源）
行业分类、市值、换手率	akshare;行业用申万或证监会分类,口径写死
连续跌停事件识别	由日频行情自建（收盘价触及跌停判据写死）
分析环境	Python: pandas、statsmodels/linearmodels(聚类 SE、事件研究)、自写 Hill 估计;纯 CPU
对照基准（**仅校验对比,不计入贡献**）	Wan–Zhang (2022)、IRFA (2023) 的 DID 系数

6 · 方法路径

1. 搭环境，建立 2016 – 2026 创业板/主板日频面板与制度标注。
2. 完成 PSM 匹配与平衡性检验，画事件研究动态系数图验证平行趋势。
3. 在原文窗口复现 Wan – Zhang 系数，完成硬门槛校验与偏差归因。
4. 估计 1/3/5 年窗口的波动率 DID（H1），跑两类安慰剂与置换推断。
5. 估计 Hill 尾指数与连续跌停频率的 DID（H2），做阈值敏感性。
6. 异质性：按散户关注代理（换手率、龙虎榜频率）分组，对接已有文献的机制说法。
7. 交叉校验：用合成控制法对创业板指数层面重估总量效应，与个股 DID 比对方向。

7 · 新颖性边界

> - 本课题不声称发现"放宽限价提高波动"——该结论归 Wan – Zhang (2022) 及 IRFA/PLOS One (2023) 等;DID/PSM 方法为标准工具。
> - 本项目贡献（主结论）：(1) 事件后 5 年长窗口的动态效应路径——短期文献结论是否收敛此前无人回答;(2) 尾部厚度与连续跌停形态转移这一新评价维度。两者合并回答"放宽限价的代价是暂时的还是结构性的"。
> - 历届丘奖对照：该赛道 DID 类获奖论文多为宏观/社会议题（2021 铜奖 COVID 垃圾 DID、2025 铜奖新农保道德风险等），无金融市场制度评估类;与 2022 提名奖"交通限制与经济下行"仅共享方法框架，对象完全不同。

8 · 决策门槛（go / no-go）

- **第 6 周末**：平行趋势检验必须通过。若创业板与匹配主板组事件前趋势显著背离（联合 F 检验 $p < 0.10$），**降级**：改用创业板内部"高/低涨跌停触及频率"分组的三重差分（DDD），或以 2019-07 科创板（上市即

20%）新股 vs 创业板老股做替代对照，主结论框架（限价宽度 × 波动/尾部）保留。

- **第 10 周末**：硬门槛复现须达标;若量级偏差 > 30% 且归因指向退市股缺失，主口径改为"事件时点在市固定面板"并在结论中明示覆盖边界。
- **第 24 周末**：若 H2 尾指数 DID 的自助法 CI 宽到无法分辨任何合理效应，把功效模拟写成研究内容，主结论退守 H1 + 连续跌停频率（计数型，功效更高）。
- 适配前提:适合对因果推断感兴趣的学生;方法链条长（PSM+DID+置换+尾指数），建议 2 – 3 人组队分工。

E08 · Brexit 的产出成本：合成控制法对 Born et al. (2019) 的复现与 2019–2025 延展估计

`优先级:中` · `族:宏观计量与政策评估(合成控制)` · `数据:OECD/IMF/世界银行(免费)` · `算力:极低` · `技能:宏观计量`

1 · 研究问题

> 用 Born et al. (2019) 的合成控制（synthetic control）设定复现"脱欧公投至 2018 年英国实际 GDP 损失约 2%"的结果后，把样本延展到 2025 年，在剔除或约束 COVID 混杂的多种口径下，英国产出相对其"合成对照体"（doppelganger）的累计缺口是多少、置信度几何?

2 · 研究背景与空白

合成控制法用捐赠池（donor pool）国家的加权组合拟合处理前的英国经济路径，处理后组合的走势即为"没有脱欧的英国"反事实。Born, Müller, Schularick & Sedláček (2019, Economic Journal)《The Costs of Economic Nationalism》用 30 个 OECD 国家季度实际 GDP 估计：至 2018 年公投已致英国产出损失约 1.7 – 2.5%，并做了安慰剂推断。此后：Springford（Centre for European Reform）的定期更新至 2022 年估计缺口约 5%（方法为匹配组合，被批评对捐赠池选择敏感）;高盛 2026-06 报告以 doppelganger 法估计缺口约 5 – 6%（非学术、方法细节不全公开）。

空白在时间覆盖与推断严格度：2019（正式脱欧）—2021（贸易合作协议生效）—2025 的完整政策期缺少**学术级推断**（in-space 安慰剂、RMSPE 比、捐赠池敏感性、留一法）的公开估计;现有更新多为智库/投行图表。最大方法障碍——COVID 对英国与捐赠池的不对称冲击——本身就是研究内容：能否分辨"脱欧缺口"与"COVID 缺口"是本课题要正面回答的问题，答案是"不能完全分辨、只能给出界"也是有效结论。

3 · 可检验假设

- H1：复现口径下（样本截至 2018Q4），自建管线的产出缺口估计与 Born et al. 的点估计差 $\leqslant$ 0.5 个百分点。
- H2：延展至 2025Q4 后，在"剔除 2020 – 2021 观测"与"含 COVID 但以捐赠池 COVID 严重度协变量约束"两种口径下，累计缺口点估计均 $\geqslant$ 3%，且英国的 RMSPE 比在捐赠池安慰剂分布中位于前 10%。若两口径给出的界宽到覆盖零，结论改述为"COVID 后合成控制无法可信识别脱欧成本"——同为有效且有价值的方法论结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：自写合成控制求解器（约束二次规划）先在 Abadie – Diamond – Hainmueller (2010) 加州控烟经典算例上复现原文权重与缺口路径（缺口曲线逐点偏差 $\leqslant$ 5%，公开数据与代码可得）;再完成 H1 的 Born et al. 复现。**此步不过关，后续全部结论无效。**
2. 推断口径写死：不用常规 t 检验;用 in-space 安慰剂的 RMSPE 比排名 p 值、in-time 安慰剂（伪公投日 2012）、捐赠池留一法（逐国剔除重估）三件套，全部主图必报。
3. 预测变量集与 Born et al. 一致并另报两套替代集（仅 GDP 滞后;加投资/开放度/通胀），三套结果并列。
4. 报效应量：季度缺口路径 + 累计缺口（百分点与货币量级两种表述）+ 敏感性区间，不报单一点估计。

5. 结构性断点：对英国-合成体缺口序列做 2016Q3/2020Q1/2021Q1 三个先验断点的检验，区分公投、COVID、TCA 生效三阶段贡献。
6. 可复现性：数据快照 + 求解器 + 全部图表脚本开源，CPU 重跑 < 10 分钟。

5 · 数据与工具

用途	来源 / 工具
30 国季度实际 GDP（1995Q1–2025Q4）	OECD Data Explorer（免费 API/CSV）;IMF IFS 交叉核对
预测变量（投资率、贸易开放度、通胀、人口）	世界银行 WDI + OECD,免费
COVID 严重度协变量（超额死亡、严格度指数）	Our World in Data / Oxford Stringency Index,免费下载
经典算例校验数据（**仅校验,不计入贡献**）	Abadie et al. (2010) 加州控烟公开数据(synth 包自带)
对照基准（**仅对比,不计入贡献**）	Born et al. (2019) 原文图表;CER Springford 更新;高盛 2026 报告(非学术,只作旁证)
分析环境	Python: scipy.optimize(自写求解器为主)、pysyncon 作对照(能力边界需核实);纯 CPU,秒级

6 · 方法路径

1. 自写合成控制求解器，在加州控烟算例上完成第一重硬门槛校验。
2. 下载并整理 30 国季度 GDP 与预测变量面板，处理基期与季调口径差异（写死规则）。
3. 复现 Born et al. 至 2018 的缺口路径（H1），偏差归因。
4. 延展至 2025，跑"剔除 COVID 观测"与"COVID 协变量约束"两口径，估计 H2。
5. 执行三件套推断（in-space、in-time、留一法）与三套预测变量集敏感性。
6. 做三阶段断点分解，尝试把缺口归因到公投预期/正式脱欧/TCA 三段（明确标注归因的假设性）。
7. 交叉校验：用增广合成控制或矩阵补全法（pysyncon 若支持,需核实;否则自写岭回归增广版）重估，比对主结论稳定性。

7 · 新颖性边界

- 本课题**不**声称发现"脱欧有成本"——该结论与方法设定归 Born et al. (2019);合成控制法归 Abadie 等;至 2022 的智库更新归 Springford/CER，2026 年 doppelganger 更新归高盛（非学术）。
- 本项目贡献（主结论）：2019 – 2025 政策全期的**学术级推断**延展估计，特别是 COVID 混杂下"可识别界"的正面刻画——这是智库图表不做、学术文献尚缺的部分。估出显著缺口与证明不可识别，两种结果都成立。
- 历届丘奖对照：该赛道有合成控制近亲方法获奖先例（2022 提名奖"沈阳赛艇赛经济效应:回归控制法"），但无国家级贸易政策对象;2020 提名奖 Brexit 论文做的是英国公司债定价，与本课题的宏观产出对象不重叠。

8 · 决策门槛（go / no-go）

- **第 4 周末**：加州控烟算例复现必须通过（纯软件问题，不通过只能修复）。

- **第 8 周末**：Born et al. 复现（H1）若偏差 > 0.5 个百分点，先核对 GDP 修订版本（用其发表时的 vintage 数据,OECD 实时数据库可得,需核实入口）;第 12 周仍不达标则**降级**：以"当前 vintage 数据下的重新估计"替代"精确复现"，明示数据修订为偏差来源，主结论框架不变。
- **第 20 周末**：若两种 COVID 口径的缺口估计符号相反或 CI 覆盖零，主结论切换为预注册的方法论结论（H2 的后半支），论文重心移到"合成控制在共同大冲击下的失效条件"——该降级路径在开题时即写入研究设计。
- 适配前提:适合对宏观与因果推断感兴趣、编程要求中等的学生;全程数据免费且干净，是 10 题中数据风险最低的一题，但要接受结论可能是"不可识别"的心理预期;英文写作素材（国际议题）对总决赛答辩有利。

E09 · 低风险异象在 A 股 2021–2026 的样本外独立检验：量化基金扩张后波动效应是否衰减

优先级:中低 · 族:因子异象独立检验 · 数据:akshare(免费) · 算力:低 · 技能:实证资产定价

1 · 研究问题

> 已被多篇文献确认的 A 股低波动异象（低波动组合月均超额约 0.9%、t≈3，样本至 2020 年前后）在 2021-01 – 2026-06 的纯样本外窗口中是否仍达 Harvey – Liu – Zhu 的 t > 3.0 门槛？其驱动源（波动率 vs 贝塔 vs 彩票偏好 MAX）的相对贡献在量化私募规模扩张后是否改变？

2 · 研究背景与空白

低风险异象指低波动/低贝塔股票的风险调整收益系统高于高风险股票，与 CAPM 预测相反。国际源头是 Frazzini & Pedersen (2014, JFE) 的 betting against beta（BAB，杠杆约束解释）与 Ang 等的特质波动率之谜。中国证据已相当充分：Blitz 等《The Volatility Effect in China》(Journal of Asset Management, 2021) 用 2000 – 2020 样本测得按三年波动率排序的低波组合月超额 0.90%（t=2.99），并指出驱动源是**波动率而非贝塔**;PBFJ 2021 系列论文进一步证明 A 股贝塔异象主要由彩票偏好（MAX 效应）与特质波动驱动，而杠杆约束类的 betting against correlation 因子 alpha 不显著。

2021 年以来中国量化私募管理规模从约数千亿扩张到万亿级（公开报道口径不一，正文须以中基协可核数据为准），低波类策略被大规模执行——按 McLean – Pontiff 逻辑，异象应随套利资本进入而衰减。

空白在时间覆盖：已发表样本基本止于 2020;2021 – 2026 恰是量化扩张、微盘股风格极端化（2023）与 2024-02 崩盘的窗口，低波异象在此窗口的表现是"文献结论 + 套利资本假说"的一次干净的样本外判决。衰减与不衰减都是有效结论。

3 · 可检验假设

- **H1**：2021-01 – 2026-06 窗口中，按 36 个月波动率排序的低-高波动多空组合（剔除市值最小 30%，价值加权）月均收益相对 Blitz 等的 0.90% 衰减 ⩾ 50%，且 Newey – West t < 3.0。
- **H2（机制）**：控制 MAX（月内最大日收益）后低波 alpha 的缩减比例在新旧样本间显著不同——若 MAX 的解释力下降而异象整体衰减，与"量化资本套利掉行为性错定价"一致。H1 不成立（异象完好）则支持"限卖空约束未变、异象难被套利"解释，同为有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用自建管线在 Blitz 等 (2021) 原样本区间（2000 – 2020）复现其核心结果：低波组合月超额收益 0.90%（偏差 ⩽ 0.15 个百分点）、t 值同向且偏差 ⩽ 25%（数据源差异归因：退市覆盖、复权、剔除规则）。**此步不过关，后续全部结论无效。**
2. 统计口径：月度组合收益检验一律 Newey – West（滞后 6，敏感性 3/12）;alpha 相对 CAPM 与 CH-3 两套基准都报（CH-3 因子取自 Stambaugh 主页或 E02 同款管线）;报点估计 + 95% CI + 年化夏普。
3. 多重检验：排序变量仅预注册 4 个（总波动、贝塔、特质波动、MAX），Bonferroni 校正;引用显著性一律用 t > 3.0 门槛。

4. 可执行性口径：另报一版含交易成本的净收益（单边 15 bp 佣金+印花税+冲击成本保守假设 20 bp,双套假设都报）与月均换手率;**策略净收益不显著本身是有效结论**，本课题不以"可赚钱"为目标。
5. 幸存者偏差：退市股缺失按年披露;主口径用"组合形成时点在市股票"，另用沪深 300+中证 500+中证 1000 成分并集做稳健性。
6. 前视偏差：波动率仅用截至组合形成月末的历史数据，管线单测覆盖。
7. 可复现性：脚本开源，CPU 全流程 ⩽ 1 小时。

5 · 数据与工具

用途	来源 / 工具
全市场日/月频行情、市值（2000—2026）	akshare，免费;与 E02/E05 共用管线,退市覆盖问题同样处理
CH-3 因子（基准调整,**仅对比,不计入贡献**）	Stambaugh 主页公开序列(更新时点需核实)或 E02 管线自建
量化私募规模背景数据（仅描述性背景）	中国证券投资基金业协会公开披露,免费;口径限制须注明
对照基准（**仅校验对比,不计入贡献**）	Blitz 等 (2021) 原文表;Frazzini—Pedersen (2014) 方法定义
分析环境	Python: pandas、statsmodels;纯 CPU

6 · 方法路径

1. 复用/搭建全市场月频特征管线（与 E02 共享代码基），写前视偏差单测。
2. 在 2000 – 2020 复现 Blitz 等核心表，完成硬门槛校验与偏差归因。
3. 在 2021 – 2026 样本外窗口重估四个排序变量的多空组合收益与 alpha（H1）。
4. 做 MAX 控制的双重排序与 Fama – MacBeth 回归（Newey – West 修正），估计机制贡献变化（H2）。
5. 计算换手率与两套成本假设下的净收益。
6. 子样本稳健性：剔除 2024-02 崩盘月、按市值分层、等权 vs 价值加权。
7. 交叉校验：用独立的成分股并集样本重跑主检验，比对结论。

7 · 新颖性边界

- 本课题**不**声称发现低风险异象——国际归 Frazzini – Pedersen (2014)/Ang 等，中国证据归 Blitz 等 (2021)、PBFJ (2021) 等，"驱动源是波动率与 MAX 而非贝塔"的结论也已发表，不得据为己有。
- 本项目贡献（主结论）：2021 – 2026 纯样本外窗口的独立判决 + 机制构成变化的再估计。该窗口在任何已发表论文样本之外，且承载了明确的经济学预测（套利资本进入→衰减），是"新样本独立检验"差异化轴的标准执行。
- 历届丘奖对照：2025 总决赛入围"Momentum Investing: ML vs Markowitz"（墨尔本中学）做动量+ML 组合优化，无严格样本外推断与多重检验校正;本课题以推断严格度与"不追求赚钱"的定位与其区分。注意:A 股动量本身缺失（文献共识），故本题选低波而非动量作为检验对象。

8 · 决策门槛（go / no-go）

- **第 8 周末**：硬门槛复现须达标。若偏差 > 0.15 个百分点且归因指向退市股缺失，**降级**：复现窗口改为 2010 – 2020（退市影响小），2000 – 2009 段放弃，主结论（2021 – 2026 样本外）不受影响。

- **第 14 周末**：若 2021 – 2026 窗口所有排序变量的 CI 宽到无法分辨 50% 衰减（约需 60+ 个月观测，本窗口 66 个月，功效临界），预注册的补救：把窗口前推至 2018-01（与旧样本部分重叠段单独标注），并报重叠/非重叠分段结果。
- **第 20 周末**：H2 机制检验若因双重排序后组合内股票数 < 30 而不稳定，降级为 Fama – MacBeth 回归单一口径，明示局限。
- 适配前提:与 E02 高度共享数据管线,适合作为 E02 团队的姊妹课题或二选一;单独做时数据工程量与 E02 相当。**注意与 E02 不可同时作为同一团队的两篇提交**（同视角、同管线,评审会视为一题拆分）。

E10 · 机器学习横截面收益预测在 A 股的降配复现与 2021–2026 样本外再检验（Leippold–Wang–Zhou 框架）

`优先级:低(高风险高回报)` · `族:因子模型×机器学习` · `数据:akshare+GKX公开数据集` · `算力:中(CPU 数小时–数天)` · `技能:机器学习+数据工程(强)`

1 · 研究问题

> 在特征集从 1000+ 压缩到 30 个、模型压缩到 5 类的 CPU 可行配置下，Leippold – Wang – Zhou (2022) 报告的 A 股机器学习收益可预测性（含"流动性类特征最重要、大盘与国企更可预测"两条结构性结论）在其样本截止（2020）后的 2021-01 – 2026-06 窗口中还剩多少？月度样本外 R^2 与多空组合收益衰减各是多少?

2 · 研究背景与空白

Gu, Kelly & Xiu (2020, RFS) 确立了用机器学习做横截面收益预测的评估框架：以月度样本外 R^2（R^2_oos）与预测排序组合收益为准绳，比较 OLS、LASSO、树模型、神经网络。Leippold, Wang & Zhou (2022, JFE)《Machine learning in the Chinese stock market》把该框架搬到 A 股（约 2000 – 2020），核心发现：流动性类特征重要性居首（与美股相反）、散户主导使短期可预测性更强、大盘股与国企在长期限上高度可预测、树模型与神经网络优于线性基准。GKX 的美国特征数据集与代码公开（Dacheng Xiu 主页），LWZ 的中国数据依赖 CSMAR/Wind，不可免费获得——这决定本课题必须自建降配特征集。

空白在时间覆盖：2021 – 2026 覆盖量化大发展、2023 微盘风格极端化与 2024-02 崩盘、2024-09 政策脉冲——ML 可预测性是否已被套利资本吞噬是开放且结论正负都成立的问题。检索未见以 2021 后 A 股为样本外窗口、完整执行 GKX 口径推断的公开英文论文（LightGBM 应用类 preprint 存在但口径不严;第 8 周须再核）。

3 · 可检验假设

- **H1**：降配管线在 2021 – 2026 样本外窗口的最优模型月度 R^2_oos > 0 但相对 2010 – 2020 验证期衰减 $\geqslant$ 50%;预测十分位多空组合（剔除最小 30% 市值,价值加权,月频调仓）毛收益 $t < 3.0$。
- **H2（结构性结论复检）**：特征重要性排序中流动性类仍居前三——若被反转（如基本面类上升），说明散户主导结构在量化时代已变化。H1 的反向结果（可预测性完好）同样有效，指向 A 股套利容量约束。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：先用 GKX 公开的美国特征数据集与公开代码口径，复现其 OLS-3 与 LASSO 的月度 R^2_oos（原文量级约 0.1 – 0.4%）：符号一致、与原文表偏差 $\leqslant$ 0.2 个百分点、模型排序一致。**此步不过关，A 股部分全部结论无效。**（GKX 数据约 2 – 4 GB,CPU 上跑线性与树模型可行;神经网络降为 1 – 2 层小网络并明示与原文差异。）
2. 时间划分写死：训练 2005 – 2015、验证 2016 – 2020、测试 2021 – 2026,滚动扩窗年度重拟合;**绝不随机划分**,并附一次随机划分对照量化泄漏引起的虚高。
3. 统计口径：R^2_oos 用 GKX 定义（分母为零基准）;多空组合收益 Newey – West t;模型间比较用 DM 检验;全部组合收益报净成本版（口径同 E09 第 4.4 条）;多重检验：模型 × 期限的检验族做 FDR q=0.10。
4. 特征集预注册：30 个特征清单（估值、动量/反转、流动性、波动、基本面五类各约 6 个）在跑任何测试集之前冻结并存档，防止特征窥探。

5. 幸存者与前视偏差：同 E02/E09 口径:财务特征滞后至公告日。
6. 可复现性：GKX 复现脚本 + A 股管线开源:A 股全流程 CPU ⩽ 24 小时（约 4000 股 × 200 月 × 30 特征 ≈ 2×10^7 行,LASSO/LightGBM 分钟 – 小时级,小网络数小时）。

5 · 数据与工具

用途	来源 / 工具
GKX 美国特征数据集与代码（**仅用于管线校验,不计入本项目数据贡献**）	Dacheng Xiu 主页(dachxiu.chicagobooth.edu)公开下载;约 2–4 GB,免费(可得性需在第 2 周核实)
A 股行情、市值、换手、财务（2005–2026）	akshare 为主、tushare 免费额度交叉核对;退市覆盖问题按年披露
CH-3 因子基准调整（**仅对比**）	Stambaugh 主页或 E02 管线
模型库	scikit-learn(OLS/LASSO/弹性网)、LightGBM(树)、PyTorch CPU(小网络);全部开源免费
对照基准（**仅对比,不计入贡献**）	GKX (2020) 表 1;LWZ (2022) 表(原文量级,如月度 R^2_oos 与特征重要性排序)

6 · 方法路径

1. 下载 GKX 数据集，完成第 4.1 条硬门槛复现（这一步同时是全套评估代码的联调）。
2. 冻结 30 特征清单并存档（预注册文件带哈希与日期）。
3. 用 akshare 构建 A 股月频特征面板，做缺失/极值处理（规则写死：横截面分位缩尾、缺失中位数填补并加缺失指示）。
4. 训练 5 类模型，在验证期调参（超参网格预先写死），测试期一次性评估。
5. 计算 R^2_oos、多空组合毛/净收益、DM 检验与 FDR 校正（H1）。
6. 用置换重要性与 SHAP（树模型）估计特征类重要性排序（H2），与 LWZ 原文排序对比。
7. 交叉校验：以"仅线性模型 + 仅前 800 大市值股"的极简配置重跑，验证主结论不依赖复杂模型或小盘股数据质量。

7 · 新颖性边界

- 本课题**不**声称提出新预测方法——评估框架归 Gu – Kelly – Xiu (2020)，A 股结论归 Leippold – Wang – Zhou (2022);"流动性最重要"等结构性发现是 LWZ 的，不得据为己有。降配（30 特征、5 模型）与原文（千级特征、11 模型）的差距必须在论文中定量声明，本课题检验的是"降配版可预测性"而非原文全量结论。
- 本项目贡献（主结论）：2021 – 2026 纯样本外窗口的可预测性衰减估计 + 结构性结论（特征重要性、大盘/国企可预测性）的独立复检。衰减与否两个方向都回答"中国 ML 定价文献的结论是否已被市场自身消化"。
- 历届丘奖对照：2025 入围"Momentum ML vs Markowitz"与 2022 铜奖 LASSO-BiLSTM 汇率预测均属 ML 预测类但无预注册特征集、无 R^2_oos 口径、无多重检验校正;本课题以 GKX 级评估纪律为差异化轴。

8 · 决策门槛（go / no-go）

- **第 2 周末**：核实 GKX 数据集当前可得性。若下载入口失效，**降级**：硬门槛改为用 Ken French 25 组合复现 GKX 附录的组合级预测结果，或以 OpenAP（Chen – Zimmermann 开放资产定价数据,免费）替代做管线校

验，主框架不变。

- **第 10 周末**：GKX 复现硬门槛须达标;不达标先查收益对齐与缺失处理，第 14 周仍失败则整题**终止并转向 E09**（共享大部分 A 股管线,沉没成本最小）——这是 10 题中唯一设"终止转向"而非"降级"的门槛，选题时须知情。
- **第 20 周末**：A 股特征面板若财务特征可用率 < 70%，降级为"价格类特征(动量/反转/波动/流动性)single 集"（约 18 个特征），LWZ 对比范围相应缩窄并明示。
- **第 32 周末**：若测试期所有模型 R^2_oos 的 CI 覆盖零且组合收益不显著，这就是 H1 的"完全衰减"结论——按预注册写成主结果，不得回头改特征集或超参（管线哈希存档即为此设）。
- 适配前提:仅适合编程强、能承受"复现失败即转向"风险、且团队 ⩾ 2 人的学生;工时预算是 10 题之最（数据工程约占 50%）。放在 E10 是因为数据工程失败概率与工作量都最高，而非科学价值最低。

E11 · 高中排课问题的混合整数规划求解：以 XHSTT 国际基准最优性差距与本校真实约束为双重判据

`优先级：最高` · `运筹优化族` · `纯计算/零实验` · `CPU 轻量` · `编程+离散建模取向`

1 · 研究问题

用免费开源求解器构建的混合整数规划（Mixed-Integer Programming, MIP）排课模型，在 XHSTT 国际标准测试集上能把最优性差距（optimality gap）压到多小？把同一模型迁移到本校真实排课约束后，相对现行人工课表能在硬约束零违反的前提下把软约束罚分降低多少？

2 · 研究背景与空白

排课（timetabling）是运筹学的经典 NP-难问题。XHSTT 是国际排课竞赛（ITC-2011）确立的统一 XML 格式，收录 11 国 38 个真实高中算例，每个算例带公开的已知最优解或最好解与下界，是极少数"中学生就能拿到、且答案可核对"的组合优化基准。

已有工作：Kristiansen、Sørensen 与 Stidsen（Journal of Scheduling, 2015）给出首个覆盖任意 XHSTT 算例的两阶段 MIP 精确方法，用商业求解器证明了 4 个已知解的最优性并刷新 9 个最好解；ITC-2011 优胜方法以元启发式为主，此后 MaxSAT 与约束规划也被证明有效。这些工作都用商业求解器或专用代码。

空白在时间与工具维度：开源求解器（HiGHS、CBC、OR-Tools CP-SAT）近三年性能大幅提升，但"纯开源工具链 + 个人笔记本在 XHSTT 上能达到什么 gap"没有系统的公开报告；且把基准模型落到一所中国高中的真实约束（走班、合班、教师跨年级）并与人工课表定量对照的公开案例几乎没有。技术门槛清晰、答案可核对、结论正负都成立，适合一年期课题。

3 · 可检验假设

- H1：在 ⩾10 个 XHSTT 中小规模算例上，笔记本 + 开源求解器在单例 ⩽2 小时限时内可将平均最优性差距压到 ⩽15%，其中至少 3 例达到已知最优。
- H2：迁移到本校约束后，模型解在硬约束零违反前提下，软约束罚分比现行人工课表低 ⩾20%；若人工课表已近最优（差距 <5%），该"人工已近优"结论同样成立并可量化。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用自建管线重跑 XHSTT 至少 3 个带已证最优解的小算例（如 Kristiansen et al. 2015 表中已证最优的实例），复现其最优目标值，偏差 = 0（整数目标值须完全一致）；另在一个可手工枚举的 4 班×5 课玩具算例上与穷举解对照。此步不过关，后续全部结论无效。
2. 报告口径预先写死：每个算例报 `（可行解罚分，已知下界，gap%）` 三元组；gap 相对官方公布下界计算，不得只报罚分。
3. 求解器对照用双成本口径：等墙钟时间（2 小时）与等迭代预算两种都报；至少对照 HiGHS 与 CP-SAT 两个求解器。
4. 本校算例须给出完整约束清单（硬/软分列、权重表预先写死并说明依据），人工课表罚分用同一评分脚本计算。
5. 随机重启类算法报 ⩾10 种子的均值 ± 标准差与 95% 置信区间。

6. 全部模型代码、XHSTT 解析器与评分脚本开源，第三方可一键重跑。

5 · 数据与工具

用途	来源 / 工具
国际基准算例	XHSTT archive（GitHub/官方站点公开 XML，38 个真实算例，含最好解与下界；仅用于校验与对比，不计入本项目数据贡献）
官方评分核对	HSEval 在线评估器（对 XHSTT 解评分；可用性需核实，不可用则按官方规范自写评分器并以公开解校验）
求解器	HiGHS（MIT 许可）、Google OR-Tools CP-SAT、CBC；全部纯 CPU、pip 安装
建模	Python + PuLP / OR-Tools 原生 API
本校数据	教务处现行课表与约束（班级、教师、场地；须匿名化教师姓名，经学校同意）

规模上限：中等 XHSTT 算例约 10^5 二元变量，笔记本 16GB 内存内单次求解 ≤2 小时；超大算例（如巴西全国实例）标注**超出范围**。降规模版本：只做单年级（≈10^4 变量，分钟级）。

6 · 方法路径

1. 装环境，写 XHSTT XML 解析与评分脚本，用官方公开解通过校验（第 1 条验收）。
2. 实现基础 MIP 模型（分配变量 + 冲突约束 + 软约束罚分线性化），在玩具算例上与穷举对照。
3. 在 ⩾10 个算例上跑双求解器、双成本口径对照，记录 gap 曲线。
4. 核实 HSEval 可用性与本校约束的可形式化边界（哪些"潜规则"无法写成约束须明示舍弃）。
5. 采集本校约束建模求解，与人工课表同口径评分对照。
6. 做求解时间-规模敏感性扫描（算例规模按变量数分档），给出"笔记本可解规模"经验边界。
7. 独立交叉校验：对最好的 2 个解用另一求解器验证可行性与目标值。

7 · 新颖性边界

> 本课题不声称提出新算法，不声称刷新 XHSTT 任何最好解（若碰巧刷新是意外收获须另行核实）。已有工作：Kristiansen et al.（2015）已给出通用 MIP 精确方法与商业求解器结果；ITC-2011 各决赛方法已发表。本项目贡献是（i）纯开源工具链 + 消费级 CPU 在 XHSTT 上的系统 gap 报告——这是可复现性维度的贡献；（ii）一个中国高中真实算例的完整形式化与定量对照——样本维度的独立检验。丘奖历届获奖中运筹类有 2020 优胜"武汉疫情资源最优配置的非线性规划"与 2023 优胜"O2O 零售宅配成本最小化"，均为物流/应急场景，无排课与基准对照工作，差异清晰。"人工课表已近最优"与"可显著改进"两种结果都是有效结论，但须给出 gap 误差棒证明有能力分辨。

8 · 决策门槛（go / no-go）

- **第 3 周末**：XHSTT 解析器 + 评分脚本对公开解校验通过。未通过→继续调试至第 5 周，仍未通过则**降级 A**：放弃 XHSTT 通用格式，改用文献中定义清晰的单一算例格式（如 ITC-2007 大学排课三赛道之一，格式简单一个数量级），保留"开源求解器 gap 报告 + 本校算例"主结论框架。
- **第 10 周末**：至少 5 个算例拿到 gap ⩽30% 的可行解。达不到→**降级 B**：缩小到小规模算例子集 + 单年级本校算例，主结论从"gap 报告"收窄为"笔记本可解规模边界研究"，框架不变。

- 第 20 周末：本校约束数据到手（须第 4 周先向教务处口头确认可行性，不要边做边发现）。拿不到→用 XHSTT 中最接近中国走班制的算例替代，本校部分降为附录讨论。
- 已知困难：软约束权重的设定有主观性——把"权重敏感性扫描（拉丁超立方 $\geqslant 50$ 组权重）下结论是否翻转"写成研究内容而非障碍。
- 预算裁剪顺序：先砍求解器对照（保 HiGHS 单求解器），再砍算例数量（保 5 个），最后砍本校算例；主结论（开源工具链 gap 定量报告）在全部裁剪下仍成立。

E12 · 城市公共急救设备（AED）选址的最大覆盖模型：以现有布点覆盖率差距与 LP 下界为判据

优先级：高 · 运筹优化族 · 公开地理数据 · CPU 轻量 · 编程+地理信息取向

1 · 研究问题

用最大覆盖选址模型（Maximal Covering Location Problem, MCLP）对一座中国城市（或城区）的自动体外除颤器（AED）布点做优化：在与现状相同的设备数量下，最优布局能把"N 分钟步行可达人口覆盖率"比现状提高多少个百分点？覆盖率提升对需求分布假设（人口密度 vs 人流强度代理）的敏感性有多大？

2 · 研究背景与空白

院外心脏骤停（OHCA）抢救的黄金时间约 4 分钟，AED 选址是设施选址（facility location）理论的标准应用。MCLP 由 Church & ReVelle（1974）提出，是整数规划教科书模型，中小规模可精确求解。

已有工作：Chen 等（ISPRS IJGI, 2023）用手机信令数据 + MCLP 对深圳 AED 做时空优化；Springer HCMS（2024）用 MCLP 变体优化志愿者响应系统的 AED 布点，报告重新布局可把加权覆盖率从 36% 提高到 49%。方法完全公开，属成熟框架。

空白在样本与评价维度：已发表工作集中于少数城市且多依赖不公开的信令/病例数据；用**纯公开数据**（OpenStreetMap + WorldPop 人口栅格 + 政府开放数据）对一座尚无此类分析的城市做"现状 vs 最优"的定量差距核算，并系统报告需求代理选择带来的结论敏感性，是学生可完成的独立检验。结论正负都成立：若现状已接近最优（差距 <5 个百分点），同样是有价值的审计结论。

3 · 可检验假设

- H1：在设备数与现状相同的约束下，MCLP 最优布局的 4 分钟步行覆盖人口比现状布局高 ⩾10 个百分点。
- H2：用夜间人口（居住）与日间人流代理（POI 密度加权）两种需求口径，最优选址集合的重合度（Jaccard 指数）⩾0.5；若 <0.5，说明结论强依赖需求假设，此依赖性本身作为主结论报告。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：在一个 5 需求点 × 4 候选点的手工可枚举算例上，自建 MCLP 代码与穷举解完全一致；再复现 Church & ReVelle（1974）原文 55 节点标准算例（Swain 数据集，文献可查）的最优覆盖值，偏差 = 0。此步不过关，后续全部结论无效。
2. 每个规模的解都报与 LP 松弛下界的差距；精确解报"已证最优"，启发式解报 optimality gap ⩽ 具体数值。
3. 覆盖判定用路网步行距离（OSM 路网 + 4.8 km/h 步速），不得用直线距离；须做一次"直线 vs 路网"对照量化直线距离高估覆盖率的幅度。
4. 敏感性口径预先写死：覆盖半径（3/4/5 分钟）、需求代理（两种）、候选点集合（两种密度）做全因子扫描 ⩾12 组，报告主结论翻转与否。
5. 现状 AED 位置清单须注明来源与抓取日期；缺失率估计写入局限性。
6. 全部代码 + 数据处理脚本开源，附一键重跑说明。

5 · 数据与工具

用途	来源 / 工具
路网与候选点	OpenStreetMap（Geofabrik 免费下载中国分区 pbf；OSMnx 提取步行路网）
人口需求栅格	WorldPop（100m 分辨率免费栅格）；中国第七次人口普查街道级数据（国家统计局公开）
现状 AED 位置	城市 AED 电子地图/政府开放数据平台（深圳、上海、杭州均有公开 AED 地图，具体接口**需核实**；不可得则换城市或用"120 急救站"替代对象）
日间人流代理	OSM POI 密度加权（免费）；百度/高德热力数据不采用（不公开）
求解	Python + PuLP/OR-Tools + HiGHS，纯 CPU
方法对照基准	Chen et al. 2023（深圳）结果——仅用于校验与对比，不计入本项目数据贡献

规模上限：需求点 ≤5,000（栅格聚合后）、候选点 ≤800 时 MIP 变量约 4×10^6 稀疏系数，笔记本单次求解分钟级；全市 10 万级需求点**超出范围**，降规模用 500m 栅格聚合（≈2,000 点，秒级）。

6 · 方法路径

1. 装环境（OSMnx + HiGHS），跑通 Swain 55 节点算例完成校验。
2. 选定城市，抓取 OSM 路网/POI 与 WorldPop 栅格，构建需求点-候选点步行距离矩阵。
3. 核实现状 AED 清单可得性与完备性（第 4 周门槛，见块 8）。
4. 求解等设备数 MCLP 与递增设备数的覆盖-数量边际曲线（每加 10 台的边际覆盖增益）。
5. 做全因子敏感性扫描与两种需求口径对照，计算选址集合 Jaccard 指数。
6. 独立交叉校验：用贪心启发式（提交近似比 1−1/e 保证）复算，与精确解差距 ≤ 理论界。
7. 汇总"现状-最优差距 + 敏感性"双结论，绘制布点建议图。

7 · 新颖性边界

> 本课题不声称提出新选址模型（MCLP 1974 年已发表），不声称掌握真实 OHCA 病例分布（用人口/人流代理并明示局限）。已有工作：深圳信令数据时空优化（2023）、欧洲志愿者响应系统布点（2024）均已发表且用了不公开数据。本项目贡献是主结论层面的两件事：（i）对一座未被分析过的城市用纯公开数据完成"现状 vs 最优"覆盖差距审计——换样本的独立检验；（ii）需求代理选择对选址结论的敏感性量化——前人评估覆盖率，本项目评估数据假设对结论的偏置，属评价维度转换。丘奖历届无 AED/设施选址获奖（2020 优胜为疫情资源配置的非线性规划、2024 优胜为应急救援博弈模型，对象与方法均不同）。"现状已近优"与"可大幅改进"两种结果都构成有效结论。

8 · 决策门槛（go / no-go）

- **第 4 周末**：确认目标城市现状 AED 位置清单可得（公开地图可导出 ≥200 台且带坐标）。不可得→**降级 A**：换有公开清单的城市（深圳 AED 地图公开性最强，需核实导出方式）；仍不可得→**降级 B**：把研究对

象换成"社区养老服务设施/急救站"选址（政府开放数据平台普遍有清单），MCLP 框架与双结论结构完全保留。

- **第 8 周末**：距离矩阵构建完成且 500m 栅格全流程 ≤30 分钟可重跑。超时→固定用 500m 栅格并把 100m 栅格标注为超出范围。
- **第 16 周末**：主优化 + 12 组敏感性完成。若敏感性显示主结论翻转→翻转本身升格为主结论（"AED 选址结论强依赖需求假设"），框架不变。
- 已知困难：OSM 中国区 POI 完备性参差——把"OSM 完备性对结论的影响"写成稳健性小节（抽样 50 个 POI 与地图实况核对，报告缺失率）。
- 预算裁剪顺序：先砍递增设备数曲线，再砍第二城市对照（若有），保"等设备数差距 + 敏感性"两条主结论。

E13 · PISA 2022 城乡学业差距的跨国 Oaxaca–Blinder 分解：以可解释份额的国别异质性为判据

优先级：高 · 教育经济学实证族 · 公开微观数据 · CPU 轻量 · 统计+计量取向

1 · 研究问题

在 PISA 2022 数据中，城乡学生数学成绩差距有多大比例可由家庭社会经济地位（ESCS）与学校资源等可观测因素解释？这一"可解释份额"在 20 个以上参与经济体之间如何变化，且能否被国家层面的城镇化率或教育支出结构预测？

2 · 研究背景与空白

PISA 是 OECD 三年一轮的 15 岁学生测评，微观数据完全公开，含学生家庭背景、学校位置（社区规模分类）与资源问卷。Oaxaca – Blinder 分解（Oaxaca 1973; Blinder 1973）把组间均值差拆成"禀赋差异（可解释）"与"系数差异（不可解释）"，是劳动与教育经济学的标准工具。

已有工作：Ramos 等（IZA DP 6515）用 PISA 微观数据分解哥伦比亚城乡差距；泰国（PISA 2009）、秘鲁（2018，禀赋解释 82.6%）均有单国研究；2025 年已有用 PISA 2022 做柬埔寨城乡差距的 RIF-分解论文（International Journal of Educational Development）。

空白在样本与时间覆盖：单国研究多、**跨国系统比较少**，且 PISA 2022 是疫情后首轮、城乡差距可能被疫情不对称冲击放大——用同一方法学横跨 20+ 经济体做 2018 vs 2022 两轮对照的公开研究未见（检索"PISA 2022 urban rural Oaxaca-Blinder cross-country"仅命中单国研究，未找到不等于不存在，写作时须再核）。方法公开、数据公开、结论正负都成立，是典型的"换样本独立检验"型课题。

3 · 可检验假设

- H1：多数经济体（≥60%）中，ESCS 与学校资源等禀赋差异可解释城乡数学差距的 50% 以上。
- H2：可解释份额与国家城镇化率负相关（城镇化率越高、不可解释份额越大），Spearman 相关在 20+ 经济体样本上 $|\rho| \geq 0.4$；若不显著，"城乡差距结构无跨国规律"同样是有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用自建管线（正确处理 10 个 plausible values 与 80 个 BRR 重复权重）复现 OECD 官方报告中 ≥5 个经济体的数学平均分与标准误，均值偏差 ≤0.5 分、标准误偏差 ≤5%；再复现 IZA DP 6515 哥伦比亚分解的方向性结论（禀赋份额同号且相对偏差 ≤10%，若该文用旧轮数据则在同轮数据上复现）。此步不过关，后续全部结论无效。
2. 统计口径预先写死：所有点估计带 BRR 标准误；分解结果给自助法（bootstrap ≥500 次，与 BRR 联合）95% 置信区间；显著性检验报效应量（差距以 PISA 分与标准差单位双报），不以 p 值单独下结论。
3. 城乡定义预先写死并做敏感性：主口径 SC001（<3,000 人 vs >100,000 人），次口径（<15,000 vs >15,000）对照，报告结论是否翻转。
4. 入选经济体标准预先写死：城乡两组各 ≥30 所学校，达标者全部纳入，不得事后挑选。
5. 缺失值处理方案预先写死（问卷缺失用 OECD 提供的处理惯例，删除率 >20% 的经济体剔除并报告）。
6. 全部清洗与分析脚本开源，从原始 SPSS/SAS 文件到图表一键重跑。

5 · 数据与工具

用途	来源 / 工具
学生/学校微观数据	OECD PISA 2022 与 2018 公开数据库（oecd.org 免费下载 SPSS/SAS 格式，2022 学生文件约 1–2 GB）
官方均值核对	PISA 2022 Results Vol. I 附表（免费 PDF/Excel；仅用于校验，不计入数据贡献）
国家层面变量	World Bank WDI（城镇化率、生均教育支出，免费 API/CSV）
分析	R（`intsvy` 或 `Rrepest` 包处理 PV 与 BRR 权重，能力边界**需核实**：若不支持分解则手写循环）+ `oaxaca` 包；或 Python + statsmodels
方法对照	IZA DP 6515、柬埔寨 2025 文——仅用于校验与对比，不计入本项目贡献

计算量：单经济体回归 + 500 次 bootstrap 在笔记本分钟级；30 经济体全流程 <3 小时，纯 CPU 无压力。

6 · 方法路径

1. 下载 PISA 2022 数据，装 R 环境，复现 5 个经济体官方均值与标准误（第 1 条验收）。
2. 按预写标准筛选经济体，构建城乡二分变量与协变量集（ESCS、学校师生比、短缺指数等，清单预注册在 OSF 或 GitHub 时间戳）。
3. 核实 `intsvy` / `Rrepest` 对分解 + 复杂权重的支持边界；不支持则手写 PV×BRR 双循环。
4. 逐经济体做 Oaxaca – Blinder 分解（含 RIF 分位数版本作为延伸），提取可解释份额及置信区间。
5. 做城乡定义与协变量集的敏感性对照。
6. 第二阶段：可解释份额对国家层面变量的秩相关与散点分析（明示仅为相关，不做因果声称）。
7. 独立交叉校验：抽 3 个经济体用 Python 独立实现复算，份额偏差 ≤2 个百分点。

7 · 新颖性边界

> 本课题不声称提出新分解方法，不做任何因果推断声称（分解是会计恒等式，系数差异不等于歧视/低效）。已有工作：哥伦比亚、泰国、秘鲁、柬埔寨等单国分解已发表（后者已用 PISA 2022 + RIF 方法）。本项目贡献是主结论层面的（i）同一方法学横跨 20+ 经济体的系统比较与 2018→2022 两轮变化——时间与样本覆盖的独立检验；（ii）可解释份额的跨国结构分析。丘奖历届相邻工作：2023 铜奖"隔代抚养与儿童学业（CFPS）"、2021 优胜"教育公平与生育选择"均为中国单一数据源的因果/机制研究，无跨国分解工作，差异清晰。若 H2 不显著，"无跨国规律"结论有效，但须给出置信区间证明检验力足够（20+ 样本可检出 $|\rho| \geqslant 0.6$，预先声明该边界）。

8 · 决策门槛（go / no-go）

- **第 3 周末**：官方均值复现通过。未通过多为权重处理错误→限期 2 周修复，第 5 周仍未通过则**降级 A**：改用 OECD 官方 `Rrepest` 全托管流程重做（牺牲部分自主性但保留全部主结论框架）。
- **第 8 周末**：确认达标经济体 ⩾20 个。若 <20（城乡样本量普遍不足）→**降级 B**：改为"2018 vs 2022 两轮 × 10 个经济体"的时间对照设计，主结论从横截面结构转为疫情前后变化，分解框架不变。
- **第 20 周末**：主分解 + 敏感性完成；RIF 分位数延伸视进度取舍（首个可砍项）。
- 已知困难：SC001 学校位置为校长自报的社区规模，非严格城乡行政口径——把口径敏感性写成研究内容；中国大陆未参加 PISA 2022，本研究不含中国，须在引言明示并讨论对政策外推的限制。

- 预算裁剪顺序：先砍 RIF 分位数分解，再砍 2018 轮对照，最后砍第二阶段跨国相关；核心主结论（多经济体可解释份额及其区间）在全部裁剪下成立。

E14 · 动力学财富交换模型中税收再分配规则的比较：以对 WID 实测基尼系数与顶端份额的联合拟合为判据

优先级：高 · 基于主体建模族 · 模拟+公开数据校准 · CPU 轻量 · 编程+统计物理取向

1 · 研究问题

在动力学财富交换模型（kinetic wealth exchange model）中加入不同的税收再分配规则（比例税 vs 按实际税制标定的累进税），哪一种能在同一组参数下**同时**拟合世界不平等数据库（WID）中 2 – 3 个国家的基尼系数与前 10%/前 1% 财富份额？拟合所需的再分配强度参数与该国实测宏观税负的差距有多大？

2 · 研究背景与空白

动力学交换模型把经济体视为大量随机配对交换财富的主体，是统计物理与经济学交叉的经典 ABM：纯随机交换收敛到指数分布（Boltzmann – Gibbs，Drăgulescu & Yakovenko 2000，**有解析解**）；引入储蓄倾向得到类 Gamma 分布（Chakraborti & Chakrabarti 2000）；Bisi、Toscani 等证明了含再分配算子的模型在连续交易极限下收敛到 Gamma 密度且形状因子由再分配权重决定（**又一个解析锚点**）。

已有工作：税收与再分配对 Pareto 指数的影响已被多篇论文研究（如 Entropy 2023 关于极端不平等模型稳定性的数值研究；Toscani 等的再分配理论）。但这些工作绝大多数停留在"改规则→看分布形状变化"的定性层面。

空白在评价维度：**很少有工作把模型输出与具体国家的实测不平等指标做联合定量校准**，并检验"拟合出的再分配参数是否与该国真实税负同量级"——这是模型与现实的双向对表。解析极限完备、数据免费、算力极小，且"模型无法联合拟合"同样是有效结论（说明该类模型缺失关键机制），适合一年期课题。

3 · 可检验假设

- H1：含累进再分配的模型可在单一参数组下同时把基尼系数拟合到实测值 ±0.03、前 10% 份额 ±5 个百分点，而比例税模型无法同时满足两者（至少对高不平等国家如美国失败）。
- H2：拟合出的有效再分配率与该国实测"税收+转移支付占 GDP 比重"（OECD/WID 口径）符号一致且同一数量级（比值在 0.3 – 3 之间）；若数量级不符，该失配量化本身是主结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：（i）零储蓄、无税模型的稳态分布与解析指数分布做 KS 检验不拒绝（N=10^4 主体、10^4 次/主体交换后，p>0.1，多种子稳定）；（ii）复现 Drăgulescu – Yakovenko 2000 图 1 的指数温标 T=⟨m⟩，偏差 ⩽2%；（iii）复现含再分配模型的解析 Gamma 形状因子，偏差 ⩽5%。任一不过关，后续全部结论无效。
2. 模拟统计口径预先写死：每个参数点 ⩾20 个随机种子，报基尼与份额的均值 ± 标准差与 95% 置信区间；稳态判定用滑动窗口基尼漂移 <0.001/10^3 步的显式收敛判据。
3. 参数敏感性用全局方法：对（储蓄倾向、税率、再分配靶向参数）做 Sobol 一阶与总指数（Saltelli 采样 ⩾1,024 组）；不得只做 OAT。
4. 校准目标预先写死：以 WID 2022 年（或最新可得年）美国、法国、（中国，可得性需核实）三国的基尼 + 前 10% + 前 1% 份额为目标向量，拟合用同一损失函数（标准化 L2），禁止逐国换损失函数。
5. 有限规模效应检查：N=10^3/10^4/10^5 三档规模扫描，主指标随 N 的漂移 <1% 方可下结论。

6. 全部代码开源（NumPy/Numba），单参数点运行时长与总机时写入附录。

5 · 数据与工具

用途	来源 / 工具
实测不平等指标	World Inequality Database（wid.world，免费 CSV/API：财富与收入基尼、顶端份额，分国分年）
宏观税负对照	OECD Revenue Statistics / World Bank WDI（免费；中国税收转移口径可比性**需核实**，不可比则换英国/德国）
累进税制标定	各国法定税率表（OECD Taxing Wages 免费出版物）
模拟框架	Python + NumPy + Numba（自写核心循环，约 200 行；不依赖重型 ABM 框架）
解析对照	Drăgulescu & Yakovenko (2000)、Toscani 等再分配论文——仅用于校验与对比，不计入本项目贡献

算力：N=10^4 主体 × 10^8 次交换的单次运行在 Numba 下约 1 – 3 分钟（笔记本实测为准）；Sobol 1,024 组 × 20 种子 ≈ 2×10^4 次运行，总机时约 3 – 7 天可分批跑，纯 CPU 可行。N=10^6 全国尺度**超出范围**。降规模版本：Sobol 降为 256 组 + 10 种子（<1 天）。

6 · 方法路径

1. 实现基础交换模型，通过三项解析校验（第 1 条验收）。
2. 加入储蓄倾向与两种税收再分配算子（比例 / 按税率表分段累进），单元测试财富守恒（每步总量漂移 <10^-10）。
3. 下载 WID 目标数据，写统一损失函数与网格 + 局部细化的拟合流程。
4. 核实中国口径可比性与 WID 财富（非收入）序列覆盖（第 6 周核实，不边做边发现）。
5. 对三国分别拟合两种规则，输出可行参数域与最优拟合及置信区间。
6. 做 Sobol 全局敏感性与有限规模扫描。
7. 独立交叉校验：对最优参数点用纯 Python（非 Numba）慢速实现复算基尼，偏差 <0.5%。

7 · 新颖性边界

> 本课题不声称提出新模型（交换核与再分配算子均已发表），不声称模型机制是现实财富分布的因果解释（它是最简代理模型，须在结论中明示）。已有工作：指数/Gamma 稳态、税收改变 Pareto 指数、高税率下双峰分布等均已发表。本项目贡献是评价维度的转换：前人评估"规则改变分布形状"，本项目评估"规则 + 单一参数组能否**联**合命中实测多指标，及拟合参数与真实税负的数量级对表"——把模型从定性演示推到可证伪的定量校准。丘奖历届 ABM/统计物理类获奖极少（2022 铜奖为三方演化博弈、2024 铜奖为元宇宙交互数学框架，均非财富分布），差异清晰。"两种规则都拟合不上"是有效结论，须报告最小可达损失及其置信区间以证明不是优化失败。

8 · 决策门槛（go / no-go）

- **第 4 周末**：三项解析校验全部通过。KS 检验反复失败多为稳态未达或抽样相关性问题→限期 2 周排查；第 6 周仍未过则**降级 A**：缩减到纯复现 + 收敛性与有限规模研究（可复现性贡献定位），保留"模拟-解析对照"框架。

- **第 10 周末**：单次运行时长实测 ≤5 分钟。若 >5 分钟→N 降到 3×10^3 并加大种子数，Sobol 降为 256 组（降规模版本，主结论保留）。
- **第 22 周末**：至少一国完成双规则拟合。若两规则均严重失配（损失下界远超阈值）→失配即主结论，补一节"缺失机制讨论"（收入流、资产回报异质性），框架不变。
- 已知困难：WID 的财富份额本身是估算值——把"目标数据不确定性传播到拟合参数"写成研究内容（对 WID 值 ±1 标准误扰动重拟合）。
- 预算裁剪顺序：先砍第三国，再砍累进税的分段精细度（降为两段），最后砍 Sobol 规模；主结论（双规则联合拟合对比）在裁剪下成立。

E15 · 高中选课分配中 Boston 机制与延迟接受机制的模拟对比：以真实偏好数据下的可操纵性与排名福利为判据

优先级：中高 · 博弈与机制设计族 · 模拟+校内匿名调查 · CPU 极轻 · 编程+博弈论取向

1 · 研究问题

用本校（或合作学校）选修课/走班分配的真实匿名偏好数据驱动模拟，Boston 机制（Boston mechanism, BM）与学生最优延迟接受机制（Deferred Acceptance, DA）在平均偏好排名、正当嫉妒（justified envy）比例与可获利操纵空间上差多少？在"只能填 k 个志愿"的受限名单下，DA 的策略无关性（strategy-proofness）损失多快？

2 · 研究背景与空白

学校选择是机制设计的旗舰应用：Abdulkadiroğlu & Sönmez（AER 2003）证明 BM 可操纵、DA 策略无关但可能非帕累托有效；波士顿 2005 年真实改制。Chen & Sönmez（JET 2006）及"从 Boston 到上海到 DA"系列实验量化了真话汇报率排序 DA>上海>BM；Decerf & Van der Linden（JET 2021）给出可操纵性的理论比较；受限名单下 DA 失去策略无关性也已有刻画。理论与实验文献均很成熟。

空白在样本维度：已发表定量结果基于美欧学区数据或实验室诱导偏好；**用中国高中真实选修课偏好结构（高相关的热门课集中度）驱动的机制对比**未见公开报告，且热门课集中度恰是决定两机制差距大小的关键结构变量——相关性结构不同，结论幅度可能完全不同。方法全公开、可在小算例上精确核对、结论正负都成立。

3 · 可检验假设

- H1：在实测偏好分布下，BM 中存在可获利操纵的学生比例 ⩾30%，而无约束 DA 为 0（理论保证，作为管线正确性的内部检验）且受限名单 k=3 时 DA 的该比例升至 >10%。
- H2：偏好集中度（前 3 门热门课获得的首选份额）每上升 10 个百分点，BM 相对 DA 的平均排名福利差距扩大——方向为正、用模拟回归系数及置信区间量化；若差距对集中度不敏感，该稳健性是有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：（i）自建 BM/DA 代码在 Abdulkadiroğlu & Sönmez (2003) 原文全部手算例上输出与论文完全一致的匹配；（ii）在 ⩽4 学生 × 3 课程的随机小算例上与穷举验证 DA 无正当嫉妒、无约束 DA 零可获利操纵（⩾1,000 个随机算例全部通过）。此步不过关，后续全部结论无效。
2. 模拟口径预先写死：每个偏好情景 ⩾500 次重抽样（对调查偏好做自助抽样 + 加噪扰动），报三项指标的均值 ± 标准差与 95% 自助置信区间。
3. 操纵判定用可复现判据：逐一枚举每个学生的全部 k-志愿单替代（课程数 ⩽20 时可精确枚举），"存在严格改进即计为可操纵"，不用启发式。
4. 偏好数据采集合规：匿名问卷、知情同意书、经学校批准；只问课程志愿排序，不采集姓名/成绩/任何敏感信息；样本量目标 ⩾200 份（覆盖 ⩾2 个年级），并预先声明：本调查用于**标定偏好分布**而非统计推断个体行为，故样本量要求以"热门课份额估计的标准误 ⩽3 个百分点"（n≈200 时二项标准误约 3.5%）为准，达不到则扩大到合作班级。
5. 结果双口径：实测偏好分布 vs 独立同分布均匀偏好对照，量化"真实相关结构"改变结论的幅度。
6. 全部代码与匿名化数据（聚合分布，非个体原始表）开源。

5 · 数据与工具

用途	来源 / 工具
偏好分布	本校选修课匿名志愿调查（自采，n≥200；须学校同意 + 知情同意；若学校已有历史选课汇总统计可申请匿名聚合版，可行性**需核实**）
课程容量	本校教务公开的选修课容量表
机制与理论核对	Abdulkadiroğlu & Sönmez (2003) 原文算例；Chen & Sönmez (2006) 实验结果——仅用于校验与对比，不计入本项目贡献
模拟	Python（自写 BM/DA/受限 DA，各 <100 行）+ NumPy
偏好生成模型	多项 logit（以调查份额拟合效用尺度），拟合用 statsmodels

算力：500 学生 × 20 课的单次匹配毫秒级；500 重抽样 × 全学生操纵枚举（每人 $\leqslant C(20,3)\cdot 3!$ 志愿单）单情景分钟级，全项目总机时 <10 小时，纯 CPU 无压力。规模上限：操纵精确枚举在 $k\leqslant 4$、课程 $\leqslant 25$ 内可行，$k\geqslant 5$ **超出范围**（改抽样枚举并报覆盖率）。

6 · 方法路径

1. 实现 BM/DA/受限 DA，通过原文算例与随机小算例穷举校验（第 1 条验收）。
2. 设计匿名问卷并走学校审批 + 知情同意流程（第 3 周前启动，见块 8）。
3. 采集偏好数据，拟合多项 logit 偏好生成模型，报告拟合优度。
4. 跑双机制 × {无约束, k=3, k=4} × 双偏好口径的全对照，输出三项指标及置信区间。
5. 做集中度扫描：人工调节 logit 温度参数生成集中度梯度，估计 H2 的响应曲线。
6. 对操纵者画像做描述统计（操纵收益分布），明示不做个体因果推断。
7. 独立交叉校验：DA 结果用现成开源实现（如 `matching` Python 库，能力**需核实**）复算一致。

7 · 新颖性边界

> 本课题不声称任何新机制或新定理（BM/DA 性质均为已发表定理），不声称实验发现（本课题是模拟研究，调查仅用于标定分布）。已有工作：理论比较（2003/2021）、实验室实验（2006 及后续）、学区实证均已发表；丘奖历届相邻工作必须点名：2020 银奖"无限市场双边匹配"（纯理论）、2022 铜奖"数据交易市场稳定匹配算法"（算法设计）。本项目差异轴是换样本 + 换结构变量：用中国高中真实偏好相关结构驱动、并把"偏好集中度→机制差距"的响应曲线作为主结论——已有文献比较机制优劣，本项目量化"结论幅度对偏好结构的依赖"。两机制差距很小（结构不敏感）同样是有效结论，须给置信区间证明可分辨 2% 的排名福利差。

8 · 决策门槛（go / no-go）

- **第 3 周末**：学校书面同意调查。未获同意→**降级 A**：放弃自采数据，改用已发表偏好分布参数（如 Chen – Sönmez 实验的诱导偏好结构 + 公开学区申报数据的集中度统计）标定，全部模拟框架与 H2 保留，仅失去"本校样本"卖点。
- **第 8 周末**：回收问卷 $\geqslant$150 份且热门课份额标准误 $\leqslant$4%。不足→追加班级或合并两年级；仍不足则按降级 A 的分布参数补齐并如实报告。
- **第 6 周末**：机制代码通过全部穷举校验（硬闸，不过不进入正式模拟）。

- **第 24 周末**：主对照完成。若 BM/DA 差距与文献方向相反→优先排查代码（回到校验），确认无误后"反向结果 + 结构解释"升格为主结论。
- 已知困难：问卷申报偏好未必是真实偏好——明示这是标定用途的局限，并做加噪敏感性（±10% 份额扰动下主结论不变才下结论）。
- 预算裁剪顺序：先砍 k=4 情形，再砍集中度扫描点数（保 5 点），最后砍操纵收益分布分析；主结论（双机制三指标对比 + 集中度方向）在裁剪下成立。

E16 · 全球粮食贸易网络的冲击级联模型：以 2022 年黑海冲击的样本外预测秩相关为判据

优先级：中 · 网络经济学族 · 公开贸易数据 · CPU 轻量 · 编程+网络科学取向

1 · 研究问题

在 FAO/BACI 双边粮食贸易网络上构建的冲击级联（cascade）模型，用 2021 年前数据标定后，能否**样本外**预测 2022 年黑海供给冲击（俄乌冲突）下各进口国小麦进口量的实际变化排序（秩相关 ρ 达到多少）？模型对再分配规则与库存缓冲假设的敏感性有多大?

2 · 研究背景与空白

粮食贸易网络是网络经济学"冲击传导"理论的可检验场景。已有工作：Heslin 等（Environ. Res. Lett. 2019）在 FAO 小麦/玉米贸易网络上建立含库存缓冲的级联模型，模拟主产国减产的传导；Acemoglu 等（2012）奠定生产网络冲击放大的理论；2022 年后已有多篇论文用多层网络模拟俄乌冲突对全球食品可得性的冲击（arXiv 2210.01846 等）。

空白在评价维度与时间覆盖：绝大多数级联研究停在"模拟情景"层面，**用真实发生的 2022 冲击做严格的样本外验证（预测排序 vs 实测排序）的公开工作很少**——而这恰是对"模型不自说自话"要求的直接回应：2022 年实测进口变化是天然的检验集。数据全免费、模型可在小网络上手算核对、"模型预测失败"同样是有效结论（量化该类模型的真实预测力上限）。

3 · 可检验假设

- H1：以 2016 – 2021 平均网络标定的级联模型，对 2022 年小麦进口下降幅度前 30 大进口国的预测排序与实测排序的 Spearman ρ ⩾0.5。
- H2：加入库存缓冲（USDA PSD 期末库存数据）比无库存版本显著提高 ρ（自助法置信区间不重叠）；若无改善，"库存数据在年度尺度上无预测增益"是有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：（i）在 4 国玩具网络上，级联代码与手算逐步传导结果完全一致（⩾3 个手算情景）；（ii）复现 Heslin et al. (2019) 的一个核心情景结果（如主产国 X% 减产下最受影响进口国名单的前 5 名重合 ⩾4 个，或关键数值相对偏差 ⩽10%，以原文可提取数据为准）。此步不过关，后续全部结论无效。
2. 样本外纪律预先写死：2022 年数据在模型标定全程封存，仅在最终验证阶段使用一次；验证指标（Spearman ρ、前 10 名命中数）预先注册（GitHub 时间戳）。
3. 混杂剥离：实测 2022 变化须先扣除各国进口的 2016 – 2021 线性趋势（用趋势外推残差作为"冲击响应"），并报告"扣趋势 vs 不扣"双口径。
4. 敏感性用全局方法：对（再分配规则 3 种、库存动用比例、需求刚性参数）做拉丁超立方 ⩾200 组扫描，报 ρ 的分布及主结论翻转比例。
5. 随机成分（比例再分配的扰动）报 ⩾50 种子均值 ± 标准差。
6. 全部数据管道（原始 CSV→网络→结果）开源一键重跑。

5 · 数据与工具

用途	来源 / 工具
双边贸易流	FAOSTAT Detailed Trade Matrix（免费 CSV，年度、分商品）；交叉核对用 BACI/CEPII（免费下载，HS 6 位，2022/2023 覆盖**需核实**最新年份）
产量与库存	FAOSTAT 产量（免费）；USDA PSD Online（免费 CSV，含期末库存与贸易年度口径）
实测冲击响应	2022–2023 年小麦进口量（同上来源，封存至验证阶段）
网络与模拟	Python + NetworkX + NumPy（级联核心自写，<300 行）
方法对照	Heslin et al. 2019；arXiv 2210.01846——仅用于校验与对比，不计入本项目贡献

算力：约 180 节点 × 10^3 – 10^4 条边的网络，单次级联毫秒级；200 组 LHS × 50 种子 ≈10^4 次运行 <1 小时，纯 CPU 轻松。规模上限：HS 6 位全商品多层网络（10^5 边 × 多层）**超出范围**；降规模版本：单一商品（小麦）单层网络，即本课题主设定。

6 · 方法路径

1. 下载 FAOSTAT/USDA 数据，清洗出 2016 – 2022 小麦贸易网络与库存表（镜像流不一致时用出口方申报并报告差异率）。
2. 实现级联模型（减产→出口削减→按规则再分配→进口国短缺→二阶传导），通过玩具网络手算与 Heslin 复现校验（第 1 条验收）。
3. 预注册验证方案，封存 2022 数据。
4. 用 2016 – 2021 数据标定基线参数（拟合历史小型冲击年，如 2010 俄罗斯出口禁令的传导，作为标定集）。
5. 跑 2022 黑海情景（俄乌出口削减幅度用 UN/FAO 公开报告数字），输出预测排序。
6. 开封 2022 实测数据，计算 ρ 与命中数；跑 LHS 全局敏感性。
7. 独立交叉校验：用简单基准（"进口依赖度排序"，即无级联的一阶暴露）对照，报告级联模型相对一阶基准的增益——增益为零则如实报告。

7 · 新颖性边界

> 本课题不声称提出新级联算法，不声称预测未来粮价（只做量的传导，不做价格）。已有工作：Heslin et al. 2019 已建立同类模型并做情景模拟；2022 年后的多层网络冲击论文已模拟俄乌情景。本项目差异轴是评价维度转换：前人做"情景模拟"，本项目做**以真实事件为检验集的样本外预测力审计**，并以"级联模型相对一阶暴露基准的增益"为主结论——这直接回答"复杂模型是否值得"这一方法论问题。丘奖历届相邻工作须点名：2025 铜奖"全球价值链多阶段增加值引力模型"（引力模型、无级联与样本外验证）、2022 优胜"中国贸易冲击与美国创新"（计量实证），方法与判据均不同。ρ 很低（模型失败）是有效结论，但须证明一阶基准同样失败或成功，以区分"级联无用"与"问题不可预测"。

8 · 决策门槛（go / no-go）

- **第 5 周末**：FAOSTAT 2022 年小麦矩阵完整性核实（覆盖前 30 进口国 ⩾90%）。不完整→用 USDA PSD 年度口径替代（国家级进口总量足够，失去双边细节但验证指标不受影响）。
- **第 8 周末**：Heslin 复现通过。原文数据无法提取到可比精度→改为"玩具网络手算 + 守恒量检查（贸易量守恒漂移 <10^-8）"作为硬门槛，并在文中说明复现受限原因。
- **第 16 周末**：2010 俄罗斯禁令标定集上 ρ⩾0.4。达不到→**降级**：放弃样本外预测定位，转为"级联模型敏感性与结构脆弱性分析"（识别系统关键节点 + 全局敏感性），保留网络与模型框架，主结论改为结构性而非预测性。
- 已知困难：2022 年进口变化混杂了价格、汇率与政策禁令——扣趋势 + 与一阶基准对照只能部分剥离，须在局限性中定量讨论（列出同期主要出口禁令清单，FAO 公开）。
- 预算裁剪顺序：先砍玉米第二商品扩展（若已启动），再砍 LHS 组数（保 100 组），最后砍 2010 标定集（改用文献参数）；主结论（样本外 ρ + 相对基准增益）在裁剪下成立。

E17 · 双向口头拍卖的课堂实验复现与信息条件处理：以价格向竞争均衡收敛速度的预注册效应量为判据

优先级：中 · 行为经济学实验族 · 校内人类被试（硬合规） · 零算力 · 实验设计+统计取向

1 · 研究问题

在中国高中生被试的双向口头拍卖（double oral auction）市场中，成交价是否像 Smith（1962）报告的那样在 3 – 5 个交易期内收敛到竞争均衡（收敛系数 α 降到 5% 以下）？公开完整报价簿相对只公布最新成交价，是否加快收敛（预注册的处理效应）？

2 · 研究背景与空白

Vernon Smith（JPE 1962）的双向拍卖实验是实验经济学奠基工作：给买卖双方私人保留价格，口头竞价，价格惊人地快速收敛到理论均衡。后续文献（Ikica 等，IER 2023 系统再检验；"Fast Convergence in the Double Oral Auction" 2015 理论解释）确认该结果对被试数量、文化背景甚至极端收益不平等都稳健。信息条件（报价簿透明度）对收敛的影响在实验室有研究但结论幅度依情境而异。

空白在样本维度：这一"实验经济学最稳健结果"在**中国高中生、非货币激励（小额奖品/积分）情境**下的预注册复现未见公开报告；且该复现天然满足行为实验的教学合规性（无欺骗、无敏感信息、任务即课堂游戏）。复现类课题的价值判断已由 Ikica et al. 2023 的再检验传统背书：稳健性检验本身是被承认的贡献形态。收敛失败（若发生）同样是重要结论。

3 · 可检验假设

- **H1**：每场市场第 4 交易期起，Smith 收敛系数 α =（成交均价对均衡价的均方根偏差/均衡价）⩽5%（Smith 原文各市场后期 α 多在 1 – 3%）。
- **H2**：完整报价簿组比仅最新成交价组更快收敛：第 2 交易期 α 的组间差 ⩾3 个百分点（预注册方向与幅度）；若无差异，"信息条件在该情境不敏感"为有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：正式实验前，用 Smith（1962）原文图 1 的供需结构做 2 场试点，复现"后期成交价落在均衡价 ±10% 带内且逐期偏差单调下降"的定性-定量模式；同时用计算机模拟零智能（Zero-Intelligence, Gode & Sunder 1993）交易者在同一供需结构下的成交价分布，验证分析脚本对模拟数据能正确算出 α（与手算一致到小数点后 3 位）。任一不过关，正式实验不启动。
2. **检验力预先算死（拒绝"随便找几十个同学填问卷"）**：独立统计单元是**市场**（不是学生个体，被试间互动使个体观测不独立）。按文献中信息处理对早期 α 的效应量 d≈1.2（大效应，需在预注册中给出所引数值来源）计，α=0.05、power=0.8 的双样本检验需每组 ⩾12 场独立市场；每场 8 – 10 名学生 × 5 交易期 ×2 组 = 约 200 人次。若只能组织每组 6 场，则可检出效应下限为 d≈1.7，须在预注册与论文中明示该检验力边界，且 H2 退化为估计（报效应量与置信区间）而非检验。
3. 伦理硬合规：学校书面批准；学生与监护人书面知情同意；全程匿名编号、不采集姓名/成绩/家庭信息；激励用小额文具或课堂积分（不用现金）；数据保存与销毁方案写入方案。以上任一不满足即不启动。
4. 预注册：假设、α 计算式、剔除规则（如串通、中途退出场次作废）、分析脚本在首场实验前上传 OSF（免费）并锁定。

5. 统计口径：主检验为组间 Mann – Whitney U（市场级 α），报效应量（rank-biserial）与自助 95% 置信区间；逐期 α 曲线报均值 ± 标准差；不报个体级 p 值。
6. 全部匿名化成交记录、脚本与预注册链接公开。

5 · 数据与工具

用途	来源 / 工具
实验设计蓝本	Smith (1962) 原文供需结构；Ikica et al. (2023) 稳健性设定——仅用于校验与对比，不计入本项目贡献
实验实施	纸笔 + 白板报价（零成本）；或免费开源 oTree 本地部署（局域网、笔记本即服务器；oTree 双向拍卖现成模板可用性**需核实**，无则纸笔方案兜底）
零智能模拟基准	自写 Python 脚本（<150 行），Gode & Sunder (1993) 规则
被试	本校 2 个年级志愿学生（目标 200 人次；跨班组织以保证市场间独立）
分析	Python + SciPy/statsmodels

算力：零智能模拟 10^4 场秒级，纯 CPU 忽略不计。实验组织是真正的资源约束：24 场 × 40 分钟 ≈ 12 课时，须提前与学校敲定。

6 · 方法路径

1. 写零智能模拟与 α 分析脚本，完成脚本校验；设计实验手册与知情同意书。
2. 走学校审批与监护人同意流程（第 2 – 6 周，见块 8）。
3. 跑 2 场试点（不计入正式数据），修订流程并完成方法学校验，随后锁定预注册。
4. 分批组织 24 场正式实验（每周 2 – 3 场，约 10 周），双人记录成交数据并当日录入核对。
5. 按预注册脚本分析：逐期 α 曲线、组间检验、与 Smith 原文及零智能基准三方对照。
6. 稳健性：剔除规则敏感性（含/不含边缘场次）与市场规模（8 人 vs 10 人）分层。
7. 独立交叉校验：第二名成员对 20% 场次的原始纸面记录独立重新录入，错误率 <1%。

7 · 新颖性边界

> 本课题不声称发现新行为规律，不声称推翻 Smith——复现 + 一个预注册处理是明确的定位。已有工作：Smith 1962 原始结果；Ikica et al. 2023 已系统再检验收敛与不对称性；信息透明度效应在实验室市场文献中已有研究（结论幅度情境依赖）。本项目贡献：（i）该旗舰结果在中国高中生 + 非货币激励样本上的首次预注册复现（换样本独立检验，方法完全公开）；（ii）信息条件效应在该新样本上的预注册估计。与零智能基准的三方对照可区分"收敛来自制度还是来自人的理性"。丘奖历届无实验室实验类获奖（2022 金奖"聪明钱羊群行为"为二级市场实证），本课题若执行严格将是该赛道少见的实验形态。收敛失败或处理无效应都是有效结论，检验力边界已预先写明。

8 · 决策门槛（go / no-go）

- **第 6 周末（硬闸）**：学校书面批准 + 同意书回收 ⩾120 人。未达→**降级 A**：缩为单组纯复现设计（12 场，删 H2 保 H1），检验力声明相应改写；若批准彻底不可得→**降级 B**：转为纯计算课题"零智能与强化学习交

易者在双向拍卖中的收敛机制模拟"（Gode – Sunder 复现 + 学习规则扩展），保留 α 收敛主框架，明示无人类数据。

- **第 10 周末**：试点 2 场完成且流程单场 ⩽40 分钟。超时→简化为 4 交易期并重做检验力核算。
- **第 24 周末**：正式场次完成 ⩾16 场。不足 24 场→按实际场次报可检出效应下限，H2 转为估计口径（预注册中已写明该分支）。
- 已知困难：学生课后串通或重复参加会破坏市场独立性——写入剔除规则与场次隔离设计（不同班级、不同供需参数微扰），并把"串通迹象检测"（期内价格方差异常）写成数据质量小节。
- 预算裁剪顺序：先砍市场规模分层，再砍处理组（保纯复现），伦理与预注册永不裁剪；主结论（新样本收敛系数及其区间）在裁剪下成立。

E18 · 居民屋顶光伏采纳的 Bass 扩散与基于主体模型对比：以美国 Tracking the Sun 分区县采纳曲线的留出期预测为判据

优先级：中低 · 环境与能源经济族 + 基于主体建模族 · 公开数据校准 · CPU 中量 · 编程+数据拟合取向

1 · 研究问题

对居民屋顶光伏（rooftop PV）的县级累计采纳曲线，含同伴效应（peer effects）的基于主体模型（agent-based model, ABM）相对经典 Bass 扩散模型，在"用 2018 年前数据标定、预测 2019 – 2023 留出期"的任务上均方根误差能降低多少？同伴效应参数在县之间的异质性能否被收入与政策变量解释？

2 · 研究背景与空白

技术扩散的 Bass 模型（1969）有**闭式解析解**（创新系数 p、模仿系数 q），是一切扩散 ABM 的天然解析锚点。光伏采纳 ABM 已有成熟文献：Zhang 等（AAAI 2015）用 ABM 预测圣迭戈屋顶光伏采纳；加州部署预测（LBNL/OSTI 报告）与新加坡、社区微电网等 ABM 均已发表；同伴效应（邻近已装机数影响采纳概率）被反复证实。Kiesling 等对"经验立地的扩散 ABM"的批评性综述（arXiv 1608.08517）明确指出该领域痛点：**多数 ABM 校准后从不做留出期预测检验**。

空白在评价维度：正面回应上述批评——把 Bass（2 参数）与 ABM（多参数）放在**同一留出期预测任务**上同台对比，回答"ABM 的复杂度是否换来预测力"，用完全公开的 LBNL Tracking the Sun 逐系统安装数据（200 万+条记录、含安装日期与邮编）在多个县上系统执行。数据免费、Bass 解析解保证可校验、"ABM 不比 Bass 好"是有效且重要的结论。

3 · 可检验假设

- H1：在 ⩾20 个县上，含同伴效应 ABM 的留出期 RMSE 中位数比同数据标定的 Bass 模型低 ⩾15%；若不低，"简单模型足够"为主结论。
- H2：逐县拟合的模仿强度参数与县收入中位数（ACS 公开数据）的秩相关 $|\rho| \geqslant 0.3$，方向为正（高收入社区同伴效应更强，与已发表加州证据方向一致）。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：（i）ABM 在关闭异质性与空间结构的极限下退化为 Bass 过程——模拟均值曲线与 Bass 解析解逐点偏差 ⩽2%（N=10^4 主体、20 种子均值）；（ii）Bass 拟合代码在 Bass (1969) 原文耐用品数据（论文表格可得）上复现原文 p、q 估计，相对偏差 ⩽5%。此步不过关，后续全部结论无效。
2. 时间序列纪律：严格按时间划分——2007 – 2018 标定、2019 – 2023 留出；并做一次"随机划分 vs 时间划分"对照量化随机划分对 RMSE 的高估/低估幅度。
3. 模拟口径：每县每参数点 ⩾20 种子，报 RMSE 均值 ± 标准差与 95% 置信区间；模型对比用双成本口径（等参数自由度不可行时，报自由度差异并用 AIC/BIC 与留出 RMSE 双判据）。
4. 参数敏感性：对 ABM 的（同伴半径、模仿权重、收益敏感度、异质性方差）做 Sobol 指数（Saltelli ⩾512 组）；县选择做敏感性（按规模分层抽样两套县集，结论须一致）。
5. 县入选标准预先写死：累计安装 ⩾1,000 套、时间序列 ⩾12 年，达标县全取或分层随机抽 20 个，不得挑好拟合的县。
6. 全部管道开源；单县单次 ABM 运行时长写入附录。

5 · 数据与工具

用途	来源 / 工具
逐系统安装记录	LBNL Tracking the Sun（emp.lbl.gov 免费公开数据集，>2×10^6 条，含安装日期、邮编、规模；最新版覆盖年份**需核实**）
潜在采纳者基数	美国 Census/ACS 县级自有住房户数（免费 API）
收入与电价	ACS 县收入中位数（免费）；EIA 州电价（免费）
政策变量	DSIRE 激励数据库（免费查询，结构化导出难度**需核实**，不可得则用州级虚拟变量）
建模	Python + NumPy/Numba 自写 ABM；Bass 拟合用 SciPy 非线性最小二乘
方法对照	Zhang et al. AAAI 2015、LBNL 加州预测报告——仅用于校验与对比，不计入本项目贡献

算力：单县 N≤10^5 主体（按户数缩放代理比例 1:10）、月步长 200 步、单种子约 10 – 60 秒（Numba）；20 县 × Sobol 512 × 20 种子为上限口径，实际用"主分析全County + Sobol 只在 3 个代表县"控制在总机时 ≤5 天。全美 3,000 县 × 全 Sobol **超出范围**；降规模版本：10 县 + 256 组 Sobol（<1 天）。中国分省数据（国家能源局分布式光伏统计）粒度与连续性**需核实**，作为可选外推讨论不作主数据。

6 · 方法路径

1. 实现 Bass 拟合并复现原文估计；实现 ABM 并通过退化极限校验（第 1 条验收）。
2. 下载 Tracking the Sun，清洗到县×月累计采纳面板（邮编→县映射用 HUD 免费对照表）。
3. 按预写标准选县；封存 2019 – 2023 留出段。
4. 逐县标定 Bass 与 ABM（标定只用 2018 前数据；ABM 校准用拉丁超立方 + 局部细化）。
5. 开封留出段，计算双模型 RMSE 分布与配对差异检验（Wilcoxon，报效应量）。
6. 跑 3 个代表县的 Sobol 敏感性与参数-收入相关分析（H2）。
7. 独立交叉校验：Bass 用 R `diffusion` 包（能力**需核实**）复算 p、q 一致性 ≤2%。

7 · 新颖性边界

> 本课题不声称提出新扩散机制，不声称 ABM 结构是采纳行为的因果模型。已有工作：光伏采纳 ABM（圣迭戈 2015、加州、新加坡等）与同伴效应实证均已发表；"扩散 ABM 缺预测验证"的批评已由综述提出。本项目贡献是把该批评落实为系统审计：同一数据、同一留出任务上 Bass vs ABM 的预测力对比及其跨县分布——评价维度从"拟合优度"转换为"样本外预测增益"，并以多县系统执行替代单城市案例（样本维度扩展）。丘奖历届相邻工作须点名：2025 入围"不确定光照下光伏最优定价机制"（机制设计，非扩散）、2021 金奖"燃油车监管的税收与配额理论"（理论模型）；无扩散预测审计类工作。"ABM 无增益"结论有效且对政策模拟界有警示价值，须报配对置信区间证明可分辨 10% 的 RMSE 差。

8 · 决策门槛（go / no-go）

- **第 4 周末**：Tracking the Sun 下载成功且 ≥20 县达标。数据入口变更或达标县不足→**降级 A**：改用州级月度装机（EIA-861M，免费且稳定），县级异质性分析（H2）降为州级（n≈15 个高装机州），双模型对比框架不变。

- **第 8 周末**：ABM 退化极限校验通过。反复不过多为时间步长/配对方式偏差→改连续时间 Gillespie 化实现；第 11 周仍不过则**降级 B**：放弃 ABM，改为"Bass 及其 3 个解析扩展（Generalized Bass、含价格项）"的留出预测对比，保留"复杂度 vs 预测力"主结论框架。
- **第 20 周末**：单县全流程（标定+留出评估）≤2 小时。超时→代理比例降到 1:50 并重做规模收敛检查（三档代理比例主指标漂移 <2%）。
- 已知困难：留出期含 2020 – 2022 疫情与利率冲击，两模型都会失准——把"结构性冲击期的模型脆弱性"写成分段分析（2019 vs 2020 – 2023 分开报），是研究内容而非障碍。
- 预算裁剪顺序：先砍 H2 相关分析，再砍 Sobol 代表县数（保 1 县），最后砍县数到 10；主结论（双模型留出 RMSE 对比）在裁剪下成立。

E19 · 银行间网络传染的 Eisenberg–Noe 清算模型：以最大熵与最小密度两种网络重构下的传染损失区间为判据

优先级：低（高风险高回报） · 网络经济学族 · 公开监管数据 · CPU 轻量 · 数学+金融取向

1 · 研究问题

用欧洲银行管理局（EBA）透明度演练公开数据重构欧洲银行间风险敞口网络后，Eisenberg – Noe 清算模型给出的系统性传染损失，在最大熵（maximum entropy）与最小密度（minimum density）两种极端网络重构假设下的区间有多宽？哪些银行的系统重要性排序对重构假设稳健？

2 · 研究背景与空白

Eisenberg & Noe（Management Science 2001）证明银行间债务清算向量的存在唯一性（不动点定理），是系统性风险的基准模型，小网络可**手算解析核对**。因双边敞口数据保密，文献用银行级总量重构网络：最大熵法（Upper & Worms 2004）被 Mistrulli（2011）证明会**低估**传染（假设完全连接）；Anand 等提出最小密度法作为另一极端；Upper（2011, JFS）综述确立"重构假设主导结论"这一方法论问题。近期工作（JRFM 2025）已把最大熵重构扩展到 EBA 2018/2021/2023 数据的多国框架。

空白在评价维度与时间覆盖：多数研究报告单一重构下的点估计；**用两种极端重构给出传染损失的区间（bound）、并识别对假设稳健的排序结论**，在最新一轮 EBA 公开数据（2024/2025 演练，覆盖需核实）上系统执行的公开工作未见。模型有解析锚点、数据免费、"区间宽到结论不可用"本身是重要的方法论结论。

3 · 可检验假设

- H1：两种重构下系统总传染损失（外生冲击设定同一）相差 ⩾3 倍——量化"重构假设不确定性"主导结论的幅度。
- H2：尽管损失水平差异大，前 10 大系统重要性银行（按传染贡献排序）在两种重构下的重合度 ⩾60%（Jaccard），即**排序比水平稳健**；若排序也不稳健，该否定结论同样有效并更具警示性。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：（i）自写 Eisenberg – Noe 不动点求解器在原文（2001）示例与 ⩽4 银行手算网络（⩾3 个情景）上输出与解析解完全一致的清算向量（逐分量偏差 <10^-10）；（ii）最大熵重构代码（RAS/IPF 算法）在文献玩具矩阵上复现行列和约束（残差 <10^-8）并与已发表小例一致。此步不过关，后续全部结论无效。
2. 传染指标口径预先写死：报（总损失/系统总资产、违约银行数、各行 DebtRank 式传染贡献）三指标；外生冲击情景预先写死（单行资产减记 X%、主权债减记情景各一，X 取文献常用值并做 3 档敏感）。
3. 重构不确定性再抽样：在两极端之间用带约束随机网络采样（如固定行列和的随机矩阵 ⩾200 个样本）报损失分布，不止两个端点。
4. 数据口径：银行间敞口用 EBA 模板中"金融机构敞口"科目（科目映射须列表公开，映射选择做一次敏感性）；资本用 CET1。
5. 全部结果附与 Upper (2011) 综述中同类研究量级的对照表（仅对比，不计入贡献）。
6. 代码 + 数据处理管道开源一键重跑。

5 · 数据与工具

用途	来源 / 工具
银行级敞口与资本	EBA Transparency Exercise 公开 CSV（eba.europa.eu 免费下载，约 100–150 家银行、逐半年；最新轮次与"对金融机构敞口"字段粒度**需核实**——这是本课题第一核实项）
清算模型	自写 Python 不动点迭代（Eisenberg–Noe 原文算法，<100 行）
网络重构	自写 RAS/IPF（最大熵）与最小密度贪心算法（Anand et al. 2015 伪码公开）
随机网络采样	NetworkX + 自写约束采样
方法对照	Upper (2011)、Mistrulli (2011)、JRFM 2025 多国重构——仅用于校验与对比，不计入本项目贡献

算力：150 银行网络的清算不动点毫秒级；200 个随机网络 × 多情景 <10 分钟，纯 CPU 极轻。本课题风险在**数据字段与金融概念理解**，不在算力。规模上限：全球 G-SIB 跨洲网络（需 FDIC+EBA+APRA 拼接）**超出范围**；降规模版本：只做单一国家子网络（如德国 ~20 家）。

6 · 方法路径

1. 实现并校验 Eisenberg – Noe 求解器与两种重构算法（第 1 条验收）。
2. 下载最新 EBA 数据，核实敞口字段粒度并建立科目映射表（第 5 周硬核实）。
3. 构建银行级（总资产、CET1、对金融机构敞口行列和）输入表，做量纲与守恒自检。
4. 跑两极端重构 + 约束随机采样 200 网络，得损失分布与区间。
5. 对每个网络样本跑预写冲击情景，计算三指标与银行排序。
6. 分析排序稳健性（Jaccard、Kendall τ）与 H1/H2 判定。
7. 独立交叉校验：抽 1 个随机网络样本用现成实现（如 `sysrisk` /NetworKit 相关模块，可用性**需核实**）或第二名成员独立代码复算清算向量。

7 · 新颖性边界

> 本课题不声称提出新清算模型或新重构算法（全部已发表），不声称估计"真实"传染风险（真实双边数据保密，本课题恰恰把这一不可知性量化为区间）。已有工作：Eisenberg – Noe 2001、最大熵及其低估问题（Mistrulli 2011）、最小密度（Anand et al.）、EBA 数据多国最大熵框架（JRFM 2025）。本项目贡献是评价维度转换 + 时间覆盖：前人报告单一重构点估计，本项目在最新一轮 EBA 数据上报告**重构不确定性传播后的损失区间与排序稳健性**——把"该领域结论有多可信"本身作为研究对象。丘奖历届相邻工作：2020 优胜"G20 股票收益有向树网络"（收益相关网络，非债务清算），差异清晰。区间过宽（结论不可用）与排序稳健两种结果都有效。

8 · 决策门槛（go / no-go）

- **第 5 周末（第一硬闸）**：确认 EBA 公开模板含可用的"对金融机构/信贷机构敞口"银行级字段。若字段过粗或已停更→**降级 A**：改用各国央行公布的银行间市场总量 + 文献中意大利/德国已发表网络统计构造校准网络（保留区间框架，明示数据为合成校准）；或**降级 B**：转为纯方法研究"重构算法对传染估计的偏差：在文献公开的真实网络（如 e-MID 已发表统计）上做受控对比"。两条路径都保留"重构不确定性区间"主结论。
- **第 12 周末**：两种重构 + 清算全流程在真实数据上跑通且守恒自检通过。

- 第 20 周末：随机采样 200 网络完成。若约束采样器混合不佳（区间明显窄于两端点）→如实报告采样局限，主结论退回两端点区间 + 敏感性。
- 已知困难：会计科目到"银行间敞口"的映射有裁量空间——把映射敏感性（两种映射方案对照）写成研究内容；金融概念门槛高，仅适合已修过或愿自学公司金融基础、且接受"数据字段核实可能否决主数据源"风险的学生。
- 预算裁剪顺序：先砍第二冲击情景，再砍采样数（保 100），最后按降级 A/B 收缩；"区间 + 排序稳健性"框架全程保留。

E20 · 通胀预期形成的自适应学习模型：以疫情后通胀急升期（2021–2024）纽约联储调查微观数据的样本外拟合为判据

优先级：低（风险最高） · 行为宏观建模族 · 公开调查微观数据 · CPU 轻量 · 计量+宏观取向

1 · 研究问题

常数增益自适应学习（constant-gain adaptive learning）模型对家庭通胀预期的刻画，在 2021 – 2024 美国通胀急升与回落期是否仍优于静态理性预期与朴素适应性预期基准（以对纽约联储 SCE 调查中位数预期的样本外 RMSE 为判据）？拟合出的学习增益参数在通胀急升前后是否发生可检测的跳变？

2 · 研究背景与空白

预期形成是行为宏观的核心战场。自适应学习（Evans & Honkapohja 2001 框架）假设主体像计量经济学家一样递归更新参数；Branch（Economic Journal 2004）用密歇根调查证明主体在理性/适应/朴素预测器间按表现切换；后续文献（Boston Fed WP 12-19；IMF WP 2023/19、2024/25）反复确认学习模型对调查预期的拟合优于理性预期，且约 40% 个体不能拒绝理性、20%+ 最像学习者。2021 年起的通胀急升提供了 40 年未有的强激励环境——文献（IMF 2024）已注意到预期分布拉宽。

空白在时间覆盖：Branch 类"预测器竞争 + 学习增益估计"的框架在**完整的急升-回落周期（2021 – 2024/25）SCE 微观数据上的系统重估**未见公开发表（IMF 2024 覆盖到疫情中段且未做预测器竞争结构估计；须在开题时再检索确认）。这是标准的"换样本（新时期）独立检验"：急升期是对学习模型最严苛的压力测试，增益跳变与否都是有效结论。风险在于计量实现门槛高，故列为 E20。

3 · 可检验假设

- H1：常数增益学习模型对 SCE 一年期中位数预期的 12 个月滚动样本外 RMSE 比理性预期基准（专业预测者 SPF 中位数作为理性代理）低 ⩾15%，比朴素基准（上期实际通胀）低 ⩾10%。
- H2：分段估计的学习增益在 2021Q2 前后显著上升（自助 95% 置信区间不重叠）——即高通胀环境使主体"更快遗忘旧数据"，与学习理论的注意力预测一致；若无跳变，"增益结构稳定"为有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：（i）自写递归最小二乘（RLS）/常数增益滤波在模拟数据（已知真参数的 AR 过程 + 已知增益的学习主体）上恢复真参数，偏差 ⩽5%（100 次蒙特卡洛）；（ii）复现 Branch (2004) 或一篇指定后续论文的核心表（预测器份额或增益估计），关键数值相对偏差 ⩽10%（若原文数据不可得，则在密歇根公开时间序列上复现其定性模式并明示受限）。此步不过关，后续全部结论无效。
2. 时间序列纪律：全部模型比较用滚动样本外（training window 预先写死为 10 年），并做一次"全样本内拟合 vs 滚动样本外"对照量化过拟合幅度；禁止用未来数据选超参。
3. 口径预先写死：预期变量用 SCE 一年期通胀预期中位数与四分位距（分布信息），实际通胀用 CPI 同比；月度频率；断点检验用 Bai – Perron 或预先指定的 2021Q2 单断点 + 自助推断，报效应量（增益变化幅度）非仅 p 值。
4. 稳健性：密歇根调查中位数序列（免费公开）作平行样本复做主分析，两数据源方向一致才下强结论。
5. 全部估计给自助法（block bootstrap ⩾1,000 次）置信区间。
6. 代码开源；SCE 微观数据按其使用条款处理（能否再分发需核实，不能则只发布处理脚本）。

5 · 数据与工具

用途	来源 / 工具
家庭预期微观/分布数据	NY Fed Survey of Consumer Expectations（newyorkfed.org 免费下载汇总与微观文件；微观文件的滞后发布窗口与使用条款**需核实**——第一核实项）
平行样本	Michigan Surveys of Consumers 公开时间序列（免费；微观数据需 ICPSR 权限，**需核实**，主分析不依赖它）
实际通胀与宏观量	FRED（CPI、失业率等，免费 API）
理性代理	Philadelphia Fed SPF（免费下载）
估计	Python + statsmodels/自写 RLS 滤波（核心 <200 行）；断点推断自写 bootstrap
方法对照	Branch (2004)、IMF WP 2024/25——仅用于校验与对比，不计入本项目贡献

算力：全部为低维时间序列滤波与 bootstrap，笔记本分钟级—小时级，纯 CPU 无压力。本课题风险全部在**计量实现与概念理解**（状态空间、学习滤波、断点推断），不在数据量与算力；降规模版本：只用公开中位数序列（不碰微观文件）做全部主分析。

6 · 方法路径

1. 写模拟-恢复测试（第 1 条验收 i），吃透常数增益 RLS 的实现细节。
2. 下载 SCE/密歇根/SPF/CPI，对齐月度面板；核实 SCE 微观条款（第 4 周）。
3. 复现 Branch 类核心结果（第 1 条验收 ii）。
4. 估计三类模型（常数增益学习 / 理性代理 / 朴素），跑滚动样本外比较。
5. 做 2021Q2 断点的增益分段估计与 bootstrap 推断。
6. 密歇根平行样本复做；分教育/年龄组的异质性分析（SCE 公开分组序列，作延伸）。
7. 独立交叉校验：用 Kalman 滤波等价形式复算常数增益估计，参数一致 ≤2%。

7 · 新颖性边界

> 本课题不声称提出新预期理论，不声称结构因果识别（模型比较是预测力竞赛 + 参数刻画，非因果）。已有工作：学习模型优于理性预期的证据（Branch 2004 及后续）、疫情期预期分布拉宽（IMF 2024）均已发表。本项目贡献是时间覆盖的独立检验：把预测器竞争 + 增益估计框架推进到完整的 2021 – 2025 急升-回落周期并检验增益跳变——这是该理论"注意力随通胀环境变化"预测的一次干净的新样本检验。丘奖历届相邻工作须点名：2025 优胜"中国菲利普斯曲线平坦化：央行沟通与通胀预期"（中国、菲利普斯曲线视角，非学习模型结构估计），差异在对象（预期形成机制本身）与方法（滤波估计 + 样本外竞赛）。H1 失败（学习模型不再占优）是有效且有趣的结论；H2 无跳变同样有效，均须给置信区间证明检验力。

8 · 决策门槛（go / no-go）

- **第 4 周末**：SCE 数据下载并对齐成功；微观条款核实完毕。微观不可用→立即锁定"公开中位数 + 四分位距序列"降规模版本（主结论框架完整保留，失去个体异质性延伸）。
- **第 8 周末（最硬闸）**：模拟-恢复测试通过。此为计量能力的试金石：若第 10 周仍未通过，说明学生数学准备不足，**整体降级**为 E20-lite："三种简单预期规则（朴素/加权移动平均/固定增益）对 SCE 中位数的样本

外预测比较 + 急升期误差分解"——不含滤波结构估计，保留"样本外竞赛 + 时期对比"主框架。

- **第 16 周末**：Branch 类复现通过（偏差 ≤10%）。原文数据不可得且定性复现也失败→重检实现；仍失败则以模拟校验为唯一硬门槛并在论文明示。
- **第 30 周末**：主竞赛与断点分析完成，密歇根平行样本启动。
- 已知困难与选择前提：本题需要状态空间/滤波的自学（约 6 – 8 周投入），仅适合数学最强、且接受"复现关卡可能通不过"风险的学生；调查中位数与个体预期的加总偏差是已知局限，写入讨论。
- 预算裁剪顺序：先砍异质性分组延伸，再砍密歇根平行样本，最后按 E20-lite 收缩；"样本外预测竞赛 + 急升期结构变化"主结论线全程保留。

附录 A · 倒排时间表（52 周，通用骨架）

> 终点是 **2027 年 9 月 15 日提交截止**（按 2026 赛季日程外推，须以当年官方公告为准）。第 0 周为 2026 年 9 月第一周。每个阶段都有可测的硬门槛，不是任务描述。

周次	阶段	硬门槛（可测，不达标即触发降级）
0—2	选题定稿	从本方案选定 1 条主路线 + 1 条备选路线；确认指导老师；核对当年官方参赛指南的学科研究范围
2—6	环境与方法学校验	**跑通该路线验收标准第 1 条的方法学校验**。这是全程最重要的门槛：不过关不得进入下一阶段
6—10	数据/器材到位	数据完成下载并通过完整性检查；或器材到齐且装置重复性达标（实验类 RSD ≤ 阈值）
10—14	第一次 go/no-go	按各路线写明的阈值判定。不达标立即切换到该路线的具名降级路径，**不要拖到第 20 周**
14—28	主体研究	完成主实验/主计算的全部参数点；中期产出可展示的核心图表 3 张
28—34	第二次 go/no-go	主结论方向已明确（含"差异不显著"这一有效结论）；误差棒已能证明有分辨该差异的能力
34—40	交叉校验	用第二种独立方法重算核心量，两者一致或差异有可解释来源
40—46	中文初稿	全文完成；图表定稿；代码与数据整理为可一键重跑的仓库
46—50	英文稿与查重	**完成英文研究报告与查重报告**（总决赛强制英文，且提交材料含查重报告）
50—52	提交与答辩准备	系统提交（9 月 15 日截止，逾期不可修改）；英文 PPT 与答辩问答演练

两个容易被忽略的时间陷阱：

其一，英文化不是最后一周的事。总决赛全程英文（研究报告、PPT、答辩），从第 0 周起就要用英文记录关键术语与方法描述，否则第 46 周会成为瓶颈。

其二，11 月 3 – 10 日的公示期会把研究报告全网公开。这意味着新颖性问题会暴露在公众视野，查重与文献边界必须在提交前就处理干净。

附录 B · 资源清单（全赛道通用底座）

用途	来源 / 工具	说明
文献检索	Google Scholar、arXiv、PubMed、Semantic Scholar、Connected Papers	**必须用英文检索**；中文检索会系统性漏掉近三年工作
全文获取	arXiv、bioRxiv、PMC、作者主页、通过学校图书馆或指导老师获取	不要使用非法下载渠道，公示期会暴露
代码与数据托管	GitHub（公开仓库）+ Zenodo（存档并获取 DOI）	可复现性是评审加分项，也是自证独立完成的证据
计算环境	Python 3.11+（conda/uv 管理）、Jupyter	环境须导出为 `environment.yml` 或 `requirements.txt`
统计与作图	numpy / scipy / statsmodels / matplotlib	图表须含误差棒；配色对色盲友好
实验记录	纸质或电子实验记录本，逐日、带日期、不得回填	评委可能追问原始记录；实验类还需按要求提交实验视频
写作	LaTeX（Overleaf）或 Word；参考文献用 Zotero 管理	英文稿建议请母语者或老师润色，但**内容必须学生本人撰写**
版本管理	Git，从第 0 周开始提交	连续的提交历史是"学生本人完成"最有力的旁证

附录 C · 红线清单

> 以下每一条，触碰即可能导致取消参赛/获奖资格，或在评审中被直接判负。

资格红线 1. 指导老师不得来自公司、培训机构或任何谋利机构。组委会有理由相信学生受培训机构指导完成研究报告，即保留取消资格的权利。 2. 不得跨校组队；同一学科每人只能报名一次；同一篇论文不得报两个学科。 3. 参赛信息须据实填写；报告提交截止后不得更改团队与指导老师信息。 4. 曾参加或拟参加其他竞赛、已投稿或发表的，**必须在报名时如实申报**。

学术诚信红线 5. 不得将他人（含指导老师、高校课题组）已完成的工作写成本人贡献。研究报告须明确区分"他人已发表成果""公开数据库/事件表"与"本项目贡献"。 6. 公开数据库、已发表事件表、预训练模型只能作为对照与输入，**不得计入学生的数据或方法贡献**。 7. 不得给出未标注为推测的物理/因果外推。讨论部分可以推测，但必须明示。 8. 提交材料含查重报告；公示期报告全网公开，抄袭与过度重合会被公开发现。 9. AI 使用政策以当年官方参赛规则为准（公开公告中未见明文，**须核实**）。无论政策如何，研究报告的科学内容必须由学生本人完成；若使用 AI 辅助，按当年规则要求致谢。

方法学红线（不违规，但会被评委问倒） 10. 没有方法学校验的结果不可信——所有路线的验收标准第 1 条都是这个门槛。 11. 不平衡分类报 accuracy、时间序列随机划分、幂律只做双对数最小二乘、模拟不做收敛性检查：这四项是评审最常见的追问点，方案里已逐条写死正确口径。 12. "差异不显著"是有效结论，但必须有误差棒证明有能力分辨该差异；没有误差棒的"无差异"等于没有结论。

附录 D · 本方案的使用方式

1. **先读赛道导语**，判断这个赛道的评委口味是否匹配学生的能力结构。跨赛道选题（例如数学好的学生去做经济金融建模的理论建模族）常常比在本赛道硬拼更有胜算。
2. **按可行性顺序从上往下读**，不要从最精彩的那条开始。编号越小越稳。
3. **选定一条主路线后，必须再选一条备选路线**，且两条最好落在不同的族——这样第一次 go/no-go 触发时有真正的退路。
4. **第 2 – 6 周的方法学校验是全案的生死线**。 这一步跑不通，说明学生的技能栈与该路线不匹配，此时换题的成本远低于第 20 周换题。
5. **本方案的赛制信息取自官方公告（yau-awards.com，2020 – 2026 全部公示文章），但规则每年会改**。 报名开放后必须重新核对当年的《参赛规则》与六个学科各自的《参赛指南》，尤其是研究范围、评审标准与 AI 使用政策。