# 反事实框架与主流计量方法

连玉君　曹昊煜

兰州大学

2022 年 12 月 23 日

# 目录

# 主要内容

- 如何理解计量经济学的基本框架
- 反事实架构与主流计量经济方法
- 计量经济学与高质量论文的组织

1. 理解计量经济学
   - 经济建模
   - 估计方法
   - 统计推断
2. 反事实与主流计量方法
3. 计量经济学与高质量研究

# 线性建模思想

- 当我们面对两组数据 Y 和 X，如何更好地使用 X 去预测 Y 是计量经济学的“元问题”
- 预测优劣的判别需要特定的标准，通常使用的标准是均方误差 $\mathrm{MSE}(g) = \mathrm{E}(Y - g(X))^2$
- 最小化均方误差的最优解是条件期望

$$
\begin{aligned}
\mathrm{E}(Y|X) &= \mathrm{argmin}_g \ \mathrm{MSE}(g) \\
&= \mathrm{argmin}_g \ \mathrm{E}(Y - g(X))^2
\end{aligned}
$$

- 因此，使用条件期望来预测 Y 在 MSE 意义下是最优的

# 线性建模思想

- 在确定了用于预测的函数后，我们可以建立回归方程

$$Y = \mathrm{E}(Y|X) + \varepsilon$$

- 仅仅知道这些信息并不足够，因为条件期望的形式是未知的
- 考虑到形式上的约简性和解释的便利性，我们一般会假定条件期望符合线性形式，由此得到通常使用的回归方程

$$Y = X'\beta + \varepsilon$$

- 由此预测问题最终归结于一组线性参数的估计

# 线性建模思想

- 当 Y 为连续情形时，一般假定其服从正态分布，线性模型用来刻画条件期望
- 但如果 Y 为离散形式，正态分布的假定很可能是不满足的
- 仍然使用线性模型难以令人相信建模的合理性

# 非线性建模范例: Logit 模型

- 最常见的情形是被解释变量是 0-1 变量的情形
- 如果我们坚持使用线性模型来刻画条件期望，那么可以得到著名的线性概率模型

$$\mathrm{P}(Y = 1|X) = \mathrm{E}(Y|X) = X'\beta$$

- 概率总是在 0-1 之前取值，但线性函数却不满足这一性质
- 因此我们要先刻画概率，再对概率的参数建立线性模型

# 非线性建模范例: Logit 模型

- 若假定 Y 取 1 的概率服从逻辑分布，则该模型称为 Logit 模型，其中参数 $u$ 代表对数几率比

$$\mathrm{P}(Y=1)=\Lambda(u)=\frac{\exp(u)}{1+\exp(u)}$$

- 进一步对对数几率比进行参数建模，则有 $u=X'\beta$，则可以得到 Lgoit 模型的完整形式

$$\mathrm{P}(Y=1|X)=\Lambda(X'\beta)=\frac{\exp(X'\beta)}{1+\exp(X'\beta)}$$

- 理解视角 1: Y 取 1 的概率服从依赖于 X 的逻辑分布
- 理解视角 2: 对数几率比的函数形式是 X 的线性函数

# 非线性建模范例: Poisson 模型

- 在部分情况下，Y 的取值只能是非负的，如专利个数等。此时我们同样无法直接使用线性模型来刻画概率，更好的选择是泊松分布。

$$\mathrm{P}(Y=y)=\frac{e^{-\lambda}\lambda^{y}}{y!}$$

- 进一步对参数 $\lambda$ 建模，该参数的含义是 Y 的期望，为保证其非负性，使用指数函数形式。

$$E(Y|X)=\lambda=\exp(X'\beta)$$

- 理解视角 1: Y 取 y 的概率服从依赖于 X 的泊松分布
- 理解视角 2: 条件期望函数形式是 X 线性组合的指数变换

# 建模的基本思路

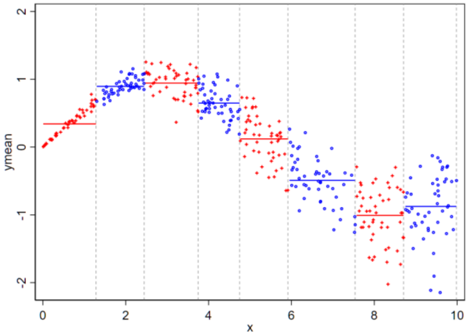
- 实证分析中更偏爱对条件期望建模是基于 MSE 的优化问题决定的
- 多数情况下，线性模型就可以较为简洁地刻画出经济变量间的关系
- 如果需要建立非线性的模型，那么需要使用“两步建模”的思想
- 到目前为止，我们仅仅是在讨论如何构建一个合理的经济模型，还没有涉及到参数的估计问题

# 参数估计

- 基本的参数估计包括 OLS，MLE 和 GMM 三类
- 三者都属于极值估计，一定情况下 GMM 可以转化为另外两种估计方式
- 无论是线性还是非线性模型，归根结底都是估计一组参数

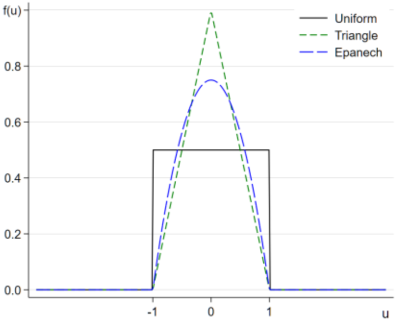
# 非参数估计

- 参数估计的缺陷在于对于函数形式的假定过于严格
- 非参数估计的主要思想是使用局部信息尽可能逼近数据
- 使用非参数估计需要明确局部的范围和权重的大小



# 非参数估计

- “局部”的范围由带宽决定，点的重要性由权重函数决定
- 常见的权重函数是核函数 $K(\frac{x_i - x}{h})$



# 非参数估计

- 核估计

$$\hat{m}_g(x) = \sum_i Y_i \cdot K\left(\frac{x_i - x}{h}\right) \Big/ \sum_i K\left(\frac{x_i - x}{h}\right)$$

- 局部线性回归

$$\min_{b_0, \mathbf{b}_1} \left\{ \sum_i \left[Y_i - b_0 - \mathbf{b}_1 \left(\mathbf{x}_i - \mathbf{x}\right)\right]^2 \cdot K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right\}$$

- 局部多项式回归

$$\min_{b_0, \mathbf{b}_p} \left\{ \sum_i \left[Y_i - b_0 - \sum_{p=1}^{P} \mathbf{b}_p (\mathbf{x}_i - \mathbf{x})^p\right]^2 \cdot K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right\}$$

# 标准误

- 无论是参数估计还是非参数估计，Estimator 总是随机的，我们只不过得到了当前样本下的**一次估计**而已
- 标准误正是刻画这种样本不确定性的指标，其形式依赖于我们对样本信息的假设
- 标准误的选择往往对应于差异化数据假设和对干扰项的理解

# 标准误

- 如果样本是 iid 的，那么标准误的形式非常简单

$$se(\hat{\beta}) = \frac{\hat{\sigma}^2}{SST}$$

- 如果存在异方差性，我们需要对标准误进行调整，即 White 稳健标准误

$$\hat{V} = n^{-1}\sum_{t=1}^{n} X_t X_t' e_t^2 = \frac{\boldsymbol{X}' D(e) D(e)' \boldsymbol{X}}{n}$$

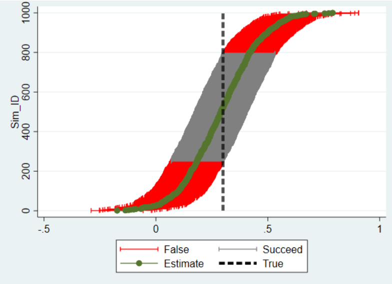
- 如果样本存在组内的相依性，那么仅使用上述的标准误可能存在很大问题

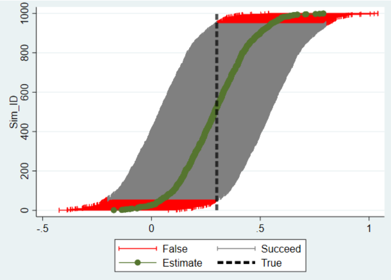
# 蒙特卡洛模拟：聚类问题

- 数据生成过程

$$
\begin{aligned}
&Y_{ij} = \beta_0 + \beta_1 X_{ij} + e_{ij}, \ \ i = 1, 2, \cdots, 240; \ \ j = 1, 2, \cdots, 30 \\
&X_{ij} = \eta_{ij} + \mu_j \\
&e_{ij} = \alpha_{ij} + \varepsilon_j \\
&\eta_{ij} \sim N(0, 1), \ \ \alpha_{ij} \sim N(0, 1), \ \ \mu_j \sim N(0, 4), \ \ \varepsilon_j \sim N(0, 4) \\
&\beta_0 = 0.5, \ \ \beta_1 = 0.3
\end{aligned}
$$

# 异方差与聚类稳健标准误



# 计量经济学四条主线

1. 建模: 对什么建模, 建立的模型形式是什么?
2. 估计: 使用参数估计还是非参数估计?
3. 推断: 是否需要进行聚类调整?
4. 数据: 截面数据、时间序列还是面板数据?

1 理解计量经济学

2 反事实与主流计量方法
- 潜在结果与因果参数
- 选择性偏误
- 两种识别思想
- 四种主流因果推断方法

3 计量经济学与高质量研究

# 潜在结果

- 如果存在“平行宇宙”，那么因果效应只需要比较两个世界中同一个个体受到处理与否导致的差异即可
- 但事实上，我们只能观测到一个个体的一个状态，处理组只能看到受处理的结果，控制组只能看到不受处理的结果
- 我们将同一个个体的两种结果分别表示为 $Y_{0i}$ 和 $Y_{1i}$，那么：

$$Y_i = Y_{0i} + D_i(Y_{1i} - Y_{0i})$$

# 因果参数

- 一般而言个体的因果效应存在异质性，即每个人的因果效应为:

$$TE_i = Y_{1i} - Y_{0i}$$

- 绝大多数情况下，我们无法处理个体的因果效应，因此我们转而考虑平均的因果效应，特定的因果参数有三个

$$\begin{aligned}
\mathrm{ATE} &= \mathrm{E}(Y_{1i} - Y_{0i}) = \mathrm{E}(Y_{1i}) - \mathrm{E}(Y_{0i}) \\
\mathrm{ATT} &= \mathrm{E}(Y_{1i} - Y_{0i}|D_i = 1) = \mathrm{E}(Y_{1i}|D_i = 1) - \mathrm{E}(Y_{0i}|D_i = 1) \\
\mathrm{ATU} &= \mathrm{E}(Y_{1i} - Y_{0i}|D_i = 0) = \mathrm{E}(Y_{1i}|D_i = 0) - \mathrm{E}(Y_{0i}|D_i = 0)
\end{aligned}$$

# 因果参数

- 有时候我们也可以获取一些协变量，那么我们转而计算条件因果参数

$$
\begin{aligned}
\mathrm{ATE}(\mathbf{x}) &= \mathrm{E}(Y_{1i} - Y_{0i}|\mathbf{x}) = \mathrm{E}(Y_{1i}|\mathbf{x}) - \mathrm{E}(Y_{0i}|\mathbf{x}) \\
\mathrm{ATT}(\mathbf{x}) &= \mathrm{E}(Y_{1i} - Y_{0i}|D_i = 1, \mathbf{x}) \\
&= \mathrm{E}(Y_{1i}|D_i = 1, \mathbf{x}) - \mathrm{E}(Y_{0i}|D_i = 1, \mathbf{x}) \\
\mathrm{ATU}(\mathbf{x}) &= \mathrm{E}(Y_{1i} - Y_{0i}|D_i = 0, \mathbf{x}) \\
&= \mathrm{E}(Y_{1i}|D_i = 0, \mathbf{x}) - \mathrm{E}(Y_{0i}|D_i = 0, \mathbf{x})
\end{aligned}
$$

- 简单地应用迭代期望率，可以将条件因果效应转换为无条件的因果效应

$$\mathrm{E}(Y) = \mathrm{E}[\mathrm{E}(Y|X)]$$

# 选择性偏误

- 如果潜在结果与处理变量独立，那么组间均值差就是因果效应的一致估计
- 如果独立性并不满足，那么会出现选择性偏误

$$
\begin{aligned}
& \mathrm{E}(Y \mid \mathbf{x}, D=1)-\mathrm{E}(Y \mid \mathbf{x}, D=0) \\
= & \mathrm{E}\left(Y_1 \mid \mathbf{x}, D=1\right)-\mathrm{E}\left(Y_0 \mid \mathbf{x}, D=0\right) \\
& +\left[\mathrm{E}\left(Y_0 \mid \mathbf{x}, D=1\right)-\mathrm{E}\left(Y_0 \mid \mathbf{x}, D=1\right)\right] \\
= & \underbrace{\left[\mathrm{E}\left(Y_0 \mid \mathbf{x}, D=1\right)-\mathrm{E}\left(Y_0 \mid \mathbf{x}, D=0\right)\right]}_{\textit{Selection Bias}}+\mathrm{ATT}(\mathbf{x})
\end{aligned}
$$

- 如果我们直接比较组间均值，得到的是 ATT 的有偏估计

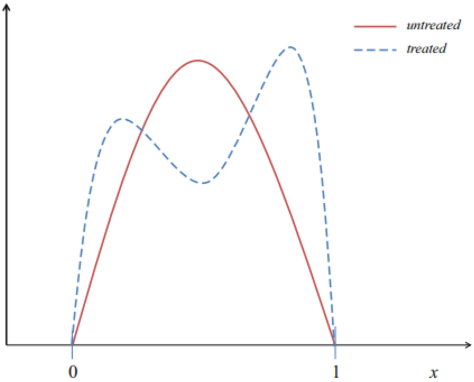
# 选择偏误的分解

- Heckman(1998) 将选择性偏差分解为三个部分

$$
\begin{aligned}
\mathrm{B}_1 &= \int_{S_{1\mathrm{x}}} \bar{y}_{01}\, dF(\mathrm{x}, D=1) - \int_{S_{0\mathrm{x}}} \bar{y}_{00}\, dF(\mathrm{x}, D=0) \\
&= \underbrace{\int_{S_{1\mathrm{x}}-S_{\mathrm{x}}} \bar{y}_{01}\, dF(\mathrm{x}, D=1) - \int_{S_{0\mathrm{x}}-S_{\mathrm{x}}} \bar{y}_{01}\, dF(\mathrm{x}, D=0)}_{\mathrm{B_A}} \\
&\quad + \underbrace{\int_{S_{\mathrm{x}}} \bar{y}_{00}\, [dF(\mathrm{x}, D=1) - dF(\mathrm{x}, D=0)]}_{\mathrm{B_B}} + \underbrace{\int_{S_{\mathrm{x}}} \bar{y}_{01}\, dF(\mathrm{x}, D=1) - \int_{S_{\mathrm{x}}} \bar{y}_{00}\, dF(\mathrm{x}, D=1)}_{\mathrm{B_C}}
\end{aligned}
$$

- 选择性偏误的三个来源分别是**弱平衡、弱重叠和不可观测的影响因素**
- $B_A$ 和 $B_B$ 一般称为显性偏误，$B_C$ 称为隐性偏误

# 平衡与重叠



# 何谓识别？

- 传统计量经济学和因果推断中，对于“识别”的理解存在一定的差异
- 传统计量经济学中，识别问题主要指的是能否满足一定的代数条件以确保参数的计算
- 在因果推断计量经济学中，识别问题主要指的是估计得出的参数是否具有因果含义
- 传统计量中保证识别的条件一般是一组代数条件，而因果推断中的识别条件是对分配的假设

# 选择机制与识别策略

- 如何正确地识别因果效应取决于我们对分配机制的理解
- 如果选择机制是明确的，那么不存在隐性偏误，因为此时满足条件独立性

$$(Y_1, Y_0) \perp D | \mathbf{x}$$

- 如果选择机制并不清晰，那么我们无法通过加入协变量的方式消除选择性偏误，需要更加复杂的识别假设
- 因此，基本的识别策略可以分为“依可观测变量的选择问题”和“依不可观测变量的选择问题”

# 主要的识别方法

- 依可观测变量的选择
  - 回归调整方法
  - 倾向得分匹配
- 依不可观测变量的选择
  - 双重差分法
  - 断点回归法

# 回归调整法

- 回归调整需要的假设是条件均值独立 $\mathrm{E}(Y_1|\mathbf{x}, D) = \mathrm{E}(Y_1|\mathbf{x})$ 和 $\mathrm{E}(Y_0|\mathbf{x}, D) = \mathrm{E}(Y_0|\mathbf{x})$，或者

$$\mathrm{E}(Y_0|\mathbf{x}, D=1) = \mathrm{E}(Y_0|\mathbf{x}, D=0)$$
$$\mathrm{E}(Y_1|\mathbf{x}, D=0) = \mathrm{E}(Y_1|\mathbf{x}, D=1)$$

- 因果参数识别

$$\begin{aligned}\mathrm{ATE}(\mathbf{x}) &= \mathrm{E}(Y_1|\mathbf{x}) - \mathrm{E}(Y_0|\mathbf{x}) \\ &= \mathrm{E}(Y_1|\mathbf{x}, D=1) - \mathrm{E}(Y_0|\mathbf{x}, D=0) \\ &= \mathrm{E}(Y|\mathbf{x}, D=1) - \mathrm{E}(Y|\mathbf{x}, D=0)\end{aligned}$$

- 回归调整方法的主要问题仍然是对条件期望建模

# 回归调整法

- 使用回归方法来得到 ATE 的估计需要进行建模
- 对反事实建立一个一般化的参数模型

$$Y_1 = \mu_1 + g_1(\mathbf{x}) + e_1$$
$$Y_0 = \mu_0 + g_0(\mathbf{x}) + e_0$$

- 其中 $\mathrm{E}(e_g|\mathbf{x}) = 0$，该模型隐含了

$$\mathrm{E}(Y_1|\mathbf{x}, D = 1) = \mathrm{E}(Y|\mathbf{x}, D = 1) = \mu_1 + g_1(\mathbf{x})$$
$$\mathrm{E}(Y_0|\mathbf{x}, D = 0) = \mathrm{E}(Y|\mathbf{x}, D = 0) = \mu_0 + g_0(\mathbf{x})$$

- 以及当两个模型的函数形式相同时，因果参数

$$\mathrm{ATE} = \mathrm{E}(Y_1|\mathbf{x}) - \mathrm{E}(Y_0|\mathbf{x}) = \mu_1 - \mu_0$$

# 回归调整法

- 将反事实与实际观测关联在一起

$$Y = \mu_0 + D(\mu_1 - \mu_0) + g_0(\mathbf{x}) + D\left[g_1(\mathbf{x}) - g_0(\mathbf{x})\right] + e$$

- 因此当两个模型函数形式相同时，需要使用的回归方程是

$$Y = \alpha + \beta \cdot D + g_0(\mathbf{x}) + e$$

- 此时我们说参数 $\beta$ 识别了因果参数，而应用的识别假设是条件独立假设

# 倾向得分匹配

- 匹配方法依赖的假定有三个
  - 条件均值独立。$\mathrm{E}(Y_g|D,\mathbf{x}) = \mathrm{E}(Y_g|\mathbf{x})$
  - 重叠假定。倾向得分满足 $0 < p(\mathbf{x}) = \Pr(D = 1|\mathbf{x}) < 1$
  - 平衡性。$\{(D \perp \mathbf{x})|\mathrm{Matching}\}$，即匹配后组间特征是平衡的
- 因果参数的识别

$$
\begin{aligned}
\mathrm{ATT} &= \mathrm{E}(Y_1|D=1,\mathbf{x}) - \mathrm{E}(Y_0|D=1,\mathbf{x}) \\
&= \mathrm{E}(Y_1|D=1,\mathbf{x}) - \mathrm{E}(Y_0|D=1,\mathbf{x}) \\
&\quad + [\mathrm{E}(Y_0|D=0,\mathbf{x}) - \mathrm{E}(Y_0|D=0,\mathbf{x})] \\
&= \mathrm{E}(Y_1|D=1,\mathbf{x}) - \mathrm{E}(Y_0|D=0,\mathbf{x})
\end{aligned}
$$

- 匹配方法实际上假定了匹配后，每一个分组内都是一个近似的随机试验

# 倾向得分匹配

- 反事实的定义

$$\widehat{Y}_{0i} = \begin{cases} Y_i & \text{if } D_i = 0 \\ \sum_{j \in \mathrm{C}(i)} h(i,j) Y_j & \text{if } D_i = 1 \end{cases}$$

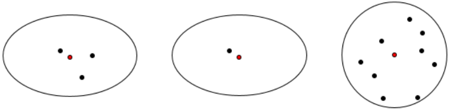
$$\widehat{Y}_{1i} = \begin{cases} \sum_{j \in \mathrm{C}(i)} h(i,j) Y_j & \text{if } D_i = 0 \\ Y_i & \text{if } D_i = 1 \end{cases}$$

- 估计因果效应

$$\widehat{\mathrm{ATT}} = \frac{1}{N_1} \sum_{i \in \{D=1\}} \left( Y_i - \widehat{Y}_{0i} \right) = \frac{1}{N_1} \sum_{i \in \{D=1\}} \left( Y_i - \sum_{j \in C(i)} h(i,j) Y_j \right)$$

# 倾向得分匹配

- 匹配方法的基本思想是使用特定点附近样本的加权平均估计反事实
    - 多么接近才能称之为“附近”?
    - 对于不同的样本如何赋予权重?
- 不同的匹配方法可以有不同的定义



# 倾向得分匹配

- 常见匹配方法的控制组选择与权重计算

表 4: 倾向得分匹配方法

| 匹配方法 | $C(i)$ | $h(i,j)$ |
|---|---|---|
| 一对一匹配 | $\{\text{Singleton } j : \min_j \|p_i - p_j\|\}$ | $1$ |
| M 近邻匹配 | $\{\text{First } j : \min_j \|p_i - p_j\|\}$ | 1/M |
| 卡尺匹配 | $\{j : \|p_i - p_j\| < r\}$ | $\frac{1}{N_{C(i)}}$ |
| 核匹配 | 全部控制组个体 | $\frac{K_{ij}}{\sum_{i\in C} K_{ij}}$ |
| 局部线性匹配 | 全部控制组个体 | $\frac{K_{ij}L_i^2 - K_{ij}\widehat{\Delta}_{ij}L_i^1}{\sum_{j\in C}\left(K_{ij}L_i^2 - K_{ij}\widehat{\Delta}_{ij}L_i^1 + r_L\right)}$ |
| 岭匹配 | 全部控制组个体 | $\frac{K_{ij}}{\sum_{j\in C} K_{ij}} + \frac{\widetilde{\Delta}_{ij}}{\sum_{j\in C}\left(K_{ij}\widetilde{\Delta}_{ij}^2 + r_R h\left\|\widetilde{\Delta}_{ij}\right\|\right)}$ |
| 分层匹配 | 全部控制组个体 | $\frac{\sum_{b=1}^{B} \mathbf{1}[p(\mathbf{x}_i)\in I(b)]\cdot\mathbf{1}\left[p\left(\mathbf{x}_j\right)\in I(b)\right]}{\sum_{b=1}^{B} \mathbf{1}\left[p\left(\mathbf{x}_j\right)\in I(b)\right]}$ |

# 依可观测变量选择问题

- 回归调整和倾向得分匹配都旨在解决显性偏误的问题
- 二者依赖的核心假设实质是是相同的，都是给定协变量的取值后，组内近似于随机试验
- 这一假设是很强的，因为翻译成传统计量的语言，其意味着控制住所有协变量后，D 不存在内生性问题
- 多数情况下我们对处理的分配机制并不清楚，隐性偏误总是存在的。此时因果的识别需要依靠其他设定

# 双重差分法

- 同质性处理效应假设，对于 $2 \times 2$ 设计

$$\mathrm{ATE}(s,t) = \mathrm{E}(Y_{1st} - Y_{0st} | s,t) = \delta$$

即平均因果效应被假定为常数

- 平行趋势假设

$$\mathrm{E}(Y_{0st}|s,t) = \gamma_s + \lambda_t$$
$$\mathrm{E}(Y_{1st}|s,t) = \gamma_s + \lambda_t + \delta$$

也就是不论任何地区或时间，潜在结果的变化趋势是相同的，差异仅仅体现在截距上

- 无预期效应，即处理前未受到处理的反事实不会受到未来处理与否的影响

# 数值示例

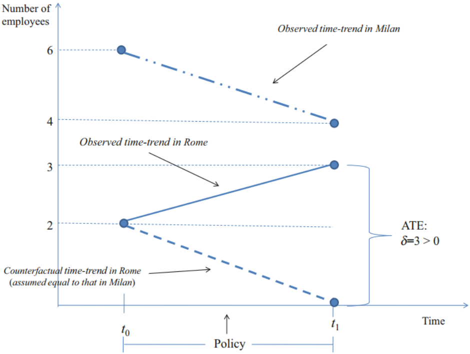
- 令处理组为 R，控制组为 M，那么待估计的 ATT 为

$$ATT_R = \mathrm{E}(Y_{1Rt_1}) - \mathrm{E}(Y_{0Rt_1})$$

- 根据潜在结果的建模

$$\mathrm{E}(Y_{1Rt_1}) = \gamma_R + \lambda_{t_1} + \delta = 3$$
$$\mathrm{E}(Y_{0Rt_1}) = \gamma_R + \lambda_{t_1} = ?$$

- 无法直接识别出因果参数



# 数值示例

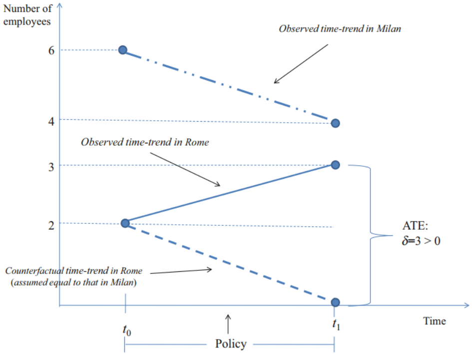
- 对于处理组

$$\mathrm{E}(Y_{1Rt_1}) = \gamma_R + \lambda_{t_1} + \delta = 3$$
$$\mathrm{E}(Y_{0Rt_0}) = \gamma_R + \lambda_{t_0} = 2$$

- 从而因果效应可以使用下面的式子计算

$$\delta = 1 - (\lambda_{t_1} - \lambda_{t_0})$$

- 使用控制组计算趋势变化



# 数值示例

- 对于控制组

$$\mathrm{E}(Y_{0Mt_1}) = \gamma_M + \lambda_{t_1} = 4$$
$$\mathrm{E}(Y_{0Mt_0}) = \gamma_M + \lambda_{t_0} = 6$$

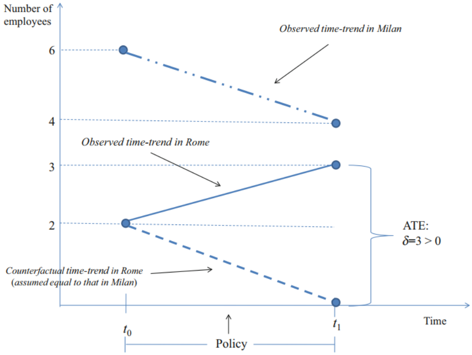
- 从而趋势变化为

$$\lambda_{t_1} - \lambda_{t_0} = 4 - 6 = -2$$

- 最后可以识别出因果效应

$$ATT_R = 1 - (-2) = 3$$

- 关键假定是平行趋势假定和同质性因果效应假定!!



# 双重差分法

- 根据平行趋势假设，对潜在结果建模

$$
\begin{aligned}
Y_{0st} &= \gamma_s + \lambda_t + e_{0st} \\
Y_{1st} &= \gamma_s + \lambda_t + \delta + e_{1st} \\
Y_{st} &= Y_{0st} + D_{st}(Y_{1st} - Y_{0st})
\end{aligned}
$$

- 做一些替换，可以得到观测值方程

$$
Y_{st} = \gamma_s + \lambda_t + D_{st} \cdot \delta + e_{st}
$$

- 此时称 $\delta$ 识别了因果效应 ATT，使用结果变量对**双向固定效应和处理变量**进行 OLS 估计即可得到因果效应的估计

# 检验平行趋势假定

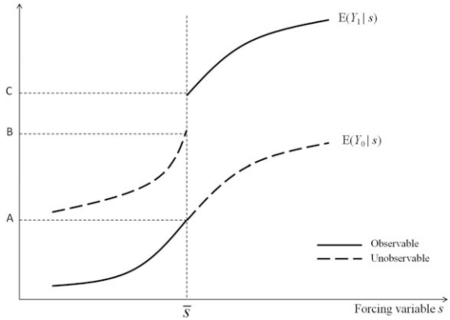
- 平行趋势假定是无法直接检验的
- 在多期情形下可以使用“事件研究法”来考察动态效应

$$Y_{it} = \alpha_0 + \sum_{p=-b, p\neq -1}^{T} \tau_p D_{i,t=p} + \varepsilon_{it}$$

- 使用政策开始前一期为基期，并纳入所有政策开始前后的时期数

# 断点回归

- 局部随机性：如果“标准”是相当随机的，那么断点的邻域内，样本将非常相似，可以作为彼此的反事实估计
- 连续性：断点两侧反事实结果的条件期望是连续的



# 清晰断点

- 驱动变量 (Forcing variable) s 完全决定了分配机制

$$D_i = \mathbf{1}(s_i \geq \bar{s})$$

- 因果参数定义为断点两侧的极限值差

$$\begin{aligned} \mathrm{ATE}_{\mathrm{SRD}} &= \mathrm{E}\left(Y_1 \mid s = \bar{s}\right) - \mathrm{E}\left(Y_0 \mid s = \bar{s}\right) \\ &= \lim_{s \downarrow \bar{s}} \mathrm{E}(Y \mid S = s) - \lim_{s \uparrow \bar{s}} \mathrm{E}(Y \mid S = s) \end{aligned}$$

- 清晰断点回归的识别假设是反事实条件期望的连续性

# 识别问题

- 在断点两侧的局部范围内 $\forall\ i \in \{\bar{s} - h < s_i < \bar{s} + h\}$ 进行建模，截距差即为因果效应的估计

$$
\begin{aligned}
Y_i &= \alpha_L + \delta_L(s_i - \bar{s}) + \varepsilon_{L,i} \\
Y_i &= \alpha_R + \delta_R(s_i - \bar{s}) + \varepsilon_{R,i}
\end{aligned}
$$

- 此时 $\mathrm{ATE} = \alpha_R - \alpha_L$。为了获得因果效应的统计推断，将两个方程合为一个

$$
Y_i = \alpha_L + \mathrm{ATE} \cdot D_i + \delta_L\left(s_i - \bar{s}\right) + \left(\delta_R - \delta_L\right) \cdot D_i \cdot \left(s_i - \bar{s}\right) + \varepsilon_i
$$

- 此时我们说 $D_i$ 的参数识别了因果效应

# 其他非参数估计

- 可以使用核估计简单地比较断点两侧样本的加权平均

$$\widehat{\text{ATE}}_{\mathrm{SRD}} = \frac{\sum_{i\in\{R\}} K\left(\frac{s_i-\bar{s}}{h}\right) Y_i}{\sum_{i\in\{R\}} K\left(\frac{s_i-\bar{s}}{h}\right)} - \frac{\sum_{i\in\{L\}} K\left(\frac{s_i-\bar{s}}{h}\right) Y_i}{\sum_{i\in\{L\}} K\left(\frac{s_i-\bar{s}}{h}\right)}$$

- 使用断点两侧的样本进行局部多项式回归

$$Y_i = \alpha_L + \mathrm{ATE}\cdot D_i + \sum_{p=1}^{P} \delta_{L,p}\left(s_i - \bar{s}\right)^p + D_i \sum_{p=1}^{P}\left(\delta_{R,p} - \delta_{L,p}\right)\cdot\left(s_i - \bar{s}\right)^p + \varepsilon_i$$
$$\forall\ i \in \{\bar{s} - h < s_i < \bar{s} + h\}$$

- 使用交叉验证等方法确定最优带宽

# 因果分析的基本思想

- 定义因果效应并确定识别假设
- 对反事实进行建模，刻画出反事实的函数形式，除了倾向得分匹配外，其他方法都对反事实进行了参数建模
- 建模后再次推导因果效应的形式，并将反事实与观测值联系起来得到回归方程，明确哪一个参数代表因果效应
- 使用参数或非参数方法估计，并进行统计推断

1 理解计量经济学

2 反事实与主流计量方法

3 计量经济学与高质量研究
- 如何构建计量经济学基础
- 评审专家的目光：以一篇论文为例
- 论文成功的要素

# 计量经济学基础构建

- 注重传统理论计量的学习
  - 全面掌握计量经济学基础，可以参考伍德里奇的《计量经济学导论》
  - 全面学习理论计量经济学的基本范式，可以参考洪永淼老师或者 Hayshi 的《计量经济学》
  - 从全面学习掌握到专题化学习，增加非参数和贝叶斯计量经济学的学习
- 注重因果推断理论的学习
  - 不必完全掌握因果推断的全部细节，但基本的逻辑框架必须搭建
  - 学习主流的基本因果推断教材，如《Mostly Harmless Econometrics》《Econometric Evaluation of Socio-Economic Programs》《Causal Inference: TheMixtape》
  - 通过追踪前沿文献和最新命令来实现方法的积累
- 注重计量经济学软件工具
  - 打好 Stata 的基础，连老师的课程是首选
  - 多写代码，而不是仅仅用眼睛看别人的代码
  - 通过学习软件的帮助文档提高命令的积累

# 评审专家关注的要点

- 核心要点 1: 创新的效力
- 核心要点 2: 理论建模的含义
- 核心要点 3: 研究设计
- 核心要点 4: 度量问题
- 核心要点 5: 数据清洗
- 核心要点 6: 结果解释及其自洽性

# 我心中论文成功的要素

- 可信是关键，在实证开始之前就让人相信你的故事
  - 尝试必要的数理建模
  - 提取明确的特征事实
- 明确研究设计及其对应的假设，必要的时候提供一些检验增强对假设的信心
- 做好数据清洗和变量度量，避免前后“两张皮”
- 细致地阐述机制，可信的机制是论文成功的必要条件
- 小心地组织全文的证据，并不是越多越好
- 提高行文间的逻辑性

# Thank You