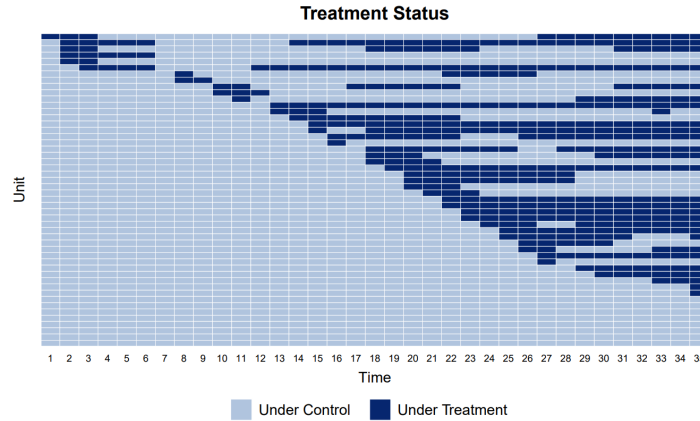


面板数据因果推断实操指南

刘安然 香樟经济学术圈 2022-05-11



图片来源: https://yiqingxu.org/papers/english/2022_fect/LWX_slides.pdf

原文信息: Liu, Licheng and Wang, Ye and Xu, Yiqing, A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data, *American Journal of Political Science*, forthcoming. Available at SSRN: <https://ssrn.com/abstract=3555463>.

推文作者: 刘安然, 中国人民大学农业与农村发展学院 2020 级硕士, 研究方向为发展经济学和劳动经济学等。邮箱: liuanran123@ruc.edu.cn。

在使用长面板数据进行因果推断时, 研究者常用的双重固定效应 (Two-way Fixed Effects, TWFE) 模型往往会遇到严格外生性假设不成立、函数形式假设不灵活等问题。由刘力成、王也和徐轶青三位学者合作发表于 *American Journal of Political Science* 的论文 “A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data” 介绍了一个反事实估计框架和一系列的检验工具来给研究 “松绑”, 以期更有效地进行因果分析。

一、引言

双重固定效应模型是研究者分析长面板数据——亦被称作时序-截面数据 (Time Series Cross-Sectional Data)——时常用的因果推断方法。然而, 使用双重固定效应模型会遇到一系列问题。

首先, 双重固定效应模型所要求的严格外生性假设难以得到满足。该假设需要不存在随时间变化的混淆因素 (Time-varying Confounders), 并且过去的结果不会影响当期的处理状态 (No Feedback)。第二, 双重固定效应模型的函数形式假设不够灵活。当处理变量为离散变量时, 该方法要求处理效应恒定不变 (Constant Treatment Effect) 且只影响当期结果 (No Carryover Effects)。倘若上述假设不能成立, 如处理时点不同且处理效应具有异质性时, 则可能会出现 “负权重” 问题; 此时, 双重固定效应将无法准确估计出处理组的处理效应 (Average Treatment Effect on the Treated, ATT)。

作者介绍了一个反事实估计框架来在一定程度上解决上述问题。该框架将处理组的观测值视为缺失值，然后利用控制组的信息，来补齐（Impute）的反事实。这种方法被称为反事实估计法（Counterfactual Estimation）。文章介绍了三种基于反事实估计法的估计量，包括：固定效应反事实估计量（The Fixed Effects Counterfactual Estimator, FEct）、交互固定效应反事实估计量（The Interactive Fixed Effects Counterfactual Estimator, IFect）和矩阵补全估计量（The Matrix Completion Estimator, MC）。下文将进一步介绍三种估计量的构造方式及使用时的注意事项。此外，作者还为长面板分析提供了一系列分析检验工具，包括动态效应可视化处理、安慰剂检验、事前趋势检验和延续效应检验等。最后，作者结合两篇已有文献的实证研究工作展示了上述框架和方法的具体应用过程。

二、反事实估计量

本部分作者介绍了估计框架与估计策略，并讨论了三种反事实估计量。

（一）估计框架

首先，作者考虑了一个包含 N 个单位、 T 个时期的平衡面板数据（该框架对非平衡面板同样适用）。 D_{it} 表示处理状态，取值为 0 或 1，分别对应潜在结果变量 $Y_{it}(0)$ 和 $Y_{it}(1)$ 。 \mathbf{X}_{it} 为外生控制变量向量， \mathbf{U}_{it} 则代表不可观测的因素， ε_{it} 为残差项。不失一般性，定义 $\delta_{it} = Y_{it}(1) - Y_{it}(0)$ 。则有如下三个假设：

假设 1：（函数形式） $Y_{it}(0) = f(\mathbf{X}_{it}) + h(\mathbf{U}_{it}) + \varepsilon_{it}$ ，其中 $f(\cdot)$ 和 $h(\cdot)$ 为已知的参数函数。

经典的两组两期 DID 方法是假设 1 的一个特例形式：

$$Y_{it}(0) = U_{it} + \varepsilon_{it} = \alpha_i + \xi_t + \varepsilon_{it}, \quad t = 1, 2$$

其中， α_i 和 ξ_t 分别为个体固定效应与时间固定效应。此时处理组的平均处理效应为 $ATT = \mathbb{E}[\delta_{it} | D_{it} = 1, C_i = 1]$ 。其中， $\delta_{it} = Y_{it}(1) - Y_{it}(0)$ （上文定义）。估计处理效应需要样本在观测时间段内处理状态至少改变过一次，即存在 t, t' 使得 $D_{it} = 0, D_{it'} = 1$ 成立，此时 $C_i = 1$ ；反之 $C_i = 0$ 。由于难以估计从未接受处理的样本和始终为处理状态的样本的处理效应，因此删去这部分样本。

此外，在实证研究中，不少研究者关注在冲击发生后第 s ($s > 0$) 期的处理效果，即：

$$ATT_s = \mathbb{E}[\delta_{it} | D_{i,t-s} = 0, \underbrace{D_{i,t-s+1} = D_{i,t-s+2} = \dots = D_{it} = 1}_{s \text{ periods}}, C_i = 1], \quad s > 0.$$

为了便于下文展开关于诊断检验（Diagnostic Tests）的讨论，定义 $ATT_s = 0, \forall s \leq 0$ 。

假设 2：（严格外生性）

$$\varepsilon_{it} \perp\!\!\!\perp \{D_{js}, \mathbf{X}_{js}, \mathbf{U}_{js}\} \quad i, j \in \{1, 2, \dots, N\} \quad s, t \in \{1, 2, \dots, T\}.$$

需要注意的是，平行趋势假设隐含于严格外生性假设中：

$$\mathbb{E}[Y(0)_{it} | \mathbf{X}_{it}] - \mathbb{E}[Y(0)_{is} | \mathbf{X}_{is}] = \mathbb{E}[Y(0)_{jt} | \mathbf{X}_{jt}] - \mathbb{E}[Y(0)_{js} | \mathbf{X}_{js}], \quad \forall i, j, \forall t, s$$

而当 \mathbf{U}_{it} 为更一般的形式时，假设 2 可表示为：

$$\mathbb{E}[Y(0)_{it} | \mathbf{X}_{it}, \mathbf{U}_{it}] - \mathbb{E}[Y(0)_{is} | \mathbf{X}_{is}, \mathbf{U}_{is}] = \mathbb{E}[Y(0)_{jt} | \mathbf{X}_{jt}, \mathbf{U}_{jt}] - \mathbb{E}[Y(0)_{js} | \mathbf{X}_{js}, \mathbf{U}_{js}], \quad \forall i, j, \forall t, s$$

即条件于可观测外生变量和不可观测因素，控制组个体 i 和个体 j 从第 s 期到第 t 期的平均潜在结果

变化相同。而这进一步引出了假设 3。此外，严格外生性排除了预期效应（Anticipation Effect）的可能性。

假设 3: (低维可分解) $h(\mathbf{U}_{it})$ 存在低维度的分解: $h(\mathbf{U}_{it}) = L_{it}$, 且 $\text{rank}(\mathbf{L}_{N \times T}) \ll \min\{N, T\}$ 。例如, $\mathbf{L} = \mathbf{\Lambda}\mathbf{F}$, 其中 $\mathbf{\Lambda}$ 为 $N \times r$ 的因子载荷矩阵 (Matrix of Factor Loadings), \mathbf{F} 为 $r \times T$ 的因子矩阵, 且 $r \ll \min\{N, T\}$ 。

该假设允许我们对 \mathbf{U}_{it} 进行设定。当不可观测的混杂因素 \mathbf{U}_{it} 存在时, 处理组的分配状态依赖于可观测的控制组结果, 即我们的分析是在一种缺失并非随机的特殊情况下进行的。而假设 3 使得我们可以通过控制 \mathbf{U}_{it} 来摆脱这汇总依赖性, 因此该假设可以被理解作为一种可行性假设。三个假设彼此之间有着密切的联系: 严格外生性 (假设 2) 取决于正确的函数设定形式 (假设 1), 且只有在满足可行性条件 (假设 3) 时才能消除估计偏差。

列出三个假设后, 作者进一步介绍了文章的估计策略。给定两组分别在控制条件和处理条件下得到的观测值 $\mathcal{O} = \{(i, t) | D_{it} = 0\}$ 与 $\mathcal{M} = \{(i, t) | D_{it} = 1\}$, 其中 \mathcal{O} 表示“已观测到的” (Observed), \mathcal{M} 表示“缺失的” (Missing)。尽管研究者选取的具体模型可能不尽相同, 但是其估计过程大抵均包含如下几个步骤:

第一步, 在未接受处理的观测值子集 \mathcal{O} 上对 Y_{it} 拟合, 得到 \hat{f} 和 \hat{h} 。该步骤依赖于 $f(\mathbf{X}_{it})$ 和 $h(\mathbf{U}_{it})$ 的函数形式假设和 \mathbf{U} 的低秩表达。

第二步, 通过 \hat{f} 和 \hat{h} 估计出每个接受处理的观测值的反事实结果 $Y_{it}(0)$ 。例如: $\hat{Y}_{it}(0) = \hat{f}(\mathbf{X}_{it}) + \hat{h}(\mathbf{U}_{it}), \forall (i, t) \in \mathcal{M}$ 。

第三步, 通过 $\hat{\delta}_{it} = Y_{it} - \hat{Y}_{it}(0)$ 对每个接受处理的观测值 $(i, t) \in \mathcal{M}$ 来估计其个体处理效应 δ_{it} 。

第四步, 根据研究兴趣对 δ_{it} 取平均值进行估计。

该估计策略没有将早处理者的观测值用作晚期处理组的控制组, 因此从根源上避开了负权重问题。此外, 这一方法尽可能地利用了数据信息, 并且没有对函数形式施加更强的假设, 因而更有效率。

(二) 三种估计量

1. 固定效应反事实估计量 (The Fixed Effects Counterfactual Estimator, FEct)。

在双重固定效应模型的基础上对 $Y_{it}(0)$ 进行插补, 即:

$$Y_{it}(0) = \mathbf{X}'_{it}\beta + \alpha_i + \xi_t + \varepsilon_{it}, \quad \forall (i, t), \quad \text{且有}$$

$$\sum_{D_{it}=0} \alpha_i = \sum_{D_{it}=0} \xi_t$$

易知在两组别、两时期、无协变量的经典 DID 条件下, FEct 估计量即为 DID 估计量。此外, FEct 与 Borusyak、Jaravel 和 Spiess 等三位学者提出的“有效估计量”是等价的。

2. 交互固定效应反事实估计量 (The Interactive Fixed Effects Counterfactual Estimator, IFect)。

当存在不可观测的随时间变化的混杂因素时, FEct 估计量得到的结果将是有偏的。而 IFect 使用因子增强模型对控制组的潜在结果进行建模:

$$Y_{it}(0) = \mathbf{X}'_{it}\beta + \alpha_i + \xi_t + \lambda'_i f_t + \varepsilon_{it}, \quad \forall (i, t)$$

换言之, $f(\mathbf{X}_{it}) = \mathbf{X}'_{it}\beta$, $h(\mathbf{U}_{it}) = \alpha_i + \xi_t + \lambda'_i f_t$ 。当模型设定正确时, IFect 的估计结果满足一致性。

IFEct 与徐老师之前的工作 gsynth 非常接近。

3. 矩阵补全估计量 (The Matrix Completion Estimator, MC)。

MC 将因果推断问题视作补全一个有缺失值的 $N \times T$ 矩阵。MC 假设 $[h(\mathbf{U}_{it})]_{i=1,2,\dots,N,t=1,2,\dots,T}$ 的 $N \times T$ 矩阵可以被更低阶的矩阵 \mathbf{L} 估计, 即

$$\mathbf{Y}(0) = \mathbf{X}\beta + \mathbf{L} + \boldsymbol{\varepsilon}$$

其中, \mathbf{Y} 是 $N \times T$ 的矩阵, \mathbf{X} 是 $N \times T \times k$ 的协变量, $\boldsymbol{\varepsilon}$ 是 $N \times T$ 的误差项矩阵。类似 IFEct, \mathbf{L} 可以被表示为两个低秩矩阵: $\mathbf{L} = \mathbf{A}\mathbf{F}$ 。不过与 IFEct 不同的是, MC 没有直接估计 \mathbf{A} 和 \mathbf{F} , 而是通过解决下列最小化问题来直接估计 \mathbf{L} :

$$\hat{\mathbf{L}} = \arg \min_{\mathbf{L}} \left[\sum_{(i,t) \in \mathcal{O}} \frac{(Y_{it} - L_{it})^2}{|\mathcal{O}|} + \lambda_L \|\mathbf{L}\| \right]$$

需要注意的是, IFEct 和 MC 的主要区别在于分解残差矩阵时对奇异值进行正则化的方式, 两种估计量何者表现更好视具体情形而定。此外, 在选择调谐参数 (Tuning Parameters) 和推断方法 (Inferential Methods) 时也应慎重。本部分更完整的设定及相关性质证明请见原论文及论文附录。

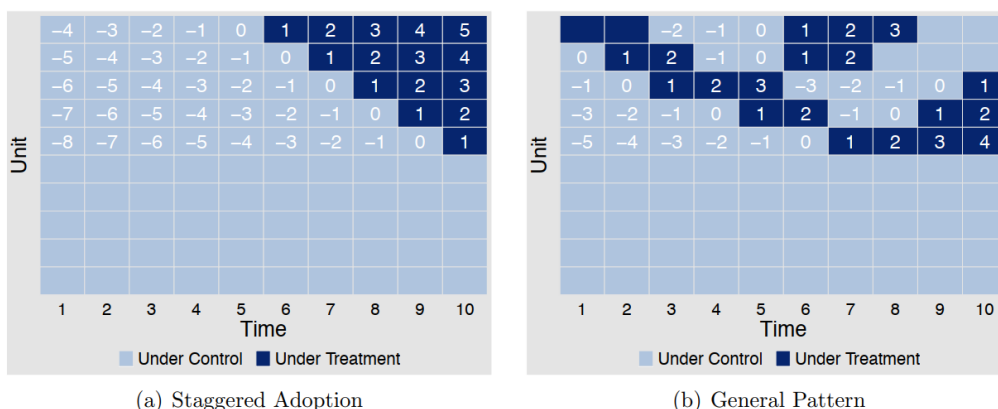
三、诊断方法

本部分作者介绍了一些诊断工具以帮助研究者分析设别假设的有效性, 包括动态处理效应图示和一系列统计检验 (安慰剂检验、事前趋势检验、延续效应检验) 等。

(一) 动态处理效应图示

在使用 TSCS 数据进行应用时, 研究者通常会绘制 “动态处理效应” (Dynamic Treatment Effects) 图示, 即处理变量与一组时期虚拟变量之间交乘项系数图。本文作者借助反事实估计量优化了动态处理效应图示这一方法, 绘制了处理组个体 Y_{it} 和 $\hat{Y}_{it}(0)$ 差异的平均值而非交乘项系数。与传统方法相比, 该方法主要有以下两个优点: 首先, 它放宽了处理效应不变的假设; 其次, 由于已经从 $\hat{\delta}_{it}$ 中减去个体未受处理的平均值, 因此无需再另外选择基准组。换言之, 基准组被设定为个体在分离时间效应后的未处理组平均值。下图展示了在渐进处理 (Staggered Adoption) 和更一般情况下的处理状态图示。

FIGURE 3. ESTIMATING THE DYNAMIC TREATMENT EFFECTS



Note: The above figure shows the treatment status with two hypothetical examples: (a) staggered adoption and (b) a general panel treatment structure. Numbers correspond to time relative to the onset of a treatment. Several cells in (b) are not assigned numbers because left or right censorship of data makes their relative positions to a treatment uncertain.

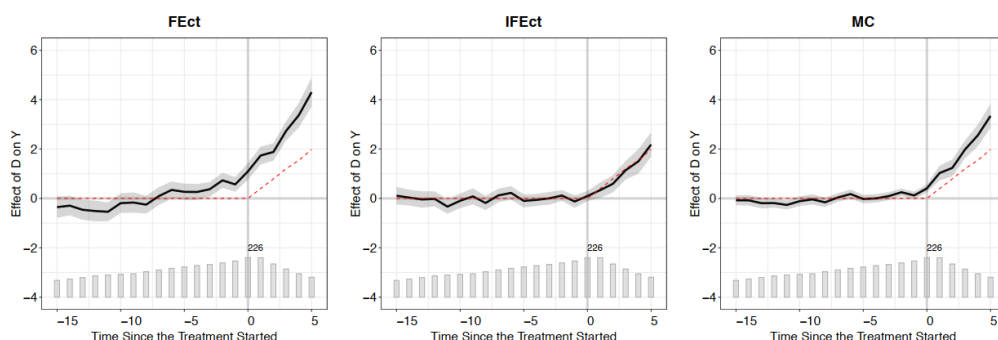
为了便于进一步检验分析，作者模拟生成了一个 200 个单位、共 35 个时期的面板数据，数据生成过程（Data Generating Process, DGP）与两个潜在因子 f_{1t} 和 f_{2t} 有关：

$$Y_{it} = \delta_{it}D_{it} + 5 + 1 \cdot X_{it,1} + 3 \cdot X_{it,2} + \lambda_{i1} \cdot f_{1t} + \lambda_{i2} \cdot f_{2t} + \alpha_i + \xi_t + \varepsilon_{it}$$

处理效应的异质性体现为： $\delta_{it} = 0.4s_t + e_{it}$ 。对于每个个体单位，处理状态都可能反复（可能由未处理变为已处理，也可能由已处理变为未处理）。

下图为使用 FEct、IFEct 和 MC 三种估计量对模拟数据进行估计后得到的动态处理效应，其中红色虚线为真正的 ATT。由于数据根据 IFEct 模型生成，因此 IFEct 对于该数据的估计效果应该最为准确，下图中间的结果也印证了这一点。而 FEct 的估计结果显示出了较强的事前趋势，估计出的 ATT 也与真实值相差甚远；MC 的估计出的 ATT 也有一定偏差。

FIGURE 4. DYNAMIC TREATMENT EFFECT FOR THE SIMULATED EXAMPLE



Note: The above figure shows the dynamic treatment effects estimates from the simulated data using three different estimators: FEct, IFEct, and MC. The bar plot at the bottom of each panel illustrates the number of treated units at the given time period relative to the onset of the treatment (the number decreases as time goes by because there are fewer and fewer units that are treated for a sustained period of time). The red dashed lines indicate the true ATT.

动态处理效应图示是一个强大的可视化工具，可以从直观上帮助研究者判断识别假设的合理性。并且，这种基于反事实估计得到动态处理效应，避免了 Sun & Abraham (2021) 等学者指出的由于效应异质性造成的估计偏差，且比他们提供的估计方案更简便易行。

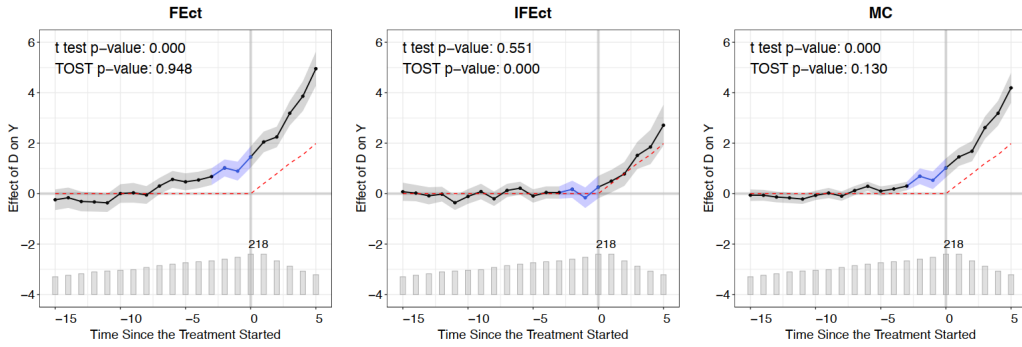
不过仅用图示难以分析出识别假设究竟是因何失效，因此还需要再进行其他检验。接下来，文章进一步介绍了一系列正式检验这些假设含义的统计方法与工具。

(二) 安慰剂检验

安慰剂检验的基本思想如下：我们假设处理比真实发生时期提前 S 期发生，然后使用反事实估计量对这段时期的平均处理效应进行估计。如果满足识别假设，那么这一虚构的平均处理效应（Fake ATT）应当几近于零。在进行 Fake ATT 是否显著异于 0 这一相等性检验时，作者介绍了均值差异法（Difference-in-means, DIM）和一种等效检验（Equivalence Test）的变体方法这两种方式。DIM 是进行此类检验的常用方法，如 t 检验或 F 检验，原假设为差异等于 0；然而当样本量较小时，很容易出现不能拒绝原假设这一情况，这意味着可能在差异实际不等于 0 的情况下而未能拒绝原假设（第 II 类错误）。而等效检验的原假设则为差异不等于 0，这意味着只有在明确拒绝原假设时才算通过了安慰剂检验——在小样本情况下，该种设置得到的结论通常更为保守可靠。不过，这种等效检验的方式需要研究者预设等值范围（Equivalence Range）以进行检验，因此需要研究者对处理效应大小有一定了解或预期。

安慰剂检验的优势之一是在模型错误设定结果仍较为稳健并且不会过度拟合，因为其主要依赖于安慰期 $Y(0)$ 的样本外预测情况。下图为使用 FEct、IFEct 和 MC 三种估计量进行安慰剂检验的结果。文中使用的两种检验方法具体为 t 检验和双单侧检验（Two One-Sided Test, TOST）。可以看到，IFEct 估计结果通过了 DIM 检验（ t 检验 p 值大于 0 很多，不能拒绝差异等于 0 的原假设）和等效检验（TOST 的 p 值非常接近于 0，拒绝有显著差异的原假设），而 FEct 和 MC 则未能通过检验。

FIGURE 5. PLACEBO TESTS FOR THE SIMULATED EXAMPLE



Note: The above figure shows the results of the placebo tests based on three different estimators: FEct, IFEct, and MC. The bar plot at the bottom of each panel illustrates the number of treated units at the given time period relative to the onset of the treatment. The red dashed lines indicate the true ATT. Three pretreatment periods ($s = -2, -1, 0$) serving as the placebo are rendered in blue. The p -values for the t test of the placebo effect and for the TOST are shown at the top-left corner of each panel. The equivalence range is set as $[-0.36\hat{\sigma}_\varepsilon, 0.36\hat{\sigma}_\varepsilon]$.

（三）两类拓展延伸

完成上述分析后，作者将安慰剂检验拓展应用于检验是否存在事前趋势（Pretrend）和是否存在延续效应（Carryover Effect）。

1. 事前趋势检验

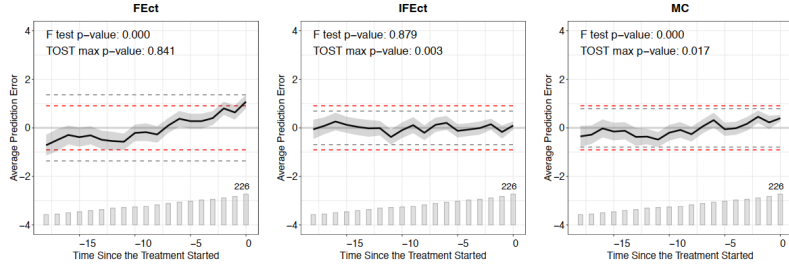
当潜在的时变混淆因素具有周期性或者在冲击发生之前未体现时，安慰剂检验可能很难捕捉到它。此时，考虑对处理前各个时期残差平均值为 0 的一系列假设进行联合检验，即对 $ATT_s = 0, s \leq 0$ 进行 F 检验。由于检验事前趋势也是一种相等性检验，所以作者也进行了等效检验，原假设如下：

$$ATT_s < -\theta_2 \text{ or } ATT_s > \theta_1, \quad \forall s \leq 0$$

其中 $[-\theta_2, \theta_1]$ 为等效范围。

借助安慰剂检验的基本思想，作者使用不断去掉一期（Leave-One-Period-Out）然后通过其他数据来估计该期值的方法，得到冲击发生前每个时期的平均样本外预测误差（Average Out-of-sample Prediction Error）。下图为使用 FEct、IFEct 和 MC 三种估计量进行等效检验的结果。灰色虚线为计算出来的最小范围（Minimum Range），即可以拒绝原假设的最小界限；红色虚线为等效范围。当灰色虚线落在红色虚线内时，可以认为通过了检验。由图可以看出，FEct 未能通过检验。IFEct 和 MC 均通过了该检验，不过 MC 的 F 检验的 p 值非常接近 0，模型很可能存在设定偏误。

FIGURE 6. TESTS FOR NO PRETEND: THE SIMULATED EXAMPLE

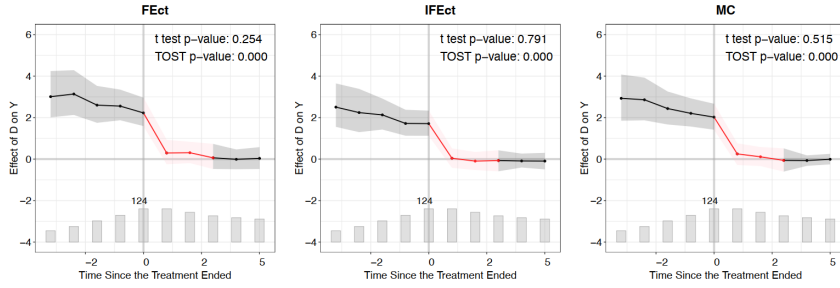


Note: The above figure shows the results of the equivalence tests based on three different estimators: FEct, IFect, and MC. Pretreatment average prediction errors and their 90% confidence intervals are drawn. The red dashed lines mark the equivalence range, while the gray dashed lines mark the minimum range. The bar plot at the bottom of each panel illustrates the number of treated units at the given time period relative to the onset of the treatment.

2. 延续效应检验

在该部分检验中，与前文不使用处理之前的几期不同，作者隐藏了处理之后的几期来预测 $Y_{it}(0)$ 。如果延续效应不存在，则应当期望平均预测偏误在这些时期接近于0。作者同样使用了DIM和等效检验两种方法。下图展示了使用FEct、IFect和MC三种估计量进行延续效应检验的结果。本文的数据生成过程中不涉及延续效应，而三种估计量的结果也均显示没有延续效应。

FIGURE 7. TESTS FOR NO CARRYOVER EFFECTS USING THE SIMULATED EXAMPLE



Note: The above figure shows the results of the tests for no carryover effects based on three different estimators: FEct, IFect, and MC. The bar plot at the bottom of each panel illustrates the number of treated units at the given time period relative to the end of the treatment. Three periods after the treatment ends are rendered in pink. The p -values for the t test of the carryover effect and for the TOST are shown at the top-right corner of each panel.

值得注意的是，如果延续效应检验不通过，并不意味着反事实估计方法无效。如果研究者发现在处理结束后的有限时间窗口中处理效应仍然存在，可以去掉这部分时期然后再来估计处理效果（类似前述检验）。此外，研究者可以将处理发生的定义改为“接受过处理则 $D_{it} = 1$ ，否则 $D_{it} = 0$ ”。此时情形被转化为渐进接受处理过程（Staggered Adoption Process），延续效应假设不再必要。

表1总结了这些检验方法。为了识别前述假设，研究者可以考虑使用传统的DIM方法（倘若统计效力足够）或者等价方法（倘若已经对估计效应大小有一定了解）。值得注意的是，无论采用何种方法，直观的可视化方式始终是避免无效识别假设带来推断错误的第一道防线。

TABLE 1. DIAGNOSTIC TESTS SUMMARY

	Placebo test		Testing (no) pretrend		Testing (no) carryover effects	
	t test	TOST	F test	TOST	t test	TOST
Null	$ATT^p = 0$	$ ATT^p > \theta$	$ATT_s = 0, \forall s \leq 0$	$ ATT_s > \theta, \exists s \leq 0$	$ACOE = 0$	$ ACOE > \theta$
If Rejecting the Null	Invalidate Assumptions	Support Assumptions	Invalidate Assumptions	Support Assumptions	Invalidate No Carryover	Support No Carryover
Equivalence threshold θ		$0.36\hat{\sigma}_\epsilon$ or eff		$0.36\hat{\sigma}_\epsilon$ or eff		$0.36\hat{\sigma}_\epsilon$ or eff

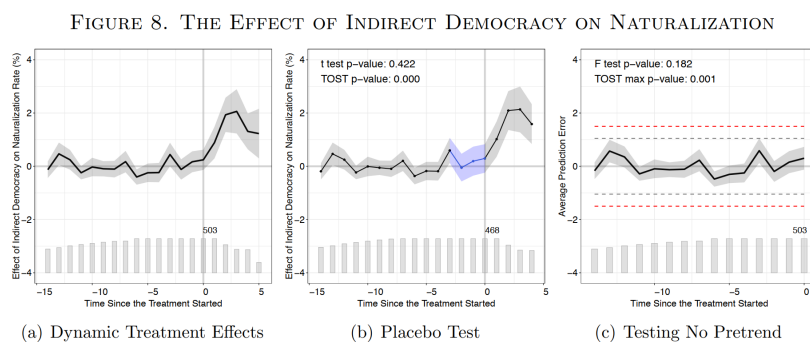
Note: Both the t and F tests are conventional difference-in-means tests, testing against the null of no difference. “Assumptions” refers to Assumptions 1-3 as a whole. $\hat{\sigma}_\epsilon$ is the standard deviation of the residuals after twoway fixed effects are partialled out using untreated data only. ATT^p denotes the average placebo treatment effect on the treated. $ACOE$ denotes the average carryover effect. “eff” represents an effect size that researchers deem reasonable.

四、实证示例

作者使用两个政治经济学研究实例来进一步对上述估计框架进行说明。作者首先使用 FEct 估计量来进行分析。倘若 FEct 估计量可以通过直观检测 (Eyeball Test) 和其他诊断检验, 那么通常无需再使用其他复杂的估计量; 倘若 FEct 估计量不能通过检测, 此时再考虑 IFect 和 MC 估计量。以下两个例子的安慰剂检验中设定 $S = 3$ 。

(一) 直接民主与入籍率

Hainmueller & Hangartner (2015) 使用广义 DID 研究了瑞士从直接民主转变为间接民主是否会提高各市少数民族移民的入籍率。结果变量为 i 市第 t 年的少数民族入籍率, 处理变量为入籍资格是否由全民公投决定 (全民公投意味着直接民主, 由官员代表民众投票意味着间接民主)。文章的数据集为 1991 年至 2009 年这 19 年间 1211 个瑞士城市数据。原文作者采用双重固定效应模型估计得出, 当决策由全民公投转为选举官员决定时, 入籍率平均增加了 1.339%。本文作者使用 FEct 估计量进行分析后得到的结果为 1.767, 略高于原始估计值。动态处理效应结果和安慰剂检验如下图所示。在本例中, FEct 结果与传统的多重固定效应模型的结果基本相同。然而, FEct 这类的反事实估计量允许研究以更便捷、更透明的方式来检验识别假设的有效性。



Note: The above figure shows the results from applying FEct to data from Hainmueller and Hangartner (2015). The left panel shows the estimated dynamic treatment effects using FEct. The middle panel shows the results from a placebo test using the “treatment” in three pretreatment periods as a placebo. The right panel shows the results of an equivalence test for no pretrend, in which the red and gray dashed lines mark the equivalence range and the minimum range, respectively. The bar plot at the bottom of each panel illustrates the number of treated units at a given time period relative to the onset of the treatment.

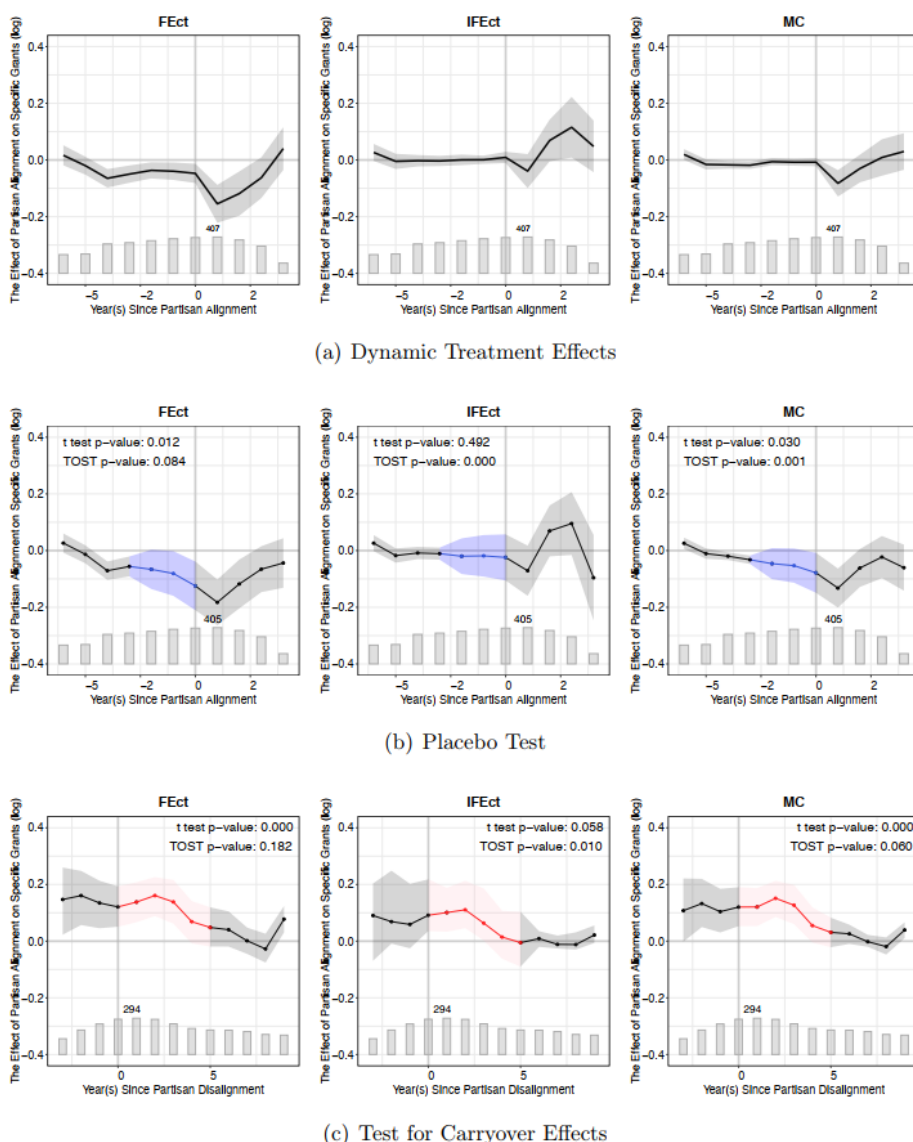
(二) 同党派与拨款分配

第二个例子来自 Fourinaies & Mutlu-Eren (2015), 作者研究了英国地方议会和中央政府同属一个党派是否会给当地带来更多拨款。结果变量为分配给地方议会的人均补助金的对数, 处理变量为执政党是否控制了地方议会。数据集为 1992 年至 2012 年的 466 个地区的数据。倘若直接使用不加入任何趋势控制的多重固定效应模型, 则会估计出负向的政策效果; 而当原文作者控制了地区线性时间趋势后, 估计结果显示与中央政府同属一个政党增加了地方收到来自中央的特定拨款 (峰值出现在三年之后)。

作者使用 FEct、IFect 和 MC 三种估计量进行了估计, 结果如下图 (a) 所示。FEct 估计结果显示事前趋势偏离了 0 值, 即部分识别假设条件可能没有得到满足; 然而, IFect 和 MC 的事前结果则非常接近 0 值。下图 (b) 显示了安慰剂检验结果, IFect 和 MC 均通过了安慰剂检验; 不过 IFect 的估计结果看起来更加平稳, 并且更加接近 (a) 的情形。下图 (c) 为延续效应检验, 显示处理效果大约持续至党徒结盟结束的五年之后。根据 IFect 的结果, 正向的延续效应至少在党徒结盟结束后的三年内持续存在。无论是原文作者的分析还是本文作者的检验, 都表明多重固定效应模型中如果不加入这些趋势交乘项控制, 其估计结果是有偏误的; 而 IFect 的估计结果与原文控制特定线性时

间趋势后的结果基本一致，在操作与估计上要简便很多，省去了加入过多交乘项可能对估计速度带来的麻烦。

FIGURE 9. THE EFFECT OF PARTISAN ALIGNMENT ON GRANT ALLOCATION



Note: The blue dots in (b) represent the periods used in the placebo tests. The red dots in (c) represent the periods used in the tests for no carryover effects.

五、结论

本文介绍了在放松同质性假设（Homogeneity Assumption）与考虑不可观测的随时间变化的混淆变量（Unobserved Time-varying Confounders）时，应用于长面板数据的反事实估计框架。该框架通过补齐处理组的反事实来对处理组的平均处理效应进行估计。此外，文章提供了一系列便于操作的检验方法以供研究者更便捷地判断识别假设的有效性。

文章在这一框架下具体讨论了 FEct、IFEct 和 MC 三种估计量。作者将这些估计量放在同一框架下进行讨论，为研究者分析、比较与应用这些方法提供了极大便利。表 2 对比了三种估计量和其他方法，可以看出这三种估计量有如下优势：它们避免了异质性处理效应下的负权重问题；适于一般性的面板结构数据，且无需舍弃部分数据；可以灵活地加入时变协变量，并有一系列的检验方法。

此外，IFEct 和 MC 还适用于包含可分解的时变混杂变量的问题。

TABLE 2. COMPARISON OF METHODS

	DID	wDID	DID _M	TWFE	FEct	IFEct/MC
Accommodate heterogeneous treatment effects	x	x	x		x	x
Allow treatment reversal			x	x	x	x
Condition on time-invariant covariates		x				
Condition on time-varying covariates				x	x	x
Use most available data	x	x		x	x	x
Easy-to-implement diagnostic tests	x			x	x	x
Condition on $U_{it} = \lambda'_i f_t$						x

Note: wDID, DID_M, TWFE represent weighted DID (Strezhnev 2018; Callaway and Sant’Anna 2021), DID multiple (de Chaisemartin and d’Haultfoeuille 2020), TWFE respectively. $U_{it} = \lambda'_i f_t$ represents decomposable time-varying confounders.

最后，作者推荐大家使用可视化手段和统计检验来判断识别假设的有效性，并且提供了七条建议供大家参考：

1. 绘制数据处理状态图，考察处理过程是否满足严格外生性假设；如果观察到不满足的迹象，则考虑使用基于序列可忽略性（Sequential Ignorability）的方法。
2. 以时间序列的形式绘制结果变量图示以辨别异常值和例外值（必要时转换数据的展示形式）。
3. 从最简单的估计量 FEct 着手绘制动态处理效应图，检查图示结果并进行一系列检验，包括使用传统的 DIM 方法（如 t 检验或 F 检验）或等效检验。
4. 倘若 FEct 没有通过安慰剂检验或事前趋势检验，则考虑使用更复杂的估计量，如 IFEct 和 MC，并再次进行检验。
5. 如果所选估计量未能通过延续效应检验，那么考虑去除处理后的几个时期重新进行估计，并再次进行检验。
6. 估计出处理效应后，可以考虑对进一步对子样本进行分析，以了解是哪一部分群体带来了最终效果。
7. 有效地向读者报告文章的发现，比如使用图示等较为直观的方法。

此外，作者提供了两个代码包 panelView 和 fect（R 和 Stata 均有相关命令）来帮大家进行上述分析。详情可参见徐轶青老师个人主页：<https://yiqingxu.org/>。