

# 计量和因果推断·强基班

连享会 [主页](#) || [视频](#) || [推文](#) || [知乎](#) || [Bilibili 站](#)



## 计量和因果推断·强基班

- **嘉宾：** 司继春 (上海对外经济贸易大学)
- **时间：** 2023 年 11 月 5/19/26 (三个周日)
- **时段：** 上午 9:00-12:00, 下午 14:30-17:30, 17:30-18:00 答疑
- **课件/软件：** 提供全套 Stata 代码和数据 (开课前一周发送)
- **方式：** 在线直播 + 20 天回放
- **预读资料：** <[在线查看](#)>, (或) <[打包下载](#)>
- **报名链接：** <https://www.wenjuan.com/s/JFflZje/#>
- **助教招聘：** <https://www.wjx.top/vm/YDoRHBe.aspx#>
- **PDF课纲：** [https://file.lianxh.cn/KC/lianxh\\_EF.pdf](https://file.lianxh.cn/KC/lianxh_EF.pdf) | [网页版](#)

## 目录

### 计量和因果推断·强基班

1. 课程导引：基础为何重要？
2. 主讲嘉宾简介
3. 课程详情
4. 预读材料
5. 报名和缴费信息
  - 缴费方式
6. 听课指南
  - 6.1 软件和课件
  - 6.2 实名制报名
7. 诚聘助教

# 1. 课程导引：基础为何重要？

最近十年中，各类因果推断方法层出不穷。令人欣喜的同时，也让很多人无所适从。虽然借助 Stata, R, Python 等软件很容易「跑出」结果，但很多人应该都经历过如下「困境」：

- **选择困难症。** 在诸多新近提出的 DID 模型中，我该用哪一个？用 `csdid`，`drdid`，`jwddid`，`flexpanelddid`，`did_multiplegt`，还是 `wooldid`？有人会说：哪个「好」用哪个？然而，此时会遭遇导师和审稿人的盘问：为什么用 A，而不用 B？A 为什么好？识别的假设条件是什么？有什么缺陷？
- **有劲没处使。** 使用 `ssc install wooldid` 安装了 `wooldid` 命令，但除了会照猫画虎地执行帮助文件 (`help wooldid`) 中 Example 部分给出的例子，其它基本上看不懂，更不用说伍老兄的原文了 (Wooldridge, 2021)。结果是：纵然手头有好工具，有程序包，有范例代码，但就是没有底气去用它，因为自己预感到在论文中写不清楚，也担心在组会或答辩中「下不了台」。此种情境下，并非我们不努力，而是有一种「有劲没处使」的无力感。
- **一开始就懵了。** 在各类有关因果推断的论文或教材中，经常看到「条件独立」假设： $(Y_1; Y_0) \perp D \mid \mathbf{x}$ ，作者说这就是意味着  $E(Y_1 \mid \mathbf{x}, D) = E(Y_1 \mid \mathbf{x})$ ，以及  $E(Y_0 \mid \mathbf{x}, D) = E(Y_0 \mid \mathbf{x})$ ，接下来就开始讲「回归调整 (RA)」，「逆概率加权 (IPW)」以及「双重稳健 (DR)」之类的识别和估计方法。问题在于，很多人看到「 $(Y_1; Y_0) \perp D \mid \mathbf{x}$ 」时就已经懵掉了，接着便纳闷  $E(Y_1 \mid \mathbf{x}, D) = E(Y_1 \mid \mathbf{x})$  中的  $D$  怎么突然就没了，由此引申出的各种估计方法自然也是云里雾里。

事实上，上述问题的根源在于我们没有清晰地理解「基本概念」，比如条件期望，条件独立，无偏性以及 [FWL 定理](#) 等。一旦掌握了这些基本知识，后续的 [Lasso 方法](#)，[双重机器学习](#) 等听起来很高深的东西都不再困难，因为它们都是基础知识和理论的组合和延伸而已。庄子所言「水之积也不厚，则其负大舟也无力」也正是这个意思。

至于 Top 期刊中的论文，越来越强调清晰的识别，而识别策略与计量工具密不可分，如果无法清晰理解每种计量方法的识别假设，那么就会误用甚至滥用计量模型。同时，想要规范地进行实证分析，并在此基础上合理使用工具来表达想法、讲好故事，同样也需要对原理和技术细节的深入理解，如方法选取、模型设定、标准误的处理等。

为此，本次课程将不以具体实证方法为主要目的，而是通过对基础的巩固，以新的视角重新审视我们之前学习和使用的方法，以期化解上文提到的各种「困境」。

「**自信**」源于对问题的深刻理解，外加努力。有了自信心和基础储备，就能自己去「拱」那些更难的东西了，此时「努力」才会有成效。

## 2. 主讲嘉宾简介

---



**司继春**，上海财经大学博士，目前任教于上海对外经贸大学统计与信息学院，主要研究领域为微观计量经济学、产业组织理论，成果见诸 Journal of Business and Economic Statistics、《中国人口科学》、《系统工程理论与实践》等期刊。司老师专长于机器学习，尤其是基于机器学习的因果推断前沿方法，有多个大型数据分析项目的实战经验。业余时间里，司老师也经常在知乎上耐心作答，用通俗的语言普及统计和计量知识。他的知乎专栏名为「**慧航**」，关注者逾 31w，获赞超过 17w。他总能抽丝剥茧，把复杂的问题讲得清清楚楚。

## 3. 课程详情

本课程主要分为六个模块：

- T1 介绍「条件期望」这一工具，这是我们接下来所需要使用的最重要的「语言」；
- T2 为统计工具的基础，在条件期望的基础上引入OLS，并在此基础上对预测、拟合问题做进一步探讨；
- T3 聚焦于反事实框架，我们将在这里系统介绍因果推断的基础概念，包括参数的定义、偏误的来源，并使用前两个模块的工具介绍常见的因果识别的工具；
- T4 以因果推断的视角重新审视OLS回归，并介绍线性回归的技术细节；
- T5 将继续引入面板数据，并以新的视角回顾经典的双重差分模型以及事件研究法；
- T6 在因果推断的视角下介绍内生性问题以及工具变量的识别问题。

### T1. 条件期望

条件期望 (conditional expectation) 这一概念在几乎所有数据科学中都占有非常重要的位置。本质上，条件期望即均方误差意义下的最优预测，无论是经典的OLS还是最近流行的机器学习方法，本质上都可以看作是对条件期望的逼近。更重要的是，很多计量经济学方法和工具是使用条件期望这一工具表达和推导的，为此掌握好条件期望这一工具是深刻理解诸多计量经济学方法的前提。本节将从条件期望的定义出发，详细介绍条件期望的通俗理解以及各种性质，并以简单的例子帮助大家掌握条件期望这一「语言」，让部分理论文章不再是「天书」。主要包括：

- 条件期望的定义
- 条件期望的直观解释
- 条件期望的性质
- 小试牛刀1：辛普森悖论
- 小试牛刀2：机器学习中的偏差-方差权衡
- 建议预读：(参见后文「预读资料」)
  - [Cer2022](#) , Chap 1, 2
  - [Hansen2021](#) , Chap 2

### T2. 线性回归与拟合

进一步，我们从条件期望的角度引入最常见的拟合和预测工具：OLS。我们首先从条件期望与OLS之间的关系入手，逐渐引入OLS的统计性质，并讨论条件期望估计中的函数形式问题。最后，我们还将从预测的角度讨论模型选择的一般标准，为机器学习和因果推断奠定良好的统计工具基础。最后，权重通常也是实证中常见的策略，我们也将讨论回归中权重的使用，并在加权最小二乘的基础上引入非参数和半参数回归。主要包括：

- 从条件期望到OLS：估计
- 使用OLS估计条件期望的前提假设和统计性质
- 条件期望中的函数形式问题
- 分步回归的思想与意义
- 预测与模型选择标准
- OLS中权重的使用
- 非参数回归与半参数回归
- 建议预读：

- Cer2022 , Chap 1, 2
- MHE2018 , Chap 3

### T3. 反事实框架：偏误来源、影响及应对

因果推断是计量经济学的核心内容之一，而因果是通过反事实的框架定义的。本节将回顾因果推断的基础内容，包括Rubin因果模型以及其中常见的定义，并分析在估计平均处理效应、处理组平均处理效应等问题时可能存在的偏差。最后，我们将介绍在无混淆分配假设下的识别方法和相应的估计方法。我们将充分利用条件期望这一工具对处理效应的识别进行分析，同时利用拟合部分的内容详细介绍如何在识别的基础上进行更进一步的估计。主要内容包括：

- Rubin因果框架
- 因果推断中关心的识别参数及其区别
- 因果推断中的偏误来源
- 无混淆分配假设
- 无混淆分配假设下的识别
- 无混淆分配假设下的估计和推断
  - 回归调整法
  - 匹配法
  - 逆概率加权法
- 建议预读：
  - Cer2022 , Chap 1, 2
  - MHE2018 , Chap 2

### T4. 线性回归：因果推断视角

实证中，基于线性回归的因果推断尤为流行，然而在使用时往往需要注意大量细节，这主要是由于我们此时使用线性回归不再以预测为目的，而是以解释和因果推断为目的。为此，本节主要结合以上介绍的因果推断基本概念，重新审视线性回归，并介绍在无混淆分配假设下控制变量的选择、固定效应的控制等问题，以及如何使用线性回归建模处理效应异质性问题。主要内容包括：

- 无混淆分配下的推断：控制变量选取
- 无混淆分配下的推断：固定效应
- 模型设定：异质性与非线性性
- 使用线性回归分析实验数据的实例
- 标准误的处理：抽样与模型的不同角度
- 建议预读：
  - Cer2022 , Chap 1
  - MHE2018 , Chap 2
- 参考文献：
  - Abadie, A., S. Athey, G. W. Imbens, J. M. Wooldridge, 2023, When should you adjust standard errors for clustering?, *The Quarterly Journal of Economics*, 138 (1): 1-35. [-Link-](#), [-PDF-](#), [PDF2](#), [Replication](#), [-cited-](#)

## T5. 面板数据和倍分法

面板数据在当前的实证研究中是被最广泛使用的数据形式，当前理论文献中关于面板数据中因果效应的识别和估计问题也是研究热点之一。本节将主要从面板数据的基础概念出发，包括面板数据中各种外生性假定、模型设定的假设以及模型之间的关系进行系统梳理，从而能够熟练掌握经典的面板数据处理方法。此外，我们还将综合使用以上条件期望、因果推断的基础内容，从理论的层面重温双重差分模型，并进一步介绍双重差分模型等新进展的识别理论和估计方法。主要包括：

- 面板数据的基本概念及外生性概念
- 差分估计量、固定效应与随机效应
- 简单双重差分模型回顾
- 事件研究法与平行趋势假设的检验
- 交错DID的识别问题：Callaway and Sant'Anna (2021) 以及 Sun and Abraham (2021)
- 建议预读：
  - Hansen2021 , Chap 17, 18
  - Cer2022 , Chap 5
- 参考文献：
  - Callaway, B., P. H. C. Sant'Anna, 2021, Difference-in-differences with multiple time periods, *Journal of Econometrics*, 225 (2): 200-230. [-Link-](#), [-PDF-](#), [PDF2](#), [Slides](#), [csdid](#), [-csdid2-](#)
  - Sun, L., S. Abraham, 2021, Estimating dynamic treatment effects in event studies with heterogeneous treatment effects, *Journal of Econometrics*, 225 (2): 175-199. [-Link-](#), [-PDF-](#), [PDF2](#), [-Slides-](#), [Stata codes](#), [Replication](#)
  - Roth, J., P. H. C. Sant'Anna, A. Bilinski, J. Poe, 2023, What's trending in difference-in-differences? A synthesis of the recent econometrics literature, *Journal of Econometrics*, 235 (2): 2218-2244. [-Link-](#), [-PDF-](#), [DID 最新综述](#)

## T6. 内生性：因果推断下的工具变量

当无混淆分配假设不满足时，就出现了内生性问题，而工具变量是处理内生性问题的一个常见方法。然而在处理效应识别和估计的背景下，很多时候工具变量的使用需要更多的假设。本节主要讨论在因果推断的背景下，工具变量的识别策略以及具体的估计方法、诊断方法。主要包括：

- 传统内生性的成因
- 遗漏变量的诊断
- IV 估计方法：Wald估计量与2SLS
- 因果推断下的 IV：局部平均处理效应
- 实验中的局部平均处理效应
- 断点回归中的局部平均处理效应
- 面板数据与工具变量：模糊双重差分
- 工具变量估计的诊断
- Bartik (shift-share) 工具变量
- 建议预读：

- [Cer2022](#) , Chap 3, Sec 4.1
- [MHE2018](#) , Chap 4
- 参考文献:
  - Goldsmith-Pinkham, P., I. Sorkin, H. Swift, **2020**, Bartik instruments: What, when. Why, and how, **American Economic Review**, 110 (8): 2586-2624. [-Link-](#) , [-PDF-](#) , [PDF2](#) , [Replication](#) , [-cited-](#)
  - Borusyak, K., P. Hull, X. Jaravel, **2022**, Quasi-experimental shift-share research designs, **Review of Economic Studies**, 89 (1): 181-213. [-Link-](#) , [-PDF-](#)

## 4. 预读材料

- [Hansen2021](#) | **Hansen** B E . 2021. Econometrics. Princeton University Press. [Data and Contents](#) , [PDF](#) , [-PDF2-](#)
- [MHE2008](#) | **Angrist**, J. D., J.-S. Pischke. Mostly harmless econometrics: An empiricist's companion[M]. Princeton, NJ: Princeton University Press, **2008**. [-Blogs-](#) , [-PDF-](#) , [Data-Codes=R-Stata-Python](#) , [-Slides-](#)
- [Cer2022](#) | **Cerulli**, G. Econometric evaluation of socio-economic programs theory and applications[M]. Springer, 2022. [-Link-](#) , [-Website-](#) (登陆学校图书馆账号可以下载 2e PDF), [PDF-1e](#)
  - 共 6 章, 介绍了反事实框架, 回归调整法 (RA), 逆概率加权 (IPW), 匹配, 多期 DID, 合成控制法等。基本上涵盖了主流的因果推断方法。
- [Chan2022](#) | **Chan**, F., L. Mátyás. Econometrics with machine learning[M]. Springer, 2022. [-Link-](#) . 登陆学校图书馆可以下载 PDF 全本.
- [Woold2010](#) | **Wooldridge**, J. M. Econometric analysis of cross section and panel data, 2nd ed[M]. Cambridge, MA: MIT press, 2010. [-PDF-](#) , [-Link-](#) (内附 Stata codes, Slides 和相关资料)

附：计量基础和 Stata 实操

- [Baum2006](#) | **Baum**, C. An introduction to modern econometrics using stata[M]. Stata Press, 2006. [-Link-](#) , [PDF](#) , 计量基础和 Stata 实操
- [Acock2018](#) | **Acock**, A. C. A gentle introduction to stata (6e)[M]. Stata Press, 2018. [-Website-](#) , [Data-Codes](#) , [-Answer-](#) , [-PDF-4e](#) , Stata 基础、数据处理等

## 5. 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** (含报名费、材料费)：3300 元/人 (全价)
- **优惠方案：**
  - 三人及以上团购/专题课老学员：75 折，2475 元/人
  - 学生 (需提供学生证/卡照片)：75 折，2475 元/人
  - 连享会会员：7 折 2310 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](mailto:wjx004@sina.com)
  - 电话 (微信同号)：王老师 18903405450 ; 李老师 18636102467

**报名链接：** <https://www.wenjuan.com/s/JFflZje/#>

或 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 (晋商银行股份有限公司太原南中环支行)
- **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：微信扫码支付



**温馨提示：** 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。

## 6. 听课指南

---

### 6.1 软件和课件

**听课软件：**支持 手机，ipad，平板以及 windows/Mac 系统的笔记本，**但不支持台式机**

#### 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。
- 一个账号绑定一个设备，且听课电脑不能外接显示屏，请大家提前准备好自己的听课设备。
- 本课程为虚拟产品，**一经报名，不得退换。**
- 为保护知识产权，课程不允许以任何形式录屏及传播。

### 6.2 实名制报名

本次课程实行实名参与，具体要求如下：

- 高校老师/同学报名时需向连享会课程负责人 **提供真实姓名，并附教师证/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证明**；
- 报名即默认同意「**连享会版权保护协议条款**」。

## 7. 诚聘英才

- **名额：** 10 名
- **任务：**
  - **A. 课前准备：** 协助完成 3 篇介绍 Stata 或 Python 或 R 语言 和计量经济学基础知识的文档，风格类似于 [lianxh.cn](http://lianxh.cn) ；
  - **B. 开课前答疑：** 协助学员安装课件和软件，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、尽职，熟悉 Stata 或 Python 或 R 语言 的基本语法和常用命令，能对常见问题进行解答和记录。
- **特别说明：** 往期按期完成任务的助教可联系连老师直录，优先考虑熟悉 Python 和 R 的申请者。
- **截止时间：** 2023 年 9 月 25 日 (将于 9 月 27 日公布遴选结果于连享会主页 [lianxh.cn](http://lianxh.cn))。

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

或扫码填写助教申请资料：



**课程主页：** <https://www.lianxh.cn/>

