

连享会 · 2026 寒假班

Update: 2025.12.5 , PDF 课纲

连享会 · 2026 寒假班

课程概览

- A. 课程概要
- B. 授课教师简介
- C. 课程特色

1. 初级班

- 1.1 课程导引
- 1.2 专题介绍 (初级班)
 - A1. Stata 简介和数据处理
 - A2. Stata 程序基础
 - A3. 实证分析可视化
 - A4. 线性回归分析
 - A5. 模型设定和解释
 - A6. 面板数据模型
 - A7. 论文复现
 - A8. 自行研读-提供复现数据和代码和中文精要

2. 高级班

- 2.1 课程导引
- 2.2 专题介绍 (高级班)
 - B1. 结构变化和异质性分析
 - B2. 动态面板模型
 - B3. 面板 VAR 模型
 - B4. 面板 ARDL 模型：长期关系
 - B5. 样本选择与自选择
 - B6. 交互固定效应和因子模型

3. 论文班

- C1. RCT 对因果识别的启示：论述内生性与溢出效应
- C2. 事件冲击与即时反应：利用调查时间随机性的准实验识别
- C3. 从冲突到教育差距：Cohort-DID 与地理 IV 的综合识别策略
- C4. 交互式 IV (Shift-share IV)：从识别直觉到排他性假设的论证

4. 报名和缴费信息

缴费方式

5. 听课指南

6. 助教招聘

说明和要求

课程概览

A. 课程概要

班次

- 初级班：2026.1.28-29 / 1.31-2.1 (4 天)，连玉君
- 高级班：2026.2.3-5 (3 天)，连玉君
- 论文班：2026.2.7-9 (3 天)，郭士祺

安排

- **时间**：2026 年 1 月 28 - 2 月 9 日
- **方式**：网络直播 + 回放
- **授课教师**：连玉君 (初级+高级) || 郭士祺 (论文班)
- **PDF 课纲**：<https://KC.lianxh.cn/PX.pdf>，[参考文献和预读资料](#)
- **课程主页**：<https://www.lianxh.cn/PX.html> (答疑和板书)
 - **Note**: 预习资料、常见问题解答等都将通过该主页发布。
- **报名链接**：<https://www.wjx.top/vm/weMgRkL.aspx#>
- **助教招聘**：<https://www.wjx.top/vm/mYvepXR.aspx#>

回放安排：每班 30 天回放

- **初级班**：2.2-2.14 (13 天)；2.24-3.12 (17 天)
- **高级班**：2.6-2.14 (9 天)；2.24-3.16 (21 天)
- **论文班**：2.10-2.14 (5 天)；2.24-3.20 (25 天)
- **多班 组合**
 - **初级 + 高级**：2.6-2.14 (9 天)；2.24-4.15 (51 天)，共计 60 天
 - **初级 + 论文**：2.2-2.6 (5 天)；2.10-2.14 (5 天)；2.24-4.14 (50 天)，共计 60 天
 - **高级 + 论文**：2.10-2.14 (5 天)；2.24-4.19 (55 天)，共计 60 天
 - **全程班**：2.10-2.14 (5 天)；2.24-5.19 (85 天)，共计 90 天

提示：以上日期均为 2026 年。

B. 授课教师简介



[连玉君](#)，西安交通大学经济学博士，中山大学岭南学院副教授，博士生导师。已在 China Economic Review、Energy Economics、Stata Journal、Pacific-Basin Finance Journal、经济研究、管理世界、经济学(季刊)、金融研究、统计研究等期刊发表论文 70 余篇。目前已完成 Panel VAR、Panel Threshold、Two-tier Stochastic Frontier 等计量模型的 Stata 实现程序，并编写过几十个小程序，如 [lianxh](#)，[ihelp](#)，[sftt](#)，[winsor2](#)，[xtbalance](#)，[bdiff](#)，[getiref](#) 等。连玉君老师团队一直积极分享 Stata 应用中的经验，开设了 [连享会-博客](#)，[连享会-知乎](#) 等专栏，已在微信公众号 ([连享会](#)，ID: [lianxh_cn](#)) 分享推文 1000 余篇，各平台阅读量逾 3000 万人次。



[郭士祺](#)，上海交通大学安泰经济与管理学院副教授。瑞士日内瓦国际关系与发展高等研究院博士，西南财经大学经济学硕士、四川大学金融学学士，曾在哈佛大学和联合国大学世界发展经济研究所访学。主要从事发展经济学、应用微观经济学领域研究。研究成果见诸《经济研究》(3 篇)、The Economic Journal ([EJ](#))、Journal of Development Economics ([JDE](#)，2 篇)、Games and Economic Behavior ([GEB](#))、Journal of Law and Economics ([JLE](#))、American Journal of Agricultural Economics ([AJAE](#))、Journal of Comparative Economics ([JCE](#)，2 篇)、Journal of Population Economics 等期刊。

C. 课程特色

这次寒假班，我们期望帮大家把三件事做扎实：把数据和代码跑顺、把模型和假设讲清、将证据与写作串联成一条能让审稿人信服的逻辑链条。课程的设计思路是：先把基本功夯牢，再进入高阶方法，最后用顶刊论文拆解把「研究设计的思考方式」练出来。

初级班 (扩充为 4 天)：前两天系统讲 Stata 数据处理、编程和可视化，同时示范如何借助 AI 工具把实证流程做得更高效、更少踩坑；后两天回到线性回归、交互项和面板模型，讲原理也讲用法，并通过两篇论文的复现，把“从数据到回归到解释”的完整流程走一遍。除此之外，我还准备了 3 篇论文的复现数据、代码和中文精要，大家可以用“精读 + 复现”的方式把课堂内容吃透，这些材料也很适合作为模板，迁移到自己的研究里。

高级班 (6 个专题)：很多同学不是不会回归，而是卡在更常见、也更棘手的几类问题上：非线性关系怎么处理、动态效应怎么识别、长期均衡关系怎么刻画、样本选择和内生性怎么说清楚。高级班就按这四类痛点拆成六个专题，每个专题都精选 2-3 篇带复现代码的论文，带大家看清楚：同一种方法在不同场景下怎么变形、怎么落地、怎么把关键假设和证据写得更稳。

论文班 (全新内容)：这次论文班是我们最想强调的部分。我们邀请到上海交通大学郭士祺副教授主讲。郭老师会带大家拆解 4-5 篇顶刊论文，其中有 3 篇出自他本人之手。论文班的目标也很明确：不是“听完觉得懂了”，而是让大家学会一套设计因果识别策略的思路和框架，能在研究设计和论文写作中把最核心的问题讲清楚：识别威胁源自哪里、如何进入估计并扭曲解释；你的识别策略依赖哪些关键条件；哪些证据能真正支撑这些条件；哪些不确定性需要在论文里坦诚而规范地呈现。学完之后，你会更清楚自己论文该怎么搭结构、怎么补证据、怎么回应质疑。

1. 初级班

- **时间**：2026 年 1 月 28-29 日；1 月 31 日-2 月 1 日 (4 天)
- **方式**：网络直播 + 30 天回放
- **授课嘉宾**：连玉君 (中山大学)
- **授课安排**
 - **授课方式**：幻灯片+Stata17 实操演示，全程电子板书+Stata 演示截图，课后以 PDF 形式分享给学员
 - **授课时间**：上午 9:00-12:00，下午 14:30-17:30 (17:30-18:00 答疑)。
 - **全程答疑**：由 10 位经验丰富的同学组成的助教团队会在课程群中全程答疑，并对答疑接龙文档进行详细的记录和分类，公布于 [课程主页](#)。
- **课程详情**：<https://www.lianxh.cn/PX.html>
 - **板书和答疑**：<https://gitee.com/arlionn/PX/wikis>
 - **PDF 课纲**：<https://KC.lianxh.cn/PX.pdf>
 - **预读资料**：[点击下载参考文献](#)
- **报名链接**：<https://www.wjx.top/vm/weMgRkL.aspx#>

1.1 课程导引

实证分析中，最伤神和耗时的事情莫过于研究设计和数据处理。我教过几千名学生，不少同学会在发了小论文或顺利毕业时发来信息，分享喜悦之情。闲聊中，我最常听到感慨是：基础打不好，很痛苦！自己之前总想着尽快上手，早点做出结果，但最终会发现，若没有扎实的基础，以及对计量和 Stata 整体架构的认识，后续的学习进度总感觉力不从心，进展缓慢。

本次初级班扩充为 4 天，以便「慢火细熬」，带着大家 (尤其是初学者) 把基本概念、基本方法和基本工具讲透。

在**内容安排**上，基本上遵循了由浅入深，循序渐进的原则。

A1-A2 两讲 依序介绍 Stata 的基本用法、数据处理、程序编写和实证分析的基本流程。学习这些内容无需太多的计量经济学基础，但对于提高实证分析能力和分析效率，大有裨益。本讲中，我会以两篇文章为实例，说明 Stata 的基本语法结构，并对数据处理过程中的关键问题进行介绍，如离群值的处理、文字变量的处理等。就我个人的经验而言，数据处理能力的高低直接决定实证分析的效率，而对于离群值等问题的处理是否妥善会直接影响全文结果的稳健性，是多数人不够重视但却至关重要的问题。此前有不少学完了高级班的同学又回炉初级班，便是感悟到了这一点。

A3 讲介绍实证分析中的各种可视化呈现方法。学习本讲的目的有二：其一，在实证研究过程中，借助各种密度函数图、散点图、柱状图，可以让我们对变量的分布、关系有更为直观的理解，有助于加深对隐藏在数据背后的规律，长期积累下来会逐渐形成直觉；其二，目前主流期刊越来越强调结果呈现的可读性和直观性，各种可视化工具也应运而生，最为典型的是，多期 DID、RDD、Bunching、合成控制法等因果推断方法，以及稳健性检验、敏感性分析、安慰剂检验等都会借助图形来直观呈现高密度信息。

A4-A5 介绍文献中使用频率最高的线性回归模型，包括 OLS 的原理、结果的解释，以及虚拟变量和交互项的使用等。对于这些内容的深刻理解和熟练掌握，构成了后续，多种主流实证模型的基础，例如，目前文献中广泛使用的固定效应模型 (FE)，倍分法 (DID)，断点回归设计 (RDD) 等方法，本质上就是在传统的线性模型基础上，增加一些虚拟变量或交互项，配合巧妙的研究设计，来实现对不可观测的个体效应的控制，以及对政策效应的估计。

很多人会觉得 OLS 很简单，但 Top 期刊中使用最多的仍然是 OLS，如何合理的构建模型、解释结果便成为实证分析中必须掌握的。我精选了大家经常面临的几个专题并结合论文进行讲解，包括：虚拟变量的使用、交互项的使用和解释、分组回归的合理设定和假设检验。我会重点强调对条件期望函数和 FWL 定理的解读，这构成了理解因果模型、面板模型以及机器学习中多种方法的基础。首经贸的一个博士生发信息给我：「连老师，我发现只要把你初级里面的虚拟变量相关的知识完全掌握，很多复杂的方法就都好理解了，甚至可以自己解决问题。」，我的回复是：「那看来你是把相关的东西基本搞明白了，我每次上初级班的时候会花很多时间讲虚拟变量和交互项，这构成了双重差分、断点回归、时间中断分析、面板数据模型等一系列模型的重要基础。」

A6 介绍固定效应模型 (FE)，涉及基本的 FE，TWFE 模型和进阶的高维固定效应模型、交互固定效应等内容。在讲解这些模型的基本思想和估计方法的过程中，我会将重点放在模型含义、使用场景和条件上来。例如，对于同一笔数据而言，何时采用 OLS 进行估计，何时采用 FE 估计？不同的方法之间有何差异和关联？结果背后的经济含义如何解读？掌握这些方法有助于大家合理控制内生性问题，以便得到更为可信的结论。更为重要的是，本讲介绍的**高维固定效应**、**交互固定效应**是理解各类面板因果推断模型 (如 DID，回归控制法等) 的基础，也是进一步学习动态面板、面板 VAR 等模型的基础。

A7 用一整天的时间，拆解 2 篇论文。它们基本上涵盖了前几讲中最核心的内容。我希望大家在学习各种计量方法和 Stata 工具的同时，能够逐渐形成「研究设计」的意识，养成良好的「编写代码习惯」，并能在动手摸数据的过程中形成对数据和变量的直觉。

1.2 专题介绍 (初级班)

A1. Stata 简介和数据处理

- 实证分析的基本流程和编程习惯
- 数据的导入/导出、转换及合并
- 离群值、文字变量的处理
- 基本统计分析
- do 文档、帮助文件和外部命令
- 如何借助 AI 工具提高编程效率？
 - Jupyter Notebook 和 Copilot 的使用
 - 如何编写高质量的提示词 (prompt)？
- 实证分析文档结构演示
 - Wang, Z., Zhang, T., Ren, X., & Shi, Y. (2024). AI adoption rate and corporate green innovation efficiency: Evidence from Chinese energy companies. *Energy Economics*, 132, 107499. [Link](#), [PDF](#), [Google](#), [cited](#), [Replication](#). 该文只有一个 `.do` 文档和一个 `.dta` 文档，可以一键执行到底。
 - Cepni, O., Şensoy, A., & Yılmaz, M. H. (2024). Climate change exposure and cost of equity. *Energy Economics*, 130, 107288. [Link](#), [PDF](#), [Google](#). [Replication](#). 该文的复现文档包含多个文件夹，分类放置了 `.do` 文件，数据文件和输出结果。数据处理过程涉及多个数据文件的合并。

A2. Stata 程序基础

- Stata 程序的基本架构和调用
- 局域暂元和全局暂元 (local, global)
- 控制语句 (条件语句、循环语句)
- 利用返回值进行结果输出和可视化
- Stata 中的各类函数

A3. 实证分析可视化

- 为什么要可视化？
- Stata 绘图命令的架构
- 直方图与密度函数图： `histogram` , `kdensity` , `biplot`
- 分仓散点图： `binscatter` , `binscatter2`
- 系数及系数差异的可视化呈现： `coefplot`
- 调节效应、倒 U 型关系及边际效应的可视化
- 面板数据、多个控制变量、高维固定效应模型的可视化
- 长期与短期关系的可视化
- 范文：2 篇

A4. 线性回归分析

- 条件期望函数：OLS, MLE 与 MM 的关系
- 线性概率模型
- OLS 估计和系数含义
- FWL 定理 (Frisch-Waugh-Lovell)
- 假设检验和统计推断
- 稳健性标准误：Bootstrap、Jackknife、聚类调整
- 结果输出与呈现

A5. 模型设定和解释

- 控制变量：选取、含义、可视化
- 变量缩放
- 取对数：弹性与半弹性
- 虚拟变量与固定效应
- 交互项、平方项、高阶项与调节效应
- 因子变量与边际效应分析
- DID, RDD 与 RKD

A6. 面板数据模型

- 何谓个体效应？FE v.s. RE
- 高维固定效应模型
- 异方差和序列相关 (Bootstrap、Cluster 调整标准误)
- 面板模型中的非时变变量和宏观变量如何分析？
- 实证分析中的常见问题
- 参考文献
 - Cameron, C. A., D. L. Miller, 2015, A practitioner's guide to cluster-robust inference, *Journal of Human Resources*, 50 (2): 317-372. [Link](#) , [PDF](#)
 - Correia, S. 2016. `reghdfe` , Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator. Working Paper. [PDF](#) , [Examples](#)
 - A7 中讲解的论文会综合应用上述模型。

A7. 论文复现

本专题讲解 2 篇论文。目的有二：一是把前面几讲学习的方法应用到实战中，二是通过论文复现来了解实证分析的整体流程和细节。

本着由浅入深的逻辑，我先介绍 Wang et al. (2024)，该文研究了 AI 技术使用率与企业绿色创新效率的关系。复现文档只有一个 dofile 和一个处理好的 .dta 数据文件。虽然 dofile 只有 200 多行，但涵盖了数据处理、描述性统计分析、回归分析、稳健性检验等多个环节，适合初学者建立对实证分析全过程的认识。

本讲的第二篇论文是 Cepni et al. (2024)，该文研究了气候变化暴露对企业股权资本成本的影响。复现文档包含多个文件夹，分类放置了 dofile，数据文件和输出结果。数据处理过程涉及多个数据文件的合并、变量生成、转换、缩尾处理等；实证分析部分则使用了高维固定效应模型、交互项、DiD 等分析方法；稳健性检验中，作者做了 13 种测试。这篇论文中没有使用非常复杂的计量模型，但涵盖了实证分析中的多个重要环节，适合初学者学习和模仿。

- Wang, Z., Zhang, T., Ren, X., & Shi, Y. (2024). AI adoption rate and corporate green innovation efficiency: Evidence from Chinese energy companies. *Energy Economics*, 132, 107499. [Link](#), [Google](#), [cited](#), [Replication](#)
- Cepni, O., Şensoy, A., & Yılmaz, M. H. (2024). Climate change exposure and cost of equity. *Energy Economics*, 130, 107288. [Link](#), [PDF](#), [Google](#), [Replication](#).

A8. 自行研读-提供复现数据和代码和中文精要

考虑到学员们的研究背景非常多元化，我提供 3 篇论文供大家自行研读。这些论文都是我此前在课堂上讲授过的。有些侧重数据处理和可视化 (如 Sherman and Tookes, 2022, JF)，有些则体现出严谨而又巧妙地研究设计，综合运用多种计量方法展开论证 (如 Akcigit et al., 2022, QJE)。

这些论文的代码都经过了 my 的反复调试和修改，添加了详细的注释。每篇论文都配有一份「中文精要.pdf」，里面总结了论文的核心内容、方法和结论，帮助大家理解论文的主要贡献和方法细节。对于一些论文中没有详细介绍的估计和检验方法，我会补充相关的理论介绍和 Stata 实操建议，以便大家更好地理解和应用。

大家在学习和复习过程中 (课程提供 30 天的回放)，可以随时在课程群中提问，我和助教团队会尽力解答大家的问题，帮助大家更好地掌握这些论文中的方法和技巧。

- Sherman M G, Tookes H E. Female representation in the academic finance profession. *Journal of Finance*, 2022, 77(1): 317-365. [Link](#), [cited](#), [PDF](#), [Replication](#)
- 该文研究了金融学术圈中的性别失衡现象。2009-2017 年，美国排名前 100 的商学院中的金融教师中，女性仅占 16.0%。性别失衡表现在几个方面：其一，在控制了研究能力后，更多的女性在排名较低的机构中任职，晋升为正教授的可行性相对较低，伴以薪酬较低。其二，女性发表的论文数量较少，但质量上不存在差异。其三，女性多与同性合作，表明她们的社会关系网较小。时间序列数据表明，上述性别差距正在缩小。
- 该文没有使用任何复杂的回归方法，仅使用了固定效应模型。但在统计分析和结果可视化方面做了很多工作，是 Stata 入门学习的绝佳范本。
- 方法：
 - 各种统计分析，列表和图形呈现： `egen` , `foreach`
 - OLS, 高维固定效应模型, 交互项, 因子变量： `reghdfe`
 - 结果可视化： `coefplot`
 - 结果输出： `estadd` , `estout` , `esttab`
- Akcigit, U., J. Grigsby, T. Nicholas, S. Stantcheva, 2022, Taxation and innovation in the twentieth century, *The Quarterly Journal of Economics*, 137 (1): 329-385. [Link](#), [PDF](#), [Appendix](#), [cited](#), [Replication](#)
 - 简介：文章研究了美国公司税和个人税对创新的影响。作者将发明人数据库、公司税率数据库，以及州级个人所得税和其他经济数据关联起来，从宏观和微观两个层面估计了税收对创新 (数量、质量、发生地等) 的影响。文中采用了多种识别策略，得到了非常一致的结论：(1) 高税率对创新的数量和发生地具有负面影响，但不会影响平均创新质量；(2) 州级层面的「税收-创新产出」弹性很大；(3) 公司税主要影响受雇发明者 (相对于自由职业者) 的创新产出和跨州流动性；而个人所得税则会对整体创新数量和发明人的流动性产生影响。
 - 方法：

- 高维固定效应、长差分、交互固定效应
 - 实证结果可视化：分仓散点图、长期效应
 - 交互项、合成控制法、事件研究法
- Bao, L., Huang, D., & Lin, C. (2025). Can Artificial Intelligence Improve Gender Equality Evidence from a Natural Experiment. *Management Science*. [Link](#), [PDF](#), [PDF-wp](#), [Appendix](#), [Google](#), [Replication](#).
 - **简介：**该文研究的问题是：当教师队伍的性别结构与观念短期难改时，能否借助被设计为“性别中性”的 AI 教学，改善学习效果并缩小既有性别差距？作者利用一次疫情隔离触发的自然实验：2021 年 1 月当地疫情反弹导致 16/36 名教师被隔离，机构以 AI 复盘替代部分人类教师，形成处理组与对照组，并追回冲击前后两个学期的训练数据(2 月春节无训练)。结果显示，AI 训练显著提高胜率与落子质量，并减少错误、重大错误及错误幅度；同时，女生的进步更快，干预后约 5 个月性别差距明显收敛乃至被消除。
 - 方法：
 - 双重差分 (DID)、三重差分 (DDD)
 - 内生性应对：样本选择偏误、遗漏变量敏感性分析、纠偏多变量联合检验
 - 稳健性检验、异质性分析部分都做的非常细致
 - 因果识别论证思路严谨、复现文档结构清晰



2. 高级班

- **时间**：2026 年 2 月 3-5 日 (3 天)
- **方式**：网络直播 + 30 天回放
- **授课嘉宾**：连玉君 (中山大学)
- **授课安排**
 - **授课方式**：幻灯片+Stata17 实操演示，全程电子板书+Stata 演示截图，课后以 PDF 形式分享给学员
 - **授课时间**：上午 9:00-12:00，下午 14:30-17:30 (17:30-18:00 答疑)。
 - **全程答疑**：由 10 位经验丰富的同学组成的助教团队会在课程群中全程答疑，并对答疑接龙文档进行详细的记录和分类，公布于 [课程主页](#)。
- **课程详情**：<https://www.lianxh.cn/PX.html>
 - **板书和答疑**：<https://gitee.com/arlionn/PX/wikis>
 - **PDF 课纲**：<https://KC.lianxh.cn/PX.pdf>
 - **预读资料**：[点击下载参考文献](#)
- **报名链接**：<https://www.wjx.top/vm/weMgRkL.aspx#>

2.1 课程导引

高级班 是初级班的深入，涵盖目前主流分析方法和研究设计框架。高级班采用「庖丁解牛」的方式讲解目前 Top 期刊中使用的计量方法和模型，突出「研究设计」和方法的「应用背景和条件」，并提供一些扩展思路。

高级班共包括六讲：

- **B1** 是结构变化和异质性分析专题，介绍面板门限模型和基于非参的函数系数模型，帮助大家理解和分析变量间的非线性关系和异质性效应。
- **B2-B4** 是动态关系和长期关系专题，分别介绍动态面板模型、面板 VAR 模型和面板 ARDL 模型。它们构成了分析变量间动态关系和长期均衡关系的主要工具，在很多论文中也会采用这些方法来应对内生性问题。
- **B5** 介绍 Heckman 选择模型，帮助大家理解和应对样本选择偏误和自选择偏误问题。这在多数实证分析中都是不可避免的问题。
- **B6** 介绍交互固定效应模型和因子模型，以及由此延伸出的回归控制法 (RCM)、合成控制法 (SCM) 和合成 DID (SDID) 等方法。这些方法具有高度的灵活性和可设计性，是政策评价的有力工具。

课程题要：

- 结构变化和异质性专题 (B1)
 - 面板门槛模型
 - 面板变系数模型
- 动态关系和长期关系专题 (B2-B4)
 - 动态面板模型
 - 面板 VAR 模型
 - 面板 ARDL 模型
- 样本选择和自选择偏误 (B5)
 - Tobit 和双栏模型
 - Heckman 选择模型

- 面板和异质性 Heckman 模型
- 交互固定效应和因子模型 (B6)
 - 回归控制法
 - 合成控制法
 - 合成 DID 方法

2.2 专题介绍 (高级班)

B1. 结构变化和异质性分析

结构变化现象广泛存在于经济研究中，主要体现为变量之间的非线性关系，如变量 x 对变量 y 的影响随 z (经济环境或政策力度等变量) 的变化而不同。当这种边际影响具有非线性或结构跳跃特征时，如「J 型，S 型或倒 U 型」关系时，传统方法 (如交互项或分组回归) 不再适用。

实证分析中不考虑非线性关系会导致模型误设偏误和遗漏变量偏误，二者都是内生性问题的来源。因此，在论文考虑非线性关系是「内生性问题」小节中常用的应对方法。此外，在论文的「进一步分析」或「机制分析中」部分，也可以考虑使用这些非线性模型来刻画变量间的复杂关系，从而深入挖掘隐藏在数据背后的经济机制。

本专题介绍面板门限模型 (Panel Threshold Model, PTM) 和基于非参的函数系数模型 (Functional Coefficient Model, FCM) 两种计量方法。前者通过识别门限值来刻画变量间的非线性关系，后者则通过平滑函数来捕捉变量间的连续变化关系，也是分析 x 对 y 之间异质性影响的有力工具。

为了帮助大家理解这些模型的经济含义和应用场景，我会结合以下几篇最新发表的论文来进行讲解：

- **静态和动态面板门限模型** (Lee et al., 2025)：识别门限效应，揭示绿色数字金融在不同气候风险水平下如何缓解能源贫困。
- **非参数系数函数模型** (Du et al., 2021; Du et al., 2024)：灵活捕捉环境规制、产业集聚对绿色创新和生产率的结构性变化。本次专题讲解的代码适用于非平衡面板中，大大扩展了该模型的应用范围。

在讲解上述模型的基本思想、估计和检验方法的同时，我会梳理这些模型中核心参数的经济含义，进而讨论模型的适用场景，激发大家的迁移灵感。

在应用上述非线性模型时，最大的挑战不是代码实现，而是如何准确地识别和理解非线性特征的产生机制：从众多备选变量中，确定几个关键的 z 变量，并论述它们的经济含义，讲个好故事。这直接决定了研究的深度与可发表性。显然，这并非易事。

论文复现：

- Lee, C.-C., Li, M., Li, X., & Song, H. (2025). More green digital finance with less energy poverty? The key role of climate risk. *Energy Economics*, 141, 108144. [Link](#), [PDF](#), [Google](#). [Replication](#)
- Du, K., Cheng, Y., & Yao, X. (2021). Environmental regulation, green technology innovation, and industrial structure upgrading: The road to the green transformation of Chinese cities. *Energy Economics*, 98, 105247. [Link](#), [PDF](#), [Replication](#), [Google](#).
- Du, C., Cao, Y., Ling, Y., Jin, Z., Wang, S., & Wang, D. (2024). Does manufacturing agglomeration promote green productivity growth in China? Fresh evidence from partially linear functional-coefficient models. *Energy Economics*, 131, 107352. [Link](#), [PDF](#), [Google](#). [Replication](#)

B2. 动态面板模型

多数经济行为的调整都会因为「调整成本」的存在而表现出粘性，即只能实现部分调整；同时，即使基于理性预期模型进行决策，但由于信息存在滞后，变量在时序上就会表现出较强的相关性。这些都构成了动态关系的理论基础，也使得动态面板模型广泛出现于经济增长、公司金融、国际贸易、劳动经济学等领域。本讲介绍该模型的估计和检验方法，尤其是实际分析中的几个主要陷阱。

- 一阶差分 GMM 估计 (FD-GMM)
- 系统 GMM 估计 (SYS-GMM)
- 序列相关检验
- 过度识别检验 (Sargan 检验)
- 模型设定常见问题(弱工具变量问题)
- 应用场景讨论：应对内生性、刻画动态关系、估计长期影响系数

论文复现和避坑指南：

- Cepni, O., Şensoy, A., & Yılmaz, M. H. (2024). Climate change exposure and cost of equity. *Energy Economics*, 130, 107288. [Link](#), [PDF](#), [Google](#). [Replication](#)
- Flannery, M. J., & Rangan, K. P. (2006). Partial adjustment toward target capital structures. *Journal of Financial Economics*, 79(3), 469–506. [Link](#), [PDF](#), [Google](#). [cited-3500 次](#).
- Acemoglu, D., Johnson, S., Querubín, P., & Robinson, J. A. (2008). When Does Policy Reform Work?: The Case of Central Bank Independence. *Brookings Papers on Economic Activity*, 2008(1), 351–418. [Link](#), [PDF](#), [Google](#).
- Zhang, Y., et al. (2025). Would geopolitical risks be the new driver of the energy transition? An empirical study on renewable energy technology innovation. *Energy Economics*, 141, 108100. [Link](#), [PDF](#), [Google](#).
- Wang, F., & Qu, M. (2024). The interaction of income inequality and energy poverty on global carbon emissions: A dynamic panel data approach. *Energy Economics*, 140, 108027. [Link](#), [PDF](#), [Google](#), [Replication](#).

B3. 面板 VAR 模型

面板 VAR 模型可以视为多变量动态面板模型，多用于估计和检验几个内生变量的动态关系。基于 IRF 和 FEVD，我们也可以进行预测。该模型在经济增长、能源、财政、创新等领域有广泛应用。

- VAR 模型简介
- 冲击反应函数 (IRF)
- 方差分解 (FEVD)
- 面板 Granger 检验
- 允许外生变量的 PVAR 模型
- 应用实例 (介绍 2 篇论文)

参考文献：

- Abrigo, M. R. M., & Love, I. (2016). Estimation of Panel Vector Autoregression in Stata. *The Stata Journal*, 16(3), 778–804. [Link](#), [PDF](#), [Google](#), [cited](#).
- Goodell, J. W., McGee, R. J., & McGroarty, F. (2020). Election uncertainty, economic policy uncertainty and financial market uncertainty: A prediction market analysis. *Journal of Banking & Finance*, 110, 105684. [Link](#), [PDF](#), [Google](#).
- Lee, J., Naranjo, A., & Velioglu, G. (2018). When do CDS spreads lead? Rating events, private entities, and firm-specific information flows. *Journal of Financial Economics*, 130(3), 556–578. [Link](#), [PDF](#), [Google](#).

论文复现：

- Acheampong, A. O., 2018, Economic growth, co2 emissions and energy consumption: What causes what and where?, *Energy Economics*, 74: 677-692. [Link](#), [PDF](#), [Replication](#), [cited](#)
- Donadelli, M., Grüning, P., Jüppner, M., & Kizys, R. (2021). Global temperature, R&D expenditure, and growth. *Energy Economics*, 104, 105608. [Link](#), [PDF](#), [Google](#), [Replication](#).

B4. 面板 ARDL 模型：长期关系

面板 ARDL 模型全称「面板自回归分布滞后模型」，主要用于估计变量之间的长期关系。在时间序列分析中，该模型主要用于分析具有协整关系的非平稳序列之间的长期和短期关系。在面板数据中，我们可以更好地控制各类固定效应、考虑空间相关以及异质性特征，以便分析一项政策或某个变化缓慢的变量 (如气候、制度) 对经济增长、创新、贸易等结果变量的影响。本专题介绍面板 ARDL 模型的主要设定形式，以及该模型在经济和金融领域的应用场景。我会以部分调整模型和理性预期模型为基础，介绍如何基于特定的故事背景设定长期目标函数和调整成本函数，以增强模型的可解释性，为机制分析和异质性分析提供了理论依据。

- 时间序列简介：AR 与 MA 过程
- ARDL 模型：AutoRegression Distributed Lag
- 理论基础：部分调整模型与理性预期模型
- 长期乘数与短期效应
- 模型扩展与应用场景讨论

论文复现：

- Dell, M., B. F. Jones, B. A. Olken, 2012, Temperature shocks and economic growth: Evidence from the last half century, **American Economic Journal: Macroeconomics**, 4 (3): 66-95. [Link](#), [PDF](#), [Replication](#)
- Ditzen, J. 2021. "Estimating long-run effects and the exponent of cross-sectional dependence: An update to xtdcce2". **Stata Journal**, 21 (3): 687-707. [Link](#), [PDF](#), [PDF2](#).
- Kahn, M. E., K. Mohaddes, R. N. C. Ng, M. H. Pesaran, M. Raissi, J.-C. Yang, 2021, Long-term macroeconomic effects of climate change: A cross-country analysis, **Energy Economics**, 104: 105624. [Link](#), [PDF1](#), [PDF2](#), [Replication](#), [Cited](#).
- Akcigit et al., 2022, Taxation and innovation in the twentieth century, **Quarterly Journal of Economics**, 137 (1): 329-385. [Link](#), [PDF](#), [Replication](#)

B5. 样本选择与自选择

样本选择偏误和自选择偏误是实证研究中常见的内生性问题，本质上是遗漏变量问题。样本选择偏误通常出现在因变量被截断或截堵的情况下，例如工资数据中低于某一水平的观测值未被记录，或贷款违约率数据中仅包含已获批准的贷款记录。自选择偏误则源于个体基于其特征自主选择进入样本，例如教育回报率研究中，受教育程度较高的个体可能更倾向于参与调查。在论文的「内生性问题」部分，Heckman 选择模型经常与 IV 和 DID 方法结合使用，以应对不同类型的内生性问题。

- 截断、截堵与选择性偏差
- Tobit 模型
- Heckman 选择模型
- 面板和异质性 Heckman 模型

论文复现：

- Allen, J., Clark, R., Houde, J.-F., Li, S., & Trubnikova, A. (2025). The Role of Intermediaries in Selection Markets: Evidence from Mortgage Lending. **The Review of Financial Studies**, 38(11), 3284–3328. [Link](#), [PDF](#), [PDF-wp](#), [Google](#), [Replication](#).
- Ding, J., Li, L., & Zhao, J. (2024). How does fintech prompt corporations toward ESG sustainable development? Evidence from China. **Energy Economics**, 131, 107387. [Link](#), [PDF](#), [Google](#). [cited](#), [Appendix](#), [Replication](#).

B6. 交互固定效应和因子模型

交互固定效应模型和因子模型在高维数据分析和因果推断中扮演着尤为重要的角色。以面板数据 (二维数据) 为例，归属于同一个省份或行业的公司会受到一些共同因素 (common factors) 的影响，致使它们的行为表现出一定的相关性。在模型设定和统计推断中，若不考虑这些共同因素的影响，会导致遗漏变量偏误，或有偏的标准误。在因果推断中，「反事实」的构造和估计本质上是一个预测问题。除了随机森林、人工神经网络等方法外，因子模型和交互

固定效应模型 (IFE) 虽然简单, 但却具有很好的预测和外推能力, 使其广泛应用于合成控制法、DID、RDD、回归控制法等模型。合成 DID, **SDID** 以及 [Arkhangelsky](#) and Samkov (2024) 新近提出的「次序 SDID, **Seq-SDID**」都是以 IFE 为基础的。

本讲从因子模型入手, 介绍交互效应模型的设定、估计和检验方法, 以及它们在因果推断中的典型应用。

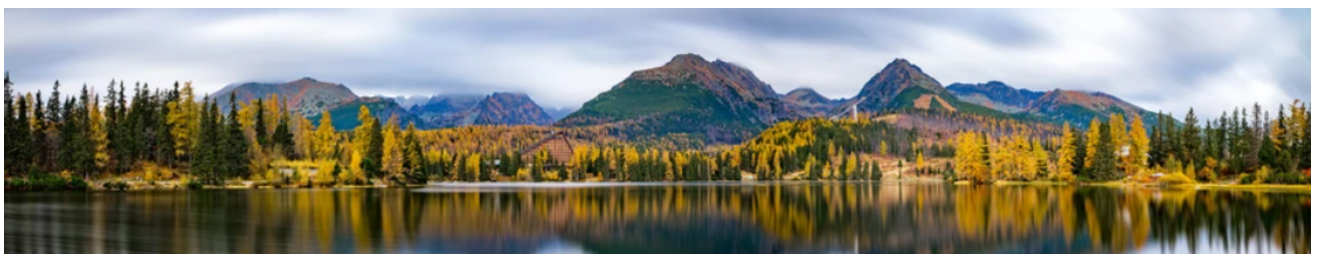
- 共同因子和因子载荷
- 面板交互固定效应模型 (IFE) 的设定和估计方法
- 包含宏观变量的高维固定效应模型 (混频数据问题)
- 应用 1: **回归控制法**
- 应用 2: **广义合成控制法**
- 应用 3: **合成 DID (SDID)**

参考文献：

- Bai, J. 2009. Panel Data Models With Interactive Fixed Effects. (2009). *Econometrica*, 77(4), 1229–1279. [Link](#), [PDF](#), [Google](#).
- Hsiao, C., H. S. Ching, S. K. Wan, 2012, A panel data approach for program evaluation: Measuring the benefits of political and economic integration of hong kong with mainland china, *Journal of Applied Econometrics*, 27 (5): 705–740. [Link](#), [PDF](#)
- Arkhangelsky D, Athey S, Hirshberg D A, et al. 2021, Synthetic difference-in-differences[J]. *American Economic Review*, 111(12): 4088–4118. [Link](#), [PDF](#), [Replicate](#), [Github](#), [Google](#).
- Clarke, D., Pailańir, D., Athey, S., & Imbens, G. (2024). On synthetic difference-in-differences and related estimation methods in Stata. *The Stata Journal*, 24(4), 557–598. [Link](#), [PDF](#), [PDF2](#), [Google](#).
- Wager, Stefan. 2024, Causal Inference: A Statistical Learning Approach. [PDF](#), [Chap 13](#)
- Yan, G., & Chen, Q. (2022). rcm: A command for the regression control method. *The Stata Journal*, 22(4), 842–883. [Link](#), [PDF](#), [Google](#).

复现论文：

- Bai, D., Hu, J., Irfan, M., & Hu, M. (2023). Unleashing the impact of ecological civilization pilot policies on green technology innovation: Evidence from a novel SC-DID model. *Energy Economics*, 125, 106813. [Link](#), [PDF](#), [Google](#).
- Zhang, Z., Wang, J., Feng, C., & Chen, X. (2023). Do pilot zones for green finance reform and innovation promote energy savings? Evidence from China. *Energy Economics*, 124, 106763. [Link](#), [PDF](#), [Google](#).
- Zhu, B., Huang, H., Liu, Y., & Zhang, N. (2025). Twin goals, one policy: Cost-effective collaborative abatement via new energy cities. *Energy Economics*, 151, 108941. [Link](#), [PDF](#), [Google](#). 使用 `sdid` 和 `ddml` 应对内生性问题
- Ko, D. G. (2025). Did the American Rescue Plan cause inflation? A synthetic control approach. *Economic Modelling*, 143, 106935. [Link](#), [PDF](#), [Google](#). [Replication](#)



3. 论文班

- **时间**：2026 年 2 月 7-9 日 (三天)
- **方式**：网络直播 + 30 天回放
- **授课嘉宾**：郭士祺 (上海交通大学)
- **授课安排**
 - **授课方式**：提供全套 Stata 程序、数据和核心论文复现资料 (开课前一周发送)。
 - **授课时间**：上午 9:00-12:00，下午 14:30-17:30 (17:30-18:00 答疑)。
 - **全程答疑**：由 10 位经验丰富的同学组成的助教团队会在课程群中全程答疑，并对答疑接龙文档进行详细的记录和分类，公布于 [课程主页](#)。
- **课程详情**：<https://www.lianxh.cn/PX.html>
 - **往期板书和答疑**：<https://gitee.com/arlionn/PX/wikis>
 - **PDF 课纲**：<https://KC.lianxh.cn/PX.pdf>
 - **预读资料**：[点击下载参考文献](#)
- **报名链接**：<https://www.wjx.top/vm/weMgRkL.aspx#>

论文班以“精讲论文”为主线，重点训练学员清晰阐述因果识别的核心问题：识别威胁源自何处、如何进入估计并扭曲结果解释；识别策略依赖哪些关键条件；哪些证据能够有效支撑；哪些不确定性需要在论文中诚实呈现。

课程将围绕 4 类在近年在政策评价中广泛使用，且具有可规范化复现性的识别路径展开：首先，从随机对照试验 (RCT) 的“理想实验”视角切入；继而讲解基于事件发生与调查访问独立性的准试验识别框架；随后讨论在复杂历史背景下如何利用队列差分 (Cohort-DID) 呈现研究证据，并通过地理工具变量 (IV) 构建更可靠的识别链；最后系统探讨交互式工具变量的潜在优势，以及如何撰写排他性假设并组织相关证据。

本课程适合已有一定实证研究经验、希望显著提升研究设计质量与论文深度的学员。课程不强调机械化的“命令清单”，而是通过对核心论文的系统拆解，帮助学员建立一套可迁移的论文写作与论证结构，包括：如何界定处理组与对照组、如何选择时间窗口与固定效应、如何逐条回应识别假设可能失效的渠道，以及如何将稳健性检验、安慰剂检验、敏感性分析与机制分析整合成一条清晰而严密的证据链。

我们期望，通过三天的学习，大家能初步掌握「识别框架→证据链→写作表达」一体化的研究设计方法。

主要方法与识别策略覆盖：

- 随机对照试验 (RCT) 及溢出效应的处理
- 基于事件冲击与调查访问时间独立性的准实验识别设计
- 队列差分 (Cohort-DID) 与地理工具变量 (IV) 构建复合识别框架
- 交互式 IV (外部冲击 × 暴露度) 的潜在优势及排他性假设的系统论证

学员背景与预备知识：

熟悉线性回归与固定效应的基本设定，理解双重差分 (DID) 与工具变量 (IV) 的基本思想；软件层面建议至少能用 Stata 或 R 复现基准回归与常见稳健性检验。

各专题详细介绍如下 (核心论文 [中文精要](#))：

C1. RCT 对因果识别的启示：论述内生性与溢出效应

讲授时长：3 小时

核心文献：

- Miguel, E., & Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1), 159–217. [Link](#), [PDF](#), [PDF2](#), [Google](#), [cited](#), [Replication Guide](#)

在因果识别中，难点往往不在于「跑出结果」，而在于把这三个核心问题讲清楚：研究究竟面临什么样的内生性问题(或识别威胁)；这些问题如何进入估计并扭曲因果解释；所采用的方法到底消除了哪一部分偏误、依赖哪些可验证的识别条件、仍留下哪些需坦诚披露的不确定性。如果这些问题说不清，论文就容易沦为「方法堆叠而论证空心」。

本讲从随机对照试验 (RCT) 的设计逻辑出发，将其视为「理想实验」，并把样本选择、遗漏变量等常见内生性问题，拆解为「理想实验」与「现实研究」之间的差距，再将这些差距映射回论文写作的关键环节：识别假设如何表达、处理组与对照组如何界定、如何区分设计层面的论证与统计层面的补充，以及如何规范呈现研究局限性，帮助大家把偏误识别转化为审稿人能够理解和接受的论证结构。

此处以 Miguel and Kremer (2004) 为精讲文献。该文以肯尼亚小学驱虫项目为例，展示了 RCT 在政策评估中的基本逻辑，并揭示当干预存在「组间溢出」效应时，研究者如何重写处理与对照的定义，重新界定要回答的因果问题，从而构建清晰的识别论证。对本课程而言，这篇论文的价值不在于「驱虫是否有效」的结论本身，而在于其提供了可迁移的研究与写作范式：把识别威胁说清楚，把因果问题界定清楚，把证据链条搭建清楚。

我们将借助一个「看似简单」的驱虫实验，循序拆解 RCT 的设计逻辑与论文表达方式：

1. 如何从政策背景与实施细节出发，明确随机分配发生在哪个层级 (学校、年级、区域)，并据此界定处理组与对照组？
2. 在存在溢出效应时，对照组为何不再「纯净」？研究者应如何重构比较对象，使「对照」仍具有因果解释性？
3. 在完成上述界定后，应如何把研究设计转写为清晰的识别策略：估计式在比较什么、关键假设是什么、证据如何逐步支撑这些假设？

本讲也将为后续章节提供一个统一参照系。之后无论讲 Cohort-DID、事件 × 调查的准自然实验设计，还是地理与交互式工具变量，我们都沿用同样的追问逻辑：如果存在一个理想的随机化实验，你会如何设计处理、对照与比较对象？现实数据和识别设计与这一「理想实验」差在哪里？这些差距哪些能够通过研究设计与证据补救，哪些需要通过限定外推与解释范围来审慎处理？

本讲的核心内容包括：

- RCT 的基本设计与因果识别逻辑
 - 围绕具体研究问题，对照「理想 RCT」与现实研究中的处理、对照与比较关系，定位内生性的来源
 - 将识别假设落到可写、可检的表述：哪些环节依赖随机分配，哪些依赖实施过程的可控性，哪些必须由证据支撑
- RCT 中典型问题及其对因果识别的启示
 - **组间溢出效应**
 - 处理可能通过同伴、地理邻近、学校网络影响未处理个体，使对照组不再反映「未受干预」状态，从而造成效应的低估或高估
 - 这一问题在观测数据研究中同样普遍，关键在于能否识别「可能受影响的范围」，并据此重构比较对象

◦ 处理组与对照组的基期不平衡

- 即使存在随机分配，有限样本下仍可能出现重要特征的偶然差异，影响估计精度与论证说服力
- 重点不在于把检验结果「做出来」，而在于解释这些差异是否影响识别，并说明研究者采取了哪些可复核的处理方式

◦ 样本流失与缺失

- 实验中的样本流失往往并非随机，可能改变最终可解释的研究对象，引入新的选择偏误
- 观测数据中的缺失与幸存偏差更为普遍，需要明确哪些结论依赖「缺失机制不影响比较关系」这一类条件，并通过稳健性与诊断性证据加以处理

C2. 事件冲击与即时反应：利用调查时间随机性的准实验识别

讲授时长：6 小时

核心文献：

- [P1] Depetris-Chauvin, E., Durante, R., & Campante, F. (2020). Building Nations through Shared Experiences: Evidence from African Football. *American Economic Review*, 110(5), 1572–1602. [Link](#), [PDF](#), [PDF2](#), [Appendix](#), [cited](#), [Replication](#), [Google](#), [Correction 2024](#).
- [P2] Guo, S., & An, J. (2022). Does terrorism make people pessimistic? Evidence from a natural experiment. *Journal of Development Economics*, 155, 102817. [Link](#), [PDF](#), [Google](#).

本讲介绍的两篇都采用了近期被广泛使用的一种「事件冲击-时间交错识别策略」：在同一地区、同一轮调查中，利用「事件发生时间」与「调查访问时间」的交错，构建近似随机的实验组与对照组，实现因果识别。事实上，[P2] Guo and An (2022, JDE) 的设计很大程度上受 [P1] Depetris-Chauvin et al. (2020, AER) 的启发。通过学习这两篇论文，有助于掌握这一类基于事件冲击的准实验研究的设计与实施思路，并进一步探索将识别策略迁移到不同情景的可能性。

这类识别策略的优势在于，「识别步骤相对规范、证据链条可按部就班搭建」。换言之，只要事件具有可核验的发生时点与地点、调查访问时间记录足够精确，研究者就可以较为系统地完成研究设计：先界定时间窗口与分组规则，再用平衡性与不可操纵性证据支撑「近似随机」这一核心假设，随后给出对应的固定效应设定与稳健性检验，并在此基础上讨论机制与解释边界。因此，这类研究的关键更多在于“精心设计与规范论证”，而不是靠运气或灵感偶然碰到一个只能做一次的机会；在数据条件满足的前提下，普通研究者也可以通过训练与复现，把这一类识别策略熟练应用到自己的问题上。

本讲的核心内容包括：

A. 识别框架：事件发生时间 × 调查访问时间

- 在同一地区与同一调查轮次内，事件发生前被访问的个体作为对照组，事件发生后被访问的个体作为处理组。
- 识别依赖于局部条件：在选定时间窗口内，访问顺序与事件发生近似独立，且不被系统性操纵。

B. 基准设定与固定效应：把比较限制在“同一情境内部”

- 课堂以统一的基准式解释固定效应与识别逻辑的对应关系，并讨论“该控制什么、不该控制什么”。基准回归式为：

$$y_{irst} = \beta \cdot \mathbf{1}\{t \geq T_r\} + X'_{irst}\theta + \alpha_{r,s} + \delta_{t,s} + \varepsilon_{irst}$$

其中, t 为访问时间, T_r 为地区 r 的事件发生时间; X_{irst} 为控制变量向量。 $\alpha_{r,s}$ 用于吸收地区在同一调查轮次内的时间不变差异, $\delta_{t,s}$ 为调查日(或周)固定效应, 用于吸收同一轮调查内的共同时间冲击。

- 识别应来自“同一地区、同一轮调查中事件前后很短时间差”的对比, 而非跨地区或跨轮次的结构性差异。

C. 时间窗口与稳健性证据: 让“近似随机”更可信

- 说明窗口选择的权衡: 窗口越窄可比性越强, 但样本量下降; 窗口越宽混入其他变化的风险上升。
- 常用稳健性检验策略: 更换窗口长度、事件前伪断点检验、事件后效应衰减检验, 以及不同地理半径的敏感性检验。

D. 关键假设与诊断: 写作中必须正面回应的疑点

- 可比性: 事件前后样本在可观测特征上是否相似, 差异是否具有实质意义
- 不可操纵性: 访问时间是否可能受调查员安排、道路封锁或治安变化影响, 从而导致系统性“回避”或“补访”
- 信息扩散与外溢: 事件是否提前泄露或跨地区传播, 使对照组被部分暴露; 对应的处理方式与结论边界如何表述

E. 适用场景与方法迁移: 从复现到选题

本讲按如下思路帮助大家建立选题和研究设计的迁移路径:

- [P1] Depetris-Chauvin et al. (AER 2020) → 基于足球比赛构建准实验;
- [P2] Guo & An (JDE 2022) → 迁移到恐怖袭击事件;
- [P3] An, Guo & Jiang (JCE 2025) → 迁移到中国对非洲援助项目 (对比战争或恐袭冲击, 识别援助项目的“准事件属性”)
- 你有什么新想法?

扩展阅读:

- An, J., Guo, S., & Zhao, R. (2026). The leverage of terrorists on democratic regimes: Evidence from natural experiments in Sub-Saharan Africa. *Journal of Law and Economics*, Forthcoming. [Link](#), [PDF](#), [Google](#).
- An, J., Guo, S., & Jiang, H. (2025). Foreign-assisted infrastructure and local employment: Evidence from China's aid to Africa. *Journal of Comparative Economics*, 53(1), 118–138. [Link](#), [PDF](#), [Google](#).
- Balcells, L., & Torrats-Espinosa, G. (2018). Using a natural experiment to estimate the electoral consequences of terrorist attacks. *Proceedings of the National Academy of Sciences of the United States of America*, 115(42), 10624–10629. [Link](#), [PDF](#), [Appendix](#), [Google](#).

C3. 从冲突到教育差距: Cohort-DID 与地理 IV 的综合识别策略

讲授时长: 6 小时

核心文献: Guo, S. (郭士祺) (2020). The legacy effect of unexploded bombs on educational attainment in Laos. *Journal of Development Economics*, 147, 102527. [Link](#), [PDF](#), [Google](#). [背景介绍](#)

这篇论文是郭士祺老师独立完成的。论文从选题、识别策略、数据处理到机制论证, 都有很多值得分享和讨论的地方。通过对这篇论文的深入剖析, 大家不仅可以学习到具体的实证方法, 还能了解一个实证研究者在整个研究过程中所经历的思考和挑战。借助这篇论文, 本讲介绍如何在复杂历史背景下构建可信的识别策略, 并展示多源数据整合、度量构建与机制论证的系统方法。

队列双差分 (Cohort-DID) 的应用很普遍，初学者往往花费大量的时间在「选方法、选命令」上，却忽略了识别策略真正的起点：历史背景是否支持「外生冲击」的叙事；谁是被冲击的人群 (队列如何界定)；冲击强度为何具有空间差异 (地理维度为何合理)；以及利用数据和可视化图表来支撑这些假设。本讲选择 Guo (2020)，是因为它提供了一套「可拆解、可复用」的因果识别范本：将 Cohort-DID、工具变量 (IV)、稳健性、机制与写作组织成一个整体，层层递进地向读者展示「识别过程是如何被搭建起来的」。

本讲的核心内容包括：

A. 识别策略：Cohort-DID × 地理距离 IV 的高阶应用

- **Cohort-DID 的完整呈现**：论文同时展示描述性证据与回归分析，通过地图、分组趋势等可视化证据强化论证，帮助学员理解如何系统呈现“事前趋势”与“识别效应”，以及这些证据各自承担什么论证功能。
- **距离 IV 的构建与论证**：依托历史资料与轰炸地理分布，构建基于距离 (或地形相关距离) 的工具变量。重点不只讲“怎么做”，更讲“为什么可相信”，以及排他性假设可能面临哪些威胁、论文如何逐条回应。
- **近似外生工具变量的补充思路**：当排他性条件存在瑕疵时，如何用敏感性分析的方式，评估“需要多大程度的违背才会推翻结论”，并把这一类结果规范地写进论文的稳健性与局限性部分。
- **OLS 与 IV 的对比讨论**：通过比较普通最小二乘 (OLS) 与 IV 的结果差异，回到“内生性来自哪里”这一根本问题展开反思，帮助学员形成研究直觉：差异是由测量误差、遗漏变量、选择机制，还是由局部平均效应的解释边界所驱动。

B. 多源数据整合与复杂指标度量

- **多样化数据融合**：包括普查、微观家户调查、行政登记数据等，并结合地理信息系统 (GIS) 进行可视化展示，使数据背景与识别逻辑紧密结合。学员将学习“变量如何从不同来源拼接出来”，以及论文中必须交代哪些口径、匹配与可复现细节。
- **对 UXO 威胁的双重度量方法**：论文构建两套核心指标并贯穿全文检验一致性。此处的重点是：当核心概念“抽象且难以直接观测”时，如何用“多指标互证”的方式提升可信度，并把度量选择与识别假设明确对应起来。

C. 机制分析的系统框架

- **个案叙述与实证结果的互证**：结合访谈、田野信息和历史记载，使机制论证更具可信度与情境感。重点在于“材料如何服务于可检验假设”，而不是把背景当作装饰。
- **全面排除替代机制**：系统列出可能的解释路径，并采用多种策略逐一排除。课堂会强调「替代机制清单化」的写作方法：每条替代机制对应什么可观测预测、用哪类数据检验、证据不足时如何谨慎表述。
- **从背景到机制的创新构思**：展示如何从独特情境中提出具有文献价值的问题，并构造可识别、可验证的新机制，使「故事」与「识别」相互咬合。

D. 故事构建与论文写作启示

- **如何从独特背景切入文献**：论文基于战争遗留物这一特殊背景提出研究问题，展示“独特但具有普遍意义”的选题路径；同时示范如何在文中明确外推边界，避免“过度延伸和推广”。
- **从个案材料中生成研究问题**：学习如何从访谈、地方史料、田野笔记中提炼机制、构建假设，并进一步转化为可操作的变量与回归检验，这是许多学员在做原创研究时最缺的一环。

扩展阅读：

- Adhvaryu, A., Fenske, J., & Nyshadham, A. (2019). Early Life Circumstance and Adult Mental Health. *Journal of Political Economy*, 127(4), 1516–1549. [Link](#), [PDF](#), [Appendix](#), [Google](#).
- Chen, Y., Fan, Z., Gu, X., & Zhou, L.-A. (2020). Arrival of young talent: The send-down movement and rural education in China. *American Economic Review*, 110(11), 3393–3430. [Link](#), [PDF](#), [Google](#), [cited](#).

- Guo, S., Gao, N., & Liang, P. (2024). Winter is coming: Early-life experiences and politicians' decisions. *The Economic Journal*, 134(657), 295–321. [Link](#), [PDF](#), [Google](#).
- Huang, Z., & Dong, X. (2025). When the levee breaks: The impact of floods on educational outcomes in China. *Journal of Development Economics*, 174, 103450. [Link](#), [PDF](#), [Google](#).
- Le, D. T., Pham, T. M., & Polachek, S. (2022). The long-term health impact of Agent Orange: Evidence from the Vietnam War. *World Development*, 155, 105813. [Link](#), [PDF](#), [Google](#).
- Liang, P., & Guo, S. (2015). Social interaction, Internet access and stock market participation—An empirical study in China. *Journal of Comparative Economics*, 43(4), 883–901. [Link](#), [PDF](#), [Google](#).
- Nunn, N., & Wantchekon, L. (2011). The slave trade and the origins of mistrust in Africa. *American Economic Review*, 101(7), 3221–3252. [Link](#), [PDF](#), [Google](#).

C4. 交互式 IV (Shift-share IV): 从识别直觉到排他性假设的论证

讲授时长: 3 小时

核心文献:

- Nunn, N., & Qian, N. (2014). US Food Aid and Civil Conflict. *American Economic Review*, 104(6), 1630–1666. [Link](#), [PDF](#), [PDF2](#), [Appendix](#), [Google](#), [Replication](#), [cited-2284](#).

这篇论文研究了一个很有挑战的识别问题: 美国粮食援助是否会加剧受援国的内战冲突? 难点在于, 援助的配置往往会对冲突风险“做出反应”, 因此简单相关性很容易混入反向因果与遗漏变量。作者的核心做法是构造一个典型的交互式工具变量: 用「美国粮食供给侧波动」(外部冲击) 与「各国对美国粮食援助的结构性暴露度」(历史依赖度或既定份额) 相互作用, 生成随时间变化、且在国家之间强度不同的外生变动。由此, 识别主要来自同一外部冲击在不同国家被不同程度“放大”的差异化暴露, 而不是来自援助在冲突发生时的政策性再分配。

本讲将以这篇论文为主线, 深入讲解交互式 IV 的“关键难点”: 排他性假设到底在排除哪些替代通道? 审稿人最常从哪些方面提出质疑? 以及你应当如何按步骤准备一套可验证的证据链来应对这些质疑。学完这一讲, 你不仅能看懂 Nunn and Qian (2014) 的整套论证结构, 更能以「外部冲击 × 暴露度」这类工具变量为基础形成你的研究设计和论证结构。

本讲的核心内容包括:

A. 从 DID 到交互式 IV: 共同的“差异化暴露”与不同的识别要求

- 共同点: 都利用“同一外部变化在不同地区/部门被不同程度地放大”来产生识别所需的变异
- 不同点: DID 依赖“无处理时的平行趋势”, 交互式 IV 依赖“工具变量仅通过内生解释变量影响结果变量”的排他性条件
- 教学重点: 如何把“差异化暴露”的直觉转换为可写进论文的识别假设与证据结构, 而不是停留在口头叙事

B. 交互式 IV 的标准构造: $S_t \times E_r$ 的含义与识别直觉

- 典型形式: $Z_{rt} = S_t \times E_r$
 - S_t : 外部供给侧冲击或宏观波动 (时间维度)
 - E_r : 本地暴露度、依赖度或结构性权重 (横截面维度)
- 直观含义: 同一外部冲击对不同地区产生不同幅度的外生变化, 从而避免“政策反应型配置”带来的内生性
- 课堂会强调: 交互式结构本身并不自动保证外生性, 它只是提供一种“可被论证的外生变动来源”

C. 排他性假设如何写: 交互式 IV 最容易被质疑的三类通道

- 排他性条件的核心表述：在控制固定效应与必要控制变量后， $S_t \times E_r$ 只能通过 D_{rt} 影响 Y_{rt}
- 三类常见质疑(也是写作中需要逐条回应的“威胁清单”)
 - 暴露度 E_r 与地区长期趋势相关，交互项继承趋势差异(需要讨论趋势控制、分组趋势、替代暴露度量等)
 - 外部冲击 S_t 与其他宏观变化同步出现，交互项捕捉了共同冲击下的异质反应(需要讨论更细时间控制、机制分解、安慰剂冲击等)
 - 外部冲击可能通过多重渠道进入结果变量，导致“工具变量不只作用于一个处理变量”(需要区分机制、扩展控制、替代结果变量检验等)

D. 经验模型：第一阶段、第二阶段与固定效应的规范做法

- 基准的两阶段设定如下：

$$D_{rt} = \pi \cdot (S_t \times E_r) + \alpha_r + \lambda_t + W'_{rt} \gamma + u_{rt}$$

$$Y_{rt} = \beta \cdot \widehat{D}_{rt} + \alpha_r + \lambda_t + W'_{rt} \theta + \varepsilon_{rt}$$
- 解释要点
 - α_r 吸收地区不变差异， λ_t 吸收共同时间冲击，识别尽量来自“同一地区在外部冲击下被差异化放大”的部分
 - 控制变量 W_{rt} 的加入原则：服务于排他性条件的论证，而不是机械扩充
 - 推断与写作：第一阶段强度、聚类标准误、以及对识别对象的解释边界都需要明确交代

E. 精讲案例：Nunn and Qian (2014) 如何把识别与证据组织成一条链

- 问题来源：粮食援助的流向可能与冲突风险相关，存在反向因果与遗漏变量
- 交互式工具变量的构造逻辑：外部供给波动 $\times$ 受援结构性暴露度
- 证据链条的组织结构：
 - 外部冲击的制度背景与外生性来源
 - 暴露度变量的定义与合理性
 - 对排他性威胁的逐条回应：稳健性检验、安慰剂检验、机制区分与替代口径

F. 案例迁移与应用：从经典范式到新问题

在 Nunn and Qian (2014) 之后，我们会简要介绍两篇近期工作，展示交互式 IV 在不同背景下的应用与排他性论证思路：

- Dreher et al. (2021, AEJ: Economic Policy)：援助冲击下的交互式工具变量构造与排他性论证
- Guo & Liu (2025, Working paper)：贸易政策冲击背景下的交互式工具变量应用
 - 基于中国加入 WTO 后 PNTR 冲击，构建城市工业部门劳动需求增加 $\rightarrow$ 农村劳动力流出的交互式 IV
 - 讨论工具变量适用性：哪些因素不会威胁排他性，哪些可能威胁，如何分析、讨论稳健性

扩展阅读：

- Guo, S., & Liu, Y. (2025). When workers leave, fires rise: Migration-induced agricultural burning and institutional solutions in China. **Working paper**. [Link](#), [PDF](#), [Google](#).
- Ash, E., Morelli, M., & Vannoni, M. (2025). More Laws, More Growth? Evidence from US States. **Journal of Political Economy**, 133(7), 2139–2179. [Link](#), [PDF](#), [Appendix](#), [Google](#).
- Dias, M., Rocha, R., & Soares, R. R. (2023). Down the River: Glyphosate Use in Agriculture and Birth Outcomes of Surrounding Populations. **Review of Economic Studies**, 90(6), 2943–2981. [Link](#), [PDF](#), [PDF2](#), [Google](#), [Replication](#).

- Gallea, Q. (2023). Weapons and war: The effect of arms transfers on internal conflict. **Journal of Development Economics**, 160, 103001. [Link](#), [PDF](#), [Google](#).
 - Dreher, A., Fuchs, A., Parks, B., Strange, A., & Tierney, M. J. (2021). Aid, China, and growth: Evidence from a new global development finance dataset. **American Economic Journal: Economic Policy**, 13(2), 135–174. [Link](#), [PDF](#), [Appendix](#), [Google](#), [Replication](#).
-

最后的话

需要特别强调的是，自我提升从来都不是件轻松的事情。因此，在开课之前，大家务必认真研读每一篇论文，了解其研究背景、研究思路、计量方法和主要结论，带着问题听课。为了便于大家快速了解本次论文班中所涉及的论文内容，我们使用 ChatGPT 生成了核心论文的 [中文精要](#)，供大家课前预习和课后复习参考。

同时，也建议大家在开课前务必掌握文献的检索方法，学会使用 [文献管理](#) 和 [笔记](#) 工具，以便追踪每篇论文的后续进展，发掘新的研究主题。



4. 报名和缴费信息

- **主办方：**太原君泉教育咨询有限公司
- **标准费用**
 - 初级班：3500 元/人
 - 高级班：3600 元/人
 - 论文班：3700 元/人
- **单班报名优惠方案：**
 - **专题课老学员单班报名：**9 折
 - **学生(需提供学生证/卡照片)：**9 折
 - **会员单班报名：**85 折
- **组合报名优惠价：**
 - **初级+高级：**5700 元/人
 - **初级+论文：**5800 元/人
 - **高级+论文：**6100 元/人
 - **全程班报名：**8800 元/人
- **温馨提示：**以上各项优惠不能叠加使用。
- **联系方式：**
 - 邮箱：wjx004@sina.com
 - 王老师：18903405450 (微信同号)
 - 李老师：18636102467 (微信同号)

报名链接：<https://www.wjx.top/vm/weMgRkL.aspx#>

📱 长按/扫描二维码报名：



缴费方式

方式 1：对公转账

- 户名：太原君泉教育咨询有限公司

- 账号：35117530000023891 (山西省太原市晋商银行南中环支行)
- **温馨提示：** 对公转账时，请务必提供「**汇款人姓名-单位**」信息，以便确认。

方式 2：扫码支付



温馨提示：

- 可以使用已经绑定公务卡的微信/支付宝/云闪付等扫码付款
- 微信转账时，请务必在「添加备注」栏填写「**汇款人姓名-单位**」信息。
- 扫码支付后，请将「**付款记录**」截屏发给王老师-18903405450 (微信同号)

5. 听课指南

听课设备：

- **支持：** 手机、iPad 等平板及 Windows/Mac 笔记本
- **不支持：** 台式机

特别提示：

- 为保护讲师的知识产权和您的账户安全，系统会自动在您观看的视频中嵌入您的「用户名」信息。
- 一个账号绑定一台设备，上课后不可切换设备，且听课电脑不得外接显示屏，请提前准备好听课设备。。
- 本课程为虚拟产品，**一经报名，恕不退换。**
- 为保护知识产权，课程严禁任何形式的录屏及传播行为。

实名制报名

本次课程实行实名参与，具体要求如下：

- 高校教师及学生报名时，需向连享会课程负责人**提供真实姓名，并附上有效的教师证或学生证图片**；
- 研究所及其他单位报名需提供 **能够证明姓名以及工作单位**的材料；
- 报名即视为同意「[连享会版权保护协议条款](#)」。

6. 助教招聘

说明和要求

- **名额：** 30 名 (初级、高级和论文班各 10 名)
- **任务：** 详情参见 [连享会助教工作指南](#)
 - **A. 课前准备：** 完成 2 篇推文，风格参见连享会主页 www.lianxh.cn，选题参见 [这里](#)；
 - **B. 开课前答疑：** 协助学员安装软件和使用课件，在微信群中回答一些常见问题；
 - **C. 上课期间答疑：** 针对前一天学习的内容，在微信群中答疑 (8:00-9:00, 19:00-22:00)；
 - **Note:** 下午 5:30-6:00 的课后答疑由主讲教师负责。
- **要求：** 热心尽责，熟悉常用 AI 工具，能够解答并记录常见问题。
- **特别说明：** 往期按时完成任务的助教，可直接联系连老师获得直接录用资格。
- **截止时间：** 2026 年 1 月 8 日 (将于 1 月 10 日公布遴选结果于连享会主页 lianxh.cn)

申请链接： <https://www.wjx.top/vm/mYvepXR.aspx#>

扫码填写助教申请资料：



课程详情： <https://www.lianxh.cn/PX.html>