

目录

目录	1
Home	2
CIP-面板数据因果推断-徐轶青 wiki!	2
News-课件更新等.md	3
FAQs	4
Day1-FAQs.md	4
CIP-面板数据因果推断-徐轶青 Day1 问答 （2022/10/15）	4
请徐老师解答的问题	4
助教回答的问题	5
Day2-FAQs	7
2022 因果推断专题 10 月 16 日课程提问	7
Day3-FAQs	15
2022 因果推断专题 10 月 22 日课程提问	15
Day4-FAQs	20
2022 因果推断专题 10 月 23 日课程提问	20
Day5-FAQs	29
2022 因果推断专题 10 月 29 日课程提问	29
Day6-FAQs	33
2022 因果推断专题 10 月 30 日课程提问	33
板书	39

Home

CIP-面板数据因果推断-徐轶青 wiki!

- 答疑文档：查看左侧菜单 FAQs
- 课件更新：查看左侧菜单 News-课件更新等
- 课程详情：[CIP课程大纲](#)

👉 点击右上角的【Fork】按钮，可以把这个项目完整复制到你的码云账号下，随时查看。

News-课件更新等.md

- 课件和课程论文
 - 徐老师课程中涉及的多数论文都可以通过 [课程主页](#) → 0.课程概览 → 备讲论文： [点击下载查看](#) 下载。
 - update: 2022/10/15 11:14

FAQs

Day1-FAQs.md

CIP-面板数据因果推断-徐轶青 Day1 问答 (2022/10/15)

全文由徐老师解答的问题、助教解答的问题两部分共同构成。

请徐老师解答的问题

Q1.为什么 Heckman two-stage 的估计结果不可信？如果 Heckman 不可信又要修正 Sample selection bias 的话，应该采用那种估计方法？

A: 不可信有两个原因：一是模型假设太强，真实事件要比设定模型复杂得多；二是缺失关键变量，关键混淆变量没有的话，模型设定的再精确、再复杂，估计也会产生误差。有些数据就是无法修正 sample selection bias，有的数据模型是不对的，我们估计出来结果也没有意义。我们只能寻找合适的数据和模型进行估计。（回答人：徐老师；整理人：王胜文）

Q2.同时性问题就等于内生性吗？

A: 是的，同时性是内生性的一种。（回答人：徐老师；整理人：王胜文）

Q3.为什么说遗漏变量、互为因果、同时性本质都是 selection 的问题？

A: 三者都是在因果推断的框架里面，都是对强 Selection on observable 假设的违反，就是进行了 selection，但是没去识别关键的变量。再解释一下什么是识别，识别和有多少数据没有关系的，因为没有关键信息，比如说教育对收入的影响，当我们无法控制能力等遗漏变量，再多的数据，我们无法识别教育和收入的因果效应。（回答人：徐老师；整理人：王胜文，屠西伟）

Q4. 老师您讲的从研究设计上来讲，倍差法这类最好，但是这种的研究不是只适用于政策研究吗？

A: 从内部效度来说，实验法肯定要比倍差法好，RDD 是其次的，面板要次于 RDD。（回答人：徐老师；整理人：王胜文）

Q5. 老师刚刚提到的传统回归异质性有些问题，那里不大理解，好多文献都是做分组异质性讨论，是 DID 做异质性分析比较好么？为什么 DID 也异质性分析呢，解释变量大部分都是 0，仅有几个 1，分组的的话有的就是仅仅使用 0 去对 y 做回归？

A: 不管是分组回归也好、交叉项回归也好，都是对异质性的探索，都是在一个线性模型的框架下面，只是把原来的 β 变成了 $\beta_0 + \beta_1 * x$ ，这前提是要允许异质性的存在，比如说贫穷省份和富裕省份效果不一样，允许男士和女士效果不一样，但是这仍然是有限的异质性，因为真正异质性是每个个体都是异质的，也就是说分组要比不分组好。但是因为数据是有限的，我们不可能无限的分组，所以在进行回归时，除了几个特殊变量，比如性别，地区东中西，其他的参数我们看做是一致的。大部分时间不是要去估计个体的因果效应，虽然存在，但是我们不直接估计每个个体因果效应，一般都是估计个体平均因果效应。第二节课会详细讲解。（回答人：徐老师；整理人：王胜文）

Q6. 老师请问单独采用 PSM 和单独采取 DID 都是为了解决非随机分组内生性问题吗？DID 比 PSM 优势在哪里，以及为什么有文章直接采用 PSM-DID，有时仅仅做 PSM 稳健性检验？

A: 在 DID 中配上 PSM 是为了让平行趋势假设更容易成立，在截面数据中用 PSM 还是基于观测量选取（Selection of observable），后面课程会详细讲解。（回答人：徐老师；整理人：王胜文）

Q7. 老师，请问协变量填补主要有哪几种方法？

A: 有多重填补法、多重填补数据分析法、敏感性分析法，等等。但是，要注意面板数据的协变量填补比较难，因为维度比较高；一般来说在截面中填补协变量是通过相关性进行填补，但是在面板数据中比较难。（回答人：徐老师；整理人：王胜文）

Q8. 非平衡面板和混合横截面的区别是什么？混合横截面是直接横截面方法处理吗？

A: 非平衡面板和混合截面还是有差别的。非平衡面板数据中还有部分人（样本个体）是有连续多次的观测数据，但混合截面数据至少说不知道数据中有没有人（个体）是重复的。混合截面处理要求更加高一点，但是其实可以加上时间效应去做，但是加个体效应就比较难。混合截面的关键是要让它的 composition（组成），就是让两个截面可以观察到的个人特征变量 composition 比较接近，比较接近之后，再去做一个前后的 difference（差分），因为你没有办法在个体层面直接根据个体的 id 去做 difference（差分），所以要用协变量来定义，这是二者的差别。混合截面要比纯截面要稍微好一点，但是它比面板要差。（回答人：徐轶青老师，整理人：陈希）

Q9. 有什么具体的 data cleaning 命令和方法吗？检验和处理老师刚刚提到的 AGE 跳跃等异常值问题的吗？

A: 一般来说，我用 Stata 做的时候，就是要找一些数据中的一些逻辑关系，然后一个一个去检验这些逻辑关系是不是符合。比如说，age 就要随年份有一个线性增长的关系；或者 bounding 的话，就不能超过 0 和 1，其实就是逻辑关系的审查。补充：Stata 数据处理：缺失值与多重补漏分析（一）：<https://www.lianxh.cn/news/716da36e2cb70.html>；Stata 数据处理：缺失值与多重补漏分析（二）：<https://www.lianxh.cn/news/c3e8128be3072.html>；Stata 数据处理：缺失值与多重补漏分析（三）：<https://www.lianxh.cn/news/3b79ea976a27e.html>（回答人：徐轶青老师、屠西伟，整理人：陈希）

Q10. 请问老师面板数据时间必须是连续的吗？比如 12,14,15,17 年的调查数据。

A: 面板数据不一定必须是连续的，只有 12、14、15、17 年的数据是非平衡面板。一般面板数据大多是非平衡面板，如果数据的缺失不是某些非随机因素造成的，也可以进行面板模型估计。（回答人：徐轶青，整理人：刘佳宁）

Q11. 请问老师 CI 一直在用，但实际含义是什么一直不知道。

A: 数理统计学中，我们将用来进行合理推断的已知数据称为样本，将被合理推断的未知数据称为参数，将这种推断方法称为参数估计。在区间估计中，需要给出一个置信区间（Confidence Interval，简称为 CI），并预测真正的参数 θ 以一定的概率存在于这一区间的方法。这一区间覆盖真实值的概率被称为置信系数。置信区间用区间的端点 $\hat{\theta}_1$ 与 $\hat{\theta}_2$ 表示， $\hat{\theta}_1 < \hat{\theta}_2$ 。当给定常数 α ($0 < \alpha < 1$)，若有 $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha$ 成立，则称 $\hat{\theta}_1$ 到 $\hat{\theta}_2$ 这一区间能够覆盖真值的概率为 $1 - \alpha$ 。这样的区间为待估参数 θ 的 $1 - \alpha$ 置信区间， $1 - \alpha$ 称为置信系数或置信水平 (confidence level，简称为 CL)， α 称为显著性水平。专题-回归分析 > Stata：如何估计置信区间？专题-Stata 入门 > 25 常见误区：P 值、置信区间和统计功效（回答人：徐轶青，整理人：刘佳宁）

Q12. 请问老师面板数据里有一年确实很严重（提取数据时导致的缺失），可以将这一年直接删除再用剩余年份数据进行 event study 分析吗。

A: 可以将这一年所有数据删掉，这没有什么问题。（回答人：徐轶青，整理人：刘佳宁）

Q13. 老师好，请问个体层面的混合截面，但是城市是平衡的，可以说是城市层面的面板吗？

A: 如果不同年份之间城市层面的个体结构保持不变，例如性别比例，此时可以生成城市层面的统计量，做城市层面的面板，相当于使用城市层面某一时间的样本代表这个城市，但需要保证样本的“可代表性”。（回答人：徐轶青，整理人：冯超楠）

Q14. 老师好，请解释一下传统模型是基于结果，但因果推断是基于设计的吗？

A: 传统模型只有在模型设定正确的情况下才能准确估计，传统模型和因果推断不矛盾，但前者很难满足。因果推断以 D（处理变量）的生成机制为出发点来选择研究设计，仅仅识别可识别的因果效应（仅识别能使用 DID 或 RD 或 IV 等的情况）。（回答人：徐轶青，整理人：冯超楠）

Q15. 老师好，请问数据时间没有包含政策时间，比如政策是 2013 年实施，但是只有 2014 与 2018 年数据，能做 DID 吗？

A: 不能。因为做 DID 需要政策实施前后的数据，政策于 2013 年实施，需要准备 2013 年前后的数据。2014 与 2018 都是政策实施后，缺乏政策实施前的数据，所以不能。（解答：徐老师；整理人：王舒瑶）

Q16. 老师好，置信区间和显著性的含义

A: 置信区间展现的是这个参数有一定概率落在测量结果的周围的程度，其给出的是被测量参数的测量值的可信程度。显著性指零假设为真的情况下拒绝零假设所要承担的风险水平。这两个概念老师第二天会画图具体讲。（解答：徐老师；整理人：王舒瑶）

Q17. 老师好，PSM-DID 很少在英文顶刊上看到，这方法有什么问题吗，可靠吗？

A: PSM 并不能起到解决样本选择偏差的作用，结果的可信度存在一定局限。（解答：徐轶青；整理人：王舒瑶）

Q18. 请问老师您做 DID 是用 diff 还是 reghdfe 比较多？声明是面板数据 xtset 中的时间用原始的 year 还是处理过的 t (就是 0 和 1)？

A: reghdfe 应用更广，比 xtreg 更好，后续课程会详细讲解。（回答人：徐轶青，整理人：王凯璇）

Q19. 请问老师怎么看待 event history analysis 事件史分析？

A: 很有必要，是 DID 的拓展。（回答人：徐轶青，整理人：王凯璇）

Q20. 请问徐老师在 Stanford 教授因果推断课程的课程大纲是否可以分享？或者目前美国大学比较前沿的因果推断的课程大纲内容是怎样的？在哪里可以看到这些信息？

A: [斯坦福因果推断课程大纲链接](#)（回答人：徐轶青；整理人：王凯璇）

Q21. 对于老师提到的反向因果问题，之前很多文章都会将解释变量滞后一期，说用以解决反向因果问题，但现在都说这种方法并不可行，请问是为什么呢？

A: 需要因果图解释，后续会详细讲解。这种解释变量滞后一期这种方法“最多”只能“减轻”反向因果关系的问题，没办法彻底解决。参见连享会：<https://www.lianxh.cn/news/c45f6113e2c2a.html> 和 <https://www.lianxh.cn/news/ca876910609c8.html>（回答人：徐轶青，刘佳宁，郭盼亭；整理人：王凯璇）

助教回答的问题

Q1. 老师提到的传统 + 新方法，是不是需要保持样本的一致性？比如非平衡面板做 TWFE，但是拓展部分就不能做 SDID 吗？SDID 需要平衡面板，需要在基础回归上删除样本，是不是就不能选择 SDID 了。

A: 样本需要保持一致性，一般需要平衡面板。非平衡面板存在一些数据缺失，这些数据缺失可能是一种自选择行为。（回答人：王凯璇；整理

人：屠西伟)

Q2. 国内做 DID 的时候很常用某一变量的中位数作为分组变量的依据。例如，劳动合同法的处理组就是高劳动密集的企业，这样的做法是否合理？如果要这样做，是不是应该用冲击前一年的劳动密集情况做分组依据？

A: 用中位数数分组，和用各种分位数分组都是可以的，相当于按照面对冲击的暴露强度，分成两组或者若干组。这相当于 Duflo (2001) 的强度差分的简化版。使用前一年的更外生一些。（回答人：王凯璇；整理人：屠西伟）

Q3. 有论文模型设定是连续变量乘以政策虚拟变量，这样设定有什么好处吗？核心解释变量为“就业人数 × 刺激就业政策”的实施时间虚拟变量。一般这么设定还需要分别加入就业人数与刺激就业政策的实施时间虚拟变量吗？

A: 一般而言，像这种全国一刀切的政策，没有办法像之前那样严格区分处理组和控制组，只能以连续变量区分处理强弱，然后再与政策前后的时间虚拟变量交乘。使用“连续变量 × 政策虚拟变量”，可以更好地捕捉连续变量增加的动态效应。一般这样设定需要分别加入就业人数与刺激就业政策的实施时间虚拟变量，但是会被固定效应所吸收。（回答：王凯璇；屠西伟；整理人：王胜文）

Q4. 徐老师好，想请教一下倾向得分匹配最合适使用的场景是什么？有的文章说用于解决样本自选择，这种说法是否正确？

A: PSM 只是通过匹配的方式在一定程度上避免了函数形式错误设定导致的偏差，并没有从根本上解决由选择偏差或遗漏变量所导致的内生性问题。具体细节可以参照连享会推文：[PSM-倾向得分匹配分析的误区](#)，[Stata+PSM：倾向得分匹配分析简介](#)。关于 PSM 解决问题的进一步理解可以参考文献：<https://doi.org/10.2308/accr-51449>。举个例子：(1) 接受某种职业技能培训会对个人收入有怎么样的影响？(2) 上周投放了某游戏广告的用户登录率更高，他们的高登录率有多大程度是由广告带来的，有多大程度是由于他们本身就是高潜力用户？（回答：陈云菲，陈希；整理人：王胜文）

Q5. 请问若 DID 中的 shock 存在反向因果的问题，那需要为该 shock 找一个 IV 吗？

A: DID 识别策略中并不要求政策冲击外生，只要平行趋势假设成立，反向因果问题不会产生问题。工具变量不是必需的，当然也可以做个稳健性检验。（回答：屠西伟）稳健性检验可以看一下这篇文章的应用：《长期护理保险、医疗费用控制与价值医疗》，通过挑选控制组的方法进行检验（回答：陈云菲，整理人：屠西伟）

Q6. 请问老师，三重差分需要做平行趋势检验吗，如何做？

A: 一般不需要。平行趋势是对因果关系是否成立进行的检验，而三重差分是对已成立的因果关系做的异质性检验。如果文章主体为 DDD，而不是将 DDD 作为 DID 异质性的小拓展。则需要进行平行趋势检验，具体做法可通过绘图或者事件研究法，参考文献：[1] 齐绍洲, 林岫 & 崔静波. (2018). 环境权益交易市场能否诱发绿色创新?——基于我国上市公司绿色专利数据的证据. 经济研究(12), 129-143; [2] Cai, X., Lu, Y., Wu, M., & Yu, L. (2016). Does environmental regulation drive away inbound foreign direct investment? Evidence from a quasi-natural experiment in China. *Journal of Development Economics*, 123, 73-85; [3] Shi, X., & Xu, Z. (2018). Environmental regulation and firm exports: Evidence from the eleventh Five-Year Plan in China. *Journal of Environmental Economics and Management*, 89, 187-200. （回答：王舒瑶，刘佳宁；整理人：屠西伟）

Q7. 视频界面自动关闭的次数较多？

A: 需要关闭杀毒软件，360 杀毒微、软杀毒都得完全关闭；也可能是播放器的问题，重新装播放器，重启一下电脑。（回答：王胜文；整理人：屠西伟）

Q8. 老师，如果因变量是多分类变量，怎么应用 DID？老师可以推荐几篇论文吗？

A: 现在对于因变量是多分类变量，还只能按照正常的 DID 估计，即线性回归的方法，目前非线性回归的方法还不成熟。（回答人：屠西伟，整理人：陈希）

Day2-FAQs

2022 因果推断专题 10 月 16 日课程提问

Q1：老师和助教好！我想问一下多期 DID 中使用企业层面的面板，但政策冲击是地区层面的，那么只控制地区 fe 和 year fe 允许吗？应该与个体 + 时间的估计结果差异较大，但平常两种方法都能看到。简而言之，固定效应的控制基本上应该控制与这个政策冲击的同一层面就够了，控制到更细的层面是更严格，但不是必要的。

答：如果政策冲击是地区层面的，只控制到地区、时间层面的固定效应是可以的，更好的选择是把数据压缩到地区层面，然后在地区层面跑回归，这是比较简单的方式。但如果你觉得想跑一个特别大的样本，需要注意cluster在地区层面，这个会有细微的差别，因为如果你是cluster的话，相当于以每个地区的企业数目进行加权，但如果你现在平均到这个地区层面，每个地区它的权重是一样的。两个差异的原因是权重不一样，例如假设上海有1万个企业，然后甘肃有1000个企业，然后两个都是省，都是地区层面的，然后假如你在企业层面跑回归，上海的权重就是甘肃的十倍。然后，如果先压缩到上海和甘肃，那上海和甘肃就是1:1，这是出现差异的原因。哪一个更正确，取决于研究背景。（回答人：徐轶青，整理人：郭盼亭）

Q2：xtreg命令有没有将估计方法自动修正为投影法？还是说对非平衡面板直接使用xtreg的fe选项，得到的结果是错的？

答：xtreg命令现在大家就不要用了，它不太好用，现在大家可以用areg或者reghdfe。xtreg的点估计没有问题，但是standard error就有问题。（回答人：徐轶青，整理人：郭盼亭）

Q3：老师可以分享一下“来自模拟的证据”的例子吗

答：我在lecture2里边有一个simulation，晓燕老师应该已经分享给大家了，那个很接近这个例子，我下周会跑一下给大家看。（回答人：徐轶青，整理人：郭盼亭）

Q4：能否请徐老师大致介绍一下 Sequential Ignorability Assumption

答：这个问题第三讲会详细讲的，请大家不要着急。基本上其实我刚才已经讲过了，意思就是明天的treatment assignment是和今天的Y的取值是有关的，例如我昨天生病了，然后医生今天给我开了新的药，这就是sequential assignment。Sequential Ignorability是说给定我可以观察到所有信息，包括你过去的身体状况，过去的Y，我是不是给药，改变我的treatment 是随机的，这叫Sequential Ignorability。但是在Sequential Ignorability情况下是不能有fixed effect的，也不能有不随时间变化的混淆变量的，这就是一个权衡。所以，要么是sequential，没有fixed effect，要么是有fixed effect，不能有sequential。（回答人：徐轶青，整理人：郭盼亭）

Q5：coverage probabilities如何在stata里面操作实现？

答：我觉得不能直接在stata里边实现，需要写simulation的代码。我没有在lecture 2做这部分，因为它比较慢，而且对大部分老师来说，这个用处不是很大，但是如果你以后想做一点点计量或者应用计量的话，这个是很首先它不难，它就是个for loop，其次，这个你需要知道这个怎么回事儿。（回答人：徐轶青，整理人：郭盼亭）

Q6：徐老师能举个例子讲讲您刚刚说的第一步画数据那里么，比如您重点看结果中的哪些问题？

答：我不知道你说的是指数据描述还是DID前边的第一步，我猜你可能是讲数据描述，数据描述就是需要给大家一个概念，最好的情况是你在没有控制fixed effect的情况，就是在横截面和时间维度上都混在一块儿的情况下，可以看到treatment 组和control 组或者X如果是连续的，有某种相关性，就可以告诉大家，虽然不能解释为因果性，但是一个粗的相关性至少是存在的，而且它不是我利用后面的统计操作或者很奇怪的模型搞出来的。然后我们再去把外面各种各样的干扰的因素去除掉之后，看核心的因果的效应是怎么样，这样会比较容易说服别人。

然后当然另外一个要点是你要用某种方式在第一步画数据那里，告诉大家你的处理变量、结果变量到底是什么分布，它是随时间变化或者是怎么变化的，这个如果能够用图画出来，会比较好。（回答人：徐轶青，整理人：郭盼亭）

Q7：刚才模拟实验里面，系数估计的真实标准差是多少？用cluster估计的标准误跟真实标准差有多大的差距？

答：这个我们下周会具体讲。（回答人：徐轶青，整理人：郭盼亭）

Q8：TWFE适用于昨天讲的哪种面板数据？能先简要总结一下这门课的几个方法哪些适用与大N小T、大T小N、大T大N吗？老师说的大N大T型的面板，一般N和T需要大于几期？

答：理论上大N小T适用，大T小N是不可以用的，但实践中大家都在用，大N大T用TWFE需要满足强外生假设。如果是浅因子模型的话，大N大T需要N、T特别大，在应用中至少超过30，最好是40、50以上。因为大数定律起作用的话，至少需要30、40以上才能够收敛，还要考虑estimate的收敛速度的问题。（回答人：徐轶青，整理人：郑晓雪）

Q9：老师之前的问题“论文系数增加一个标准差导致Y增加多少”感觉是在问 如果D是一个连续变量 如果现在得到一个他的系数beta 那么增加D的一个std.dev 会改变Y多少

答：这个说法是对的，如果D从0变为1，并不需要对D做标准化。但如果D是个连续变量，就可以对D做标准化，看D变化一个标准差Y如何变化。（回答人：徐轶青，整理人：郑晓雪）

Q10：请问徐老师，如果是三重差分模型而且是staggered 多期分批实施的，根据最新文献，这种的偏误应该用什么方法纠正，目前的方法好像都是针对双重差分。

答：这是我们第三讲和第四讲的核心内容，这并不是三重差分的偏误，三重差分是倍差法的异质性，倍差法在staggered的情况下就是会有偏误的。

我们要解释偏误的来源以及如何克服，再做三重差分就是在这个基础上做效应的异质性。（回答人：徐轶青，整理人：郑晓雪）

Q11：老师能分享一下数据描述那四幅图的code么？不太会画图，想学一下～

答：Stata绘图详细可参考 Stata绘图全解：绘图语法-条形图-箱型图-散点图-矩阵图-直方图-点图-饼

图：<https://www.lianxh.cn/news/ae36a721cfe18.html> Stata绘图：世行可视化案例-条形图-密度函数图-地图-断点回归图-散点

图：<https://www.lianxh.cn/news/96989b0de4d83.html> Stata绘图：面板数据可视化-panelvie：<https://www.lianxh.cn/news/78c21ab215c46.html>

给你的图形化个妆：Stata绘图常用选项汇总-上篇：<https://www.lianxh.cn/news/59bc1ff7d027a.html> 给你的图形化个妆：Stata绘图常用选项汇总-

下篇：<https://www.lianxh.cn/news/180daa074ef27.html>（回答人：徐轶青，整理人：郑晓雪）

Q12：徐老师能否把 clans and calamity 那篇文章，画图的R语言code也分享一下

答：给徐老师发邮件，yiqingxu@stanford.edu。（回答人：徐轶青，整理人：郑晓雪）

Q13：bootstrap使用时在更高层面聚类会提醒有重复的年份，是不可以在更高层面聚类吗？

答：可以在更高层面聚类，重复也没关系的。（回答人：徐轶青，整理人：郑晓雪）

Q14：fixed effects 中的 within estimator 和 between estimator 是各利用组内变异与组间变异，那么 two way fixed effects 利用的是什么变异？

答：在没有交错处理情况或者加权问题存在的时候，two way fixed effects 可以理解成 DID 的效果，换句话说 two way fixed effects 估计是 DID 的一种实现方式。（回答人：徐轶青，整理人：陈云菲）

Q15：徐老师能否把 clans and calamity 那篇文章，画图的 R 语言 code 也分享一下

答：需要的同学可以发邮件给老师，老师将通过邮件发送（回答人：徐轶青，整理人：陈云菲）

Q16：bootstrap 使用时在更高层面聚类会提醒有重复的年份，是不可以在更高层面聚类吗？

答：bootstrap 使用时有重复的年份不会影响估计，也可以在更高层面聚类。（回答人：徐轶青，整理人：陈云菲）

Q17：bootstrap 步骤得到的 _b 是为了应用在哪个步骤呢？

答：_b 是为了得到 coefficients,在下次课程中将详细解释 bootstrap 的原理。（回答人：徐轶青，整理人：陈云菲）

Q18：请问为什么在画动态系数图时，选取的模型是不带 control 的？

答：带 control 和 不带 control 都是可以的，带 control 可能会损失一部分样本。（回答人：徐轶青，整理人：陈云菲）

Q19 如果 Y 是 quarterly county income, control variable 可以用 annual gdp 吗

答：如果是县的收入和国家的 GDP, 可能存在同时性 (simultaneity) 的问题，因为县的收入也是国家的 GDP 的一部分。虽然中国有 3000 多个县，一个县的收入的影响程度相对较小，让同时性的问题没有那么严重，但还是有点奇怪。因此，建议将变量换成县收入增长率和国家 GDP 的增长率，或增加控制地级市/省的趋势。（回答人：徐轶青，整理人：陈云菲）

Q20：请问控制单位趋势和控制time FE相比，更加控制了哪类 counfounders 呢？

答：控制时间 time FE 的用意在于吸收时间维度上不可观测的同质性冲击的影响，即所有个体共有的时间因素，如宏观经济冲击、财政货币政策等等，假定这些因素在特定年份对不同个体的影响是一致的。控制单位趋势是为了控制不同个体的被解释变量可能存在的并且尚未被其他控制变量和 FE 所覆盖/解释的增减趋势，因为不同组别（规模、性质、政策分组、生命周期等）个体的被解释变量的时间趋势或许存在一定程度的差异，并且在控制已有的解释变量之后依然可能存在较为明显的时间趋势。（回答人：徐轶青，整理人：冀承）

Q21：论文中说系数增加一个标准差导致Y增加多少，这个如何理解，是怎么算的？

答：实证研究中，自变量 X 变化一个标准差，因变量Y变化了多少？具体而言，有两种理解：一是因变量 Y 变化了多少个单位标准差（百分比变化）= 两者系数 * 自变量标准差/因变量标准差；二是因变量Y变化了多少 = 两者系数 * 自变量标准差。（回答人：徐轶青，整理人：冀承）

Q22：平行趋势假设可否再讲一次，刚刚右上角被您的头像遮住了

答：平行趋势假设在多期情况下，可以假设 $D = 0$ （基准没有干预时），控制协变量 X，没有干预情况下的潜结果在两个时间的差对于所有的 i 都是一样的。（回答人：徐轶青，整理人：冀承）

Q23：对于这种情况X和Y标准化之后做回归是不是能够直接得到上一楼说的beta？

答：是能够直接得到beta，但不要对 X 进行标准化，对 Y 做标准化。（回答人：徐轶青，整理人：冀承）

Q24：老师，table3回归结果与原文中不一样，老师报告出来的表格的最后两行，absorb之后的

答：将代码中的 of 改为 in 就可以得到与原文相同的结果。（回答人：徐轶青，整理人：冀承）

Q25：Unique 命令不能用？答：Stata中的unique命令能够帮助研究者找出一个变量或一组变量的所有不同值(unique value)，可以正常使用。关于如何统计组内非重复值个数以及生成相应变量可以参考如下连享会推文：

- 专题：数据处理
 - Stata：变量非重复值统计-distinct
 - Stata数据处理：统计组内非重复值个数（回答人：张铭鑫）

Q26：请问第一个论文例子中动态效应的baseline是什么？答：参考徐老师（2015）文章中对文中图3的描述（即动态效应图），该文在讨论动态效应时来自第一或第二大家族的村官上任五年之前的数据合并起来作为动态效应的基准组（reference category）。（回答人：张铭鑫）

Q27：请问宗族和饥荒文章中的县的死亡率数据是从各个县的县志中来的吗？答：根据徐老师（2022）第三章第一节中对数据来源和计算方法的描述，县一级层面的死亡率（每千人中的死亡人数）通过使用人口统计数据以及地方公报数据计算得出。（回答人：张铭鑫）

Q28：为什么第二篇文章的模型里只有交叉项，没有放主效应？答：因为第二篇文章的模型中分别加入了城市（country）和年份（year）层面的固定效应，因此主效应（\$Clan\$ 与 \$Famine\$）的效应被固定效应吸收，所以回归方程中无需再加入主效应，该做法在国内主流期刊中也被普遍接受。（回答人：张铭鑫）

Q29：请问可否请老师再讲解一下第二篇文章的设计和传统did的区别呢？在徐老师（2022）的文章中，所有人都受到了饥荒的冲击，不存在不受饥荒冲击的地区，只是由于宗族的存在导致不同地区所收到的饥荒的冲击大小不同，因此宗族相当于是一个调节变量（moderator）。在传统did中，一般要求存在不受冲击影响的对照组，由此可见徐老师（2022）的文章并非传统的did。（回答人：张铭鑫）

Q30：请问什么叫赛马的交叉项？

答：赛马，顾名思义就是存在很多个因素可能同时影响被解释变量，这个时候我们可以都扔到模型中看看哪一个因素更加显著。以徐老师JDE的文章为例，文章主要讨论的是宗族和饥荒对于死亡率的影响，但是一方面宗族可能通过影响其他因素影响死亡率，另一方面存在其他的因素直接影响死亡率，所以可以通过控制赛马交叉项\$Z_c * Famine_t\$来控制上述因素，这里的交叉项可以理解为调节效应。通过控制赛马交叉项可以有效控制其他干扰因素，服务与因果推断，同时明确核心解释变量的统计显著性和经济显著性意义。（回答人：王本丞）

Q31：为什么这两篇复现的文章，前者算倍差法，后者不算，两篇的本质差别是什么？

答：第一篇文章宗族与公共品提供是一种标准的多期DID设计，核心解释变量是大姓村主任的虚拟变量，满足DID的一般性假设条件。第二篇文章宗族和饥荒对死亡率的影响虽然存在宗族和饥荒两个维度上的差异，但是并不服从DID研究设计的要求，例如处理前后的平行趋势并不能识别因果效应，因为还要其他混杂因素的影响。所以，文中的处理更多的是一种调节效应，并考虑了“horse race”来明确宗族效应的强弱。徐老师认为并非所有在两个维度具有差异的回归都可以被视为DID，现在很多称作“广义双重差分”的研究严格来看并不是DID设计。（回答人：王本丞）

Q32：第一篇文章K的基点怎么选择的？

答：在包含常数项的动态效应模型中，往往需要删除其中一期作为基准期以避免共线性问题，一般选择处理前一期或者距离处理期最远的处理前样本期。徐老师的文章中是把处理前大于5期（\$k < -5\$）bin到一起作为基期，具体的处理可以参考徐老师的代码和数据。（回答人：王本丞）

Q33：在赛马的交叉项\$X1 * X2\$ model中 为什么不把\$X1\$ \$X2\$也加入到model？

答：以\$Z_c * Famine_t\$为例，在控制双向固定效应之后，地区层面的变量\$Z_c\$被县固定效应吸收，时间层面的变量\$Famine_t\$被时间固定效应吸收，因此只需要保留交叉项，这个利用TWFE估计DID设计中只需要保留交叉项是一致的。（回答人：王本丞）

Q34：赛马的这个处理机制，还有没有什么文献还能参考一下呢？

答：提供一些相关的参考文献：

- 江艇，2022：《因果推断经验研究中的中介效应与调节效应》，《中国工业经济》第5期。
- Chen T, Kung J K, Ma C. Long Live Keju! The Persistent Effects of China's Civil Examination System[J]. The Economic Journal, 2020.
- Cao J, Xu Y, Zhang C. Clans and calamity: How social capital saved lives during China's Great Famine[J]. Journal of Development Economics, 2022, 157: 102865. (回答人：王本丞)

问：如果一个政策在2018年部分上市公司实施，2019年就在所有的上市公司全部实施了，可以用多期did吗？如果一个政策一开始就是适用于全部上市公司，没有对照组，是不是不能用did的方法识别政策效应呢？

答：问题1：所有个体最终都进入了实验组接受了干预，这种情况是可以做多期DID的；问题2：如果政策只影响全体，同时找不到其他不受处理的个体作为控制组，同时拥有政策出台前后多期数据。那么建议这样思考：政策虽然影响全体，但影响的强度不一样，依据影响强度不同构建相应指标，采用(连续变量)双重差分方法估计。可参考文献：Nunn.N and Qian.N,2011,“The Impact of Potatoes on Old World Population and Urbanization”, The Quarterly Journal of Economics, 126(2), 563-650.（回答人：陈云菲）

问：老师好，昨天您提到因果推断假设总体有限，而传统估计假设总体无限。请问两者有什么具体区别呢？或者说在有限总体假设下是如何保证估计结果的一致性呢？谢谢

答：因果识别的总体可以是有限和无限的，这里的意思是假定拥有总体的信息，能不能识别总体的因果参数。统计推断，是由于我们在实际中只能得到总体的一部分样本信息，我们能不能利用有限的样本信息去估计能够反映总体参数的信息。实际推断的样本就是有限的，因果识别是建立在总体上的，用有限的信息推断总体的信息叫做统计推断。如果保证一致性，就是我们需要借助各种因果识别的模型(DID/RDD)等等，在各自的识别假设下，去利用有限的样本一致估计出总体的因果参数。（回答人：屠西伟）

问：国内准自然实验的论文中，在进一步研究部分，那个横截面分析是什么意思？这种情景中的横截面是什么意思？谢谢！

答：国内准自然实验论文的进一步研究部分中横截面分析就是异质性分析，依据不同区域、性别、企业性质、要素密集度等不同的截面进行异质性分析。（回答人：冀承）

问：徐老师，请问差分里面控制变量的作用是为了排除自选择吗？和普通回归方法的控制变量有区别吗

答：DID与普通回归中控制变量用来在回归分析中缓解混杂变量对因果效应估计的干扰，DID中更加关注D的处理效应，对控制变量的解释不作为重点。具体关于控制变量的影响、作用和选择，可以参考连享会推文：[控制变量！控制变量！]: <https://www.lianxh.cn/news/5156229fb5147.html>[不用太关心控制变量，真的！]: <https://www.lianxh.cn/news/77d0128e722e7.html>[控制变量越多越好吗？]: <https://www.lianxh.cn/news/7baa597b33b8d.html>[Stata：控制变量与核心解释变量地位对等吗？]: <https://www.lianxh.cn/news/4eb3cb3d50dcd.html>[控制变量怎么选？大牛们的10条建议]: <https://www.lianxh.cn/news/e1324fad69f21.html>（回答人：陈云菲）

问：请老师再解释一下 TWFE的模型设定那里，D是连续变量，如何画的图。

答：先用Y对时间t回归，得到残差序列1，然后用D对时间t回归，得到残差序列2，最后以残差序列1为Y轴，以残差序列2为X轴画图（回答人：郭盼亭）

问：老师好，在双重固定模型设定那张PPT，D是连续变量时，t是不变化的影响？老师是如何把时间这个变量处理掉，不表示在坐标里？上课期间，听得不是很懂

答：当D是连续变量时，分别估计两个方程：首先使用Y对一系列时间虚拟变量进行回归，然后取残差用以时间变量对Y的解释力，即从Y中剔除时间的影响；其次，使用D对一系列时间虚拟变量进行回归，然后取残差用以消除时间变量对D的解释力。在完成上述两步之后，以D的残差做横轴，以Y的残差做纵轴，即得到消除时间变量影响之后的两者之间的关系。（回答人：张铭鑫）

问：老师您好，连续变量这个图如何证明处理效应是常数的，不是很理解，不知道是否可以再解释一下，谢谢！

答：在D是连续变量的情况下，处理效应表现为斜率，可以观察到A、B、C三个组所有点拟合线的斜率是相等的，这就证明了常数固定效应。（回答人：张铭鑫）

问：课程上说双重固定效应的OLS回归是一致估计，但D和X存在多重共线性，这不是会导致评估有偏吗？

答：D(it)和X一般不会存在多重共线的问题，在强外生性等条件满足情况下，双向固定效应的估计才是一致估计。（回答人：屠西伟）

问：2x2DID 中的两个 differences 分别来自于时间和空间。请问为什么2WFE模型也可以看做是DID？如果解释变量是一个虚拟变量，2WFE中的因果效应不就只有一个difference吗？

答：2WFE就是 two-way fixed effects，双向固定效应，包含时间和空间两个维度，分组虚拟变量（处理组=1，控制组=0）；时间虚拟变量（政策实施后=1，政策实施前=0）；交互项 表示处理组在政策实施后的效应。（回答人：陈云菲）

问：倍差法英文是DID吗？

答：是的，倍差法（Differences-in-Differences）（回答人：陈云菲）

问：昨天老师说的提到的贝叶斯合成控制论文是哪一篇？还有什么将贝叶斯方法与因果推断结合的文献吗

答：应该是徐老师发表在Political Analysis的论文：Pang, X., Liu, L., & Xu, Y. (2021). A Bayesian Alternative to Synthetic Control for Comparative Case Studies. Political Analysis, 1-20. doi:10.1017/pan.2021.22

- 专题：Stata教程
 - Stata：贝叶斯方法-bayes（回答人：王本丞）

问：确实如老师答疑时所说，将解释变量滞后一期反而结果更容易显著，但是当期却并不显著，如果审稿人让用当期的再做一次，怎么办呢？

答：如果审稿人让用当期的再做一次，可以先根据审稿人的要求在回复审稿意见中做出来结果，然后根据当期结果跟审稿人解释当期不显著的原因以及滞后一期就会显著的原因。（回答人：冀承）

问：动态效应检验，可以不用全画么？全画出来不满足检验，但如果只画前后两的，就能通过检验。或者如果遇到政策实行前后都不显著，可不可以只画政策实行前的不显著图，然后说政策实行前没有显著差异。

答：这个是不行的，需要给出完整时间，处理前后的时间都要给出。如果平行趋势假设不满足，这个时候需要去思考不满足的原因。（回答：屠西伟）

平行趋势检验的目的在于验证DID模型能够赋予回归系数因果含义的关键假设，因此在动态效应图中作者应当对所用样本区间内各期动态效应予以展示来佐证文章结果的可靠性与可靠性，展示挑选过后的结果或者仅展示部分结果都会使文章结果的信度造成一定程度的损失。更多关于动态效应检验请关注Stata：一行代码绘制平行趋势图-eventdd: <https://www.lianxh.cn/news/e715545930fcf.html> Stata：平行趋势不满足？主成分DID来帮你！-pcdid: <https://www.lianxh.cn/news/aa1435c753163.html> 多期DID之安慰剂检验、平行趋势检验: <https://www.lianxh.cn/news/0a63a4fb8eb70.html> Stata 倍分法: 不满足平行趋势假设咋办?: <https://www.lianxh.cn/news/4a297047ad9a4.html> 多期DID：平行趋势检验图示: <https://www.lianxh.cn/news/63d0c22fd7afc.html>（回答人：张铭鑫）

问：动态处理效应第一项希望可以有个简单的数据的例子

答：可以参考连享会平行趋势检验的相关推文，一行代码绘制平行趋势图-eventdd: <https://www.lianxh.cn/news/e715545930fcf.html> (回答人：陈希)

问：请问在event study上，如果shock是对所有个体同时发生的，是不是无法加时间固定效应了，谢谢

答：如果shock本身是对所有个体同时发生，就是标准的DID形式，即treat×post形式。如果shock的强度在不同个体间具有差异，那么可以利用强度(intensity) DID或连续DID的方式估计，可以加入时间固定效应进行控制。(回答人：王本丞)

问：动态效应图是不是替代平行趋势图，两个图有何差异？

答：我们传统的DID，简单理解为两组，处理时点是固定的，这个时候可以画出事前的平行趋势。在多点DID中，不同处理组，处理时点也不同，这时候难以画出之前的平行趋势图，这时候我们基于时间研究法，我们可以得到在处理发生前的各期的估计系数，例如处理前的第二期(-2).这个对应的估计系数就是在解释相对于基准组的差异，如果不显著，就说明不存在政策效应。平行趋势检验目的也是在说政策发生前没有处理效应的发生。(回答人：屠西伟)

问：关于为什么广义倍差法不是倍差法的解释不是很清楚

答：徐老师认为传统的DID可以从两个角度理解，一个是design的视角，需要满足平行趋势假定；另一个是estimator的视角，即通过回归方程估计参数。“广义”倍差法更多的是一种estimator的估计方法，而不是一种严格意义上的design。(回答人：王本丞)

问：(1) 离散(但非二元)的处理变量，被理解成同一政策但处理强度不同，还是被理解成多个处理变量(对应于多项政策)？(2) 如果将其拆分成多个二元变量，怎么解释含义？是将其中某个二元处理作为基准进行解释吗？

答：如果处理时间相同，但是处理强度不同，比如用离散型数据表示处理的强度，建议将其拆分为多个二元变量，然后都跟基准组去比较，可以参考文章《Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects - American Economic Association》(回答人：陈云菲)

问：省份的时间趋势是什么，怎么加

答：STATA中可以直接利用reghdfe的命令，在absorb选项中，加入(i.province#c.year)。可参考如下推文：Stata：时间虚拟变量还是时间趋势项?: <https://www.lianxh.cn/news/fdbdeab2066e7.html> 傻傻分不清：时间趋势项与时间虚拟变量: <https://www.lianxh.cn/news/5bbe8408904fb.html> (回答人：屠西伟，张铭鑫)

问：请问双向固定效应模型是不是永远优于控制单独的unit或者time的fixed effects models? 如果不是，在什么情况下应该选择普通的 fixed effects models?

答：在固定效应模型中，在选择双向固定效应模型或者个体(时间)固定效应时重点在于研究者对不可观测变量的假设：是否存在不随时间变化但随个体变化的不可观测因素(例如地区的宗教、气候、地理条件)会导致内生性问题？同时是否存在不随地区变化但随时间变化的不可观测因素(经济周期、时间趋势)会导致内生性问题？如果上述两种因素都存在(这往往是一个更现实的假设)，那么双向固定效应模型就是更合适的。如果研究者对于数据影响机制有非常清楚的认识以及证据能够证明只存在某一种不可观测变量而没有另一种，那么单个固定效应在一定程度上也可以接受。关于固定效应模型的更多话题可见如下推文：FE！FE！面板固定效应模型：你用对了吗？<https://www.lianxh.cn/news/6fc15a8a71d8e.html> (回答人：张铭鑫)

双向固定效应模型一般而言是优于控制单个固定效应，因为无论是时间固定效应、还是个体固定效应都不可避免与核心解释变量相关，除非能够给出强有力不控制的证据。固定效应的控制是因果识别方法中消除混杂因素的手段，是在尽力满足核心变量外生性的手段，不是一种机械性操作(回答：屠西伟)。

问：fixed effects中的within estimator和between estimator是各利用组内变异与组间变异，那么 twoway fixed effects 利用的是什么变异

答：双向固定效应，你可以理解为两步去除固定效应的估计，第一步是组内估计，去除个体固定效应，第二步利用第一步demean 掉的各变量去除组间的时间固定效应。(回答人：屠西伟)

问：如何理解did没有协变量不显著，加了协变量显著

答：在多元回归中缺失协变量时可能存在由于遗漏变量偏误所导致的内生性，因此在无协变量的条件下教程交乘项系数可能是有偏的，由此导致不显著。加上协变量后一定程度上缓解了由于遗漏变量导致的内生性问题，因此did回归系数可能发生改变。具体原因需要结合研究者对于识别策略的认知。更多内容可关注伍德里奇《计量经济学导论》第三章内容。(回答人：张铭鑫)

问：i.year#i.province和c.year#i.province操作上很常见，但经济含义上的区别是什么，控制的是哪个层面上的controls，能举个例子吗

答：i.year#i.province去除了在省份层面所有控制变量，包括省份层面随时间变量和不随时间变化的不可观测变量。c.year#i.province只是控制了各个省份的时间趋势，相当于每个省份有个时间趋势，每个省份的时间趋势都不一样。控制i.year#i.province比控制c.year#i.province更加严格(回答：屠西伟)

问：代码中村趋势那里是对应的文章中TANBLE2的village fix effect还是village linear effect？为什么控制village linear effect的列更多？

答：村趋势的中的c.year#i.province是对应的village linear effect。

问：想请问一下老师，聚类的类别最小能否到固定效应的层级？比如如果是个体固定效应，聚类能否聚类到个体层面？

答：聚类 and 固定效应本身并没有对应关系，一般而言聚类层级越高，对于相关性矩阵所施加的参数假设越少，结果越稳健。因此，如果本身控制了个

体固定效应或者双向固定效应，为了结果更稳健，应该聚类到更高层面，例如村庄或县级。

- 专题：回归分析
 - Stata：标准误！标准误！
 - Stata：聚类标准误的纠结
 - areg：允许干扰项随意相关的稳健性标准误
 - Stata：聚类调整标准误笔记
 - Stata：聚类调整后的标准误-Cluster-SE（回答人：王本丞）

问：foreach var in varlist 和 foreach var of varlist区别是什么呢

答：没有本质区别，可以理解为foreach命令的两种语法，具体可见foreach的帮助文档。（回答人：张铭鑫）

问：画动态图可以理解为平行趋势的检验吗，有了动态图就说明用did合理？

答：当前使用DID进行研究设计就必须画动态图（event study），这样的好处在于：一方面可以检验事前处理组和控制组是否满足平行趋势；另一方面可以考察在接受处理后处理效应随时间变化的趋势，例如效应持续性等。有了动态图并不能充分说明DID的合理性，平行趋势检验通过不能说明DID设计完全有效，还需要满足DID设计中的其他假设，例如SUTVA等。可以参考：黄炜，张子尧，刘安然：从双重差分法到事件研究法[J/OL]．产业经济评论. <https://doi.org/10.19313/j.cnki.cn10-1223/f.20211227.002>。（回答人：王本丞）

问：post*treat+post*treat，后面两项post和treat可以在回归中省略吗？

答：当使用双向固定效应模型时，post项与treat项可被个体固定效应和时间固定效应吸收，因此可以省略，但是若模型中没有加入双向固定效应则post项与treat项不可省略。（回答人：张铭鑫）

问：写代码时提到的聚类是什么含义？聚类与不聚类有什么区别，为什么会对结果产生比较大的影响？一般情况下都需要进行聚类吗，应该对什么变量进行聚类，会有什么样的好处？

答：聚类是对回归系数标准误的调整方式，根据徐老师的建议，应当将标准误聚类到发生变化的最小层次，例如当关键变量是地区层面的变量时应当将回归系数的标准误聚类到地区层次。关于什么是聚类，为什么要聚类以及聚类的影响可以参见如下推文：Stata：聚类标准误的纠结：<https://www.lianxh.cn/news/c9dbc1acea75a.html>（回答人：张铭鑫）

问：双向固定效应模型具体和DID是什么关系？是不是可以认为老师今天所讲的第一篇论文例子应用的是多期DID的方法？

答：传统的DID研究设计可以直接利用OLS进行估计，在面板数据的情况下我们一般使用TWFE（双向固定效应）模型，通过控制双向固定效应可以吸收一次项，只需要保留交叉项即可。可以参考文献：黄炜，张子尧，刘安然：从双重差分法到事件研究法[J/OL]．产业经济评论. <https://doi.org/10.19313/j.cnki.cn10-1223/f.20211227.002>。（回答人：王本丞）

问：请问老师可以再讲一遍partial out time的操作吗

答：当D是连续变量时，分别估计两个方程：首先使用Y对一系列时间虚拟变量进行回归，然后取残差用以时间变量对Y的解释力，即从Y中剔除时间的影响；其次，使用D对一系列时间虚拟变量进行回归，然后取残差用以消除时间变量对D的解释力。在完成上述两步之后，以D的残差做横轴，以Y的残差做纵轴，即得到消除时间变量影响之后的两者之间的关系。（张铭鑫）

问：平行趋势检验与动态处理效应有什么不同，感觉图非常像

答：平行趋势检验和动态处理效应都可以通过事件研究法得到。动态处理效应图可以反映以下两点信息：一是处理前是否满足平行趋势假定，即系数是否显著；二是处理后处理效应的持续性。（回答人：王本丞）

问：以DID为例，政策会对我的output产生影响，在这种情况下如何理解Yi独立于Di这样的处理变量条件强外生假设？

答：在这利用Rubin的潜在因果框架更有助于理解，徐老师这里强调并不是可观测的Yi条件独立于Di，而是潜在结果条件独立于处理变量Di，换句话说个体是否接受处理的潜在结果本身是独立于实际可观测的处理状态的。现在的研究中我们一般可以使用Rubin潜在因果框架定义和识别处理效应。（回答人：王本丞）

问：画平行趋势图，国外大多把t=-1作为基准组，但国内是把处理变量和每一期的交互项都放进去，这两者有什么差异，哪一个做法更科学？

平行趋势检验需要有一期作为“标杆”（基准组），然后将其他各期与这一“标杆”进行对比，从而判断处理组与控制组的差异是否随时间发生了显著变化，关于基准组的选择主要有两种：第1期（基期，最开始的一期）或者-1期（政策时点前1期）。如果选择基期作为基准组，那就是以政策第1期为“标杆”，我们需要丢掉第1期，如果选择-1期作为基准组，那就是以政策前第1期为“标杆”，需要丢掉pre1，这样可能就会显得有些突兀，突然从-2期就跳到第0期了，所以很多人都会在绘图时都会加上-1期的估计系数和置信区间。所以两种方法都是可以验证平行趋势的，可以自行选择（回答人：陈云菲）

问：如果D是一个因子分析后的综合指标，bata系数应该怎么理解

答：因子分析将原有具有单位的指标无量纲化，D这里可能只能解释变化一个单位，y变化多少。（回答：屠西伟）

问：通常要满足D与Y是满足平行趋势的，也允许x与d相关；但如果想排出个体会对D的实施有策略调整，即Xi调整，想看看D是否对X产生影响，这时候要满足平行吗？即便D对Y的影响控制了X。比如在政策正式实施前就有政策的announce，reviewer想看observed effect不受个体对xi调整的影响

答：首先X的控制变量是事前变量，意思是X对D可以产生影响，但D对X产生影响是不行的，会产生Bad control的问题。

问：老师好，我想请教与fe得到系数解释问题，由于字数限制我分为几次提出。谢谢！一方面，我看到有论文指出由于fe通过demean数据计算的，所以解释系数magnitude时需结合demean后数据的variation来解释。比如，系数是1，demean后x的sd是0.5，那么x对y的影响应该是系数demean后x的sd=10.5=0.5。另一方面，我有点困惑的是fe是估计方法，所以用fe估计的系数也是原始回归的系数，那么在解释系数影响的时候是否需要像这些文献建议的结合demean数据的variation解释呢？

答：在FE估计中的demean是time-demeaned data（时间均值），而不是x减去每一年的均值，因此，fe得到的估计系数是原始回归系数，可以参考Wooldridge《Introductory Econometrics》14章。如果您的经过了处理在解释系数的时候需要考虑demean对数据variation的影响，具体可参考文献：deHaan, E. 2021. Using and Interpreting Fixed Effects Models. Working Paper.（回答人：刘佳宁，陈云菲）

问：为什么动态设定能够检验强外生性假设？

答：动态效应可以检验事前处理组和控制组是否满足平行趋势，如果估计系数不显著表明两组并不存在趋势上的显著差异，事前的平行趋势假设得到满足。需要说明的是，强外生性假设是定义在不可观测的潜在因果框架下的，并不能借助可观测数据进行完全检验。平行趋势检验只是强外生假设的一个推论，并且真实的平行趋势同样无法检验，只能借助事前的平行趋势提供部分证据。（回答人：王本丞）

问：为什么双向固定效应模型可以被理解为事件研究法？

答：双向固定效应模型是一种估计方法，DID、交错DID或者事件研究法都是一种研究设计。在满足识别假设的条件下，我们可以使用双向固定效应模型识别我们想要的因果效应参数。双向固定效应既可以用于DID研究，也可以用于事件研究法。可以参考文献：黄炜，张子尧，刘安然，从双重差分法到事件研究法[J/OL]. 产业经济评论. <https://doi.org/10.19313/j.cnki.cn10-1223/f.20211227.002>.（回答人：王本丞）

问：我的问题是由于一些文献指出 within unit 或者 within year 的比较，所以多维固定效应是否不能完全排除 unit-level heterogeneity and time shocks，但是可以在很大程度上排除或者控制这些因素（比如在双向固定效应下，因为 twfe 相当于 two way mundlak estimator）？

答：问题1：多维固定效应最终其实是个体去均值后的结果，在加入个体固定效应后，我们将回归限制在“对X组内变异的分析”，这个过程不仅消除了随个体变化不随时间变化的不可观测变量，也消除了随个体变化不随时间变化部分可观测变量，因此不能完全得到between unit的比较。问题2：多维固定效应可以排除一定不可观测因素，但无法完全排除。例如在加入个体、时间固定效应后，我们只排除了随个体变化不随时间变化的效应以及随时间变化不随个体变化的效应；但对于随时间、个体变化的效应是不可以排除的。（回答人：刘佳宁）

我感觉你的理解是对的。TWFE还是一种很好的方法，可以很大程度上控制住confounders. 如果在模型设定的时候，允许异质性随个体和随时间都改变的话(fully saturated model)，可能会因为 parameters > observations 而导致你的回归方程估不出来。所以你说“无法完全排除个体和时间层面的处理效应异质性”也是对的。Wooldridge, J. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. Available at SSRN 3906345. 可以参考一下 wooldridge 这篇文章，感觉大佬还是很力挺 TWFE 的，尽管现在出了一大波新文献来说 TWFE 估计得beta是有偏误的。（回答人：郭楚玉）

问：有的文献中利用exp(beta)-1解释x增加一个标准差导致y增加多少，beta 是x的系数值，请问这中做法的理由是什么？

答：如果回归方程的设定形式为 $\ln(y) = \beta x + e$ ，所预计的y的精确比例变化为 $y = \exp(\beta x) - 1$ ，乘100后变成百分比变化，即x变动1单位，y变动 $100(\exp(\beta) - 1)\%$ 。具体可以参考伍德里奇《Introductory Econometrics》第6章。对于系数的具体理解需要结合回归模型的设定进行分析。关于回归系数解读的更多内容可见推文：正确解释：回归系数的解释与评估：<https://www.lianxh.cn/news/fcbcc40a9fc8d.html>（回答人：陈云菲 张铭鑫）

问：今天讲的双向固定效应模型设定中为什么没有常数项？

答：常数项本质上是一个既不随时间也不随个体变化的变量，在加入个体和时间固定效应后可以常数项可以被固定效应吸收。（回答人：张铭鑫）

问：面板数据中如果某个个体，在前后两段时间里处理变量为1，例如D随时间的变化是111000111000或者是001110001111100这种，那么在动态处理效应中应该如何编码？

答：按照你的例子，这是一种存在进入退出情形的DID设计，即同一个体在接受处理后还可以从状态1退化为状态0的情形。Goodman Bacon (2021) 在附录部分讨论了这一情形可能存在的偏误问题。另外还可以参考以下几篇文章：Andrew Goodman-Bacon, 2021. Difference-in-differences with variation in treatment timing. *JoE* 225, 254–277. <https://doi.org/10.1016/j.jeconom.2021.03.014> 徐轶青老师在imputation的视角下提供了处理方法：Licheng Liu, Ye Wang, Yiqing Xu, 2022. A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data. *American Journal of Political Science* n/a. <https://doi.org/10.1111/ajps.12723> 此外最近许文立老师发表在《财政研究》的文章中讨论了这一情形：田淑英,孙磊,许文立,范子英.绿色低碳发展目标下财政政策促进企业转型升级研究——来自“节能减排财政政策综合示范城市”试点的证据[J].*财政研究*,2022(08):79-96.DOI:10.19477/j.cnki.11-1077/f.2022.08.008.（回答人：王本丞）

问：残差的序列相关和误差的序列相关有什么区别？

答：需要区分残差和误差的区别。简单来说，误差是由于模型设定本身存在的问题，例如研究教育回报率时，如果不控制能力就会导致误差项中包含能力等因素。残差是指在使用回归方程进行估计的过程中不可被解释变量解释的部分。误差针对模型设定，残差针对回归使用的具体样本数据。（回答人：王本丞）

问：残差或者误差的序列相关，不违反twfe的强外生性假设吗

答：误差的序列相关性对于统计推断存在影响，所以在实际中我们需要对标准误进行调整。一般来说我们通过聚类标准误处理可以有效解决误差的序列相关性问题。这是从Duflo（2004）的经典文献以来的常用做法。Marianne Bertrand, Esther Duflo, Sendhil Mullainathan Reviewed work(s);

2004. How Much Should We Trust Differences-in-Differences Estimates? QJE 119, 249–275.
<https://doi.org/10.1162/003355304772839588> (回答人：王本丞)

问：第一篇文章加入省级变量和时间变量交乘项的目的是什么？

答：控制包括省份层面随时间变量和不随时间变化的不可观测变量。（回答人：陈云菲）

Day3-FAQs

2022 因果推断专题 10 月 22 日课程提问

Q1：为什么大多数工具变量不被认可？

A：因为工具变量需要满足相关性和外生性假设，而外生性假设通常很难被满足。（助教：初虹）

Q2：请问强外生假定是在保证什么结论，是面板数据的参数识别还是因果参数的识别？

A：从因果效应识别过程角度看，强外生性假设的目的是为了说明在控制可观测和不可观测因素后，处理状态与结果变量满足相互独立。因此，处理组与控制组样本满足随机分配，两组组间差分就如同随机试验一样，衡量了政策效应。从系数估计角度看，强外生性假设恰好使得估计结果满足无偏性需求。但是，这个并不是重点所在，关键在第一个。（助教：李鑫）

Q3：修改后的强外生性很像 CIA 假定，二者是等价的吗？

待补充

Q3：老师说因果推断中有异质性，但是在传统的双向固定效应模型中并没有直接涉及到异质性（需要分组或者交乘项）。那么如何理解因果推断中的异质性呢？

A：1.若是对于不同群体的不同处理效应，可以采用分组的DID回归或者三个交互项的DDD设定；2.由于处理效应本身在组间还或者不同时期可能会变化，即同质性假设不成立，最新研究已经提供了不同的解决方案，老师接下来会讲到。Chaisemartin and D'Haultfoeuille(2022)的文章提供了详细的survey。（助教：杜静玄）

Q4：广义倍差法也是一个交互效应模型，可交互效应模型不是要求将所有交互项的自变量都放进模型中吗？徐老师那篇文章只有 Clan 和 Famine 的交互项，为什么没有单独的 Clan 和 Famine呢？,,,A:Clan和Famine的单独项分别被County和时间的固定效应吸收了。

A:Clan和Famine的单独项分别被County和year的固定效应吸收了。(回答人:陈贤孟)

Q5：连续处理变量与广义倍差法之间的区别是不是连续处理变量中 $treatshock$ 中的 $shock$ 只有部分人为1，而广义倍差法中 $treatshock$ 的 $shock$ 全部人都为1？

待补充

Q6：关于动态效应图，Sances 2016 文章中仅仅对时间虚拟变量回归，但是徐老师的文章是对时间虚拟变量和 Clan 的交互项进行回归，请问为什么有这个差异？

A：首先很抱歉，由于通过Sances 2016搜索到的文章比较多，不确定具体是哪一篇。一般对时间虚拟变量做回归可能是为了求结果变量的均值，不知您提到的那篇文章是否同样是为了求两组结果变量的均值才对时间虚拟变量做回归。同样对时间虚拟变量做回归的还有陆毅老师发表在AEJ的关于Markup dispersion的文章，该文章通过对时间虚拟变量进行回归画了两组结果变量均值的时间趋势。而对交互项进行回归主要是求两组结果变量的差异。(回答人:陈贤孟)

Q7：徐老师能就刚刚您提到的“机制要清晰”这里谈谈您在写作中的经验吗？大多数的机制阐述都是通过文献回顾提出研究假说的形式，感觉都是先做出来结果再回来写的假说，有点刻意。您一般是怎么做到“机制要清晰”的呢？

待补充

Q8：被解释变量是Ypost-Ypre、解释变量是Treat的回归与被解释变量是Y、解释变量是Treat*Post的回归有什么区别呢？,第一个回归比如附图中的Bleakley (2010)。

A：dY对Treat的回归接近FD（一阶差分模型），若满足强外生性假设，FD与TWFE的参数估计应该是一致的。若不满足强外生性假设，参数估计会不一致甚至符号相反，解决办法是使用GMM或者IV估计量。Grieser和Hadlock(2018)认为，可以通过比较FE和FD估计(或者IV估计)中的FE-2SLS和FD-2SLS来判断是否违反了严格外生性假设。（回答者：杜静玄）

Q9：我看到一种说法是，协变量不受到处理变量的影响，而是影响处理变量。例如处理变量是是否入选职业培训，协变量是年龄，性别等。这些不受到处理变量影响，但会影响是否被选入职业培训。我想问这种说法是对的吗？

A：是对的。需要被控制的协变量一般是既能影响处理变量，也能影响结果变量的变量。（回答人：胡煊翊）

Q10：能详细解释一下 $aw = 1/se_occscore$ 是起什么作用？

连续DID中，早期cohort的地区数量、观测值比较少，所以估计的标准误就比较大，就会给这些观测值较少的比重。采用analytical weight，标准误越大，weight越小。（整理：宋慧慧）

Q11：老师在reg的时候，选项后面加入了 $aw=cellsize$ ，请问一下，aw的权重设定有什么用呢？我们后续研究中是否有必要用到呢？

已在微信群内解决。

Q12：请问怎么区分参数模型和非参数模型？

参数模型：通常假设总体服从某个分布，这个分布可以由一些参数确定，如正态分布。非参数模型：对于总体的分布不做任何假设，无法得到其分布的相关参数。参数模型和非参数模型中的“参数”并不是模型中的参数，而是数据分布的参数。

当你拥有许多数据而先验知识很少时，非参数学习通常很有用。（助教：巴宁）

Q13：请问在疟疾与劳动生产率的论文中，最后回归，为什么要控制队列的平方以及三次方

待补充

Q14：最后回归，检验系数与exposure关系，权重设置（ $aw = 1/se_occscore$ ）什么意思？

已在微信群内解决。

Q15：请问老师，若两个shock发生的同一个处理年份，但我只关心其中一个shock的处理效应，该怎么办？我需要把这两个shock对应的处理变量各自都放入TWFE中吗？

A:如果可以区分两个shock所针对的不同人群，那么可以区分其中一个shock的效应，但是两个shock针对的人群有可能有重叠，因此很难只关心其中一个shock（整理：宋慧慧）

Q16：请问老师，怎样去检验是否有处理组到控制组的spillover effect？即，控制组看到处理组被干预了，控制组会发生行为改变。

A：一个比较原始的处理方式，你就加一些这个special special的控制，但是这并不是最好的方式，最好的方式还是做一些principle。（解答：徐轶青老师；整理：刘源）

Q17：在今天的代码中，回归通过 $aw = weight$ 的方式赋予了权重（两次回归权重还不一样），这种具体而言是如何的（是weight越大，其样本在回归时权重越大吗），我们实际在做研究时，一般要考虑哪些权重呢

待补充

Q18：did中控制了个人固定效应后，控制变量中的性别就被omitted掉了，如果想要性别这个控制变量，要怎么办呢，是不能控制个人固定效应吗？

A：控制个体效应，性别自动就被控制住了，因为个体异质中包括性别的异质，不用再控制性别了（回答：刘源）

Q19：请问疟疾那篇文章，美国，巴西，哥伦比亚，墨西哥四张图中，除了巴西，其他国家在1900年以后的相关性的均值那条线并不等于0，有的为正的有的为负的，请问这怎么解释？

待补充

Q20：用两年的面板数据做did，并且没有前几年的数据，需要做平行趋势检验吗？

A：两期的面板数据没有办法做平行趋势检验，但是可以做事前变量的平衡性检验。（回答人：初虹）

Q21：请问若我的处理组的shock强度在时间窗口内发生了变化怎么办呢？比如shock(e.g., 补贴)初次发生时，强度为100元，第二年强度上升到200元。控制组则一直没有受到shock。

待补充

Q22：我们在学习计量基础时，会对模型内加入的X（不区分处理变量还是控制变量）的共线性做一定要求，否则会对系数含义造成影响。现在的面板研究中好像不再强调这一点了，只要不完全共线被omit就行，请问是为什么呢？

A：控制变量X本身就与处理变量D相关，如果要求X完全非共线，则R方会很低，因此不必过多关注共线性（回答：刘源）

Q23：前边的动态图是一个回归中用一系列虚拟变量来做，疟疾的这篇是以年度分组分别回归，区别是什么，什么情况下用什么方法更合适呢？

待补充

Q24：因变量和处理变量都是二元变量，能做DID吗，DID方法对因变量类型有要求吗？

A：可以，DID方法对因变量类型没有特殊要求，连续和二元变量都行。（回答人：杜静玄）

Q24：如果政策发生前实验组和对照组画出来的动态效应图中还有差异，但是反而政策快实施的时点动态效应图上都变成0了。去掉控制变量是这样的，这意味着两组都没有什么差异？政策影响不成立么？还是有什么原因和解决办法？

待补充

Q25：用coefplot画动态效应图的问题，图上就会自动把基期那一年忽略了，推文和网上找感觉也没有解决这个问题，想请教老师，用coefplot如何设置代码才能实现把基期是0的那一期也画到图上？

A：一般回归的时候，不放入代表treat乘以基期的虚拟变量那一项。如果要把0画出来的话，需要人为生成一个变量为0，然后放入回归中，同时结合coefplot的omit选项，就可以画在图上。下面的是一篇RES的文章的代码，zero表示生成了一个变量，为常数0($gen\ zero=0$)，后面 $zero=1$ ，表示基期为事前一期。具体可参照原文图1。

代码示例: `reghdfe ihsav impt_3 impt_2 zero impt0 impt1 impt2 if idobs==1&(inrange(yrsfromref,-3,2))|reform_year==.)`, `abs(id year agritaxe`

参考文献: Chari, A., Liu, E. M., Wang, S. Y., & Wang, Y. (2021). Property rights, land misallocation, and agricultural efficiency in China. *The Review of Economic Studies*, 88(4), 1831-1862.

(回答人: 陈贤孟)

Q26: 请问倍差法模型里参数的平均值具体是怎么算出来的?

建议您阅读伍德里奇《计量经济学导论》第15.3小节, 有详细的公式推导。(巴宁)

Q27: 老师好, 请问无预期假设和平行趋势假设有什么区别呢? 我似乎有些分不清

待补充

Q28: 冲击一定要是外生的吗, 看到有些文献也会用明显自选择的变量。不满足无预期假设时怎么做

自选择问题是在Rubin(1974) 因果效应模型中的问题, 解决措施需要确保处理组和控制组样本的相似性, 方法如: PSM, 熵匹配、平行趋势检验等。无预期假设是为了说明政策效果不会在干预前影响处理组和控制组。两者之间的作用目标存在差异。(助教: 李鑫)

Q29: 请问老师如何检验 无预期假设 是成立的呢?

一方面可以从逻辑上验证, 如某项政策在实施前没有提前若干期发布公告, 另一方面可以观察动态效应检验, 假如冲击发生前人们已经有了预期并提前采取行动, 那么在冲击前几期可能动态效应就已经显著了(回答人: 胡煊翊)

Q30: 如果不满足非混淆性假设, 存在同时影响处理变量和结果变量的混淆变量, 要怎么处理呢?

待补充

Q31: 老师提到的瘟疫的文章是不是截面数据, 这和Duflo在2001年研究印尼兴建小学的政策研究是类似的吗? 都叫队列DID。

A: 是的, 类似的还有Nancy Qian关于消失的女性的论文还有, 范子英老师AER上山下乡的论文(回答人: 陈贤孟)

Q32: 请问: 省份面板, 2018发生shock。但2020年开始有疫情, 疫情对于不同的省份影响不同, 所以各地疫情严重程度是一个time-varying confounder吗? 需要额外去控制Covid吗?

待补充

Q33: 动态效应的结果不满足平行趋势, 如何理解并解决呢? 例如, 处理之前两组的差距显著为负, 且差距的值一直稳定, 处理之后差距变为显著为正, 能理解为该处理逆转了发展趋势吗? did的结果还有意义吗?

待补充

Q34: 点估计是属于非参数估计吗? 比如一个点估计例子: 用样本均值代替总体均值, 不考虑总体分布, 认为总体分布和样本分布是一样的。还是说没有模型没有参数就算是非参数估计?

A: 这是两个概念。点估计是用一个数来估计目标参数, 要求样本具有代表性, 总体分布和样本分布不可能是一样的。非参数估计的识别不依赖于参数假设。参数估计: 要求函数的数学模型形式已知, 如假定研究的问题具有正态分布或二项分布, 再用已知类别的学习样本估计里面的参数, 参数检验是利用总体的信息针对参数做的假设, 但这种假定有时并不成立。非参数估计: 在不考虑原总体分布, 或者不作关于参数假定的前提下, 不假定数学模型, 直接用已知类别的学习样本的先验知识直接进行统计检验和判断分析的一系列方法的总称。以样本信息对总体分布作出推断, 针对总体分布情况做的假设。非参数估计不假定数学模型, 可避免对总体分布的假定不当导致重大错误所以常有较好的稳健性。(解答: 徐轶青老师, 整理: 梁淑珍)

Q35: 省份面板数据是季度为t单位, 问题是处理变量和结果变量都有对应的季度数据, 但控制变量只有年度数据, 请问该如何赋值控制变量的季度数据?

如果是结果变量和处理变量分属于不同的频次, 那么回归结果会由于频次差异导致估计偏误问题。解决措施可以参考股票交易的文献。但是, 如果控制变量存在频次差异问题, 这个对于结果并不存在影响。(助教: 李鑫)

Q36: 回归代码可以同时添加cluster和r吗? 有必要吗?

不能同时使用vce 和cluster, 只能使用cluster或者vce(robust), 二者都是计算聚类稳健的标准误, 从而消除异方差对估计系数标准误造成的偏误。cluster(var)是按var聚类计算标准误, 考虑了组内相关, 而vce (cluster var) 是考虑了组内相关的稳健标准误。(助教: 巴宁)

Q37: 能否介绍一下什么叫bad control? 坏的控制变量会对估计产生什么影响, 谢谢?

简要介绍的话, 坏的控制变量常见情形为X影响Y的中间变量, 具体需结合因果图进行理解。可参考连享会推文《Stata模拟: 控制变量! 控制变量! Good-Controls-Bad-Controls》(助教: 陈贤孟)

Q38: 请问如果我只有一个横截面数据, 研究学生家庭收入对其学业表现的影响(使用班级固定效应模型, 在此情况下应该在个体层面聚类还是班级

层面聚类呢？比较困惑因为这类研究并没有shock可言~

关于聚类问题最低需要聚类到截面层级，即学生层面，将家庭收入的时序相关性纳入标准误的估计中。如果需要进一步考虑班级之间的界面相关性，那么可以进一步聚类到班级层面。这里的关键在于文中分析时，讨论模型可能的问题，并根据相关问题，采用不同的措施予以应对。诸如您所研究，您可以先忽略重点班和普通班之间的异质性，首先解决家庭收入的序列相关性，然后进一步将班级之间异质性考虑在内。这样可以逐步解决模型潜在问题。（助教：李鑫）

Q39：老师提到DID只需要满足平行趋势假设，但今天提到了DID的三个假设，研究中是否需要也考察另外两个假设的推论能否成立呢？是否可以理解为，只做平行趋势这一个规定动作是不够的？

我不太清楚您的问题。如果您的问题是：SUTVA和无预期假设和平行趋势假设之间的关系的话，那么可以简单参考一下分析。SUTVA是将Rubin(1974)因果模型与观测值联系在一起。无预期假设是为了排除干预前处理组合控制组均不受到干预影响的情形，保证干预后的政策效应是政策干预的结果。而平行趋势假设是为了消除Rubin(1974)因果效应模型中的选择性偏误问题，即平行趋势检验通过，则近似认为选择性偏差为零，或者说我们无法拒绝政策效应是由政策干预所致。政策效应近似满足无偏估计。（助教：李鑫）

Q40：徐老师2022年这篇文章中提到对数据进行归一化处理，想问下老师一般什么时候对哪些数据要进行归一化处理？为什么以及如何对数据进行归一化处理呢？

待补充

Q41：treatment是连续的采用did回归中将dummy变为连续变量的做法，或是参考AEJ疟疾的做法，或是did with continuous treatment的做法都等价嘛？

这两种处理方法可以作为稳健型检验进行。（助教：李鑫）

Q42：DID不允许随时间变化的不可观测变量导致的自选择问题，请问有哪些方法可以缓解这一问题，能否推荐几篇国外TOP5期刊上的相关文献？

自选择问题主要是通过模型设定中进行有效规避。或者，您可以采用安慰剂检验的方式进行间接验证。您核心目的是为了将潜在影响因素均予以检验，说明您的结果并不会因为这些因素而产生较大变动。（助教：李鑫）

Q43：如果政策干预是在二位数行业层面，被解释变量是在企业层面，标准误聚类到二位数行业层面估计系数不显著，那么将标准误聚类到“二位数行业X年份”层面的做法是否可行？

您只需要说明使用该种聚类方式的目的是解决什么问题，或者您的问题与某篇文献相似，您参考了它的做法。（助教：李鑫）

Q44：第一讲中提到的贝叶斯估计量与因果推断方法结合的文献能否推荐几篇？

[1] Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic books. (强烈推荐，贝叶斯网络之父朱迪亚·珀尔著作，有中文版)

[2] Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1), 217-240.

[3] Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Analysis*, 15(3), 965-1056.

[3] Best, H., & Wolf, C. (Eds.). (2013). *The SAGE handbook of regression analysis and causal inference*. Sage.

（助教：巴宁）

Q45：拿到样本取值后，只有一个样本均值。在只有一个数的情况下，怎么计算分布的方差呢

A：这时候可以去计算残差，有了残差项可以推断干扰项的标准误。通过数学计算获得 β 在 N 趋近于无穷大时的表达式，再对误差项的方差做估计。（解答：徐轶青老师，整理：梁淑珍）

Q46：能简单讲下jackknife

A：可以参考 Stata 中的帮助文档。简单来说，bootstrap 是全部都是放回抽样，jackknife 是每次随机丢掉一个观测值后得到一个新的样本。（解答：徐轶青老师，整理：梁淑珍）

Q47：老师请问最低工资案例那如果固定效应加在店面层面，标准误聚类到州层面可行吗

A：在这个例子中是2*2，固定效应加在店面层面和州层面没有太大差别。标准误聚类到州层面不可行，因为只有两个州，聚类数目太少会导致程序运行不起来。（解答：徐轶青老师，整理：梁淑珍）

Q48：一直不明白自由度的含义，虽然常用

A：在线性回归中，根据样本数减去待估参数个数，要估计多少个参数就要消耗多少个自由度。举个极端例子：有100个样本，模型中的待估参数也有100个，那么就要消耗掉100个自由度，此时模型估计没有预测能力。但一般来说，实际情况中只需要估计有限个参数。（解答：徐轶青老师，整理：梁淑珍）

Q49：徐老师，我想再确认一下，省级面板数据都用bootstrap吗？

A: 我会使用 cluster 的 bootstrap，这涉及到审稿人教育的问题，不是所有的审稿人都理解这件事情。如果目标是较好的期刊，至少要做 cluster，和 cluster 之后的 bootstrap 比较一下，cluster 的 bootstrap 在省层面的标准误通常会比 cluster 的标准误大，这时候仍然显著，这可能是一个比较好的方式。(解答：徐轶青老师，整理：梁淑珍)

Q50：如果动态效应检验结果只有政策实施当年，即第一期显著，然后就不显著了，可以说政策的效果只持续了1期么？

A: 第二期不显著可能是由于样本量比较小，也有可能第二期存在效应但效应具有异质性。动态效应显著性下降是显然的，关键在于动态效应证据能否支撑平行趋势假设的正确性。如果 point estimate 都回到0，那么可以说第二期以后就没什么效应了。(解答：徐轶青老师，整理：梁淑珍)

Day4-FAQs

2022 因果推断专题 10 月 23 日课程提问

Q1：协变量的意思可以理解成控制变量吗？

A: 协变量属于控制变量的一种。

控制变量是在实验中一定要保证不变的量；协变量在实验中不一定要控制不变，或者无法使其不变，其改变对实验结果不是很敏感。（图片来自维基百科）

ANCOVA EXAMPLE

Independent Variables

(Factor)

Level of Education
(High School, College Degree,
or Graduate Degree)

(Covariate)

Number of Hours
Spent Studying

Dependent Variable

(Response)

Test Score

(解答：巴宁)

Q2：请问如果审稿人明确指出有协变量受到干预政策的影响会发生变化，这种情况要如何应对呢，是直接丢掉协变量吗？

A：您可以将斜变量作为结果变量进行回归，说明其不具有统计显著性。同时，经济显著性很低，即使存在估计偏误，那么也很小。另一种思路是，首先通过结果变量对该斜变量进行回归，取残差后，控制该斜变量的影响后，进行分析。（解答：李鑫）

Q3：怎么选取这里的协变量呢

W3

ebalance npc price2007 roa2007 roa2006 roa2005 margin2007 margin2006 margin2005 rev2007 rev2006 rev2005 so_portion gind_ind* debrati

A：这里的协变量选取的目的是为了平行趋势满足，可以是关心的被解释变量，以及其他影响被解释变量的变量。（回答人：秦利宾）

Q4：能不能后面介绍一些多期DID使用熵平衡法或者Matching+DID的例子，比如TrueX这个例子用了处理期之前的DV做加权，但如果多期的时候，每个处理组处理的时不一样怎么做匹配和平衡呢？

[Stata-Matching：肾脏交换匹配问题](#)

[Stata：广义精确匹配-Coarsened-Exact-Matching-\(CEM\)](#)

Q5：TRUEX2014这个里面作为截面来做。这个数据是面板，那是用处理前的一年和之前的数据来做截面吗？还是怎样？

A：处理前一年，或者处理前几年的数据转换成截面数据。（回答人：秦利宾）

Q6：请问，用 y , y_1 , y_2 对同一 x 回归，且 $y=y_1/y_2$ ，得到 x 的估计系数分别为 b , b_1 , b_2 ，三个回归其他部分相同，则系数是否需要满足 $b=b_1/b_2$ ，如果不是，原因是什么呢，谢谢

A：不清楚您的目的是什么。导致估计差异可能是 y_1 与 y_2 取比值后，改变了变异程度，如增加了部分离群值，或者降低了离群值，导致了结果估计不

一致问题。（李鑫）

Q7：为什么第二步DID分析没有控制一些变化的控制变量？

A：DID可以加控制变量，也可以不加，可以在一个表格汇总予以对照（唐堂）

Q8：对于ebalance和psm的区别可以理解成ebalance是按照事件前样本的roa，npc，其他控制变量的均值或方差匹配出差不多的处理组和控制组，而psm是用roa，这些变量本身的值去匹配么

A：不是，既然要做匹配DID，一定是根据事前的。（回答人：秦利宾）

Q9：徐老师2015年那篇文章fig1d，里面的三根线也是要先画三根线，再addplot加在上面吗？

A：是的。（唐堂）

Q10：TRUEx文章代码第93行，merge m：m多对多的这种合并是怎么合并的呀，是不是这种多对多合并很容易出问题了？

A：确实会出现问题，建议使用 m:1 或 1:m。（回答人：秦利宾）

Q11：请问老师，若极端情况下比如1000个个体的样本仅10的个体接受了干预，剩下990都是控制组，此时用TWFE得到的DID系数标准误会不会很大？

A：待补充

Q12：在Truex(2014)论文复现，生成了三种权重，想问一下这三种权重的区别，以及如何根据自己的研究确定不同的权重？

A：解答：三种权重是根据不同的斜变量进行平衡处理组与控制组得到的，控制的越多协变量，权重会越集中。（解答：徐轶青老师，解答：宋慧慧）

Q13：请问Truex文章中，第二步进行加权DID估计不需要加控制变量吗？

A：可加可不加（回答人：秦利宾）

Q14：请问stata中啥时候用aweight pweight以及fweight

A：weight 表示赋予观测值的不同权重。Stata 提供了 4 种不同的权重设置方式：fweight 频数权重 (frequency weight)，表示某个观测值重复出现的次数。例如，如果某个观测值的 fweight 等于 5，表示实际上有 5 个与该观测值完全相同的观测值；pweight 概率权重/抽样权重 (probability weight)，表示根据抽样方法，该样本被抽中的概率的倒数。例如，如果某个观测值的 fweight 等于 5，表示该观测值所属的子总体中每个个体被抽中的概率是 1/5，或该子总体的每一个样本能代表 5 个个体；aweight 分析权重 (analytic weight)，通常适用于数据是某个组或某个地理区域的数据均值的情形，aweight 表示计算该均值使用的个体的数量。例如，某 10 万人城市的收入均值为 2500 元，则可以设置该收入均值的 aweight 为 100000。具体的权重设定应该由回归设定决定。

（解答：杜静玄）

Q15：到底何时平衡干预前的协变量？平衡趋势不满足的时候吗？它一般是在基准的分析时用还是后面的稳健分析时用呢？

A：平衡的目的是为了说明满足了强外生性假设中的条件独立性假设。只有满足对应的假设条件，DID模型设定才合理。稳健型检验就是对一些可能的违反条件进行分析。（李鑫）

Q16：请问DID回归时设置的权重是来源于仅该回归中的outcome变量在熵平衡中得到的结果，还是说可以来源于熵平衡 一系列不同的outcome变量得到的权重？

A：主要是与该回归中outcome相关的变量。（回答人：秦利宾）

Q17：徐老师能再讲一下PSM和熵平衡法的区别吗，或者说是哪个更好呢？因为国内文献现在用PSM好像更多一点。

A：熵平衡法与PS weighting在大样本的情况下是等价的，在小样本中熵平衡法更好。（解答：徐轶青老师，整理：袁洛琪）

Q18：看到文献中会控制 处理前协变量水平*trend，请问这样与平衡存在区别吗？

A：用参数估计的方法解决平行趋势假设不成立的情况，与balancing相近，但balancing更好，因为还可以解决overlap的问题；协变量水平*trend更类似回归，平衡更类似匹配；匹配不对方程形式进行假定。（解答：徐轶青老师，整理：袁洛琪）

Q19：上周第一个例子（关于村长选举的例子）表2dependent variable mean是常数项吗？是代表小姓当选的公共品投资吗？我在output中怎么没看到这几个数？

A：待补充

Q20：实操中的小问题，希望助教帮忙解答一下，问题见下图

2.在XY2015里tw命令里option mlalign() not allowed, 是版本的问题吗? 我的是15; 还是我的命令没更新? 更新replace后没有用

```
3.在CXZ2022的tw (sc lnbook_80 lnbook_50 if CR1_high==1, msym(oh) mc(red) ) ///
(lowess lnbook_80 lnbook_50 if CR1_high==1, lc(red) xlabel(0(2)7) ) ///
(sc lnbook_80 lnbook_50 if CR1_high==0, msym(th) mc(blue)) ///
(lowess lnbook_80 lnbook_50 if CR1_high==0, lc(blue)), ///
    ytit("#Genealogies compiled after 1980") xtit("#Genealogies compiled
before 1950") ///
    legend(label (1 "High intensity of CR") label (2 "Corr = `corrcoef1'") ///
    label (3 "Low intensity of CR") label (4 "Corr = `corrcoef0'") pos(11) ring
(0))
```

为什么我导出的图的小数位数就是不对

4.[aw = cellsize]表示什么意思? 对于aw的具体含义还是不太清楚

Q21: 有老师说, 一旦方程加入企业固定效应, 则不用加入企业层面的控制变量, 是这样的吗? 感觉可以加企业层面的随时间变化的变量, 回归时可以只加企业固定效应不加企业层面的变量吗 请问如何决定是否进行blancing。如果blancing会让趋势结果更好, 是不是做did时都是做了blancing更好?

A: 企业固定效应控制的是企业层面不随时间变化的变量, 但如果没有控制企业*年份固定效应, 则需要控制其他企业层面随时间变化的变量。如果因变量是企业-年份层面的变量, 回归时最好控制企业固定效应以及一些企业层面随年份变化的变量。(解答: 杜静玄)

Q22: 请问如何决定是否进行blancing。如果blancing会让趋势结果更好, 是不是做did时都是做了blancing更好?

A: 2. stata 17可以正常运行, 建议更换stata版本

3. 先运行以下代码

```
local corrcoef = round(r(rho),0.01)

/*round: r(rho)四舍五入为两位小数 (0.01) */

/*r(rho): 存储的相关系数的值*/

/*当执行corr的命令后, stata会自动储存一些返回值, 通过return list可以查看;

r(rho)是其中的一个返回值*/
```

4. 这是在根据cellsize进行对观测值加权, aweight 表示计算该均值使用的个体的数量。例如, 某 10 万人城市的收入均值为 2500 元, 则可以设置该收入均值的 aweight 为 100000。

(袁洛琪)

Q23: 请问做匹配只能用干预前不随时间变化的协变量做匹配吗? Truex文章中用到了05年的ROA做匹配, 但是ROA是结果变量

A: 可以选择任何截面维度变量进行匹配。如果考虑时间维度, 那么试着使用干预前斜变量在时间维度的均值进行匹配。(李鑫)

Q24: 能不能讲ebalance的工作原理, 直接将前些年的结果变量放入协变量组, targets (1) 匹配后的预处理时期的均值为0, 是如何调整那些需匹配样本的均值的, 直接调整为与匹配变量相等的均值吗

A: 熵平衡法的原理介绍参见连享会推文 [ebalance: 基于熵平衡法的协变量平衡性检验](#)

(解答: 梁淑珍)

Q25: 老师讲的DID三个要求, 平行趋势、样本间独立和无预期效应, 后两个在实操中怎么做检验? 感觉大多基本都是只检验平行趋势呢

A: 样本间独立和无预期效应没办法直接检验, 多期数据可以test平行趋势、样本间独立和无预期效应的推论。(解答: 徐轶青老师, 整理: 袁洛琪)

Q26：双重固定效应的棘手之处：可行情况：2×2，且无控制变量；不可行情况：多期，或有控制变量。请问这一页ppt中无控制变量是什么意思？2.2中有控制变量是否也与DID不同？

A：解答：若只有不随时间变化的控制变量，用匹配的方法平衡掉或者控制变量*时间dummy进行平衡。（解答：徐轶青老师，整理：宋慧慧）

Q27：今天第二节，数据模拟时，采用cluster和bootstap为啥标准误接近真实值，而均值没有接近1呢？

A：测试不同的标准误估计时，常规标准误、聚类标准误、自助法标准误、聚类加自助法标准误得到的点估计值是相同的，本例中均为0.880，但是标准误有较大区别。在社会科学研究中，假设回归模型中的误差项在集群（cluster）中相关是很常见的。这也是为什么许多 Stata 估计命令提供了一个聚类选项，来实现聚类-稳健方差矩阵估计。该方法对集群内相关性和未知形式的异方差都是稳健的。当聚类的样本数量很大、聚类相当均匀（在大小和方差矩阵方面相似）、以及对照组聚类（treated clusters）的数量不太少时，基于聚类标准误的推断具有较好的效果。但当这些条件不成立时，推断可能是不太准确的。使用 bootstrap。其主要思想是通过生成大量的 bootstrap 样本，来模拟实际样本的分布。通常来说，bootstrap 数据生成过程（DGP）越接近真的 DGP，估计就越可靠。（解答：梁淑珍）

Q28：双重固定效应模型的强外生性假设和OLS的外生性（误差与X不相关）的区别

A：二者本质上都是在处理外生性问题，区别在于首先二者的模型设定不同，其次，双重固定效应模型的强外生性假设从分配角度来说其假设性更强，同时满足没有随时间变化的混淆变量、没有反馈效应、没有延续效应和预期效应、线性参数等。（解答：郑裕璇）

Q29：前边的课给我的感觉是did可以放松强外生性假设，只需要平行趋势满足就可以，这是did的优势。今天的课又去审视did的强外生性，did中强外生性的角色到底是什么？

A：did不能放松强外生性假设；did的优势是可以允许异质性。在实践中平行趋势假设+SUTVA+无预期假设基本等于强外生性假设；在数学中强外生性假设略强（解答：徐轶青老师，整理：袁洛琪）

Q30：Imai&Kim中在没有feedback和没有carryover外，还有没有ldv假设（过去y不影响现在y），这个假设与没有carryover有什么区别呢？可以通过控制过去x处理吗？

A：待补充

Q31：DID政策具有一般均衡效应怎么办，需要怎么检验

A：待补充（唐堂）

Q32：请教下PPT中满足强外生性和下面解释逻辑的关系，（例如采用交错方式）：难道是只要有交错场景就满足强外生吗，就可以用twfe准确估计？

A：强外生性假设是研究设计部分，只有满足强外生性假设条件，那么DID模型设定才是合理的。交错行DID模型的问题主要是存在异质性个体处理效应时，早处理组作为晚处理组的控制组，无法满足平行趋势检验，使得政策估计存在偏误问题。

Q33：不理解老师课上讲的怎么用非参的方式画出那6个cohort的图，能再具体讲一下怎么画？画哪些变量么？

A：待补充（解答：徐轶青老师，整理：宋慧慧）

Q34：为什么平行趋势假设不成立PPT中第三幅图看起来都挺像成立的呢？政策实施后不是开始呈现了显著的正向影响吗？不成立的预趋势影响应该怎么解决呢？更换识别方法吗？

A：政策发生前有几项显著拒绝了系数=0的假设，说明不满足平行趋势假设。可以试试先做PSM再做动态效应检验。（回答人：胡焯翊）

Q35：您好，想请问一下在因果分析里，好的被解释变量和解释变量的标准是什么？二次计算的变量适合作为被解释变量吗，例如全要素这种经过较复杂二次计算的变量，是不是一个好的被解释变量呢？

A：是否是好的被解释变量或者解释变量跟研究主题紧密挂钩，好的（被）解释变量是能解释说明研究问题，并非越复杂越好（回答人：郑裕璇）。

Q36：请问TWFE不可以在等式右边加控制变量是什么意思呢？比如有一个time varying的控制变量Xit，它不影响Dit，但Xit+1会受到Dit的影响，那需要把Xit放在TWFE的等式右边吗？

A：解答：Xit可能会影响D，也同时影响y。Xit+1不能受到D的影响（解答：徐轶青老师，整理：宋慧慧）

Q37：请教一下，第一天论文XY（2015）中是否大姓村长当选，也出现11和00的多次交错，结合这讲内容，请问直接使用TWFE有什么问题，或者需要注意什么吗？

A：- DID最新进展：异质性处理条件下的双向固定效应DID估计量（TWFEEDD）

Q38：一个政策在不同的州交错执行，不同州的政策内容会有差异（条款1,2,3）。即州1的政策只包含条款1和2，州2的政策只包含条款1，州3的政策包含全部1-3条款。如何估计各个条款的影响？

A：看你要分析该政策的哪一方面影响。（回答人：秦利宾）

Q39：政策分两年进行，第一年一部分人，第二年全部人都处理，那么treat 所有样本都是1，只有year这个变量有时间差异，treat*year只有一个变量有变化，还是DID吗？只有时间差异，没用组间差异了？

A：不是did（回答人：秦利宾）

Q40：用did研究政策A的影响，但同一年出现了其他政策B同时影响结果变量Y这种情况下怎么办呢？还可以用did么，还是说想研究的政策A和政策B对于treat组的划分是不一样的，所以同一时期有其他政策没关系

A：这个时候需要排除政策B的干扰。如果想研究的政策A和政策B对于treat组的划分是不一样的，那么可以通过删去受政策B影响的treat组后再估计政策A的影响作为稳健性检验。（解答：袁洛琪）

Q41：在有预期效应的情况下，比如进入政策两年前就观察到了Y的趋势变化。那么能够把前两年也设为D=1吗？如果有相关研究或者文献，请指教。好呢？

A：A：例如某政策冲击在实施前两年已经发布公告，并导致Y的变化，那么可以把政策前两年设为1。（胡煊翊）

Q42：老师提到的敏感性检验是稳健性检验的一种吗？主要针对DID中哪种问题？应该具体怎么实操？

A：解答：敏感性检验主要是在截面数据中做的比较多。（解答：徐轶青老师，整理：宋慧慧）

Q43：老师说的有多期数据可以做pre trend test是什么意思？就是画趋势图看平行趋势假设是否成立吗？

A：是的，就是画趋势图看平行趋势假设是否成立。（郑裕璇）

Q44：请问若在交错冲击的情况下，一个处理组早处理(t=1)，一个晚处理(t=2)，一个在t=1受到处理，但是shock强度在t=2上升，还有一个纯控制组。请问这个情况该怎么办呢？

A：解答：交错did，下一讲会讲处理方法（解答：徐轶青老师，整理：宋慧慧）

Q45：为什么要拿早期处理组和晚期处理组做比较？

A：交错行DID模型中，当处理组个体存在异质性处理效应时，早处理组作为晚处理组的控制组，平行趋势难以满足，是的估计结果存在偏误问题。（李鑫）

Q46：做多期DID的好人培根分解法怎么去在论文中做汇报，需要汇报什么统计量，知道原理但是不知道写论文怎么汇报。

A：下周讲解。

Q47：事件法和DD有什么区别呢？

A：本质上相同。事件研究法经常出现在研究证券市场发生事件冲击导致股票异常收益率时较为常见。（胡煊翊）

Q49：多期did需要做bacon分解吗？

A：需要，下一讲会详细介绍。

Q50：老师讲到今天，您能用今天的知识回头说说您为什么觉得宗族那篇文章不是一个did PPT中Bacon分解那里的权重能再讲一下吗？

A：今天的知识和宗族那篇不是一个did，宗族的那篇文章你基本可以认为是一个截面的问题，因为他不随时间变化，是固定的，包括建国以前有没有宗族，宗族多不多，这是一个截面的数据，来解释超额死亡率，你可以把这个panel压扁，把时间维度去掉。公式的右边是截面上有没有宗族，他就变成一个截面上的回归，是一个select on observable的结构，只是我们把它说成是一个广义did的结构。（解答：徐轶青老师，整理：巴宁）

Q51：老师没有太明白为什么晚处理和早处理组比较时如果是固定不变的就没有关系，如果早处理组效果更大，即使不变不也会有一个反向的结果吗？

A：建议查看回放

Q52：啥时候用aweight pweight以及fweight？

A：可以画图，把早处理组的图像向上平移和晚处理组重叠，如果早处理组存在不随时间变化的处理效应（并满足平行趋势假设），那么晚处理组和早处理组的差值仍然是处理效应；反之，如果早处理组存在随时间变化的处理效应，那么两组的差值并不等于处理效应。（秦利宾）

Q53：徐老师能讲一下做匹配的时候PSM与熵平衡法有什么区别么？感觉看到的文章中PSM+DID的比较多，一般什么时候用PSM，什么时候用熵平衡法呢？

A：熵平衡法与PS weighting在大样本的情况下是等价的，在小样本中熵平衡法更好。（解答：徐轶青老师，整理：袁洛琪）

Q54：老师，请问熵平衡里面将数据作为截面来做的时候，是将所有数据如何截面化，是否可以讲一下

A：是的，将所有希望控制的数据截面化（解答：徐轶青老师，整理：袁洛琪）

Q55：老师，请问为什么在第二步面板回归中没有加入控制变量，第一步平衡的也只是控制变量在政策发生前的平衡，在政策发生后一些时变的控制变量是否需要控制？

A: 因为控制变量不随时间变化，平衡可以用到这样的信息；只能控制政策发生前的变化，不能控制政策发生后的变量。(解答：徐轶青老师，整理：袁洛琪)

Q56: 刚才这个例子，没有看清是如何在横截面里平衡的，具体是横截面的哪一年呢？2. 到底什么时候平衡干预前的处理变量还有些不明白，平衡趋势不满足的时候么？它一般是放在论文中的基准回归的地方还是后面的稳健性等分析中？

A: 2005-2007是用于平衡；部分时候作为稳健性检验，如果有很强的信心平行趋势假设成立，可以不做平衡。相似的，如果不平衡的时候发现存在pre-trend，可以考虑先做平衡。(解答：徐轶青老师，整理：袁洛琪)

Q57: 老师，DID研究设计可以同时研究政策同时对两个因变量的影响吗？这两个因变量是具有相关性的。是简单的分别对y1、y2回归？还是应该有什么体现同时性影响的方法设计？

A: 分开做回归即可。(解答：徐轶青老师，整理：袁洛琪)

Q58: 请问老师，什么情况下建议先做(matching/ ebalance)之后再跑 面板回归呢？如果不做预处理，PTA就可以通过，是不是就可以不做预处理？还是说 都推荐做预处理，因为 预处理之后，event-study plot 感觉会长得更好看

A: 原来就很好，做robustness check；原来不好，先做平衡(解答：徐轶青老师，整理：袁洛琪)

Q59: 如果控制组数量很少，处理组数量很大，是不是就不好用熵平衡的办法了

A: 可以反过来label，处理设定为控制；这时候ATT就是反过来的政策效果(解答：徐轶青老师，整理：袁洛琪)

Q60: 如果不满足非混淆性假设，存在同时影响处理变量和结果变量的混淆变量，要怎么处理呢？

A: 截面情况下，可以选择IV或者RD；或者做敏感性检验。(解答：徐轶青老师，整理：袁洛琪)

Q61: 为什么平行趋势假设不成立的第三幅图看起来都挺像成立的呢？政策实施后不是开始呈现了显著的正向影响吗？不成立的预趋势影响应该怎么解决呢？更换识别方法吗？

A: 因为改革前已经有许多显著的地方，之后发现的差异不能轻易的说是处理造成的，也可能是时间趋势变化(解答：徐轶青老师，整理：袁洛琪)

Q62: 如果一个政策在不同的州以staggered adoption的形式出现，并且这个政策在不同州，它的内容会有差异（条款1,2,3）。比如，州1的政策只包含条款1和2，州2的政策只包含条款1，州3的政策包含全部1-3条款。请问徐老师，如何估计各个条款treatment effect？

A: 如果差异足够大，应该定义为不同的treatment(解答：徐轶青老师，整理：袁洛琪)

Q63: 动态效应的结果不满足平行趋势，如何理解并解决呢？例如，处理之前两组的差距显著为负，且差距的值一直稳定，处理之后差距变为显著为正，能理解为该处理逆转了发展趋势吗？did的结果还有意义吗？

A: Pre-trend不好最好进行匹配或者平衡(解答：徐轶青老师，整理：袁洛琪)

Q64: 如果政策发生前实验组和对照组画出来的动态效应图中还有差异，但是反而政策快实施的时点动态效应图上都变成0了。去掉控制变量是这样的，这意味着两组都没有什么差异？政策影响不成立么？还是有什么原因和解决办法？

A: 解答: 可以进行匹配有差异的结果变量(解答：徐轶青老师，整理：宋慧慧)

Q65: merge为什么是m:m，截面数据中企业id(gvkey)不应该是唯一的吗

A: 解答：其实是1:m，但m:m包括了1:m(解答：徐轶青老师，整理：宋慧慧)

Q66: 感谢徐老师分享。请教一下，第一天论文XY（2015）中是否大姓村长当选，也出现11和00的多次交错，结合这讲内容，请问直接使用TWFE有什么问题，或者需要注意什么吗？

A: 解答：效应异质性可能会造成有偏误，但其实践上问题并不严重。在没有pure control的情况下，异质性会变的非常严重。(解答：徐轶青老师，整理：宋慧慧)

Q67: 徐老师之后可以讲一些多期DID匹配或者熵平衡法应用的例子吗？之前的例子根据处理期前的DV做加权，但是多期DID因为处理组处理期都不一样，怎么根据处理期前的DV或者协变量做匹配或者加权呢？(Q31相同)

A: 解答：拆成非交错DID，分别做平衡。(解答：徐轶青老师，整理：宋慧慧)

Q68: 请问如果有充分理由认为协变量会受到干预政策影响，是否应该直接扔掉这个协变量，但这样做会不会使得估计产生偏误？

A: 解答：如果协变量是一个confounder，丢掉之后会造成强外生性假设不成立，就会造成偏误。(解答：徐轶青老师，整理：宋慧慧)

Q69: 老师，看到Imai&Kim（2019）中在no feedback（过去y不影响现在x）和no carryove（过去x不影响现在y）外，还有过去y不影响现在y的假设，这个假设与no carryover有什么区别呢？可以也通过控制过去x处理吗？

A: 解答：feedback是y影响D；这三个假设都不同。carryover可以通过控制过去的x进行处理，但是基于的是同质性假设。但同质性假设本身可能存在问题，造成偏误。(解答：徐轶青老师，整理：宋慧慧)

Q70.请问TWFE不可以在等式右边加控制变量是什么意思呢？比如有一个time varying的控制变量 X_{it} ，它不影响 D_{it} ，但 X_{it+1} 会受到 D_{it} 的影响，那需要把 X_{it} 放在TWFE的等式右边吗？

Q71.不理解老师课上讲的怎么用非参的方式画出那6个cohort的图，能再具体讲一下怎么画？画哪些变量么？

A: 解答

Q72.如果发生处理效应的反转，比如美国两党之间对气候政策的不同态度，比如奥巴马加入巴黎协定，特朗普退出，拜登重新加入，应该怎么设定模型

A: 解答：处理carryover effect，下一章会讲 (解答：徐轶青老师，整理：宋慧慧)

Q73.在有预期效应的情况下，比如计入政策两年前就观察到了Y的趋势变化。那么能够把前两年也设为 $D=1$ 吗？如果有相关研究或者文献，请指教。

A: 解答：不能，可能会做出任何结果，违法假设。如果是做实验，提前宣布了这项treatment，是可以接受的，但是观察性数据是不被接受的。
(解答：徐轶青老师，整理：宋慧慧)

Q74.请问若在交错冲击的情况下，一个处理组早处理($t=1$)，一个晚处理($t=2$)，一个在 $t=1$ 受到处理，但是shock强度在 $t=2$ 上升，还有一个纯控制组。请问这个情况该怎么办呢？

Q75.Treatment vs Pre-1964 reform states是什么呢？

A: 解答：Treatment是指always treated；Pre-1964 reform states在数据还没开始前已经被treated (解答：徐轶青老师，整理：宋慧慧)

Q75.老师好，粗略的看左边的动态系数图是否就可以判断出处理效应是随时间异质性的

A: 解答：不可以，还是要看shock之后的变化，时间的异质性总是存在的 (解答：徐轶青老师，整理：宋慧慧)

Q76.老师，那种类似"广义倍差法"中如果评估渐进政策？比如，您2022年那篇宗族与死亡率的关系中，如果饥荒冲击存在两期，或几年后又来了来一波饥荒冲击，那该怎么做识别呢？

A: 解答："广义倍差法"本质上是截面上的异质性，没有很好的识别方法。

Q77.看到文献中会控制 处理前协变量水平*trend，请问这样与平衡存在区别吗？

A: 解答:存在。有时候会用处理前协变量水平*year dummy来控制，用参数估计的方法来放松平行趋势假设。例如，男性和女性具有不同的趋势，受干预的人群中男性和女性的比例是不同的。如果分别控制男性和女性的趋势，控制之后就可以得到平行趋势假设。基于条件平行趋势的做法。类似于回归与匹配的差异。(解答：徐轶青老师，整理：宋慧慧)

Q78.我们在学习计量基础时，会对模型内加入的X（不区分处理变量还是控制变量）的共线性做一定要求，否则会对系数含义造成影响。现在的面板研究中好像不再强调这一点了，只要不完全共线被omit就行，请问是为什么呢？是因为无法解决，还是可以忽略这样的影响呢？

A: 解答：在因果推断框架中，除了干预变量和结果变量，其他的变量是为了解决混淆的问题，并不关注这些变量的系数。如果高度相关的话，去掉其中一个即可，因为他们带来的混淆效果也是相同的。(解答：徐轶青老师，整理：宋慧慧)

Q79："beta_TEFE 为基允许异质性，bata本来不就是对每个unit的beta的加权平均，每个unit的beta不就是异质性的，这里的异质性指的是什么呢
matching、PSM、balancing区别是什么。matching在Stata中如何实现呢，也是没有Y吗

延续效应彩虹那个图不明白

A：在经典的2*2的DID中，beta_TEFE是可以收敛到ATT的，但是在非经典的DID中，无法得到ATT，甚至无法得到个体处理效应的凸的加权，权重有可能是负的，因此需要做调整，将在下一讲进行讲解。

差别只是不同的estimate，然后用不同的方差进行调整，这个差别很像我用控制变量，可以用匹配或者回归的方法进行调整，不同方法各有优势，回归的方法效率高，匹配的方法假设性比较弱

matching的方法可以参考往期推文：

- 专题：Stata命令
 - Stata数据处理：模糊匹配-reclink2-matchit-strgroup
- 专题：PSM-Matching
 - Stata：模糊匹配-matchit-reclink
 - Stata-Matching：肾脏交换匹配问题
 - Stata：iematch-近邻贪婪匹配
 - Stata：终极匹配 ultimatch
 - Stata：广义精确匹配-Coarsened-Exact-Matching-(CEM)
 - Stata：模糊匹配之matchit (解答：徐轶青老师，整理：巴宁)

Q80：一般DID的研究设计中，控制变量需要与冲击的虚拟变量相乘吗？

A：这取决于你冲击的虚拟变量是否随时间变化，一般来说如果是 2×2 的DID，你的控制变量可以和冲击虚拟变量相乘，相当于把平行趋势假设放松。

(解答：徐轶青老师，整理：巴宁)

Q81：请问为什么要拿前期处理组和后期处理组作比较，两者不都是处理组吗？在twfe里D不是都是1吗？

A：TWFE就是在拿前期处理组和后期处理组进行比较，实际上是我们偷懒，我们可以将他们拆分成很多 2×2 的DID，然后其中部分DID是有问题的。如果不用TWFE，你自己去选择做比较的组，就不存在这个问题。

(解答：徐轶青老师，整理：巴宁)

Q82：老师，请问在staggered的DID中，就算是同质处理效应，但因为early group作为late group的对照组，这样是不是还是会出现偏误的？

A：如果是同质的处理效应就不会有问题，即使是这样进行对照，依然平行，当存在异质性才会有问题。

(解答：徐轶青老师，整理：巴宁)

Q83：徐老师，您2022年那篇文章中最后的机制分析中，可不可以理解为也是中介渠道分析呢，因为M是可能影响死亡率的因素，就只做 宗族*饥荒 影响M这一步呢？

A：我觉得不是中介，而是调节，差异在于中介变量是受到干预的影响，但是调节是不会收到干预影响，在这里，是否有宗族是不会受到饥荒的影响的，因为宗族在饥荒之前已经存在或不存在。

(解答：徐轶青老师，整理：巴宁)

Q84：老师，card and kruger（1994）中将Ypost-Ypre作为被解释变量、Treat作为解释变量，请问这样与用Treat、Post和Treat*Post回归有什么区别呢？

A：倍差法本质上是先进行pre-post，对每个人先做两期前后比较，得到一个差值，然后用这个差值作为被解释变量跑回归，这个结果在数学上和TWFE是没有区别的。

(解答：徐轶青老师，整理：巴宁)

Q85：您刚才提到STAGGERED DID里面的平衡分别做是什么意思？是指将每一年分别做吗？还是怎么样比较好？

A：我的意思是有一组始终没有收到处置，其他分别在2003，2005，2007收到处置，在分析2003年的处置组不考虑2005和2007两组，同样在分析2005的时候不考虑2003和2007组，分析完之后再进行加总。(解答：徐轶青老师，整理：巴宁)

Q86：老师，如果不同州的政策具有异质性，定义为不同的处理组，那要分别与控制组进行回归？画很多动态效应图么？那要怎么才能对比政策在不同州的异质性效应呢？

A：注意，政策的异质性，需要说明是政策不一致还是政策效果不一致。如果是政策不一样，那就是SUTVA不满足，需要定义为不同处理组。如果是政策效应不一致，我们将在下一讲进行分析。

我们可以允许每一组的效应的都不同，然后再进行加总。

(解答：徐轶青老师，整理：巴宁)

Q87：老师，如果画pre trend图发现有很强的预趋势该怎么处理呢？

A：在第5,6讲进行解决。(解答：徐轶青老师，整理：巴宁)

Q88：老师，请问就2013年成立的碳排放交易市场的这个政策，但是有一个PNAs的设置是将前两年的announcement也加进去了，这个是不是会有预期效应的存在呢？

A：有可能，顶刊并不代表正确。我觉得你的想法是对的。(解答：徐轶青老师，整理：巴宁)

Q89：不好意思老师，刚才表述有问题，我想问的是从event study或者dynamic did的动态系数中看到处理效应的系数是在变化的，是否就验证了处理效应的异质性

A：是的，这至少证明了因果效应在时间维度上是有异质性的，而且证明了因果效应在单位的维度上是有异质性的。

(解答：徐轶青老师，整理：巴宁)

Q90：老师，能把TWFE中的估计系数设为异质性的么，就是变成Berta_i，这样可以估计吗？

A：想法很好，但是异质性既有i又有t，这样纯粹变成it很难估计。将在下一讲说明这个问题。(解答：徐轶青老师，整理：巴宁)

Q91：徐老师好，刚您说交错did做match或者balancing，可以分成多个非交错的treatment分别做平衡或者匹配。具体操作的时候分别做完多个非交错treatment effect的平衡之后是在整体的交错DID模型中加入多个权重？还是需要分别用DID估计出多个非交错的treatment effect的结果再取平均？

A：你可以在分组matching后再按照每一组的比例做加权平均，这是可以的。(解答：徐轶青老师，整理：巴宁)

Q92：staggered did 是否调整为event study 就可以允许个体异质性？

A：不是的，event study 可以解决部分问题，但是Sun & Abraham (2021)的研究发现，随时间的异质性是不同的，那么即使是用event study 的图依然会有偏误。(解答：徐轶青老师，整理：巴宁)

Q93：controls*i.time这种控制，比如gender，指的是在male中C_T组的趋势和female中C_T组的趋势不一样吗？

A：可以，用一个函数控制协变量不同导致的异质性。(解答：徐轶青老师，整理：巴宁)

Q94：徐老师您好，我在数据预分析这里做的一直比较少，之前您讲过一般回归前会看一下变量间关系，您也针对关心哪些点做了解答，想说您能不能在简单列举几个您平时回归前做数据分析的图例。

A：如果是截面数据，建议先看一下每个数据之前的两两相关性和分布；如果是面板数据，建议看一下panel view；如果是相关性，可以先把FE给partial out，把残差去掉后再重新画一下。(解答：徐轶青老师，整理：巴宁)

专题：[面板数据](#)

- [Stata绘图：面板数据可视化-panelview](#)

Q95：老师，如果用原始数据画图不满足平行趋势假设推论，但是平衡后满足了，这样可以吗？

A：可以，这说明你的条件平行趋势满足。(解答：徐轶青老师，整理：巴宁)

Q96：老师您好，我想问一个论文写作的问题，如何能把机制分析写的不那么刻意？感觉特别不会写机制分析的论述那里，请您指教！

A：最好的机制分析，是带着经济或者其他社会科学的理论，明显的告诉你这个机制存在，或者用机制分析去讨论哪个理论更加可能是对的，这个时候做机制分析更加具备价值。(解答：徐轶青老师，整理：巴宁)

Day5-FAQs

2022 因果推断专题 10 月 29 日课程提问

Q1：请问徐老师可否帮助总结下各种did在实证运用中的具体步骤呢，这样可以更好的规范实证研究。

A：看看数据，多少单位被干预，干预和处理的分布情况。判定三种类型的设定到底是哪一种，取log做一些变换。目标的估计参数，处理和对照组的预期情况，选用其中一种情况，是否是经典DID或者交错DID，baseline的结果还是要汇报双向固定效应模型，然后选用五个之一作为稳健性检验，需要图示呈现动态效应图，一图胜千言。根据idea选用合适的研究设计，再根据设定和数据，选用合适的模型。（回答人：何庆红）

Q2：交错DID和多期DID一样吗，有什么区别？

A：交错DID英文又叫做Staggered DID,这个设定是针对进入处理时间的不同，定义不同的队列。传统经典2X2DID，有两组两期，多期是指现在不只两期，有多期。（回答人：屠西伟）

Q3：Borusyak et al. (2021) Imputation Estimator 什么时候使用？

A：在你认为你的数据，平行趋势假设满足，但是存在异质性处理效应的时可以使用。（回答人：郭楚玉）

Q4：政策的进出一般都会受到之前结果的影响，比如扶贫政策会受到该地区之前贫困情况的影响，碰到这种情况在评估扶贫政策效果时如何处理？

A：这种情况可以采用徐老师讲的插补法，对处理组的反事实情况进行估计，该方法通过用真实值减去估计得到的反事实值得到处理效应 参考文献：Liu L, Wang Y, Xu Y. A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data[J]. arXiv preprint arXiv:2107.00856, 2021.（回答人：陈云菲）

Q5：为什么iw和csdid不能估计reversal情况？

A：这两种估计量都没有考虑，政策实施后再退出这样的数据情况。都是一旦进入处理组，stay on treated。（回答人：郭楚玉）

Q6：请问插补法具体是用什么估计什么？是估计了干预的反事实吗？

A：插补法实际上是将处理的数据给拿掉，用剩下的数据去跑回归，这样我可以拿跑完回归的模型预测被拿掉的那些点，这样就可以把预测的点作为控制组。（回答人：屠西伟）

Q7：IW方法中，估计出的每个cohort怎么加权出显著性呢？是软件自动计算吗？

A：IW方法里面，作者提供了计算weights的STATA命令：eventstudyweights。参考这篇原文：p183页。（回答人：郭楚玉）

Q8：老师，如果我想算ATT，如果从手动计算的角度来看IW方法的话，它是计算出每一个cohort的effect后，然后直接平均算每个cohort的效应，然后如何计算总体的ATT呢？

A：每个cohort的ATT计算出来之后，按照cohort size (每个组样本数量) 进行不同cohort之间加权，得到总体的ATT。（回答人：郭楚玉）

Q9：请问老师，如果treatment是连续型，交错timing，并且treatment强度还会改变（比如强度从10%上升到20%，但不会从treated变回untreated）的话，用哪种方法更合适呢？

A：离散化，如果数值很多，还是要施加模型。类似于IW，先做cohort dummy，先生成组的dummy，分组去估计，再加总。参照疟疾文章。

Q10：第四讲的“异质性”是指时间维度的异质性还是unit层面的异质性，还是两者都有呢？

A：两者都有。（回答人：郭楚玉）

Q11：请问老师，如果认为可能存在处理组到控制组到spillover effect，有什么解决方法吗？

A：一种是隔离（扔掉一些样本），另一种是加上结构化的假设再进行估计。可以参考的文章：“A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series CrossSectional Data” (with Licheng Liu and Ye Wang). American Journal of Political Science, forthcoming.（回答人：徐老师，何庆红整理）

Q12：请问动态did和事件研究法的区别到底是什么，对这两者的理解一直很模糊。

A：动态双重差分法可以被视为交错双重差分法的动态效应检验，是通过检验处理组和控制组在干预前和干预后的组间均值差异变化来识别政策的动态效应。但动态双重差分法不以绝对时间为参照系，而是以干预发生时点作为相对时间参照系。而事件研究法通常只有时间维度，即干预前后，是组内pre和post的差异。（回答人：王卓）

Q13：假设处理后有3个观察期得到d1,d2,d3三个效应。这3个效应与treatpost是什么关系？treatpost是d1,d2和d3的加权吗？为什么有时候会出现，d1,d2,d3不显著，但是treat*post显著。

A：可以将treatpost的系数理解为D1，D2，D3的平均效应。这种情况相对来说是比较正常的，可以理解为因为估计treatpost系数时比估计d1，d2，d3包含的信息更多，也更容易显著。（回答人：陈希）

Q14：请问老师，控制事前不变的地区特征变量与各种时间趋势的交乘项，与直接控制各年份的连续变量有什么区别吗，或者前者有什么好处吗？

A：控制事前不变的特征变量与各种时间趋势的交互项相比于直接控制连续变量的好处是更加的外生，避免政策对事后的控制变量产生影响（回答人：屠西伟）。

Q15：类似最低工资政策，第一期发生之后，几期之后又会出现（如第三期，第五期），这种应该如何处理？

A：根据插补法进行处理(回答人：屠西伟)

Q16：比较一阶差分与双向固定效应的结果，是不是可以用来检验强外生性假设？判断差异大小的标准是什么？

A：如果强外生性假设有错，那么两种都是有偏误的。先检验有没有单位根。

Q17：双向固定效应的系数本来就是估计的处理效应的平均，是不是意味着处理效应异质性也是被允许的？

A：如果双向固定效应模型是经典DID，在强外生性假设和平行趋势假设成立的情况下，那么估计的系数就是ATT。但在交错DID或者存在reversal的一般状态下，是不可以的。原因正如Goodman Bacon那篇文章指出的，早被干预的组被错误的作为后干预组的对照组，这时估计得到的处理效应是存在异质性偏误的。（回答人：徐老师，何庆红整理）

Q18：用Interaction Weighted的方法，能否换成合成控制法（用never treated组来合成）来估计每一个cohort的处理效应？

A：二者处理的问题不一样。Interaction Weighted的方法是选择合适的控制组来处理异质性处理效应问题，不是为了解决平行趋势假设不成立的情况。而合成控制法是用来解决平行趋势不满足的情况。另外，如果合成控制法存在交错DID、多个cohort情况，可以对每个cohort采用合成控制再进行分析。（回答人：徐老师，何庆红整理）

Q19：请问因果推断框架下的双重差分，在满足了SUTVA，无预期假设，和平行趋势假设后，是否就可以允许模型中不放入时变confounder？

A：首先时变的confounder含义是会影响到系数的一致估计，他就一定会导致这些假设不满足，既然满足，那就不需要，说明你认为的confounder不是真正意义上的混淆变量，因为他没造成这些假设不满足。（回答人：屠西伟）

Q20：样本共3组，1组在整个样本期都处理，2组t=1处理，3组t=2处理，最终全部接受处理，在treat*post中，treat所有样本均是1，只有post变化，还是DID吗，这样偏误是否过大？

A：这种情况应当不适用DID方法，treat变量实际上没有任何变化（始终保持1），在STATA估计时就不会生成有效的估计（会被omit），而此时的交乘项也并没有做到“对照”的作用（因为样本中不存在对照组），所以整个回归实际上退化成了一个单独的政策强度变量对被解释变量的回归。（回答人：康瑜欣）

Q21：请问老师，在did实证文献中，有的会控制改革变量和控制变量的交乘项，这个是控制什么呢，谢谢

A：允许控制变量在政策发生前后，对Y变量的影响可以不一样。（回答人：郭楚玉）

Q22：请问老师，双重差分方法的实证结果经济含义解读如何描述比较规范，是与控制组相比，改革导致实验组变化多少吗

A：一般论文采用如下的叙述：“交乘项的系数估计值为XXX，且在X的水平上统计显著，表明XX实施后，处理组企业比控制组企业的XX水平显著XX。”（回答人：康瑜欣）

Q23：如果动态效应模型第0期就显著 但是理论上政策应该是滞后的 这种情况怎么办

A：可以选择让0点成为参考点。但严格来说，这可能是实证无法支撑理论。也可以试图重新匹配控制组和样本组来调整结果。（回答人：马惜缘）

Q24：上周提到交错DID为了满足平行趋势，可以根据不同处理期把不同的cohort和对照组做匹配或平衡，再加权平均。有没有文献作为支撑或引用呢？平行趋势需要每一个cohort都分别验证是否满足平行假定？

A：可参考经典的应用该方法的文献：Chen Yi, Fan Ziyang, Gu Xiaomin, Zhou Li-An. Arrival of Young Talent: The Send-Down Movement and Rural Education in China[J]. American Economic Review, 2020 （回答人：马惜缘）

Q25：请问插补法的结果是不是很依赖于simulation的模型设定？比如假设单位的趋势是线性还是非线性，可以算出不同的反事实结果。

A：是的。课件里有讲到三种simulation的模型设定，包括线性假定和非线性假定（二阶或者三阶）。插补法simulation的模型设定并不比倍差法模型假定强。（回答人：徐老师，何庆红整理）

Q26：为什么imputation的基期是前面各期的均值

A：明天会讲

Q27：动态效应图下面的柱状图是什么意思呢，code是什么呢

A：是对应时期下的样本量，命令是fect，详情请参看本讲中的do文件。（回答人：何庆红）

Q28：placebo test没听明白，能再讲下吗

A：这里老师讲的安慰剂检验是，如果真实的政策冲击发生在第t期，现在假设政策冲击在t-2期，采用插补法估计t和t-1期的结果变量，再与实际的t和t-1期结果变量进行比较，发现二者相差不大，说明我们假设的这种政策冲击没有效果，侧面反映了不存在时间趋势，我们估计的真实的t期的政策冲击效果是有效的，证明了我们结果的稳健性（回答人：陈云菲）

Q29：PPT中问题2：明显的事前趋势，如果画图有趋势，但回归已经被时间dummy控制了，为什么还会影响回归结果

A：说明趋势控制的不足。举个例子，比如政策发生在省份层面，你控制了一个时间的趋势，这个趋势是说对所有省份都一样，事实上各个省份的趋势不一样。这个时候你需要控制 $i.province \times c.year$ ，让每个省份都有一个时间趋势。（回答人：屠西伟）

Q30：为什么要对干预前的系数做F检验

A：因为由于事前单个时间t检验的问题是，数据量比较小，检验效力比较弱，可能事前趋势不满足平行趋势，但是由于检验效力较弱出现平行趋势，这时候可以进行事前联合检验（回答人：屠西伟）

Q31：好多微观数据库都是几年调查一次，导致可用的数据期数较少，无法画出时间跨度较多的趋势图。如2011-2018年跨度8年但事实上可能只有4期数据，这样的还需要做趋势图吗，应该怎么操作呢

A：一切did都必须进行平行趋势检验。数据无法支撑平行趋势的话，可能不适用did方法。如果事件前时间极其短，大概率是实证不稳健的，除非所研究的问题特别重要和新鲜，才能弥补这样的实证硬伤。或者这样的did是为了估计一个结构化模型里的参数，则也会放松一点实证要求（回答人：马惜缘）

Q32：看事前趋势图的时候，，每张图下方的灰色柱状图是什么指标？

A：是对应时期下的样本量。（回答人：何庆红）

Q33：请问数据缺失到什么程度是不可接受的。

A：依赖于具体问题场景。但如果是纯实证的文章的话，对于数据样本缺失容忍度很低。可以考虑加入结构化模型（回答人：马惜缘）

Q34：第四讲的论文复现工作有发表正式或者workingpaper等，可以引用吗？

A：工作论文是可以引用的，比如很多论文引用了NBER working paper。（回答人：何庆红）

Q35：请教下老师，如何生成空间滞后项？在徐老师2022年宗族那篇文章的稳健性检验中有提到这个做法。那篇文章的附件中稳健性检验部分也有很多方法，那部分也能提供个代码学习一下么？谢谢老师！

A：需要空间matrix，一种常见做法是将邻居的均值放在回归中。（回答人：徐老师，何庆红整理）

Q36：ppt里聚类到更高的层面可解释潜在的空间溢出：若是firm-level研究，但担心干预组公司会影响控制组公司的outcome，那把SE聚类到行业层面是否就可以消除这种扩散效应对DID估计的干扰？标准误的聚类解决溢出效应的原理是什么？

A：标准误聚类到高级层面，是为了避免因误差项的spacial correlation造成的低估不确定性，进而得到合理的估计。但是，为了解决spacial spillover，还需要对点估计进行调整，比如将数据压缩到更高的行业层面，在行业层面做回归，这个时候会改变点估计量，但这个点估计量是不存在偏误的。（回答人：徐老师，何庆红整理）

Q37：DID政策有一般均衡效应的话怎么去检验或者怎么去论证

A：这个问题是untestable，可以做稳健性检验、敏感性检验。也可以通过构建structure model来分解出一般均衡效应和因果效应。这个时候可以做counterfactual analysis，但结构的假设是比较强的。（回答人：徐老师，何庆红整理）

Q38：请问如果平行趋势图中，预趋势估计值为负，但不显著，干预后估计值趋势为正，比较显著。这种情况能否说明结果有好的预趋势呢？

A：预趋势估计值为负数，确定全部都不显著的话。按现在主流的分析，应该就是满足PTA。应该是好的预趋势吧。（回答人：郭楚玉）

Q39：很多研究中，动态效应图，干预前效应不显著的话，就结论说满足平行趋势假设。这样有什么问题吗？或者有其他什么检验平行趋势的方法。

A：感觉依然是目前主流方法。最近有一篇文章Roth, Jonathan. (2022). Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. American Economic Review: Insights. 指出这种方法的问题是：pretrend test 的检验效力很低（low power），即 及时你的数据存在pretrend，但是你做一个F-test的结果告诉你，这些都=0，满足平行趋势假设。但是其实是因为检验效力低，检查不出pretrends，而不是因为你的数据里确实没有pretrends。文章还指出了第二个问题，就是及时通过PTA，也可能会后政策后期的系数产生偏误。然后这个作者又写了一篇文章【More Credible Approach to Parallel Trends (with Ashesh Rambachan). 2022. Forthcoming, Review of Economic Studies】，在这篇文章里，他好像就建议大家不用那么看重pretrend test的结果了。可以看一下。（回答人：郭楚玉）

Q40：pretrend=平行趋势假设不满足，这样理解对吗？

A：有pretrend存在 = 平行趋势假设不满足。pretrend test = 检验事前趋势是否成立的检验。（回答人：郭楚玉）

Q41：staggered DID，每个cohort（进入处理的年度）中观测值很小，比如只有2-10，用什么方法比较好呢？

A：关于处理组数据较少的问题：解决方法首先采用原始数据进行估计，之后可以删减一部分控制组样本作为稳健性检验。最后，如果处理组数值过少，也可以反过来操作，将控制组作为处理组，估计反效应（回答人：陈云菲）

Q42：请问老师，您在课上展示了不满足数据缺失导致不满足平行趋势的情况，请问好的处理组干预时点的图是怎样的？

A：事前平行趋势严格满足，事后处理组呈现系数显著差异，图像不随参考点变化而发生太大变动（回答人：马惜缘）

Q43：老师，如果干预后的期数太少会有什么问题吗？比如评估一个近几年刚发生的干预，干预前数据会有二十几期，但是政策干预后却仅有一两期的数据，这样的样本如果去做TWFE会有什么偏误么？该怎么解决呢？

A：太少，直接问题是政策效果可能有延迟，出现不显著的问题。干预前的期数如果太长，可能不容易满足平行趋势假设，因为时间长，中间影响趋势的因素过多。可以画图，看看哪一段的趋势更加平行，就截取到哪里。（回答人：屠西伟）

Q44：异质性考察，是做三个三重交互项，还是直接样本分组回归，还是同时都做比较好呢？

A：DDD是可以在双重差分的基础上探讨异质性的，使用分样本回归的话，如果每个分样本都可以支持做DID的话也是可以的，但可能要考虑一下分样本后各个样本的分布问题。（回答人：陈希）

Q45：请问如何选择合理的基期？是否选择处理前一期（-1）最合理呢？什么情况下可以选择之前，比如（-2或者-3期）作为基期合理呢？

A：这个没有理上的说法，你会看到有事情(-1)，当期(0)，最远(-k)。Sun and Abraham（2021）中列举了这几种常见方法，选择(-2)/(-3)基本上不这么做，选取常见的就可。（回答人：屠西伟）

Q46：老师，did协变量应该怎么加，2*2与多期经典did是不只能加非时变量，广义did是不能加时变量，交错did加时变还是非时变，加的时候考虑哪些问题，理解为加了能得到条件于协变量的平行趋势对吗？

A：did的协变量不一定都要求有时变性。但如果有个体固定效应的话，这类非时变的变量会被个体固定效应吸收。（回答人：马惜缘）

Q47：感觉老师在结果这里讲的有点快，徐老师能再简单讲一下如果存在较强的预趋势怎么办？如果预趋势检验的不通过怎么办？怎么救？

A：这个问题会在明天合成控制法的课上详细讲解。（回答人：徐老师，何庆红整理）

Q48：请问这几种方法都能加controls吗？

A：可以。每种方法都可以参看help文件，里面有详细的添加控制变量的方法。（回答人：屠西伟）

Q49：老师，如果就是那种简单的多期的2*2形式的DID中，画平行趋势图发现有很强的预趋势，该怎么解决呢？还是直接就不能使用。

A：这个问题会在明天合成控制法的课上详细讲解。（回答人：徐老师，何庆红整理）

Q50：placebo为什么选干预前的期数，而不是干预后的期数。

A：一般选择干预前。干预后的情况比较复杂，因为干预已经发生，且不同组之间还有spillover的可能，作为安慰剂检验的concern更多，说服力更低。（回答人：马惜缘）

Q51：请问老师，课程中涉及画图的命令，能输出适合打印风格的吗？图中字体能改吗，比如宋体，新罗马等，中文期刊会要求图例字体，这个怎么实现呢？

A：可以的。绘图命令的help帮助文件里有详细的option选项供用户修改图片显示风格、字体大小等。比如，图片风格可以查看option选项里scheme(schemename)，提供了不同的风格。可以参看连享会中绘图推文。（回答人：何庆红）

Q52：今天讲的预趋势，意思是估计系数的趋势吧（从stata命令里看到是系数画图）。那PPT里坏的事前趋势和好的事前趋势是什么意思，好坏的判定是？

A：您的理解是对的。坏的事前趋势，你可以看到从事前到事后期有某种明显的趋势(线性，抛物线)。而且事前的系数很多异于0。好的事前趋势，事前的线都水平在X轴，而且大部分点的置信区间都跨过0，即我们不能拒绝这些系数=0。即处理组和控制组的Y变量随时间变化趋势没有显著区别。满足平行趋势假设。而政策发生后，系数会有一个比较明显的跳跃。（回答人：郭楚玉）

Q53：静态固定效应模型和动态固定效应模型区别在哪里？

A：对于静态固定效应，处理变量是0/1变量，系数可以理解为平均处理效应ATT。动态估计效应，处理变量是不同时期的dummy变量，估计的是不同时间相对于对照期（一般选-1期）的效应。（回答人：何庆红）

Q54：请问老师可不可以快速讲一下平行趋势假设不满足怎么处理呢？感觉很多问题的根源都是这个。

A：合成双重差分法(Arkhangelsky et al 2021)，可以帮助处理平行趋势被违反的情况。具体会在明天的课上讲解。（回答人：何庆红）

Q55：老师今天讲一般传统的2*2 DID不存在异质性偏误，那是说传统的2*2 DID不存在预趋势问题吗？

A：这里区分异质性效应和平行趋势假设是两个不同的假设。2*2 DD不存在由处理效果异质性带来的偏误，原因在于，你不可能画出把处理组当做未处理组的控制组这种方框（参考老师今天讲课的时候画方框）。而处理组和控制组在 处理组受到政策处理前，可能有很多期（参考lecture 4-p5-左上图），依然可以画事前3期的动态处理效应图出来，来判断是否有预趋势。（回答：郭楚玉）

Day6-FAQs

2022 因果推断专题 10 月 30 日课程提问

Q1:有些文章使用面板模型，引入 $A*B$ 的交乘项，但没有控制A和B，不知这个操作是否正确，或者在什么情况下使用才是正确的？

A:取决于A和B是否随时间或随个体变化。如，不随时间变化，加入后也会被固定效应吸收，这种情况下可以不放。但一般情况下，既存在时间变异又存在个体变异应该放入。（回答人：徐老师，王卓整理）

Q2:如果结果变量在时间维度上变化很大，例如变化大小在100倍以上，那么在应用SCM或FEct，是否需要结果变量取对数处理呢

A:是否取对数不应该以变化范围大小为标准，以回归意义的解释为准。（回答人：马惜缘）

Q3:如果安慰剂区域 (Placebo Region) 与y轴零值不存在交集，但是安慰剂检验的 p 值大于0.05，那这安慰剂检验是通过还是没通过呢？

A:安慰剂检验p值仅大于0.05是不够的（回答人：马惜缘）

Q4:老师昨天没有讲fect命令中的等效检验，想问下等效检验有必要做吗？如果处理前 ATTs 的 90% 置信区间会偶有超过等效边界，三种方法的等效检验都不通过，该怎么办呢？

A:经济学可以不做，在统计学和政治学中接受度较高；等效检验较为困难的是需要事前设定阈值，阈值和自己认为的估计出的因果效应的量级有关，但大多数情况下观测性数据不知道阈值，难以事前确定。（回答人：徐老师，王卓整理）

Q5:徐老师，请问staggerd did操作的步骤，如是不是要先看平行趋势。另外中国很多评比项目，如分批的乡村振兴村镇实验或文明城市评选，理论机制应该关注评选前还是评选后？

A:是的，要先看平行趋势。如果趋势检验通不过，尝试合成双重差分法和交错插补法等（回答人：何庆红）

Q6:昨天提问了能否推荐一些交错DID做匹配或者平衡的支撑文献？感谢助教同学的回答，但我刚看了一下推荐的AER的论文并未使用匹配或平衡，其数据本身在控制趋势后就满足平行趋势，所以还有没有其他文献可以参考？

A:匹配方法请参看下面两篇文献：Heyman, F., Sjöholm, F., & Tingvall, P. G. (2007). Is there really a foreign ownership wage premium? Evidence from matched employer–employee data. *Journal of International Economics*, 73(2), 355-376. Benzarti, Y., Carloni, D., Harju, J., & Kosonen, T. (2020). What goes up may not come down: asymmetric incidence of value-added taxes. *Journal of Political Economy*, 128(12), 4438-4474.（回答人：何庆红）

Q7:DID要求不能存在溢出效应，但一些顶刊论文在论证完DID的主效应，比如村里企业的扶持政策对企业的影响，还会将扶持政策对旁边村企业的影响即溢出作用作为额外的结果发现。这是否说明其主模型的结果是有问题的？

A:这些文章是有问题的。除非主模型的样本剔除了受到溢出效应的个体。（回答人：徐老师，王卓整理）

Q8:请问应该怎么理解交互固定效应呢？什么时候用aitrend、什么时候用aift呢？aift可以理解为冲击对不同地方影响不同吗？22did中如果用ai*ft是不是不能估计did系数了呢？

A:交互固定效应可以理解为推广的ai时间趋势，一般情况下加线性趋势；aift可以理解为冲击对不同地方影响不同。（回答人：徐老师，王卓整理）

Q9:MSPE是什么？

A:均方误差（回答人：何庆红）

Q10:如果所有个体都受到处理，处理强度不同。有些人用均值来分类，将大于均值的设为1，将小于均值的设为0，请问这种做法合适吗？

A:有这样处理的文章。但通常这类文章不是用于政策评估，而是讨论X变量本身的影响。且不一定分为两组，也可以分为若干组，比较最高组和最低组的差距。（回答人：马惜缘）

Q11:请问连享会有平衡法和的推文吗？谢谢

A:

- 专题：[PSM-Matching](#)
- [ebalance](#)：基于熵平衡法的协变量平衡性检验（回答人：屠西伟）

Q12:老师可否点评regression control方法？因为这个方法就是利用负权重

A:regression control不是特别稳健，需要T比N大或加正则化，没有主流方法稳定；尽量不使用负权重。（回答人：徐老师，王卓整理）

Q13:合成控制法和合成双重差分法一次只能放1个处理组，那如果有多个处理组，就只能一次一次的做吗？每次只保留1个处理组，其他处理组先屏蔽掉吗？那最后得到很多个估计系数也要一个一个分析吗？

A:合成控制法适用于仅存在一个处理组，多个控制组（合成为一个反事实控制组）的情形。部分文献会采用将多个处理组加总、平均值等操

作，但是结果稳健性存疑。如果存在多个处理组，可以尝试广义合成控制法。可以参考徐老师这篇文章：Xu Y. Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models[J]. Political Analysis, 2017, 25. (回答人：康瑜欣)

Q14:合成DID 中时间加权不太明白，时间维度是怎么确定权重的

A:时间维度上的权重是用控制组的数据来做的，即利用控制组事前的数据来估计时期的权重，以达到这个权重的预测效果可以和控制组在处理后期的结果相似。一般而言，时间上的权重会更多的赋予处理时期的前几期，这几期是对预测效果作用更大。(回答人：屠西伟)

Q15:昨天说的将处理组与控制组互换的情况， $treat \times post$ 里面的post变量的设定是冲击前为1，冲击后为0吗，还是不变？

A:是将treat替换，post不变。(回答人：屠西伟)

Q16:为什么合成控制法中，只有一个处理组就会引发平行趋势不满足的担忧？多个控制组是否可以用合成控制法？另外合成did的优势应用场景是否可以介绍下

A:不管处理组和控制组分别的个数有多少，只要你同时有控制组和处理组存在，你就需要担心：这两个组之间的结果变量的平行趋势是否满足这个问题。对于经典的合成控制法，它正好就是需要只有一个处理组，有很多的控制组。(回答人：郭楚玉)

Q17:intercept shift在model specification中在哪里呢

A:见第五讲的课件(回答人：徐老师，王卓整理)

Q18:老师，mspe是什么意思啊，在这里能说明什么？

A:这是均方预测误差， $mspe_pre$ ，就是用来说明处理前拟合程度的好坏。相关推文：

- 专题：[合成控制法](#)
- [合成控制法简介](#)
- [Stata：纠偏合成控制法介绍-allsynth](#)
- [Synth_Runner命令：合成控制法高效实现](#)
- [Stata：合成控制法程序分享](#)
- [Stata：合成控制法-synth-命令无法加载-plugin的解决办法](#)
- [合成控制法 \(Synthetic Control Method\) 及 Stata实现](#) (回答人：屠西伟)

Q19:请问两种方法都可以的情况下合成did是否比简单的合成控制法要好呢？

A:如果两种方法的适用条件都满足，那么这两种方法的结果应该是一致的。(回答人：屠西伟)

Q20:Spillover问题的后续（1）：若政策是在行业层面，而又担心行业内的spillover，则是否可以把A行业的干预组和其他行业的控制组放在一起做倍差法，得到针对A行业的ATT估计。

A:可能是可以的(回答人：徐老师，王卓整理)

Q21:Spillover的后续（2）：再对其它行业也循环做一遍。最后对估计出的每个行业的ATT 以各行业样本数占总样本的比值进行加权得到政策对全行业的ATT估计？

A:可以的(回答人：徐老师，王卓整理)

Q22:请问合成控制法的适用范围。如果处理组观察值有100个，是否要对每个观察值分别合成一个控制组呢？

A:合成控制法适用于仅存在一个处理组，多个控制组（合成为一个反事实控制组）的情形。部分文献会采用将多个处理组加总、平均值等操作，但是结果稳健性存疑。如果存在多个处理组，可以尝试广义合成控制法。可以参考徐老师这篇文章：Xu Y. Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models[J]. Political Analysis, 2017, 25. (回答人：康瑜欣)

Q23:昨日连续型处理变量在交错+变强度情况下的后续：请问老师昨天说的建立一系列cohort，对每个cohort估计ATT然后进行加权，最后是用每个cohort占总体样本的比例进行加权求和吗？

A:是的，按cohort 占总体样本的比例加权，是比较好的加权方式。满足凸组合的权重。(回答人：郭楚玉)

Q24:老师之前的课上提到强外生性假设可以推出平行趋势，但反过来不行，那么即使发现预趋势比较好，是不是也不能完全由此说明干预后实验组与对照组的差异就是由因果效应造成的？

A:外生性假设比did的平行趋势假设要求更高。对应的就是随机试验和did的区别。did当然是因果识别，满足did的假设和机制后，为何能成为因果识别，可以再好好参考下课件。有很多同学甚至老师把did假设的内涵与外生性混淆，可以根据课件着重理解。(回答人：马惜缘)

Q25:老师您好！请问在选择哪些年进入模型作为控制变量时，是不是只要能够使得处理之前的拟合程度好就可以了？还是有什么其他的选择准则？多谢！

A:这个确实有自己的主观选择，拟合程度好是必然的。但要留有一些年份以作为检验，如果把所有事前年份都作为拟合变量可能会出现过度拟合的问题。(回答人：屠西伟)

Q26:SDID可以用在多个处理组个体交错DID的情形吗？如果可以的话，当交错DID平行假定不满足时，应该做匹配和平衡比较好？还是直接用SDID？另外我看例子中SDID没有加控制变量，可以加控制变量吗？

A:可以的，SDID可以用在多个处理组个体交错DID的情形；匹配和平衡都可以试试；若数据存在混淆变量且可以由factor model解释，数据量大的情况下几种方法可以得到类似的结果。（回答人：徐老师，王卓整理）

Q27:请问如果有多个处理单位可以用SDID吗？交错的情况可以用吗？

A:可以的，SDID可以用在多个处理组个体交错DID的情形；匹配和平衡都可以试试；若数据存在混淆变量且可以由factor model解释，数据量大的情况下几种方法可以得到类似的结果。（回答人：何庆红）

Q28:请问使用合成控制法，比如研究海南自由贸易港的经济效应，被解释变量分别为GDP、对外贸易、FDI，构成三个回归方程，那么三个方程中的控制组的权重是否可以不一样？

A:可以三个权重不一样（回答人：徐老师，王卓整理）

Q29:sdid的两个例子，加州控烟的例子中事前拟合值和实际值有截距差异，西德的例子中事前拟合值和实际值没有截距差异，这是为什么？

A:两个的可比性需要使用同一个sample.（回答人：徐老师，王卓整理）

Q30:老师，CATT估计量在估计过程可能会丢失较多样本，导致估计无效，这样应该怎么办？另外，BJS（2021）和LIU等（2022）依赖于正确的模型设定，那在实证中应该如何选择正确的模型设定呢？

A:模型设定不是很大的问题；与其他模型相比，插补法没有特别大的劣势（回答人：徐老师，王卓整理）

Q31:关于sdid。第一，是否一定要平衡面板数据；第二，是否允许添加控制变量（原文好像说有办法处理）；第三，sdid是否可以允许有多个受处理的个体（原文好像说可以）；第四，关于sdid的标准误，使用哪一个。

A:1.不需要，但可能会drop一些数据，变为平衡面板；2.目前不允许；3.可以允许；4.用jackknife更好（回答人：徐老师，王卓整理）

Q32:之前好像听您说did不能添加随时间变化的控制变量，能添加不随时间变化的控制变量。这个地方不明白。

A:感觉您说的应该是 TWFE 回归模型里面，它应该是不能添加 不随时间变化的控制变量，因为会被 个体固定效应吸收掉。可以添加随时间变化的控制变量，比如把控制变量和c.year 进行交乘。（回答人：郭楚玉）

Q33:没加入控制变量，个人固定效应和时间固定效应的时候，did不显著，但是加入了控制变量还有固定效应以后did就显著了，这种结果好吗，刚开始不显著的结果可以汇报吗

A:可能的原因是：加入控制变量(假如是好的控制变量)会提高模型的拟合优度，这个时候模型的残差就小，由于标准误的估计与残差有关，就会使得标准误变小，系数就变得更加显著。可以汇报。（回答人：屠西伟）

Q34:CATT是什么

A:CATT是针对cohort（队列）的处理效应。（回答人：屠西伟）

Q35:老师，合成控制法是比较适用于只有一个处理组的情况吗？如果处理组比较多的话还适用吗？

A:合成控制法适用于仅存在一个处理组，多个控制组（合成为一个反事实控制组）的情形。部分文献会采用将多个处理组加总、平均值等操作，但是结果稳健性存疑。如果存在多个处理组，可以尝试广义合成控制法。可以参考徐老师这篇文章：Xu Y. Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models[J]. Political Analysis, 2017, 25.（回答人：康瑜欣）

Q36:谢谢老师提供这么多种方法。现实中经常遇到不知道哪种方法最合适。请问能否简单总结如何根据具体问题和数据选择方法。

A:如果能一言以蔽之我们就不需要这么多课程内容了。需要大量的文章阅读和漫长的科研训练，乃至写作和发表经验才能对于这些方法有相对系统深入的理解。不要试图naive的简化，有了课程的知识框架后，面对自己研究的具体问题，批判性思考和干中学更加重要。（回答人：马惜缘）

Q37:老师为什么我用TWFE做出来干预效应显著，但是SDID做的每个处理组都不显著呢？

A:两者假设不同，处理时数据量不同。TWFE的事前趋势没有问题的情况下可以不用SDID，SDID会删掉一些数据。（回答人：徐老师，王卓整理）

Q38:之前看到过有人说，如果说数据量小或者说干预后数据少，建议做合成控制法，想问下老师，这种说法对吗？

A:有一定的合理性。合成控制法针对的是处理个体少，T大的情况。很多研究都是集中在城市、州、省份层面的数据去做，这些宏观层面的数据相对于微观数据确实数量少很多。（回答人：屠西伟）

Q39:在违反平行趋势假设的情况下加入unit specific time trend能否解决这个问题

A:如果说违反的原因是unit specific time trend那么是可以解决的，如果说不是时间趋势上的混淆变量，那么是解决不了的。（回答人：屠西伟）

Q40:老师，潜因子法如果 $T < 20$ ，不够大会存在什么问题？

A:小于20时结果不稳健，可能还存在偏误。（回答人：徐老师，王卓整理）

Q41:老师昨天的课上讲到回归之前的设计非常重要，有这方面的参考书推荐的吗？

A:所谓的研究设计是针对D的生成机制进行研究，这里可以参考的书有：国内：《基本有用的计量经济学》；国外《基本无害的计量经济学》。（回答人：屠西伟）

Q42:老师，我看到一些文章中，下标i城市、p省份、t年份， $Y_{ipt} = X_{ipt}\beta + \alpha_i + \xi_{pt} + \epsilon_{it}$ ，请问这样与 $Y_{it} = X_{it}\beta + \alpha_i + \xi_t + \lambda_i * ft + \epsilon_{it}$ 之间区别是什么呢？

A: ξ_{pt} 指的是省份交乘年份的固定效应， $\lambda_i * ft$ 是个体交乘时间趋势。（回答人：屠西伟）

Q43:老师，传统2*2式的DID中不存在异质性问题，还可以使用插补法或潜因子法吗？

A:潜因子法，是需要许多控制组去提取主因子，如果只有两个组倒不适用。而且，如果不存在异质性问题，平行趋势由没问题，不需要把问题复杂化。（回答人：屠西伟）

Q44:老师，为什么插补法和潜因子法中都是异质性的系数，但是放入了所有处理组最后得到一个估计系数呢？这个是考虑了异质性后的平均值吗？

A:是的，因为我们估计的是平均效应，但均值的背后是异质的个体效应。（回答人：徐老师，王卓整理）

Q45:如果有一个多期的政策，请问什么情况不用多期DID，而需要用潜因子法呢？

A:当平行趋势假设不满足的时候（回答人：徐老师，王卓整理）

Q46: (问题part1) 想请问老师DID实操中的一个例子，在DID（2*T情形）动态效应的模型设定中，如果不仅加入 D_{it} ，还加入某些控制变量在处理年份当期值与各个年份的虚拟变量，这是为了保证PTA成立吗？

A:是的，这么做是为了得到conditional parallel trends（回答人：徐老师，王卓整理）

Q47: (接上一问题) 比如一篇文章研究银行贷款冲击（D）对于地方政府后期债务（Y）的影响。作者在模型设定中还加入了银行贷款冲击当年的其他控制变量（比如省份（单位）的GDP）与年份dummy的交叉项（感觉像是加入了其他的处理变量），作者给出的解释是：our specification allows for a fairly rich set of cross-province heterogeneity in the trends of debt issuance(结果变量Y) during our sample years请问如果在因果推断的框架下理解，这是为了论证满足平行趋势假设吗？

A:这种情况下需要conditional parallel trends（回答人：徐老师，王卓整理）

Q48:怎么理解Cengiz等（2019）“The Effect of Minimum Wages on Low-Wage Jobs”中的Stacked Regression Estimator？

A:相当于对每个cohort进行估计，然后平移到零点（得到的一年），然后进行stack。（回答人：徐老师，王卓整理）

Q49:抱歉，我修改了上一个提问，谢谢！看到一些文章，下标i城市、p省份、t年份，回归是 $Y_{ipt} = X_{ipt}\beta + \alpha_i + \xi_{pt} + \epsilon_{ipt}$ 。请问这样与交互固定效应之间区别是什么呢？什么时候应该选择交互固定效应呢？

A: ξ_{pt} 指的是省份交乘年份的固定效应，如果我改成 $i.peovince \# c.year$ ，这就是省份交乘年份的时间趋势，这与潜因子模型中的设定是一致的。省份交乘年份的固定效应可以吸收省份交乘年份的时间趋势，具备控制 $i.peovince \# c.year$ 的效果，同时省份交乘年份的固定效应允许不同省份在不同年份有不同的时间上的影响，而 $i.peovince \# c.year$ 只是允许不同省份有不同的时间趋势，这个时间趋势在同一省份的影响是一样的（回答人：屠西伟）

Q50:徐老师好！关于时间效应的异质性，请问合成控制法中干预前结果变量的M个线性组合怎么找的？潜因子模型三步法中，选择主因子之后是如何估计处理组的反事实的（即如何估计时间效应的异质性）？

A:原文中未提及，基本上是中后找到三个点。理论上随机找会得到类似结果，但现实中数据不稳定，很难稳定。（回答人：徐老师，王卓整理）

Q51:老师，如果用了Interactive Fixed Effects，后续想再加一个三重交互项分析异质性时，该怎么做呢？

A:Interactive Fixed Effect解释的是残差项，和可观测的三重交互项异质性是无关的。（回答人：徐老师，王卓整理）

Q52:请问之前曾讲过用平衡方法解决平行趋势不满足。而lec5，6又交了其它方法。请问是否balancing方法（如熵平衡）只限于非交错情况下的DID呢？而Lec5，6的方法 则对交错和非交错都可以用？

A:熵平衡法是用干横截面数据。（回答人：郭楚玉）

Q53:请问徐老师bacon分解的图怎么画呀？我根据bacon的代码画出来好像无法得到early v. late的结果

A:因为bacondecomp这个命令，更新了。之前是可以画出 early T vs. late C 的。现在的话，你可以去e(sumdd)里面查看，里面储存了每个小的2*2DD的结果。可以参考一下下面的命令（可能有些地方需要根据你的变量名称修改一下）

```
matrix list e(sumdd)
```



```

mat A = e(sumdd)
clear
svmat2 A, names(col) rnames(group) full

g temp = Beta * TotalWeight

* sum to three groups
preserve
collapse (sum) temp TotalWeight, by(group)
replace temp = temp/TotalWeight
list
/*
+-----+
| group temp TotalW~t |
+-----+
1. | Early_v_Late -.1867701 .1797593 |
2. | Late_v_Early 3.511967 .4299179 |
3. | Never_v_timing -5.3309069 .3903228 |
+-----+
*/
restore
(回答人：郭楚玉)

```

Q54:看到很多文章在DID的设定下还会分析冲击 (Dit) 的来源, 是来自于供给侧还是需求侧的冲击。请问在因果推断的框架下, 分析冲击的来源, 主要是为了论证PTA, 还是为了论证SUTVA成立? 或是其他?

A:可能是为了论证SUTVA。可能需要进一步进行reduce form的估计。(回答人：徐老师, 王卓整理)

Q55:老师, 今天讲的用fct命令做的IFE估计, 和用reghdfe做TWFE多控制个time*id的效应有什么区别呢? 感觉解释起来好像意思很相似呢

A:reghdfe的time (c.time) 是连续的, 时间趋势是线性的, 而ife的时间趋势是任意的。两者有相似之处。(回答人：徐老师, 王卓整理)

Q56:老师, 合成控制法可以用在公司年度层面的数据吗, 比如构造不受政策影响的样本。好像目前的合成控制法的文章都是在地区层面圣上的。

A:可以。(回答人：徐老师, 王卓整理)

Q57:做政策研究模型里面x是post, 和模型中x是post*treat, 这两种回归的结果是符号相反的这是正常的吗? (单独两个模型, 不是放在一个模型里面)

A:这种情况是有可能出现的, 如果模型中仅存在post, post前的系数表示政策干预这个时点之后, 所有样本的目标变量的平均变化方向。而模型中交乘项的估计结果显示, 干预时点之后, 处理组相对控制组的目标变量变化方向。第一种情况的模型是不能用于政策解读的, 仅仅分析了某个时点过后所有样本的平均变化。(回答人：康瑜欣)

Q58:第六讲中“Fouirnaies and Mutlu-Eren (2015)”有四页 (每页三幅图), p值都是拒绝原假设, 都是不通过检验? 老师能再讲一下这几幅图都是在做什么检验吗? 有的图code里面没有呢

A:可以理解为做F检验。用不同估计量看趋势 (残差的均值) 在0附近 (不显著)。(回答人：徐老师, 王卓整理)

Q59:老师, sdid里面vce(placebo)是啥意思啊?

A:这个可以参考help sdid中的vce(vcetype)注释, In the case of placebo, this requires at least one more control than treated unit to allow for permutations to be constructed。(回答人：何庆红)

Q60:如果平行趋势图已经更好看了, 还需要进行安慰剂检验吗?

A:还可进行安慰剂检验。参考相关推文：专题：[回归分析](#)

- [Stata：一行代码实现安慰剂检验-permute](#)
- [aoeplacebo：地理安慰剂检验](#)
- [专题：倍分法DID](#)
- [多期DID之安慰剂检验、平行趋势检验](#)
- [专题：内生性-因果推断](#)
- [安慰剂检验！安慰剂检验！](#) (回答人：屠西伟)

Q61:最后复习时候, 平行趋势假设不成立时候, 没有提到entropy balance方法。那么这个方法用来解决什么问题呢?

A:entropy balance 的方法是为了使得控制组和处理组, 各协变量的均值是相似的。目的是对控制组进行加权, 以平衡掉这些协变量的影响。(回答人：郭楚玉)

Q62:匹配法是为了处理平行趋势的问题, 今天讲的也是为了处理平行趋势的问题, 两者在处理平行趋势方面有什么区别?

A:我感觉本质上是一样的。都是为了让政策的发生更随机。实际上政策发生在某些地方(处理组),未发生在某些地方(控制组),大概率是因为处理组和控制组存在不同。用匹配和合成控制法,都是为了构造出跟处理组 足够像的控制组。因此让政策的发生显得更随机,也因此更容易满足PTA。(回答人:郭楚玉)

Q63:数据不平稳具体什么意思呢?

A:应该指时序上不存在平稳性,平稳性定义分两种:1.数据的分布形式不随时间发生改变(强平稳),2.数据的均值与方差不随时间发生改变(弱平稳,使用较多)(回答人:康瑜欣)

Q64:政策退出可能是因为 $y(t-1)$ 的影响,是否存在反馈效应,是否不可以做DID

A:你的理解是对的。如果过去的Y能影响D(t)之后的政策。那就是存在反馈效应。最严格意义上,确实不能做DID。但是还是建议你做一下动态效应图,看看事前期的PTA是否满足,若满足的话,感觉也能继续做DID。毕竟理论要求太严格了,如果都按着教科书来,大部分文章都做不了,除非自己设计实验。(回答人:郭楚玉)

Q65:Callaway (2021) 提出采用连续型变量作为treatment需要“更强的平行趋势假设”,除了常规检验方法外,是否有新的方法能对连续型处理变量情形下的平行趋势假设进行检验?有哪些最新的文献参考?

A:其实就这两年的文献才刚刚发现因为异质性处理效应的存在,用TWFE的系数估出来是有偏的。然后new alternative estimators 刚扩展到D是0-1变量,数据格式可以是staggered,或更general ($0 > 1 > 0$)的情况。对于D是连续变量的情况下,如果进行有效的估计ATT,都好像没讨论出来。推荐一篇写连续型DID的文章:De Chaisemartin, C., & D'Haultfoeuille, X. (2022). Difference-in-differences estimators of intertemporal treatment effects (No. w29873). National Bureau of Economic Research. 再推荐一篇可以避开做平行趋势假设检验的文章:Rambachan, A., & Roth, J. (2022). A More Credible Approach to Parallel Trends. Working Paper. (回答人:郭楚玉)

板书