

*The Latest Progress of Frontier Econometrics :
SDID, FDID, DDML, Parallel Trend test*

前沿计量部分最新进展：一个简明实用手册
——平行趋势检验与 *SDID, Fuzzy DID, D/DML*

2023年2月8日

厦大管院

潘王雨昂

前言：追求“实用”的手册

计量方法分类

- 按照 抽样-假设-估计推导-检验-实践（代码）等去划分计量方法的学习路径的话。针对估计方法的推导部分(诸如GMM等)，本笔记为实用起见，不做过多追求。

原因：记忆曲线理论与黑匣子理论

- 要掌握推导的阻力很大
 - 记忆曲线：任何知识点都需要回顾才能记忆，而推导是最难回顾的。比起原理会在交流，论文中被动回顾，推导方法几乎没有被动回顾的可能性，需要专门找时间回顾。这就很容易学会之后忘却。
- 要掌握推导的需求很小
 - 黑匣子理论：如同使用手机时，其内部芯片的复杂运算我们不需要知道，只需要知道输入一个指令必然会输出一个指令即可。我们在使用计量软件时，也不会知道其运算过程。

但我们依然推荐掌握推导方法

- 这里引用“司马懿”老师的观点：
- 知乎：只做经验研究的经济学研究者是否有必要掌握好高级计量中的推导？

如果一直是这个姿势水平，那是无法窥得大道的
就像编程序只会背系统自带库的接口一样，终究是在食物链的下游，核心竞争力弱，一旦系统库自己出了问题，完全没有debug的能力，挺痛苦的

仅仅只是本手册不会给出相应的估计方法罢了！

- 具体的方法在相应引用中都有提到，可以参考原文

依然是“实用手册”性质，是全力追求通俗化的

- 不推荐再去网上搜其他资料了，乃至看原文了，鱼龙混杂。完全能把人绕晕。
 - 上述论断截止时间为2023年2月8日，不排除之后有更多更优秀的解析与手册的出现。

代码不做专门讲解

- 这个是需要练习，光讲毫无意义，都在笔记中贴上了

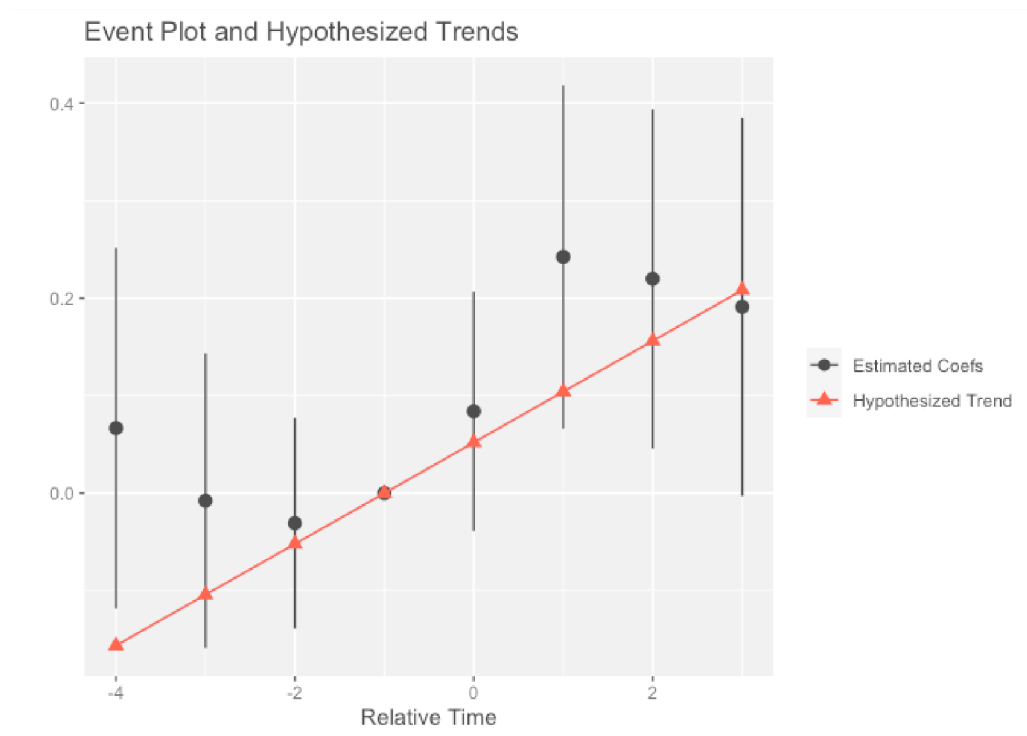
顺序

一、平行趋势检验新进展

二、双重机器学习

三、模糊双重差分

四、合成双重差分



Parallel Trend Robust Test

一、平行趋势稳健性检验

“太过于信任”的“事前趋势检验”

首先，我们需要明确一点：平行趋势假设是完全不可验证的

- 我们平常做的，更为准确的说法，叫做“事前趋势检验”或者“处理前趋势检验”
- 现在我们所说的通过事前趋势检验“证明了平行趋势”存在的说法，写在论文中是错误的。只能说“给平行趋势提供了证据”。

其次，在“事前趋势检验”满足的情况下，也依然可能存在各种问题

- “处理前趋势检验”的效力远没有我们想象的那么高：
 - 处理前趋势系数全部不显著，但平行趋势假设可能依然不满足。处理前趋势检验的效力只在50%-80%左右

具体参见：Roth J, 2022. Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends[J]. American Economic Review-Insights, 4(3): 305-322.

再次，“事前趋势检验”看起来很难通过，但实际上，也是可以操纵的

- 学者们对AER文献的复现发现，可以通过更改事前趋势的窗口期和控制变量的方法，操纵事前趋势检验。

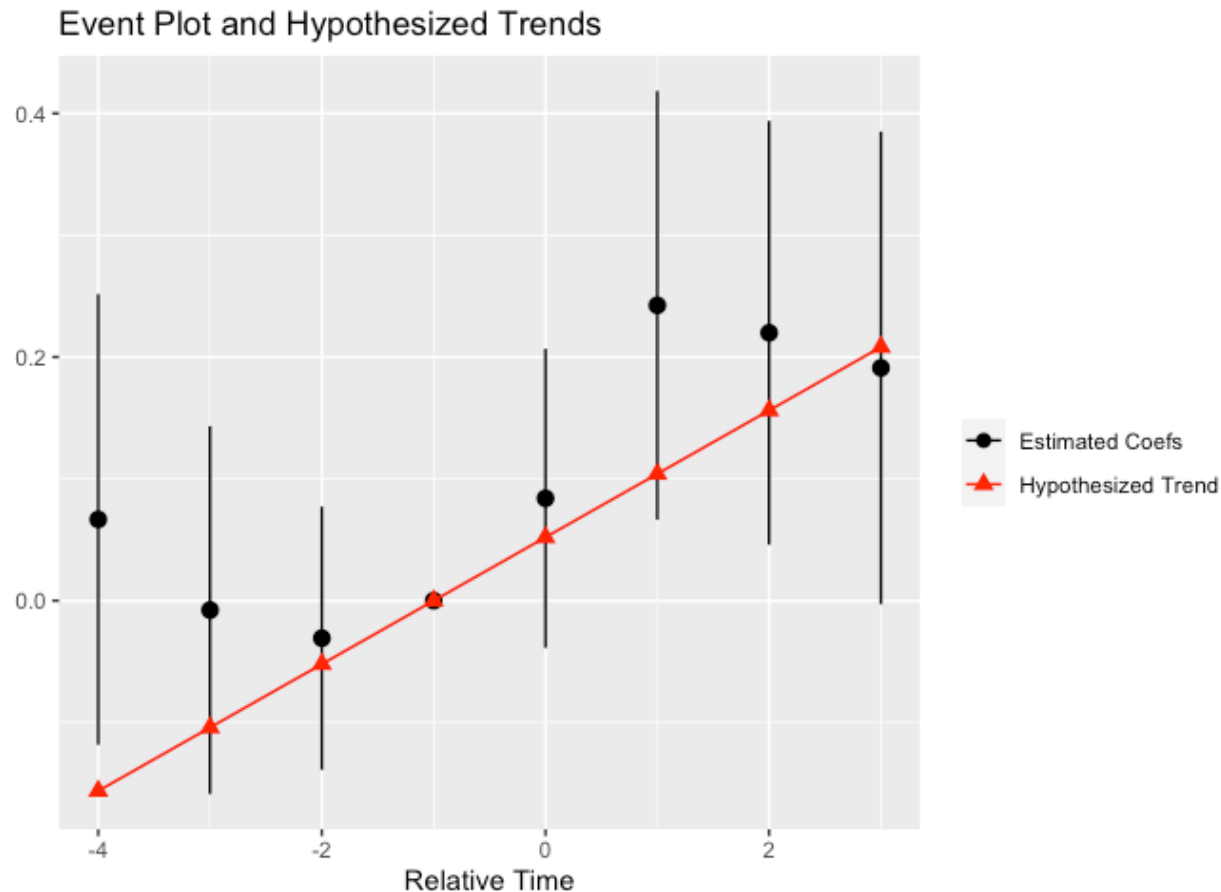
参见：Jaeger D A, Joyce T J, Kaestner R, 2020. A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing?[J/OL]. Journal of Business & Economic Statistics, 38(2): 317-326. DOI:10.1080/07350015.2018.1497510.

那怎么办呢？三种方法

- pretrend test
- honest DID
- 制度、文化背景论述

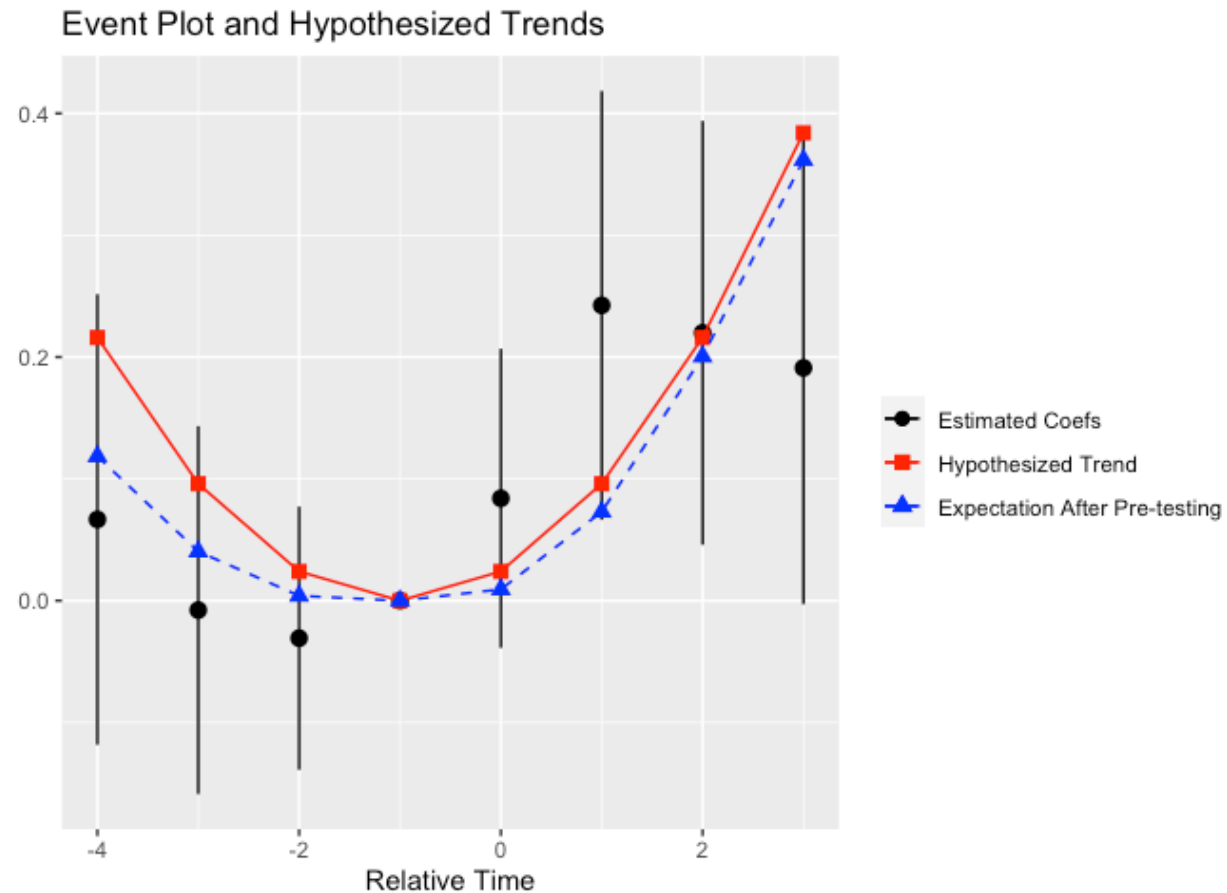
Pretrend Test

- 首先，我们看右图,这是一个典型的运用事件研究法类去检验平行趋势的图片
- 如果没有红色的线的话，实际上，右图可以说是一个比较完美的“事前趋势检验”了。
- 但很可惜，如果真实的数据趋势，是类似于红色的线一样呢？



Pretrend Test

- 第二种情况
- 如果是类似于如图所示的抛物线呢？



Pretrend Test

- 所以，简单来讲一句话：pretrend test 就是用来检验上述这几种情况是不是存在
- 你可以检验你自己设想出来的线性的，或者二次函数形式的统计效力。
- R代码包的具体使用方法参见：[连享会](#)
- 很可惜，该文章的作者，还没有编写Stata包.....

为啥不编写stata包QAQ



Jonathan Roth

jonathandroth

Follow

Assistant prof @ Brown.

👤 108 followers · 0 following

🔗 <https://jonathandroth.github.io/>

🐦 @jondr44

📖 Overview

📁 Repositories 17

📁 Projects

📦 Packages

★ Stars 2

Popular repositories

staggered

Public

● R ☆ 45 🍴 12

pretrends

Public

R package for power calculations and visualization for pre-trends tests

● R ☆ 35 🍴 8

PretrendsPower

Public

Power calculations for pre-trends testing

● R ☆ 8

staggered_stata

Public

● Stata ☆ 7 🍴 5

LinearMomentInequalities

Public

● MATLAB ☆ 5 🍴 2

jonathandroth.github.io

Public

Forked from academicpages/academicpages.github.io

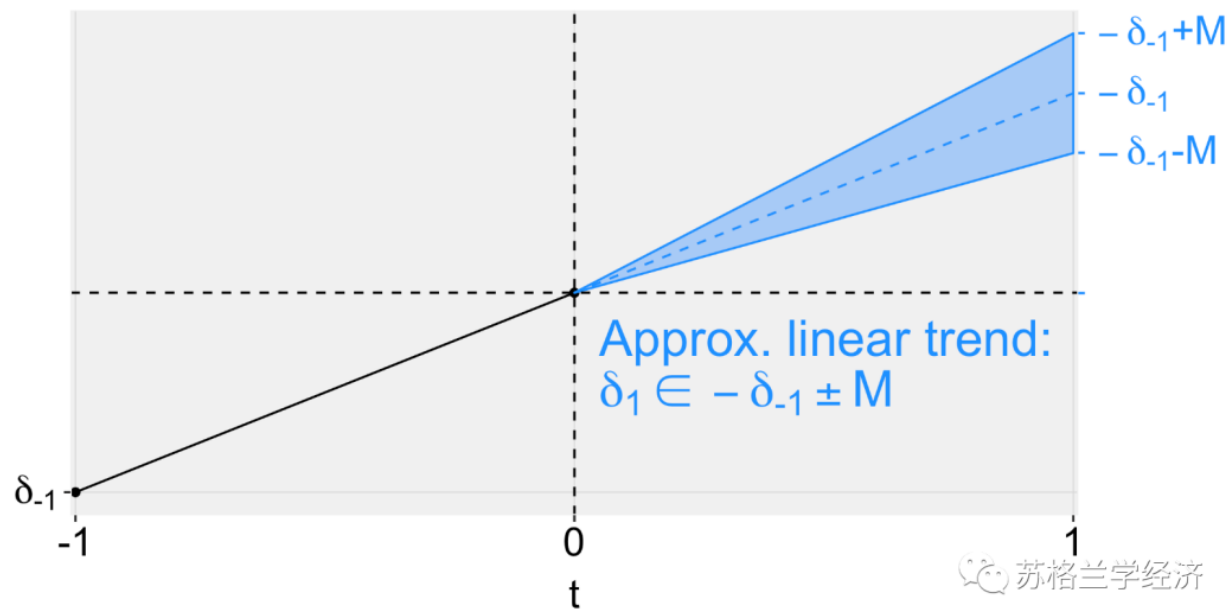
Github Pages template for academic personal websites, forked from mmistakes/minimal-mistakes

● JavaScript ☆ 3 🍴 5

61 contributions in the last year

Honest DID

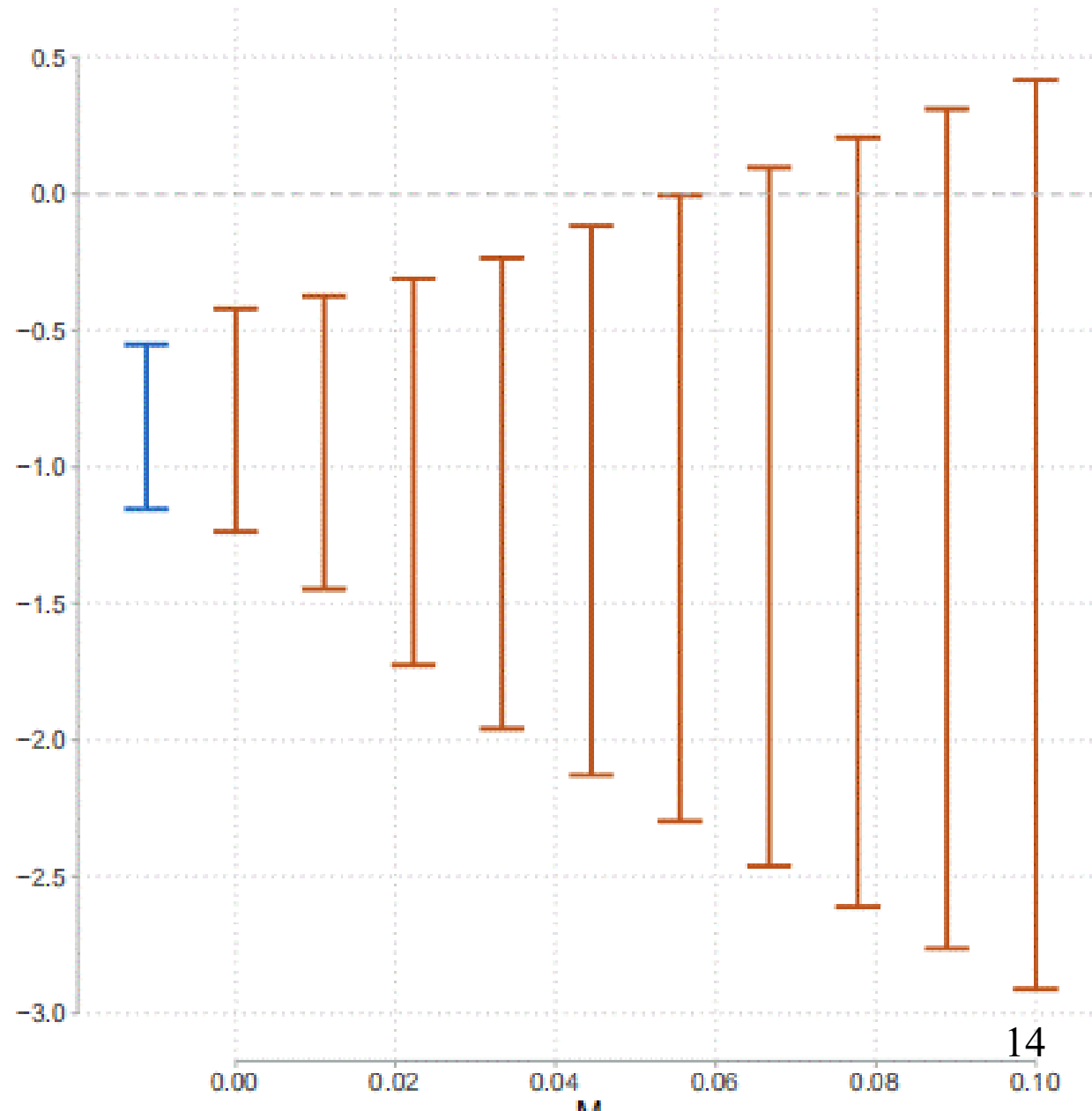
- Honest DID,实际上可以看做是平行趋势检验的敏感性分析
- 即假设事后的值比事前的值偏离个M倍。
- 然后可以给出在偏离了多少的程度下，核心解释变量的系数会从显著走向不显著
 - 下一页PPT的图片展示的更加明晰
- 右图来源[苏格兰经济学公众号](#)



Honest DID 复现相关

- 很可惜的是，目前没有对该代码的讲解，
- 只有github的原阅读文件了
 - [戳我](#)
- 右图来源，[香樟学术推文](#)

(a) Without imposing monotonicity ($\Delta^{SD}(M)$)



关键且容易被忽略的第三种方法：相似性论述

- 即通过对感兴趣的问题的制度背景、经济社会环境等因素的论述来阐释平行趋势假设成立的可信性。
 - 而不仅仅只是看图，看表说话！
 - 可以参考的文章：

[1] Bleemer Z, 2022. Affirmative Action, Mismatch, and Economic Mobility After California's Proposition 209[J/OL]. Quarterly Journal of Economics, 137(1): 115-160.

[2] Cameron L, Seager J, Shah M, 2021. Crimes Against Morality: Unintended Consequences of Criminalizing Sex Work[J/OL]. Quarterly Journal of Economics, 136(1): 427-469.

[3] Cantoni E, Pons V, 2021. Strict Id Laws Don't Stop Voters: Evidence from a Us Nationwide Panel, 2008-2018[J/OL]. Quarterly Journal of Economics, 136(4): 2615-2660.

平行趋势新进展小结

- 平行趋势不可验证，需要改掉“平行趋势检验通过”的说法
- 事前趋势检验虽说难以通过，但目前来看，即使通过，效力也很低，并且还可以人为操纵
- 可以通过pretrends检验线性或二次型函数，可以通过honest DID检验平行趋势敏感性
- 在论述DID类文章时，需要结合经济社会背景来论述平行趋势通过的可能性



Double Machine Learning

二、双重机器学习

双重机器学习

双重机器学习是一种估计方法

- 类似于OLS, GMM, 是一种估计方法, 而不是一种类似于识别出LATE的识别策略
- 所以实际上, 如果运用的好的话, 是可以和DID一起识别出ATT等的
- 可以用于解决协变量是非线性的
 - 但, 对于严格的因果推断来讲, 我们不关心协变量啊hhhh
- 数理上的好处 (可以忽略)
 - 一开始利用机器学习, 诸如LASSO回归等, 是有偏的, 双重机器学习解决了其无偏的问题 (大样本下表现为一致性), 使得其能够应用在计量上

具体推导

假设我们的回归式子如下：

$$Y = \alpha + \tau \cdot D + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$$

我们知道，得到我们需要的系数 τ 是根据如下方法得到的(FWL定理)：

第一步：利用 Y 对所有协变量进行回归，得到 Y 的余值 \hat{Y} ，即下式中的 e_1

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e_1$$

第二步：用解释变量对所有协变量进行回归，得到解释变量的余值 \hat{D} ，即下式中的 e_2

$$D = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e_2$$

第三步：用相应的余值进行回归，可以得到我们的估计量 τ

$$\hat{Y} = \tau \hat{D}$$

具体过程，随便一本计量书都有讲这个

基于OLS的推导很好理解

- 从被解释变量和解释变量中剔除掉这些我们能够观测到的协变量的混淆因子引起的变动，留下解释变量和被解释变量自身，然后看他们的关系，就是单纯的因果关系。
 - 注意，我们这里是假设所有的变量都被观测到了，且不存在严重内生性时
- DDML (Double/Debiased Machine Learning, 有时简写为DML)“双重机器学习”，或者叫“纠偏机器学习”就是一样的原理。
- 具体表达式如下：

$$Y - ML_Y(X_1, X_2, X_3) = \tau[D - ML_D(X_1, X_2, X_3)] + \epsilon$$

- $ML_Y(X_1, X_2, X_3)$ 表示利用机器学习方法去用 X_1, X_2, X_3 去预测Y
- $Y - ML_Y(X_1, X_2, X_3)$ 和 $D - ML_D(X_1, X_2, X_3)$ 是不是和我们上述的余值很像？
- 只不过估计方法用了机器学习的函数诸如LASSO等而已

- 具体步骤也和上述一样
 - 第一步：用一个机器学习回归模型来根据协变量X预测被解释变量Y，并得到被解释变量的余值；
 - 第二步：用一个机器学习回归模型来根据协变量X预测解释变量D，并得到解释变量的余值；
 - 第三步：对两个余值进行回归。此时可以使用OLS。
 - 在机器学习领域中，第三步的术语叫cross-fitting
- 前提假设
 - 需要满足 $E(D|Y) = 0$ ，也就是说，还是需要满足我们类似的严格外生性假说，诸如不存在未观测到的混杂因子。
 - 双重机器学习最主要的是解决“模型设定问题”，即协变量非线性问题（尽管目前我们并不在乎协变量），而不是因果推断，识别策略问题。
- 除开上述之外，还可以用于高维数据
 - 什么是数据维度？戳我

具体操作代码

- 连享会：获取双重机器学习的聚类标准误
- 许文立：DDML操作方法

参考论文

- (1) 王茹婷, 彭方平, 李维, 等, 2022. 打破刚性兑付能降低企业融资成本吗? [J/OL]. 管理世界, 38(4): 42-64. DOI:10.19744/j.cnki.11-1235/f.2022.0051.
- (2) Valente M, 2023. Policy evaluation of waste pricing programs using heterogeneous causal effect estimation[J/OL]. Journal of Environmental Economics and Management, 117: 102755. DOI:10.1016/j.jeem.2022.102755.

具体推导讲的比较清楚的（不讲）

- 看这个解析
 - 陈强公众号
- 其次看原文：

(1) 【主要cite论文】 Chernozhukov V, Chetverikov D, Demirer M, et al., 2018. Double/debiased machine learning for treatment and structural parameters[J/OL]. The Econometrics Journal, 21(1): C1-C68. DOI:10.1111/ectj.12097.

(2) Knaus M C, 2022. Double machine learning-based programme evaluation under unconfoundedness[J/OL]. The Econometrics Journal, 25(3): 602-627. DOI:10.1093/ectj/utac015.

DDML小结

- 目前该方法还是一个婴儿，还需要更多的发展
- 基于Chernozhukov et al.(2018)等追踪相应文献，目前的应用乏善可陈，甚至都不能表达为差强人意。绝大部分的研究还集中在如何计算相应标准误，算区间等难以应用的方面，远没有到颠覆现有范式的地步。



Fuzzy DID

三、模糊DID

简单理解模糊DID

- 简单来讲，就是不存在非常明确的分组方式，即没有“干净”的处理组和控制组。
 - 经典案例：一个地区新增报纸数量和选举的投票率。
 - 可能新增报纸会增加投票率，但是新增1份，新增2份.....新增45份的边际效应显然是不一样的。这明显不是一个连续DID。
- 用等式来理解，假设G为分组的虚拟变量，T为处理前和处理后的虚拟变量，D表示处理变量
 - 那么一般DID为：

$$D = G \cdot T$$

- 甚至呢，这个G还可以是一个连续变量，也就是我们常说的连续DID。

简单理解模糊DID

- 而模糊DID为：

$$D \text{ 不等于 } G \cdot T$$

- 回顾我们刚刚说的报纸的案例，就很好理解了
- 一般来讲，我们的D设置为0-1变量
 - 然后再为1的组内软件帮你算模糊值
- 而相应的识别可以看做为：

$$\tau = \frac{DID_Y}{DID_D}$$

- 或者：

$$\tau = \frac{E(Y_{11}) - E(Y_{10}) - (E(Y_{01}) - E(Y_{00}))}{E(D_{11}) - E(D_{10}) - (E(D_{01}) - E(D_{00}))}$$

补充

- 还是那句话，具体上式怎么算出来的，就交给软件吧。
- 具体可参考：

(1) de Chaisemartin C, D'Haultfœuille X, 2018. Fuzzy Differences-in-Differences[J/OL]. The Review of Economic Studies, 85(2): 999-1028. DOI:10.1093/restud/rdx049.

(2) de Chaisemartin C, D'Haultfœuille X, Guyonvarch Y, 2019. Fuzzy differences-in-differences with Stata[J/OL]. The Stata Journal, 19(2): 435-458. DOI:10.1177/1536867X19854019.

- 代码使用方法:连享会

FDID的其他问题

- 补充
 - 你可以在 $D=1$ 的组内人为的分组，还是以报纸数量为例，你可以按照0-1,1-2,2-45分组
 - 实际上，FDID有三个估计量，除开DID估计量之外，还有CIC，和TC估计量，其中TC能够放松时间的异质性假设（不需要处理效果对时间不变的假设）
 - 在指令中，还可以进行分位数回归

FDID的诸多劣势

- 时期限制：实际上，FDID只能计算两期的数据，多期的是加权平均，不能估计动态效应
- 计算限制：计算时，如果加入固定效应等，由于目前的Stata包的标准误计算是用的Bootstrap，可能会卡死
- 异质性限制：个体异质性不能解决
- 平行趋势限制：不能做平行趋势检验
- 可替代性限制：如果能进行分位数的话，为什么不用连续DID呢？

FDID小结

- 看起来FDID是拓宽了研究的可选择性，但实际上选择性很小，而且其替代性很强
- 掌握和理解也不是那么容易，如果没有需求，并不是十分推荐学习



Synthetic DID

四、合成控制DID

先从SC和DID开始讲起

在使用面板数据时，我们一般来讲要获得因果效应，有SC和DID等为代表的两种方法：

- SC：合成控制方法，用很多其他的控制组的个体，选取相应的权重，去合成一个处理组个体的反事实，然后进行因果计算
 - 不懂的话，补充在下一页
- DID：直接找两组相近的个体，一个受处理一个未处理，然后再计算

这两个面板数据评估，对数据有不同的要求

- SC：适用于一个或者极少数的处理组和大量的控制组，不需要满足平行趋势
- DID：适用于多个处理组和多个控制组，需要满足平行趋势

补充页：合成控制法互联网资料整合

扫盲看下面

- 合成控制法扫盲：知乎
- 合成控制法扫盲：连享会
- 合成控制法Stata实现：连享会

进阶看下面

- synth2
- scpi
- synth_Runner
- synth-命令无法加载-plugin-的解决办法

如何理解SDID

简单一句话讲明白SDID

- 简单来讲，就是把和实验组相近的个体赋予更多的权重，给和处理后更相似的处理前时期更多的权重,然后通过最优化的方法，算出来一个唯一的相应权重矩阵

还是那句话，推导就不细讲了

(1) 目前推导比较详细的互联网资料：[戳我](#)

(2) 具体原文(可供引用): Arkhangelsky D, Athey S, Hirshberg D A, et al., 2021. Synthetic Difference-in-Differences[J/OL]. American Economic Review, 111(12): 4088-4118. DOI:10.1257/aer.20190159.

SDID详述

SDID能够结合两者的优势!

- SDID能够自己“创造”平行趋势!
- SDID的方差更小，有效性更高!
- 不同于PSM-DID:
 - SDID是唯一的! 而PSM有各种各样的匹配方法，所以容易造假，目前来看，主回归用PSM-DID，经济学类期刊逢稿必拒
 - 并且相应的权重的值，也是公开可重复的，不是PSM那种黑箱
- 拥有以上特点的SDID，几乎可以说是目前因果推断的“大杀器”，而且目前应用还比较少

SDID小结

- 在TWFE-DID不满足平行趋势的情况下，无脑上SDID！大杀器！
- 并且SDID的方差更小，有效性更高，强烈推荐！
- 可惜的是，理论上可以做到SDID之后做Honest DID等，但代码支不支持就不知道了.....
- 目前没有找到利用该方法详尽讨论时间和个体异质性的

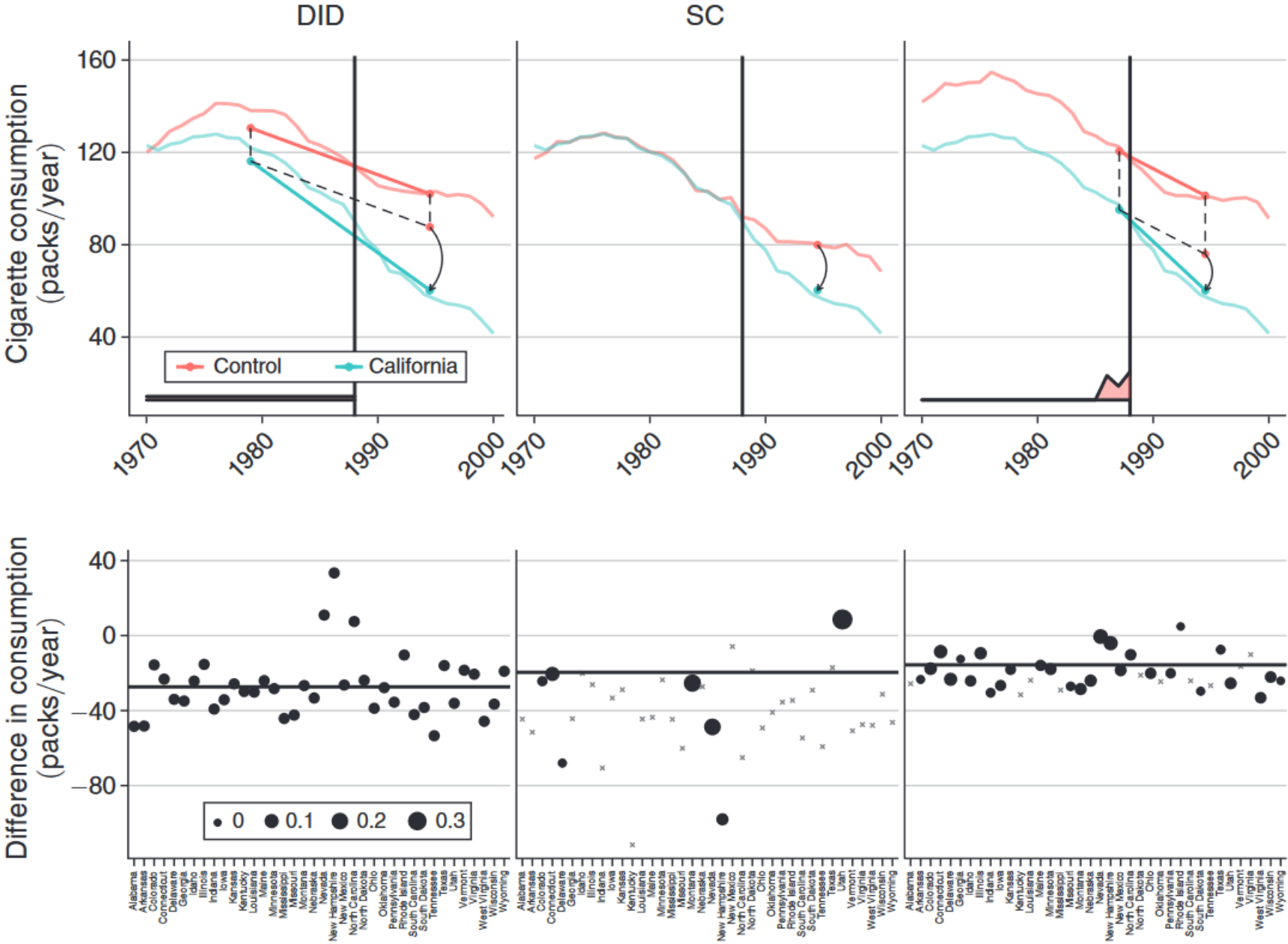
应用案例

- 经典加州控烟的数据

Abadie A, Diamond A, Hainmueller J, 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program[J/OL]. Journal of the American Statistical Association, 105(490): 493-505.
DOI:10.1198/jasa.2009.ap08746.

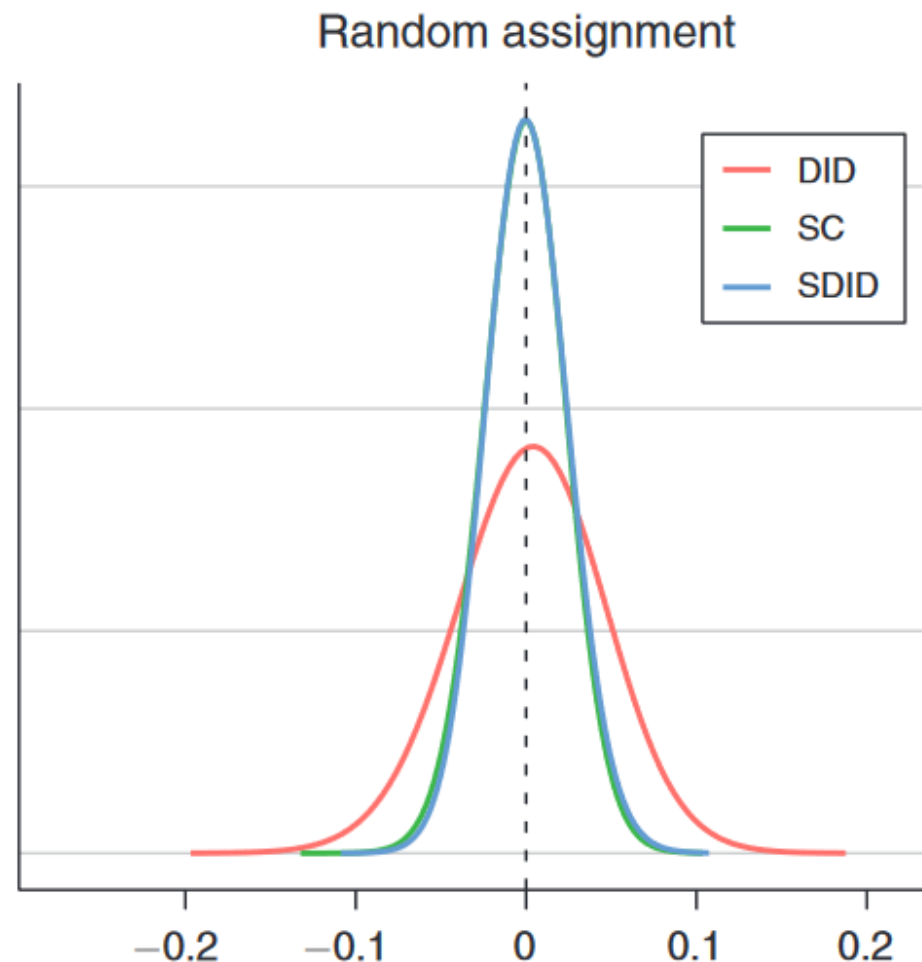
- 注意下图的线、权重

- 第二列图片中的圆圈是个体权重，×表示权重为0
- 时间权重在第一列的图片下方，第一列第一张图不同时间的权重都为1，第三张图1986-1990赋予了更高的权重，而第二章图时间未被加权



SDID的有效性更高

- 根据原文作者的模拟，一般来讲，SDID、SC和DID都是渐进一致的，且SDID的有效性更好！



SDID标准误计算（不讲）

- 第一种方法是聚类自助抽样（clustered bootstrap），将样本看做是总体，进行有放回的抽样，这种方法的主要缺点是每一次抽样需要运行一次完整的SDID算法，对于大型数据集来说，计算成本可能很高。
- 第二种方法是刀切法（leave-one-out jackknife），每次拿出一个观测个体不用，得到不同的估计量，计算标准误，这种方法只需要运行一次完整的SDID算法，减少了计算负担。以上两种方法，不适用于只有单一处理个体的SC方法。
- 第三种方法是安慰剂检验（placebo test），依次将控制组中的每个个体作为假想的处理个体，加州作为控制组（可去掉），使用SC，计算小概率事件发生的概率，这种方法适用于任何样本容量，但依赖同方差假定。

SDID代码实现参考

- 连享会
- 许文立



对比页：更深层次理解DID,SC和SDID

区别	DID	SC	SDID
是否需要满足平行趋势	需要	不需要	可以自己创造
样本条件	多个实验组和多个对照组	单个实验组和多个对照组	多个实验组和多个对照组
加权方法	可以理解为不加权	对个体进行加权	对个体和时间进行加权

如果有任何错误，欢迎随时与我联系

接受任何批评指正

pwya1998@126.com

pwya.github.io