

The Augmented Synthetic Control Method

增广合成对照方法

史长浩

中国人民大学统计学院

2022 年 12 月 14 日



中國人民大學

RENMIN UNIVERSITY OF CHINA

- ① 研究背景
- ② 准备工作
- ③ 增广合成对照方法
- ④ 校准模拟和实证分析
- ⑤ 讨论

① 研究背景

② 准备工作

③ 增广合成对照方法

④ 校准模拟和实证分析

⑤ 讨论

一个例子：评估政策干预效果

表 1: 1990-2013 年 A 国各州州内生产总值 (Gross State Product)。

	1990	1991	...	2012	2013
A 州	$Y_{1,1}$	$Y_{1,2}$...	$Y_{1,T-1}$	$Y_{1,T}$
B 州	$Y_{2,1}$	$Y_{2,2}$...	$Y_{2,T-1}$	$Y_{2,T}$
⋮	⋮	⋮	⋮	⋮	⋮
Z 州	$Y_{N,1}$	$Y_{N,2}$...	$Y_{N,T-1}$	$Y_{N,T}$

- 除 A 州外，其余各州不受相关政策干预；A 州作为试点在 2012 年末施行减税政策。
- 希望评估减税政策对 A 州经济状况的影响，即：若不进行政策干预，2013 年 A 州的 GSP 与现实相差多少？

一个例子：评估政策干预效果

表 2: 1990-2013 年 A 国各州对数生产总值 (log Gross State Product)。

	1990	1991	...	2012	2013
A 州	$Y_{1,1}(0)$	$Y_{1,2}(0)$...	$Y_{1,T-1}(0)$	$Y_{1,T}(1)$
B 州	$Y_{2,1}(0)$	$Y_{2,2}(0)$...	$Y_{2,T-1}(0)$	$Y_{2,T}(0)$
⋮	⋮	⋮	⋮	⋮	⋮
Z 州	$Y_{N,1}(0)$	$Y_{N,2}(0)$...	$Y_{N,T-1}(0)$	$Y_{N,T}(0)$

- 将表 1 问题用 [Neyman, 1923] 潜在结果的语言重新叙述，得表 2。希望识别的因果参数为 $\tau = Y_{1,T}(1) - Y_{1,T}(0)$ 。核心问题在于识别 $Y_{1,T}(0)$ ，一种想法是基于其他州 2013 年可观测的数据合成 A 州 2013 年不可观测的数据作为对照，这即是合成对照方法的思想。

文献综述

- 合成对照方法 (Synthetic Control Methods), 简称 SCM, 最早由 [Abadie and Gardeazabal, 2003] 和 [Abadie, Diamond and Hainmueller, 2010, 2015] 提出, 用于估计面板数据中单个单元的处理效应。
- 此后, 一个方向的工作是, 在保持经典 SCM 对权重严格限制的条件下发展一系列更稳健的估计方法: 在 [Abadie, Diamond, and Hainmueller, 2015] 的基础上, [Robbins, Saunders and Kilmer, 2017], [Doudchenko and Imbens, 2017] 和 [Abadie and L'Hour, 2018] 在 SCM 的优化问题中加入对权重的惩罚; [Gobillon and Magnac, 2016] 探讨了维数约简的策略和其他可以提高后续估计量性能的数据变换方法。

文献综述

- 第二个方向的工作是，放松经典 SCM 对于权重的限制，特别是对于非负性的限制：[Doudchenko and Imbens, 2017] 认为在很多问题中负权是有意义的；[Amjad, Shah and Shen, 2018] 提出了一个有趣的变体，将负权重与预处理步骤相结合；[Powell, 2018] 则允许通过 Frisch-Waugh-Lovell 型投影进行外推（extrapolation），类似地推广了经典 SCM 方法；[Doudchenko and Imbens, 2017] 和 [Ferman and Pinto, 2018] 都将截距项纳入 SCM 问题中。
- 第三个方向的工作是，修正 SCM 中的偏倚：[Abadie and L'Hour, 2018] 基于回归方法提出纠偏方法；[Kellogg et al., 2020] 提出加权平均和匹配以权衡插值和外推偏差；[Arkhangelsky et al., 2019] 提出合成双重差分估计量。

文献综述

- 在最新的工作中，一些研究者建议使用结果模型（outcome modelling）而非 SCM 式的加权平均，他们包括：[Athey et al., 2017] 的矩阵填充（matrix completion）方法，[Xu, 2017] 的广义合成对照方法，以及 [Hsiao et al., 2018] 的组合方法（combined approaches）。

- 1 研究背景
- 2 准备工作
- 3 增广合成对照方法
- 4 校准模拟和实证分析
- 5 讨论

符号与设定

- 考虑标准的 SCM 面板数据，设有 N 个单元 $i = 1, \dots, N$ 在 T 个时间点 $t = 1, \dots, T$ 的观测值，单元 i 只可能在时刻 T_0 ($T_0 < T$) 接受处理，设 W_i 为单元 i 接受处理的示性函数。
- 只考虑单个单元接受处理的情况，不失一般性，设 $W_1 = 1$ ，剩余的 $N_0 = N - 1$ 个单元被称为供体 (donors)，为了简化符号，仅考虑只有一个处理后观测的情况，即 $T = T_0 + 1$ 。本文讨论的方法很容易能拓展到多个单元接受处理及存在多个处理后观测的情况。

符号与设定

- 作 SUTVA 假定，这使得潜在结果有良定义，且保证了每个单元的潜在结果不受其他单元处理的影响，将第 i 个单元在 t 时刻在对照和处理条件下的潜在结果分别记为 $Y_{it}(0)$ 和 $Y_{it}(1)$ 。
- 作一致性假定，该假定建立了观测数据与潜在结果之间的联系，得式 (1)。

$$Y_{it} = \begin{cases} Y_{it}(0) & \text{if } W_i = 0 \text{ or } t \leq T_0, \\ Y_{it}(1) & \text{if } W_i = 1 \text{ and } t > T_0. \end{cases} \quad (1)$$

- 为了强调预处理结果是被当作协变量对待的，之后使用 X_{it} 代替 Y_{it} ，对任意的 $t \leq T_0$ 。
- 感兴趣的因果参数为 $\tau = Y_{1T}(1) - Y_{1T}(0)$ 。

符号与设定

- 数据结构如式 (2) 所示, 其中 \mathbf{X}_0 . 为控制单元预处理结果 (control unit pretreatment outcomes), \mathbf{Y}_{0T} 为控制单元结果 (control unit outcomes), \mathbf{X}_1 . 为处理单元预处理结果 (treated unit pretreatment outcomes)。

$$\left(\begin{array}{cccc|c} X_{11} & X_{12} & \dots & X_{1T_0} & Y_{1T} \\ X_{21} & X_{22} & \dots & X_{2T_0} & Y_{2T} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{N1} & X_{N2} & \dots & X_{NT_0} & Y_{NT} \end{array} \right) \equiv \left(\begin{array}{c|c} \mathbf{X}_1. & Y_{1T} \\ \mathbf{X}_0. & \mathbf{Y}_{0T} \end{array} \right). \quad (2)$$

- 之后将交替使用预处理拟合 (pretreatment fit) 和协变量平衡 (covariate balance) 两个术语。

数据生成机制的假定

- 假设 $Y_{it}(0)$ 的生成机制为 $Y_{it}(0) = m_{it} + \varepsilon_{it}$, 其中 m_{it} 为主项, $\varepsilon_{it} \sim P(\cdot)$ 为可加噪声项, 服从尺度参数为 σ 的次高斯分布 (sub-Gaussian)。本文考虑两种特殊的设定。
- (a) $Y_{iT}(0)$ 是其滞后项的线性函数, 即

$$Y_{iT}(0) = \sum_{\ell=1}^{T_0} \beta_{\ell} Y_{i(T-\ell)}(0) + \varepsilon_{iT}. \quad (3)$$

- (b) $Y_{it}(0)$ 是一些隐因子的线性函数, 即

$$Y_{it}(0) = \phi_i \cdot \mu_t + \varepsilon_{it} = \sum_{j=1}^J \phi_{ij} \mu_{jt} + \varepsilon_{it}, \quad (4)$$

其中 μ_t 是时变的因子, ϕ_i 是随单位改变的因子载荷。

合成对照方法

合成对照方法归结为求解如下约束优化问题 [Abadie, Diamond and Hainmueller, 2015]:

$$\begin{aligned} \min_{\gamma} \quad & \left\| \mathbf{V}_{\mathbf{x}}^{1/2} (\mathbf{x}_{1\cdot} - \mathbf{x}'_{0\cdot} \gamma) \right\|_2^2 + \zeta \sum_{W_i=0} f(\gamma_i), \\ \text{subject to} \quad & \sum_{W_i=0} \gamma_i = 1, \\ & \gamma_i \geq 0 \quad i : W_i = 0. \end{aligned} \tag{5}$$

其中 $\mathbf{V}_{\mathbf{x}} \in \mathbb{R}^{T_0 \times T_0}$ 是对称重要性矩阵 (symmetric importance matrix)，为了符号简单计，本文取其为单位阵； $f(\gamma_i)$ 是对权重的分散惩罚。

合成对照方法

- 该方法相较最初的 SCM 做了两个改变：（1）加入了对权重的惩罚；（2）排除了额外的协变量信息。
- 当预处理结果拟合较差时, [Abadie, Diamond and Hainmueller, 2015] 不建议使用该方法。
- 但在实际中, 预处理结果拟合较差的情况常常发生, 本文通过对 SCM 估计的修正, 将方法的适用范围进一步拓展。

- 1 研究背景
- 2 准备工作
- 3 增广合成对照方法
- 4 校准模拟和实证分析
- 5 讨论

增广合成对照方法

设 \hat{m}_{iT} 是 m_{iT} 的估计量，则定义 $Y_{1T}(0)$ 增广合成对照方法 (ASCM) 估计量为：

$$\begin{aligned}\hat{Y}_{1T}^{\text{aug}}(0) &= \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} Y_{iT} + \left(\hat{m}_{1T} - \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} \hat{m}_{iT} \right) \\ &= \hat{m}_{1T} + \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} (Y_{iT} - \hat{m}_{iT}),\end{aligned}\quad (6)$$

其中， $\hat{\gamma}_i^{\text{scm}}$ 是前文 SCM 方法得到的权重，标准的 SCM 估计是该估计的一个特例，只需令 \hat{m}_{iT} 为常数。本文将主要作为预处理结果函数的估计量，即 $\hat{m}_{iT} \equiv \hat{m}(\mathbf{X}_i)$ 。

增广合成分析方法

- 式 (6) 蕴含了两种提出 ASCM 估计的动机: (a) 一方面, 因为 \hat{m}_{iT} 是 $Y_{iT}(0)$ 的估计, 所以直觉上可将 $\hat{Y}_{1T}^{\text{aug}}(0)$ 视为对 $\hat{Y}_{1T}^{\text{scm}}(0)$ 的一次纠偏 (第一行); (b) 另一方面, 又可类比标准的双稳健估计 [Robins, Rotnitzky and Zhao, 1994], 从结果模型出发, 再重新对残差赋权 (第二行)。
- 双稳健的直观解释: 若 $\sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} Y_{iT} \xrightarrow{P} Y_{1T}(0)$, 则

$$(\hat{m}_{1T} - \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} \hat{m}_{iT}) \xrightarrow{P} 0,$$

反之若 $\hat{m}_{1T} \xrightarrow{P} Y_{1T}(0)$, 则

$$\sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} (Y_{iT} - \hat{m}_{iT}) \xrightarrow{P} 0.$$

即只要有一个模型指定正确, 就有 $\hat{Y}_{1T}^{\text{aug}}(0) \xrightarrow{P} Y_{1T}(0)$ 。

增广合成分析方法

- 基于式 (6), ASCM 的重点将是 \hat{m} 的选择, 不同的选择将得到不同的估计, 本文简要地综述了两种特别的情形:
- (a) \hat{m} 是处理前结果的线性函数:

$$\hat{m}(\mathbf{X}) = \hat{\eta}_0 + \hat{\eta} \cdot \mathbf{X}, \quad (7)$$

- (b) \hat{m} 是供体处理后结果的线性函数:

$$\hat{m}(\mathbf{X}) = \sum_{W_i=0} \hat{\alpha}_i(\mathbf{X}) Y_{iT}. \quad (8)$$

- 本文研究的核心 Ridge ASCM 被证明同时满足这两个情形。

增广合成分析方法

- (a) $\hat{m}(\mathbf{X}) = \hat{\eta}_0 + \hat{\eta} \cdot \mathbf{X}$, 代入式 (6) 得:

$$\begin{aligned} \hat{Y}_{1T}^{\text{aug}}(0) &= \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} Y_{iT} \\ &+ \sum_{t=1}^{T_0} \hat{\eta}_t \left(X_{1t} - \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} X_{it} \right). \end{aligned} \quad (9)$$

- (b) $\hat{m}(\mathbf{X}) = \sum_{W_i=0} \hat{\alpha}_i(\mathbf{X}) Y_{iT}$, 代入式 (6), 得:

$$\begin{aligned} \hat{Y}_{1T}^{\text{aug}}(0) &= \sum_{W_i=0} \left(\hat{\gamma}_i^{\text{scm}} + \hat{\gamma}_i^{\text{adj}} \right) Y_{iT}, \\ \hat{\gamma}_i^{\text{adj}} &\equiv \hat{\alpha}_i(\mathbf{X}_1) - \sum_{W_j=0} \hat{\gamma}_j^{\text{scm}} \hat{\alpha}_i(\mathbf{X}_j). \end{aligned} \quad (10)$$

岭 ASCM

- 接下来，专注于一种特别的情形， $\hat{m}(\mathbf{X}_i)$ 通过岭回归确定，即 $\hat{m}(\mathbf{X}_i) = \hat{\eta}_0^{\text{ridge}} + \mathbf{X}_i' \hat{\eta}^{\text{ridge}}$ ，其中

$$\left\{ \hat{\eta}_0^{\text{ridge}}, \hat{\eta}^{\text{ridge}} \right\} = \arg \min_{\eta_0, \eta} \frac{1}{2} \sum_{W_i=0} (Y_i - (\eta_0 + \mathbf{X}_i' \eta))^2 + \lambda^{\text{ridge}} \|\eta\|_2^2. \quad (11)$$

- 从而得到岭 ASCM 估计：

$$\hat{Y}_{1T}^{\text{aug}}(0) = \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} Y_{iT} + \left(\mathbf{x}_1 - \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} \mathbf{x}_i \right) \cdot \hat{\eta}^{\text{ridge}}. \quad (12)$$

岭 ASCM

本文证明，岭 ASCM 估计可以写为：

$$\hat{Y}_{1T}^{\text{aug}}(0) = \sum_{W_i=0} \hat{\gamma}_i^{\text{aug}} Y_{iT}, \quad (13)$$

其中

$$\hat{\gamma}_i^{\text{aug}} = \hat{\gamma}_i^{\text{scm}} + (\mathbf{X}_1 - \mathbf{X}'_0 \hat{\gamma}^{\text{scm}})' \left(\mathbf{X}'_0 \cdot \mathbf{X}_0 + \lambda^{\text{ridge}} \mathbf{I}_{T_0} \right)^{-1} \mathbf{X}_{i.}.$$

更进一步，可以证明岭 ASCM 的权重 $\hat{\gamma}^{\text{aug}}$ 为如下优化问题的解：

$$\min_{\gamma \text{ s.t. } \sum_i \gamma_i = 1} \frac{1}{2\lambda^{\text{ridge}}} \|\mathbf{X}_1 - \mathbf{X}'_0 \cdot \gamma\|_2^2 + \frac{1}{2} \|\gamma - \hat{\gamma}^{\text{scm}}\|_2^2. \quad (14)$$

这表明岭回归参数 λ^{ridge} 决定了 $\hat{\gamma}_i^{\text{aug}}$ 与 $\hat{\gamma}_i^{\text{scm}}$ 的相似程度，当 λ^{ridge} 趋于无穷时， $\hat{\gamma}_i^{\text{aug}} = \hat{\gamma}_i^{\text{scm}}$ ； λ^{ridge} 越小，外推 (extrapolation) 越严重。

额外协变量信息的利用

- 当存在额外的协变量信息时, 可以通过求解优化问题 (15) 和 (16) 将信息引入估计中。

$$\min_{\gamma \in \Delta^{N_0}} \theta_x \|\mathbf{X}_1 - \mathbf{X}'_{0,\gamma} \gamma\|_2^2 + \theta_z \|\mathbf{Z}_1 - \mathbf{Z}_{0,\gamma}\|_2^2 + \zeta \sum_{W_i=0} f(\gamma_i), \quad (15)$$

$$\min_{\eta_0, \eta_x, \eta_z} \frac{1}{2} \sum_{W_i=0} (Y_i - (\eta_0 + X'_i \eta_x + Z'_i \eta_z))^2 + \lambda_x \|\eta_x\|_2^2 + \lambda_z \|\eta_z\|_2^2. \quad (16)$$

- 其中式 (15) 是直接基于 SCM 的基础上进行调整; 而式 (16) 是选择估计函数 $\hat{m}(\mathbf{X}, \mathbf{Z}) = \hat{\eta}_0 + \mathbf{X}' \hat{\eta}_x + \mathbf{Z}' \hat{\eta}_z$, 在 ASCM 的框架下进行调整。本文证明基于式 (16) 的调整具有更良好的性质。

岭 ASCM 的理论性质

- 本文详尽地叙述了关于岭 ASCM 的理论性质，证明了其对预处理结果的拟合误差严格小于 SCM 的误差，导出了在不同数据生成机制下岭 ASCM 估计的界等。

- ① 研究背景
- ② 准备工作
- ③ 增广合成对照方法
- ④ 校准模拟和实证分析
- ⑤ 讨论

校准模拟

- 校准模拟研究 (Calibrated Simulation Studies) 的思想是：首先假定一种数据生成机制 (模型)，然后用最终感兴趣的实证分析中的数据拟合该模型，得到各参数的估计值 $\hat{\theta}$ ，基于 $\hat{\theta}$ ，生成模拟实验中模型参数的取值：

$$\theta \sim \mathcal{N}(\hat{\theta}, \hat{\sigma}^2(\hat{\theta})). \quad (17)$$

- 直观上，校准模拟中参数的取值相当于对真实问题中参数取值做了一个局部扰动 (在模型指定正确的前提下)，因此数值模拟的结果更能反映实证分析的效果。也可理解为校准模拟是对实证分析做了相应的敏感性分析。

校准模拟

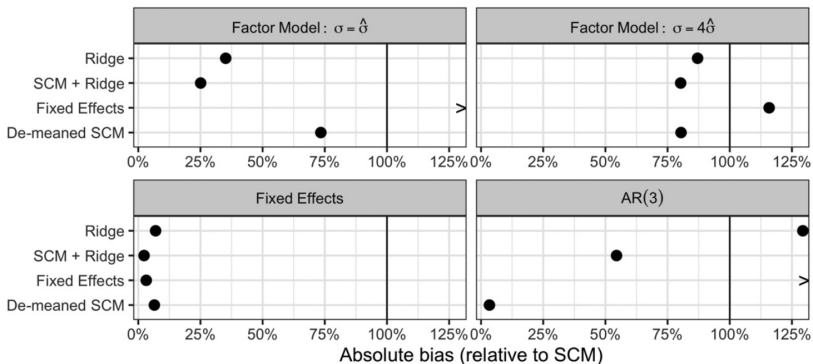


图 1: 在四种数据生成机制下, 四种方法的绝对偏差 $|Y_{1T}(0) - \hat{Y}_{1T}(0)|$ 与 SCM 绝对偏差的比值。

校准模拟

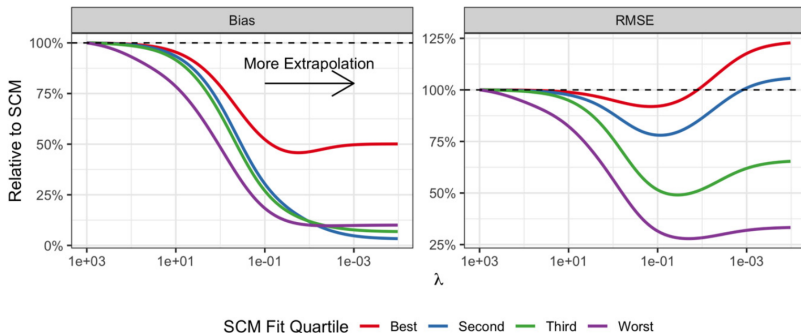


图 2: 岭 ASCM 的 Bias, RMSE 与 SCM 对应值之比随超参数 λ 的变化曲线。按照数值模拟的四分位数展示结果。

校准模拟

表 3: 95% 置信区间的覆盖率，基于 1000 次重复实验。

Method	AR(3)	Factor $\sigma =$ Model: $\hat{\sigma}$	Factor $\sigma =$ Model: $4\hat{\sigma}$	Fixed Effects
SCM	0.934	0.926	0.930	0.889
SCM + Ridge	0.932	0.950	0.936	0.939

实证分析：2012 年堪萨斯州减税政策

- 2012 年，Sam Brownback 上任堪萨斯州州长，实施了大幅的个人所得税削减，既降低了税率，又减少了抵扣扣除。这是对“供给侧”模式的一次有价值的检验，Brownback 认为：减税将增加堪萨斯州的商业活动，促进经济增长并带来额外的税收，以弥补财政收入的损失。
- 堪萨斯州随后的经济表现相对于它的邻居来说并不令人印象深刻；然而，潜在的混杂因素包括干旱和当地重要的航空工业的衰退。因此，为堪萨斯州找到一个可靠的控制是具有挑战性的，而 SCM 提供了一个潜在的解决方案。

实证分析：2012 年堪萨斯州减税政策

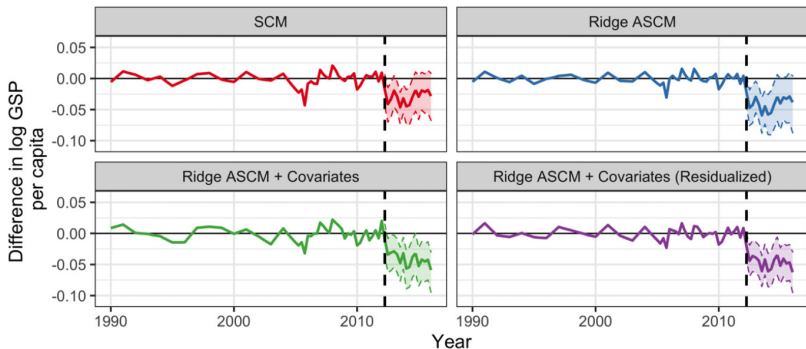
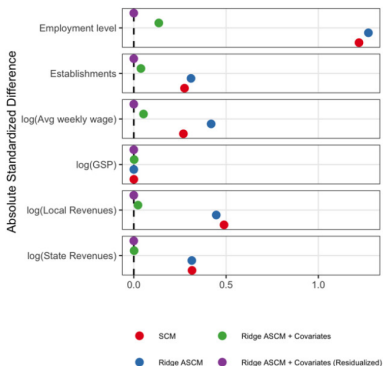
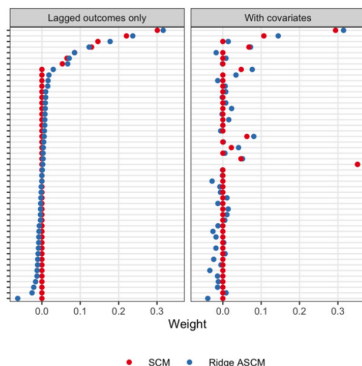


图 3: 四种方法构建的合成对照组相对于观察值之差（对于对数人均国民生产总值，log gross state product per capita）。虚线为 95% 置信区间。

实证分析：2012 年堪萨斯州减税政策



(a)



(b)

图 4: (a) 四种方法对六个协变量的平衡 (即拟合程度 $|Z_{1k} - \sum_{w_i=0} \hat{\gamma} Z_{ik}|$); (b) 单独的 SCM 和带有协变量的岭 ASCM 中供体单元的权重分布。

实证分析：2012 年堪萨斯州减税政策

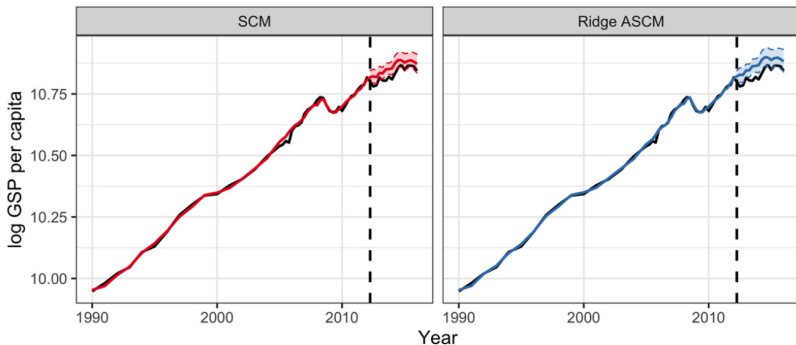


图 5: SCM 与岭 ASCM 构建的合成对照组（针对无减税的反事实曲线，红蓝色）与观察值（黑色）的对比曲线。虚线为 95% 置信区间。

实证分析：2012 年堪萨斯州减税政策

- 该政策的实施最终以失败告终，在 Wikipedia 上搜索词条“Kansas experiment”，即可得到该政策干预的后续追踪研究。
- 到 2017 年，堪萨斯州收入下降了数亿美元，导致道路、桥梁和教育方面的支出大幅削减。由于经济增长一直低于平均水平，堪萨斯州共和党立法机构投票决定废除该法案；尽管 Brownback 否决了这次判决，但立法机关成功地否决了他的否决权。

- ① 研究背景
- ② 准备工作
- ③ 增广合成对照方法
- ④ 校准模拟和实证分析
- ⑤ 讨论

讨论

- SCM 是一种在城市或州等管辖水平上评估政策影响的流行方法，但其要求预处理结果的拟合要足够好，这在实际中往往很难做到。本文推广的 ASCM 得以在预处理结果拟合较差时使用。
- 未来有以下几个方向的工作：(1) 对模型假设和数据生成机制做敏感性分析 [Soriano et al., 2020]; (2) 将 ASCM 算法推广到多个处理单元的情形 [Abadie and L'Hour, 2018, Robbins, Saunders and Kilmer, 2017], 不同时间接受处理的情形 [Ben-Michael, Feller and Rothstein, 2019]; (3) 更复杂的数据结构，如：同一单元包含多个结果序列；具有个体和总体层面结果信息的分层数据结构；离散或计数的结果等。

Thanks!