

## 第 18 章 合成控制法

第 18 章的合成控制法与第 19 章的回归控制法均主要适用于只有一个或几个处理组个体的情形。这些个体通常为地区，故也称为区域政策分析(regional policy analysis)。

### 18.1 比较案例分析

经济学家常要评估某政策或事件对于某地区的因果效应。

但有时该政策仅实施于一个地区，导致处理组只有一个“处理地区”(treated region)，故不适用双重差分法。

受到处理的地区可以是国家、省、州或城市。

使用反事实框架(counterfactual framework)，假想该地区若未受政策干预将会怎样(即反事实结果)，并与事实上受到干预的实际结果进行对比，而二者之差即为“处理效应”(treatment effect)。

但反事实结果无法观测。

常见解决方法为，寻找控制组，即在各方面与受处理地区相似却未受政策干预的其他地区，作为处理地区的反事实替身。

但通不易找到理想的控制地区，在各方面都接近于处理地区。

比如，假设某政策仅实施于北京，欲考察该政策的因果效应。

若以上海作为控制地区，但上海毕竟与北京不尽相同。

用其他一线城市(上海、广州、深圳)构成北京的控制组，比较上海、广州、深圳与北京在政策实施前后的差别。

此方法称为**比较案例分析**(comparative case study)，广泛应用于社会科学领域。

在评估政策效应时，比较案例分析详尽地研究几个未受政策冲击，但极具可比性的控制地区。

在经济学中，比较案例分析的一个经典例子为 Card (1990)。

该文研究移民对于当地劳动力市场的影响，并以 1980 年古巴移民潮对迈阿密(Miami)劳动力市场的冲击作为自然实验。

在 1980 年 5-9 月，由于古巴政经危机，约 12.5 万古巴人从马里埃尔港乘船涌入美国，即所谓“马里埃尔船运”(Mariel boatlift)；其中半数定居于迈阿密，使得迈阿密的劳动力增加 7%。

Card (1990)选择美国四个与迈阿密相似的城市构成控制组，即亚特兰大、休斯顿、洛杉矶与坦帕圣彼得堡(Tampa-St. Petersburg)。

结果发现，与未受移民潮冲击的四个控制城市相比，马里埃尔船运对于迈阿密劳动力市场的工资与失业率几乎没有影响。

与迈阿密相似的美国城市不在少数，Card (1990)并未说明为何仅选择那四个城市。

传统的比较案例分析在选择控制组时可能存在一定的主观随意性，或影响结论的客观性。

Card (1990)隐含地将这四个控制城市的结果进行了简单算术平均，即默认使用相同的权重。

若将更大权重赋予与迈阿密更相似的控制城市，则可能提高估计效率。

## **18.2 合成控制法的思想**

基于比较案例分析的局限性，Abadie and Gardeazabal (2003)提出合成控制法(Synthetic Control Method, 简记 SCM)。

回到政策实施于北京的例子。

虽然无法找到北京的理想控制地区，但可对中国的若干大城市进行适当的线性组合，以构造一个更为优秀的“合成控制地区”(synthetic control region)，并将“真实北京”与“合成北京”进行对比，故名“合成控制法”。

合成控制法的优点是，可根据数据选择线性组合的最优权重，即所谓“数据驱动”(data-driven)，从而避免研究者主观选择控制组的随意性。

Abadie and Gardeazabal (2003)的初衷是研究西班牙巴斯克地区(Basque country)恐怖活动的经济成本。

巴斯克人长期居住于巴斯克地区，拥有独特的语言与文化，历史上多次成功抵抗强敌入侵。

20 世纪 70 年代初，巴斯克地区的人均 GDP 在西班牙 17 个地区中排名第三。

但从 1974 年起，因民族独立诉求未获满足，巴斯克地区兴起大规模的恐怖活动。

图 18.1 展示了在 1968-2000 年间，巴斯克地区恐怖活动的死亡人数。

从图 18.1 可见，恐怖活动始于 1968 年，但最初规模较小，直至 1974 年才大规模爆发。

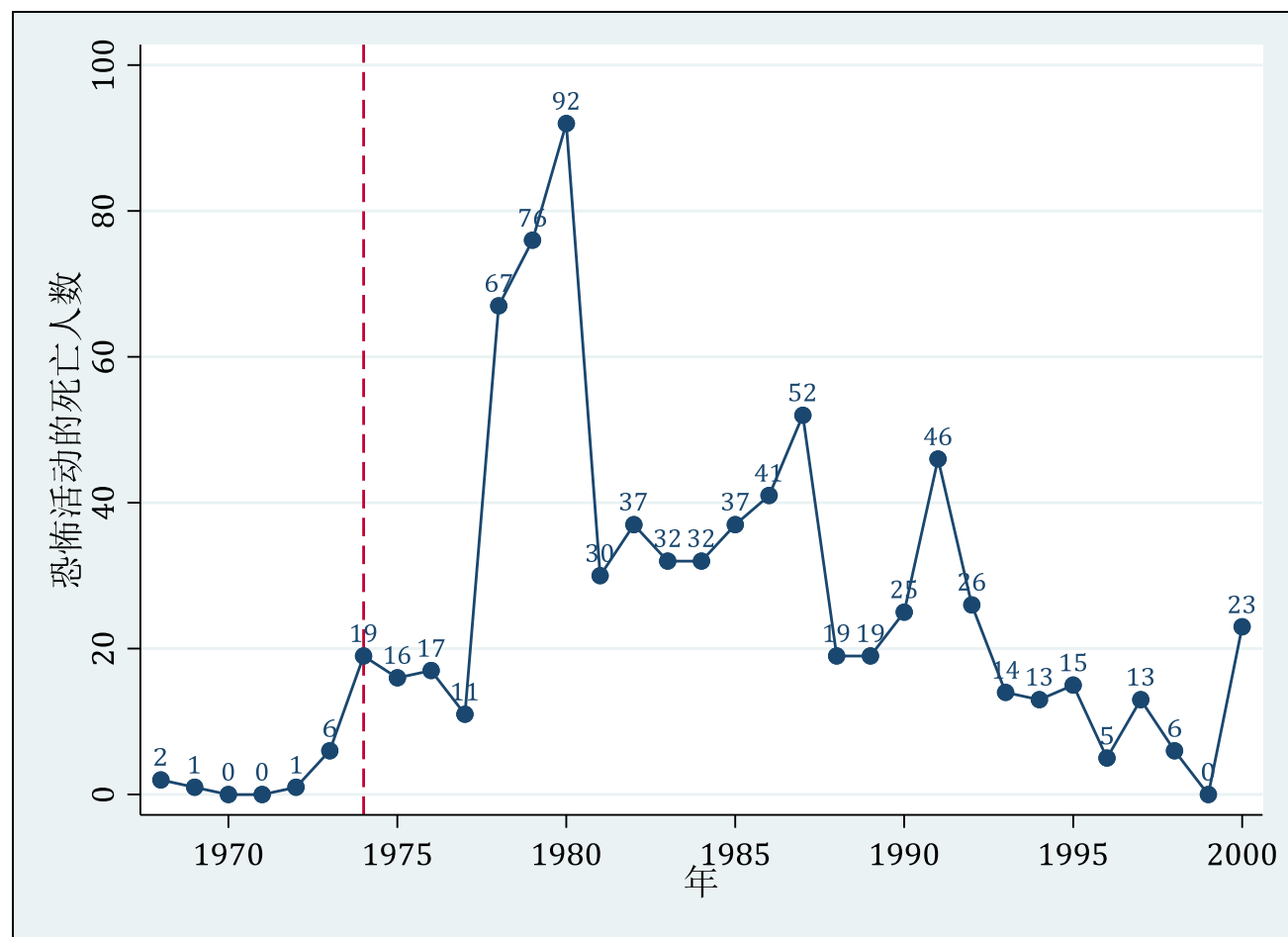


图 18.1 巴斯克地区恐怖活动的死亡人数  
数据来源：Abadie and Gardeazabal (2003) Table 1。



巴斯克地区的恐怖活动重创了巴斯克经济，至 1990 年代末，其人均 GDP 在西班牙排名已降至第六。

但 70 年代末至 80 年代初，西班牙整体经济也下行，故不易区分恐怖活动的单独效应。

在恐怖活动之前，巴斯克地区的经济增长潜力也与西班牙其他地区不尽相同。

Abadie and Gardeazabal (2003)使用西班牙其他地区的线性组合来构造合成的控制地区，并使得合成控制地区的经济特征与 60 年代末恐怖活动爆发前的巴斯克地区尽可能相似，然后把“合成巴斯克地区”(synthetic Basques country)的人均 GDP 与“真实巴斯克地区”(actual Basque country)进行对比。

## 18.3 合成控制法的算法

假设共有  $N$  个地区，其中第 1 个地区为受到恐怖活动冲击的巴斯克地区，而其余  $(N-1)$  个西班牙地区未受冲击(在此例中  $N=17$ )，构成潜在的控制组，称为捐献池(donor pool)。

一个潜在假定是，恐怖活动仅影响巴斯克地区，而未波及西班牙的其他地区(事实上恐怖活动也主要集中于巴斯克地区)。

这正是第 15 章引入的个体处理值稳定假定(SUTVA)。

将合成控制地区的权重记为  $(N-1)$  维列向量：

$$\mathbf{w} \equiv \begin{pmatrix} w_2 \\ \vdots \\ w_N \end{pmatrix} \quad (18.1)$$

其中， $w_i$ 表示第*i*个地区的权重( $i = 2, \dots, N$ )，并要求所有权重皆非负，且权重之和为 1。

权重向量 $\mathbf{w}$ 的不同取值即构成不同的合成控制地区，简称**合成控制**(synthetic control)。

结果变量为人均实际 GDP(real per capita GDP)，记为  $y$ 。影响  $y$  的解释变量或预测变量(predictors)包括投资率、人口密度、产业结构、人力资本等，共 12 个变量，均为恐怖活动之前 1960 年代某段时期的平均值，参见表 18.1。

表 18.1 恐怕活动爆发前的地区特征(20 世纪 60 年代)

变量	巴斯克地区	西班牙	“合成”巴斯克地区
人均实际 GDP (1986 USD)	5,285.46	3,633.25	5,270.80
总投资占 GDP 百分比	24.65	21.79	21.58
人口密度	246.89	66.34	196.28
产业结构(占总产出百分比)			
农林渔业	6.84	16.34	6.18
能源与供水	4.11	4.32	2.76
工业	45.08	26.60	37.64
建筑工程	6.15	7.25	6.96
市场化服务	33.75	38.53	41.10
非市场化服务	4.07	6.97	5.37

人力资本(占劳动年龄人口百分比)			
文盲	3.32	11.66	7.65
小学或无学历	85.97	80.15	82.33
高中学历	7.46	5.49	6.92
高中以上学历	3.26	2.70	3.10

注：数据来自 Abadie and Gardeazabal (2003) Table 3。

在恐怕活动爆发之前，巴斯克地区与西班牙整体的经济特征有很大不同。与西班牙全国相比，巴斯克地区的人均收入更高，投资率更高，人口密度更大，工业占总产出比重更高，且劳动力的受教育程度更高。

巴斯克地区与“合成”巴斯克地区在经济特征的各方面已较为接近。

在恐怖活动爆发之前(20 世纪 60 年代), 记巴斯克地区各预测变量的平均值为向量  $\mathbf{x}_1$  ( $K$  维列向量, 此例  $K=13$ , 下标 1 表示处理地区)。

将西班牙其他地区相应预测变量的平均值记为矩阵  $\mathbf{X}_0 = (\mathbf{x}_2 \cdots \mathbf{x}_N)$  ( $K \times (N-1)$  维矩阵, 下标 0 表示控制地区), 其中  $\mathbf{x}_i$  为地区  $i$  预测变量在 60 年代的平均值。

希望选择权重  $\mathbf{w}$ , 使得  $\mathbf{X}_0 \mathbf{w}$  尽可能地接近于  $\mathbf{x}_1$ , 即经过加权之后, 合成控制地区的经济特征应尽量接近处理地区:

$$\mathbf{X}_0 \mathbf{w} = (\mathbf{x}_2 \cdots \mathbf{x}_N) \begin{pmatrix} w_2 \\ \vdots \\ w_N \end{pmatrix} = w_2 \mathbf{x}_2 + \cdots + w_N \mathbf{x}_N \approx \mathbf{x}_1 \quad (18.2)$$

希望  $\mathbf{X}_0 \mathbf{w} - \mathbf{x}_1 \approx \mathbf{0}$ 。

考虑使用欧氏距离  $(\mathbf{x}_1 - \mathbf{X}_0 \mathbf{w})'(\mathbf{x}_1 - \mathbf{X}_0 \mathbf{w})$  (即以向量内积计算的平方和)，度量“合成地区”经济特征  $\mathbf{X}_0 \mathbf{w}$  与“处理地区”经济特征  $\mathbf{x}_1$  的接近程度。

但每个预测变量对于  $y$  的预测能力通常有大小之别，故在距离函数中应享有不同的权重。分别记  $K$  个预测变量的重要性为  $v_1, \dots, v_K$ ，并以  $K \times K$  维对角矩阵  $\mathbf{V}$  表示：

$$\mathbf{V} = \begin{pmatrix} v_1 & & \\ & \ddots & \\ & & v_K \end{pmatrix} \quad (18.3)$$

对角线元素均为非负，反映相应的预测变量对于解释  $y$  的重要性。

考虑有约束的“加权平方和”最小化问题：

$$\begin{aligned} \min_{\mathbf{w}} (\mathbf{x}_1 - \mathbf{X}_0 \mathbf{w})' \mathbf{V} (\mathbf{x}_1 - \mathbf{X}_0 \mathbf{w}) \\ s.t. \quad w_i \geq 0 \quad (i = 2, \dots, N); \quad \sum_{i=1}^N w_i = 1 \end{aligned} \quad (18.4)$$

此最小化问题的目标函数是二次函数，但由于存在非线性的非负约束，故一般需进行数值求解。

$\mathbf{w}$  的最优解依赖于对角矩阵  $\mathbf{V}$ 。记此最小化问题的最优解为  $\mathbf{w}^*(\mathbf{V})$ 。

如何选择最优的  $\mathbf{V}$ ？



选择 $\mathbf{V}$ 使得在处理前，合成巴斯克地区的人均 GDP 与真实巴斯克地区尽量接近。

记巴斯克地区在 1960-1969 年间的人均实际 GDP 为 $\mathbf{z}_1$ (10 维列向量)。

记其他控制地区在 1960-1969 年间的人均实际 GDP 为 $\mathbf{Z}_0 = (\mathbf{z}_2 \cdots \mathbf{z}_N)$ ( $10 \times (N-1)$  维矩阵), 其中 $\mathbf{z}_i$  ( $i = 2, \dots, N$ ) 为控制地区 $i$ 在 1960-1969 年间的人均实际 GDP。

希望 $\mathbf{Z}_0 \mathbf{w}$ 尽可能接近 $\mathbf{z}_1$ ，即合成控制地区的处理前结果变量尽量接近于处理地区：

$$\mathbf{Z}_0 \mathbf{w} = (\mathbf{z}_2 \cdots \mathbf{z}_N) \begin{pmatrix} w_2 \\ \vdots \\ w_N \end{pmatrix} = w_2 \mathbf{z}_2 + \cdots + w_N \mathbf{z}_N \approx \mathbf{z}_1 \quad (18.5)$$

以  $\mathbf{Z}_0 \mathbf{w}^*(\mathbf{V})$  预测  $\mathbf{z}_1$ ，然后选择  $\mathbf{V}$ ，使得均方预测误差(Mean Squared Prediction Error，简记 MSPE)最小，即最小化“预测误差平方和的平均”：

$$\min_{\mathbf{V}} \frac{1}{T_0} (\mathbf{z}_1 - \mathbf{Z}_0 \mathbf{w}^*(\mathbf{V}))' (\mathbf{z}_1 - \mathbf{Z}_0 \mathbf{w}^*(\mathbf{V})) \quad (18.6)$$

其中， $T_0$  为  $\mathbf{z}_1$  的维度，在此例  $T_0 = 10$  (1960-1969 年共 10 年)。

求解此最小化问题，即可得最优的  $\mathbf{V}^*$ ，以及构成合成巴斯克地区的最优权重  $\mathbf{w}^* = \mathbf{w}^*(\mathbf{V}^*)$ 。

Abadie and Gardeazabal (2003)发现，只有两个控制地区的权重为正，即加泰罗尼亚(Catalonia，权重 0.851)与马德里(Madrid，权重 0.149)，而其他地区的权重均为 0。

$$\text{合成巴斯克地区} = 0.851 \times \text{加泰罗尼亚} + 0.149 \times \text{马德里}$$

在得到合成巴斯克地区的最优权重之后，即可计算其人均实际 GDP 在样本期间的演化过程。

记巴斯克地区在样本期间(假设为 $T$ 期)的人均实际 GDP 为向量  $\mathbf{y}_1(T \text{ 维列向量})$ 。

记其他地区在样本期间的人均实际 GDP 为矩阵  $\mathbf{Y}_0 = (\mathbf{y}_2 \cdots \mathbf{y}_N)$  ( $T \times (N-1)$  维矩阵), 其中  $\mathbf{y}_i$  为控制地区  $i$  的人均实际 GDP。

由此可得合成巴斯克地区的人均实际 GDP 序列  $\hat{\mathbf{y}}_1$ :

$$\hat{\mathbf{y}}_1 = \mathbf{Y}_0 \mathbf{w}^* = (\mathbf{y}_2 \cdots \mathbf{y}_N) \begin{pmatrix} w_2^* \\ \vdots \\ w_N^* \end{pmatrix} = w_2^* \mathbf{y}_2 + \cdots + w_N^* \mathbf{y}_N \quad (18.7)$$

可将实际的  $\mathbf{y}_1$  与合成的  $\hat{\mathbf{y}}_1$  画时间趋势图, 即所谓缺口图(gap graph), 结果参见图 18.2。

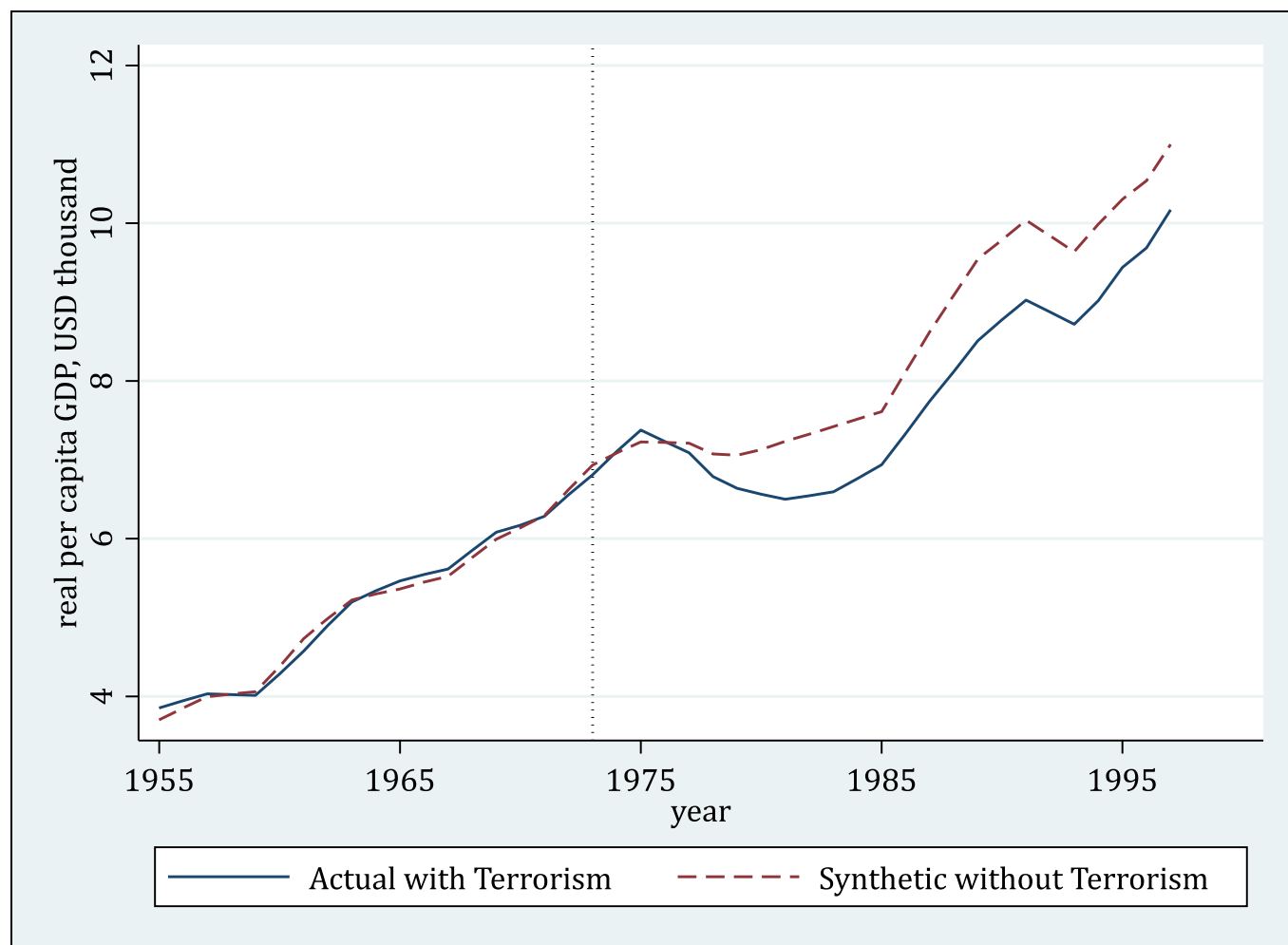


图 18.2 巴斯克与合成巴斯克地区的缺口图

在处理前(1973 年及之前), 巴斯克(实线)与合成巴斯克(虚线)的人均实际 GDP 十分接近。

即使在 1974-1976 年, 虽然大规模恐怖活动已经爆发, 二者也依然较接近。这说明恐怕活动对经济增长的影响存在滞后。

从 1977 年开始, 二者开始明显分岔; 到了 80 年代与 90 年代, 巴斯克的人均 GDP 比合成巴斯克低约 10%。

巴斯克地区恐怖活动的经济成本是大约损失 10% 的人均 GDP。

## **18.4 合成控制法的理论基础**

合成控制法主要适用于只有一个处理地区的情形。

若有多个处理地区，通常可分别对每个处理地区进行 SCM 估计。

Abadie, Diamond and Hainmueller (2010)首次证明了合成控制法的基本性质。

假设共有  $N$  个地区，其中第 1 个地区受到政策干预，而其余  $(N-1)$  个地区未受冲击，构成捐赠池(donor pool)。

记  $y_{it}$  为地区  $i$  在第  $t$  期实际观测到的结果变量，其中  $i = 1, \dots, N$ ，而  $t = 1, \dots, T$ 。

记  $y_{it}^0$  为地区  $i$  在第  $t$  期若未受政策干预的潜在结果

$y_{it}^1$  为地区  $i$  在第  $t$  期受政策干预的潜在结果。

记  $T_0$  为处理前的时期数，且  $1 \leq T_0 < T$ 。

假设政策在前  $T_0$  期对于结果变量没有影响，即对于所有  $i$  与  $t \leq T_0$ ，都有  $y_{it} = y_{it}^0$ 。

如果政策在实施之前即产生影响(比如，通过预期效应)，则可重新定义  $T_0$  为政策实际产生影响之前的那个时期。

一个潜在假定是各地区之间不会互相影响(no interference between units)，即 SUVTA 假定；特别地，控制地区的结果变量不受处理地区政策冲击的影响。



我们关心当 $i = 1$ 而 $t > T_0$ 时的处理效应:

$$\tau_{1t} = y_{1t} - y_{1t}^0, \quad t = T_0 + 1, \dots, T \quad (18.8)$$

由于 $y_{1t}$ 可观测, 故只要估计反事实结果 $y_{1t}^0$ 即可。

假设无政策干预的潜在结果 $y_{it}^0$ 由以下因子模型(factor model)所决定:

$$y_{it}^0 = \alpha_t + \mathbf{x}_i' \boldsymbol{\beta}_t + \mathbf{b}_i' \mathbf{f}_t + \varepsilon_{it} \quad (18.9)$$

其中,  $\alpha_t$ 为时间固定效应(time fixed effects)。 $\mathbf{x}_i$ 为可观测的向量(不受政策干预影响, 也不随时间而变; 比如, 干预前预测变量的平均值)。

由于  $\mathbf{x}_i$  对于  $y_{it}^0$  的作用可随时间而变，故  $\mathbf{x}_i$  的系数  $\boldsymbol{\beta}_t$  带时间下标  $t$ 。 $\boldsymbol{\beta}_t \equiv \boldsymbol{\beta}$  为其特例。

$\mathbf{b}_i' \mathbf{f}_t$  为不可观测的交互固定效应(interactive fixed effects)，即多维个体固定效应  $\mathbf{b}_i$  与多维时间固定效应  $\mathbf{f}_t$  的乘积。 $\varepsilon_{it}$  为随机扰动项。

根据因子分析(factor analysis)的术语， $\mathbf{f}_t$  为共同因子(common factors)，即不同地区所面临的共同冲击(common shocks)。

例如， $\mathbf{f}_t$  包含三个分量，即  $\mathbf{f}_t = (f_{1t} \ f_{2t} \ f_{3t})'$ ，分别表示贸易冲击(trade shock)，金融冲击(financial shock)与技术冲击(technology shock)。

$\mathbf{b}_i$ 称为因子载荷(factor loading), 表示各地区对于共同冲击 $\mathbf{f}_t$ 的反应力度可以不同, 例如 $\mathbf{b}_i=(b_{i1} \ b_{i2} \ b_{i3})'$ 。

交互固定效应可写为

$$\mathbf{b}_i' \mathbf{f}_t = (b_{i1} \ b_{i2} \ b_{i3}) \begin{pmatrix} f_{1t} \\ f_{2t} \\ f_{3t} \end{pmatrix} = b_{i1} f_{1t} + b_{i2} f_{2t} + b_{i3} f_{3t} \quad (18.10)$$

$f_{1t}$ 对于个体 $i$ 的作用系数为 $b_{i1}$ ,  $f_{2t}$ 对于个体 $i$ 的作用系数为 $b_{i2}$ , 而 $f_{3t}$ 对于个体 $i$ 的作用系数为 $b_{i3}$ 。

交互效应模型不仅允许多维的时间效应, 而且不同个体对于共同冲击可以有异质性反应(heterogeneous impacts)。

交互效应模型是双向固定效应模型(two-way fixed effects model)的推广。

例如，若 $\mathbf{b}_i = (1 \ u_i)'$ ，而 $\mathbf{f}_t = (\lambda_t \ 1)'$ ，则 $\mathbf{b}_i' \mathbf{f}_t = u_i + \lambda_t$ ，这正是双向固定效应模型。

基于因子模型(18.9)，Abadie et al. (2010)证明：

在一定的正则条件下，如果最优合成控制 $\mathbf{w}^*$ 能完美地再现处理地区的经济特征与干预前结果变量，则当处理前时期数 $T_0$ 趋向无穷大时，SCM估计量“渐近无偏”(asymptotically unbiased)，即在大样本下SCM估计量无系统性偏差。

渐近无偏的结论比“一致性”(consistency)更弱；即后者意味着前者，而反之则不然。

如果合成控制 $\mathbf{w}^*$ 只是不完美地再现处理地区的经济特征与干预前结果变量，则SCM估计量是渐近有偏的。

此偏差也称为**内插偏差**(interpolation bias)，因为此时使用内插法构造的合成控制并不是处理地区的完美的反事实替身。

在使用SCM时，一个重要的模型设定检验就是，考察合成控制的经济特征与处理前的结果变量是否与处理地区足够接近。

如果无法找到 $\mathbf{w}$ ，使得合成控制能很好地再现处理地区的经济特征以及处理前的结果变量，则不建议使用合成控制法。

由于SCM要求权重非负且权重之和为1，故不允许“外推”(extrapolation)，也就避免了外推偏差(extrapolation bias)，这是SCM的优势之一。

## 18.5 合成控制法的统计推断

如何判断 SCM 的点估计是否在统计上显著 (statistically significant)?

Abadie et al. (2010)认为，在应用 SCM 时，由于潜在的控制地区数目通常不多，故不适合使用大样本理论进行统计推断。

Abadie et al. (2010)提出使用**安慰剂检验**(placebo test)进行统计推断。此法类似于统计学的“置换检验”(permutation test)，适用于任何样本容量。

“安慰剂”(placebo)一词来自医学上的随机实验，比如要检验某新药的疗效。

可将参加实验的人群随机分为两组，其中一组为实验组，服用真药；而另一组为控制组，服用安慰剂(比如，无用的糖丸)。

一般不让参与者知道自己服用的究竟是真药还是安慰剂，以避免由于主观心理作用对实验结果产生影响，即所谓安慰剂效应 (placebo effect)。

安慰剂检验借用了安慰剂的思想。原假设是处理地区的各期处理效应为0，即

$$H_0 : \tau_{1t} = 0 \quad (t = T_0 + 1, \dots, T) \quad (18.11)$$

回到西班牙恐怖袭击的例子。

我们想知道，使用SCM所估计巴斯克地区的处理效应 $\hat{\tau}_{1t}$ ，是否完全由偶然因素所驱动(即 $\tau_{1t} = 0$ )?



若随机抽取一个控制地区进行SCM估计，能否得到类似的效应？

如果从所有地区(含处理地区与控制地区)随机抽取一个地区，则得到与巴斯克地区处理效应一样大的概率( $p$ 值)是多少？

**空间安慰剂检验(in-space placebo test)**依次将每个控制地区作为**伪处理单位(fake treatment unit)**，假设也在1974年后受恐怖活动影响，而将巴斯克地区作为控制地区对待，然后使用SCM估计每个控制地区的“安慰剂效应”(因实际上未受恐怖袭击，故不是处理效应)。

由此可得到安慰剂效应的分布，并将巴斯克地区的处理效应与之对比。

如果巴斯克地区的处理效应，比所有控制地区的安慰剂相应都更“极端”(extreme)，则认为此处理效应显著，并非完全由偶然因素所驱动。

在对某控制地区使用SCM估计安慰剂效应时，如果在处理前其合成控制的拟合效果很差(即MSPE很大)，则可能使得处理后的安慰剂效应波动也很大，故结果不可信。

可限制控制地区处理前的MSPE不超过处理地区MSPE的若干倍数，比如两倍；而仅使用符合此条件的控制地区进行安慰剂检验。

Abadie et al. (2015)进一步提出了基于时间的安慰剂检验，即所谓时间安慰剂检验(in-time placebo test)。

时间安慰剂检验假设一个伪处理时间(fake treatment time)；比如，假定政策冲击发生在第 $\tilde{T}_0$ 期，早于冲击实际发生的时期 $T_0 + 1$ ，即 $\tilde{T}_0 < T_0 + 1$ ，然后进行SCM估计。

如果所估安慰剂效应在区间 $[\tilde{T}_0, T_0]$ 很大，则处理效应的显著性就是可疑的；因为处理地区在 $[\tilde{T}_0, T_0]$ 期间实际上并未受到政策冲击。

Abadie, Diamond and Hainmueller (2015)所提出的时间安慰剂检验并无 $p$ 值，无法进行严格的统计检验。

Chen and Yan (2023)提出混合安慰剂检验(mixed placebo test)，同时使用伪处理时间与伪处理地区。

在进行混合安慰剂检验时，首先给定一个伪处理时间，并将其视为真实的处理时间，然后进行空间安慰剂检验，即可获得时间安慰剂检验的 $p$ 值。

## 18.6 合成控制法的稳健性检验

由于 SCM 要求权重非负且权重之和为 1，故 SCM 的最优权重通常是稀疏的(sparse)。

可能只有少数几个控制地区的权重为正，而其余控制地区的权重均为 0。

在西班牙恐怖活动的案例中，共有 16 个控制地区，但仅有两个地区的权重为正。

SCM 所估计的处理效应是否主要由某个权重为正的 control 地区所驱动？

Abadie et al. (2015) 提出 留一稳健性检验 (leave-one-out robustness test), 简记 LOO。

此检验每次去掉一个权重为正的 control 地区(即将该地区留在捐赠池外), 重新进行 SCM 估计, 并考察这些“留一估计”(leave-one-out estimates)是否接近于使用全部 control 地区的 SCM 估计。

虽然留一估计会损失效率, 但可考察是否 SCM 估计结果主要为单一 control 地区所驱动。

如果留一 SCM 估计均与全样本 SCM 估计类似, 则认为 SCM 估计是稳健的。

SCM 的稳健性检验还包括使用不同的协变量进行 SCM 估计，以考察 SCM 估计结果对于协变量的敏感性(sensitivity)。

## 18.7 合成控制法的 Stata 案例

Abadie et al.(2011)提供了合成控制法的Stata程序synth, 下载安装方法为

```
. ssc install synth, replace
```

其中，选择项“replace”表示可覆盖已安装的旧命令。

命令synth的基本句型为

```
. synth y x1 x2 x3 , trunit(#) trperiod(#)
cunit(numlist) xperiod(numlist) mspeperiod()
figure nested allopt sigf(#)
```

“y”为结果变量，“x1 x2 x3”为预测变量。必选项  
“trunit( # )”指定处理地区(trunit表示treated unit)。

必选项“trperiod( # )”指定政策干预开始的时期(trperiod表示treated period)。选择项“cunit(numlist)”指定潜在的控制地区(cunit表示control units), 默认为处理地区以外的所有地区。



选择项“`xperiod(numlist)`”指定将预测变量进行平均的期间，默认为政策干预前的所有时期(the entire pre-intervention period)。选择项“`mspeperiod()`”指定最小化MSPE的时期，默认为政策干预前的所有时期。选择项“`figure`”表示画结果变量的缺口图。

选择项“`nested`”表示使用嵌套的数值方法寻找最优的合成控制，比默认方法更费时，但更精确。在使用选择项“`nested`”时，若再加上选择项“`allopt`”(即“`nested allopt`”), 则比单独使用“`nested`”还更费时，但精确度进一步提高。

选择项“`sigf(#)`”指定判定算法是否收敛所用的有效数字位数，默认为“`sigf(7)`”，即精确到7位有效数字。

Abadie et al. (2010)将SCM用于研究美国加州1988年第99号控烟法(Proposition 99)的政策效应。

1988年11月加州通过了当代美国最大规模的控烟法(anti-tobacco legislation)，并于1989年1月开始生效。

该法将加州的香烟消费税(cigarette excise tax)提高了每包25美分，将所得收入专项用于控烟的教育与媒体宣传，并引发了一系列关于室内清洁空气的地方立法，比如在餐馆、封闭工作场所等禁烟。

Abadie et al. (2010)的数据为美国1970-2000年的州际面板数据。

由于Massachusetts, Arizona, Oregon与Florida的州政府在1989-2000年间也引入了正式的控烟立法，故将此四州从控制组中去掉。

还去掉了在1989-2000年间将香烟消费税提高了每包50美分及以上的州(Alaska, Hawaii, Maryland, Michigan, New Jersey, New York, Washington)，以及首都所在地华盛顿特区。

最后剩下38个州作为潜在的控制地区(即使加上去掉的州，结果也依然稳健)。

命令synth自带了加州控烟法的数据集synth\_smoking.dta。

面板变量为`state`(州)，而时间变量为`year`(年)。结果变量为`cigsale`(人均香烟消费量，包/年)

预测变量：`lnincome`(人均收入对数)、`age15to24`(15-24岁人口所占总人口比重)、`retprice`(平均香烟零售价格)、`beer`(人均啤酒消费量)。这些预测变量均为1980-1988年的州平均值。

还使用了结果变量的滞后，即1975、1980与1988年的人均香烟消费量，作为三个额外的预测变量。

使用命令`sysuse`载入数据集`synth_smoking.dta`，设为面板数据，并进行SCM估计。

```
. sysuse synth_smoking, clear
```

```
. xtset state year  
. synth cigsale lnincome age15to24 retprice beer  
cigsale(1988) cigsale(1980) cigsale(1975),  
trunit(3) trperiod(1989) xperiod(1980(1)1988)  
figure nested allopt sigf(6)
```

其中, “cigsale(1975) cigsale(1980) cigsale(1988)” 分别表示人均香烟消费在1975、1980与1988年的取值。

必选项 “trunit(3)” 表示第3个州(即加州)为处理地区; 必选项 “trperiod(1989)” 表示控烟法在1989年开始实施。

选择项 “xperiod(1980(1)1988)” 表示将预测变量在1980-1988年间进行平均, 其中 “1980(1)1988” 表示始于1980年, 以1年为间隔, 而止于1988年。

在算法方面，仅使用“`nested`”选项，并未使用最费时的“`allopt`” (以避免后续进行安慰剂检验时过于费时)。

为保证安慰剂检验的估计结果收敛，通过选择项“`sigf(6)`”限制收敛标准为六位有效数字。

若使用选择项“`nested allopt`”与默认的“`sigf(7)`”选项，所得结果类似。

此命令的输出结果包括三部分，即(1)数据设定；(2)最优化过程；(3)估计结果。

Synthetic Control Method for Comparative Case Studies	
First Step: Data Setup	
control units: for 38 of out 38 units missing obs for predictor beer in period 1980 -ignored for averaging control units: for 38 of out 38 units missing obs for predictor beer in period 1981 -ignored for averaging control units: for 38 of out 38 units missing obs for predictor beer in period 1982 -ignored for averaging control units: for 38 of out 38 units missing obs for predictor beer in period 1983 -ignored for averaging treated unit: for 1 of out 1 units missing obs for predictor beer in period 1980 -ignored for averaging treated unit: for 1 of out 1 units missing obs for predictor beer in period 1981 -ignored for averaging treated unit: for 1 of out 1 units missing obs for predictor beer in period 1982 -ignored for averaging treated unit: for 1 of out 1 units missing obs for predictor beer in period 1983 -ignored for averaging	
Data Setup successful	
Treated Unit: California Control Units: Alabama, Arkansas, Colorado, Connecticut, Delaware, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, West Virginia, Wisconsin, Wyoming	
Dependent Variable: cigsale MSPE minimized for periods: 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 Results obtained for periods: 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000	
Predictors: lnincome age15to24 retprice beer cigsale(1988) cigsale(1980) cigsale(1975)	
Unless period is specified predictors are averaged over: 1980 1981 1982 1983 1984 1985 1986 1987 1988	
Second Step: Run Optimization	
Nested optimization requested Starting nested optimization module Optimization done	
Optimization done	47

第一步的数据设定(Data Setup)过程显示，在计算处理前预测变量的均值时，有些缺失值被忽略了(ignored for averaging)；处理单位(treated unit)为加州，并罗列了全部控制单位(control units)；被解释变量为*cigsale*；最小化MSPE的区间为1970-1988年，而计算结果的区间为1970-2000年。

第二步的最优化过程则显示，使用“嵌套最优化”(nested optimization)的算法已完成。

Loss: Root Mean Squared Prediction Error	
RMSPE	1.779546



Unit Weights:	
Co_No	Unit_Weight
Alabama	0
Arkansas	0
Colorado	.148
Connecticut	.059
Delaware	0
Georgia	0
Idaho	0
Illinois	0
Indiana	0
Iowa	0
Kansas	0
Kentucky	0
Louisiana	0
Maine	0
Minnesota	0
Mississippi	0
Missouri	0
Montana	.207
Nebraska	0
Nevada	.241
New Hampshire	0
New Mexico	0
North Carolina	0
North Dakota	0
Ohio	0
Oklahoma	0
Pennsylvania	0
Rhode Island	0
South Carolina	0
South Dakota	0
Tennessee	0
Texas	0
Utah	.345
Vermont	0
Virginia	0
West Virginia	0
Wisconsin	0
Wyoming	0

Predictor Balance:		
	Treated	Synthetic
lnincome	10.07656	9.852656
age15to24	.1735324	.1736614
retprice	89.42222	89.37659
beer	24.28	24.2236
cigsale(1988)	90.1	91.3866
cigsale(1980)	120.2	120.2357
cigsale(1975)	127.1	127.0999

第三步的估计结果汇报了“均方根预测误差”(Root Mean Squared Prediction Error, 简记RMSPE)为1.780; 其中, RMSPE为MSPE的开平方。

合成控制的结果依然是稀疏的：大多数州的权重为0，只有以下五个州的权重为正，即Colorado (0.148)，Connecticut (0.059)，Montana (0.207)，Nevada (0.241)与Utah (0.345)。

在预测变量平衡(Predictor Balance)方面，加州与合成加州的预测变量均十分接近，故合成加州可很好地再现加州的经济特征。

均方根预测误差(RMSPE)为1.780，究竟算大还是小？为此，计算加州的结果变量在1989年前的样本均值：

```
. sum cigsale if state==3 & year<1989
```

Variable	Obs	Mean	Std. dev.	Min	Max
cigsale	19	116.2105	11.68303	90.1	128

在处理前加州的人均香烟消费量均值为116.211。

此命令还输出了结果变量的缺口图(gap graph), 可比较加州与合成加州的人均香烟消费量在1989年前后的表现, 参见图18.3:

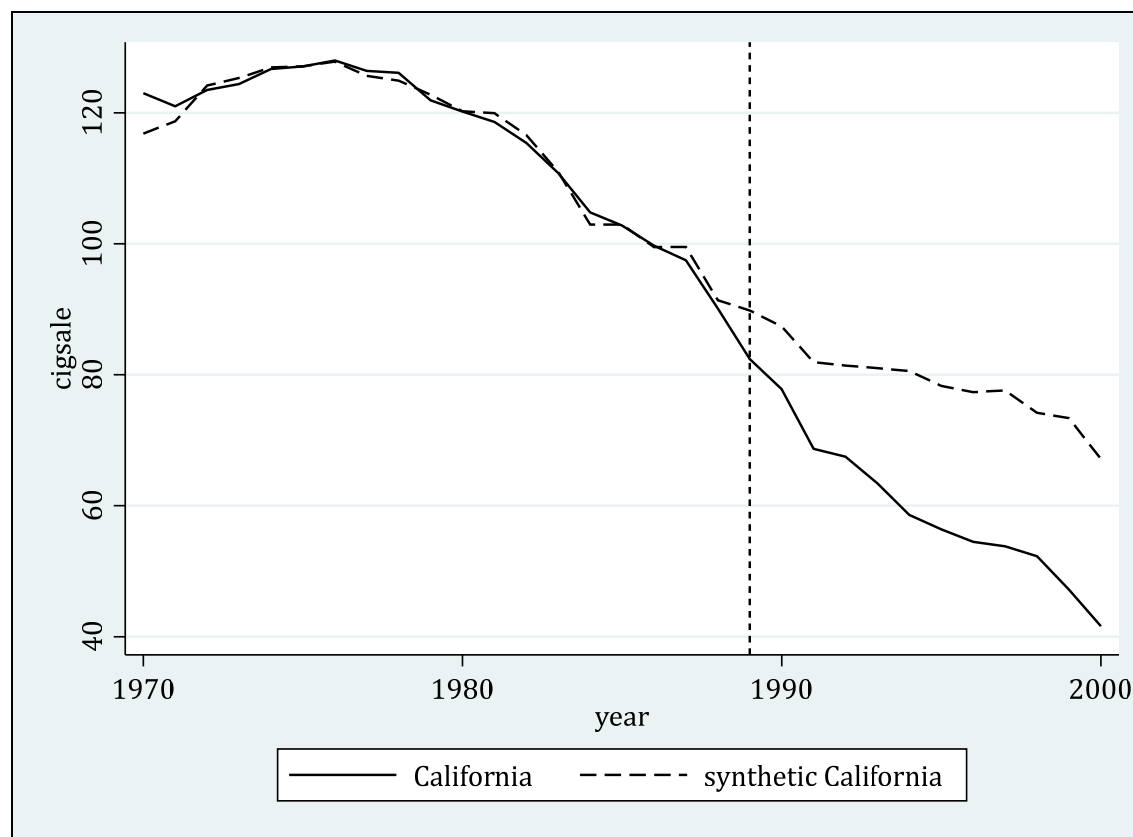


图18.3 结果变量的缺口图

命令`synth`并不提供安慰剂检验的功能。

我们转向命令`synth2`(Yan and Chen, 2023)。该命令提供了一系列安慰剂检验、稳健性检验与可视化的功能，其下载安装方法为

```
. ssc install synth2, all replace
```

其中，选择项“`all`”将安装命令`synth2`自带的加州控烟法数据集`smoking.dta`(与上文的`synth_smoking.dta`完全相同)。

由于命令`synth2`的底层算法调用`synth`，故在使用`synth2`前，须先安装`synth`。

命令synth2的句型也与synth十分类似:

```
. synth2 cigsale lnincome age15to24 retprice beer  
cigsale(1988) cigsale(1980) cigsale(1975),  
trunit(3) trperiod(1989) xperiod(1980(1)1988)  
nested sigf(6) placebo(unit cut(2)) loo
```

选择项“placebo(unit cut(2))”表示使用伪处理单位(fake treatment unit)进行空间安慰剂检验, 但仅使用“处理前MSPE”不超过“处理地区处理前MSPE”2倍的控制地区。

选择项“loo”表示进行留一(leave one out)稳健性检验。

命令synth2并不使用“figure”选项, 因为默认会画一系列的图; 但可使用选择项“nofigure”指定不画图。

此命令的输出结果包括：(1)处理前的拟合效果；(2)处理前的预测变量平衡；(3)合成控制的最优个体权重；(4)处理后的预测结果；(5)空间安慰剂检验的结果；(6)留一稳健性检验的结果。

Fitting results in the pretreatment periods:					
Treated Unit		: California	Treatment Time		: 1989
Number of Control Units		= 38	Root Mean Squared Error		= 1.77955
Number of Covariates		= 7	R-squared		= 0.97411

第一步“处理前的拟合效果”除了汇报RMSE(即RMSPE)，还显示 $R^2$ 高达0.974，更直观地展示了处理前的极佳拟合效果(命令synth并不汇报此 $R^2$ )。

Covariate balance in the pretreatment periods:						
Covariate	V.weight	Treated	Synthetic Control Value	Bias	Average Control Value	Bias
lnincome	0.0001	10.0766	9.8527	-2.22%	9.8292	-2.45%
age15to24	0.0020	0.1735	0.1737	0.07%	0.1725	-0.59%
retprice	0.0147	89.4222	89.3766	-0.05%	87.2661	-2.41%
beer	0.0083	24.2800	24.2236	-0.23%	23.6553	-2.57%
cigsale(1988)	0.0142	90.1000	91.3866	1.43%	113.8237	26.33%
cigsale(1980)	0.1923	120.2000	120.2357	0.03%	138.0895	14.88%
cigsale(1975)	0.7683	127.1000	127.0999	-0.00%	136.9316	7.74%

Note: "V.weight" is the optimal covariate weight in the diagonal of V matrix.  
 "Synthetic Control" is the weighted average of donor units with optimal weights.  
 "Average Control" is the simple average of all control units with equal weights.

第二步“处理前的预测变量平衡”，汇报了每个预测变量的权重(V.weight)，其中变量*cigsale(1975)*的权重高达0.768，而*cigsale(1980)*的权重为0.192，而其余变量的权重均很小(命令synth并不汇报变量权重)。



此表还以百分比形式，汇报了预测变量在合成加州与加州之间偏差(Bias)。

以 $\ln income$ 变量为例，由于加州的处理前均值为10.0766，而合成加州的处理前均值为9.8527，故偏差为 $(9.8527 - 10.0766) / 10.0766 = -2.22\%$ 。

结果显示，变量 $\ln income$ 的偏差绝对值最大，但也仅2.22%；说明预测变量已达到很好的平衡。

如果直接使用所有控制地区的简单算术平均，则此平均控制(average control)的偏差绝对值可达26.33%。这展示了合成控制的优势。

Optimal Unit Weights:	
Unit	U.weight
Utah	0.3450
Nevada	0.2410
Montana	0.2070
Colorado	0.1480
Connecticut	0.0590
Note: The unit Alabama Arkansas Delaware Georgia Idaho Illinois Indiana Iowa Kansas Kentucky Louisiana Maine Minnesota Mississippi Missouri Nebraska NewHampshire NewMexico NorthCarolina NorthDakota Ohio Oklahoma Pennsylvania RhodeIsland SouthCarolina SouthDakota Tennessee Texas Vermont Virginia WestVirginia Wisconsin Wyoming in the donor pool get a weight of 0.	

第三步“最优个体权重”所汇报的合成控制权重与上文synth的结果完全相同。

在汇报格式上，synth2仅汇报取正值的权重(权重为0的控制地区在注释中列出)，并按照权重从大到小排序，以便查阅。

Prediction results in the posttreatment period:

Time	Actual Outcome	Synthetic Outcome	Treatment Effect
1989	82.4000	89.8201	-7.4201
1990	77.8000	87.3789	-9.5789
1991	68.7000	81.9182	-13.2182
1992	67.5000	81.4061	-13.9061
1993	63.4000	81.0228	-17.6228
1994	58.6000	80.5678	-21.9678
1995	56.4000	78.3083	-21.9083
1996	54.5000	77.3429	-22.8429
1997	53.8000	77.6174	-23.8174
1998	52.3000	74.1877	-21.8877
1999	47.2000	73.3950	-26.1950
2000	41.6000	67.1478	-25.5478
Mean	60.3500	79.1761	-18.8261

Note: The average treatment effect over the posttreatment period is -18.8261.

第四步“处理后的预测结果”列出了实际结果、预测结果及二者之差，即处理效应；而处理后的平均处理效应为-18.826。

命令synth2还输出了预测变量平衡图、最优变量权重的条形图、最优个体权重的条形图、结果变量的缺口图，及处理效应图。

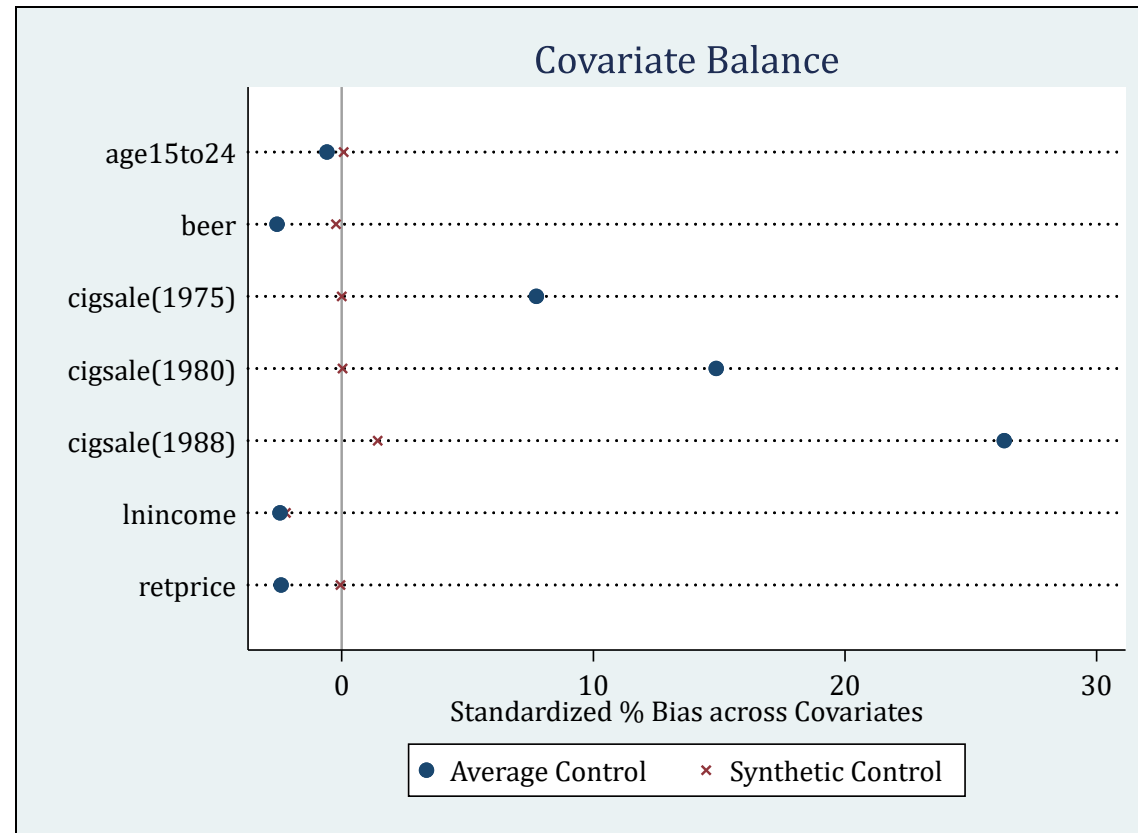


图18.4 预测变量的平衡图

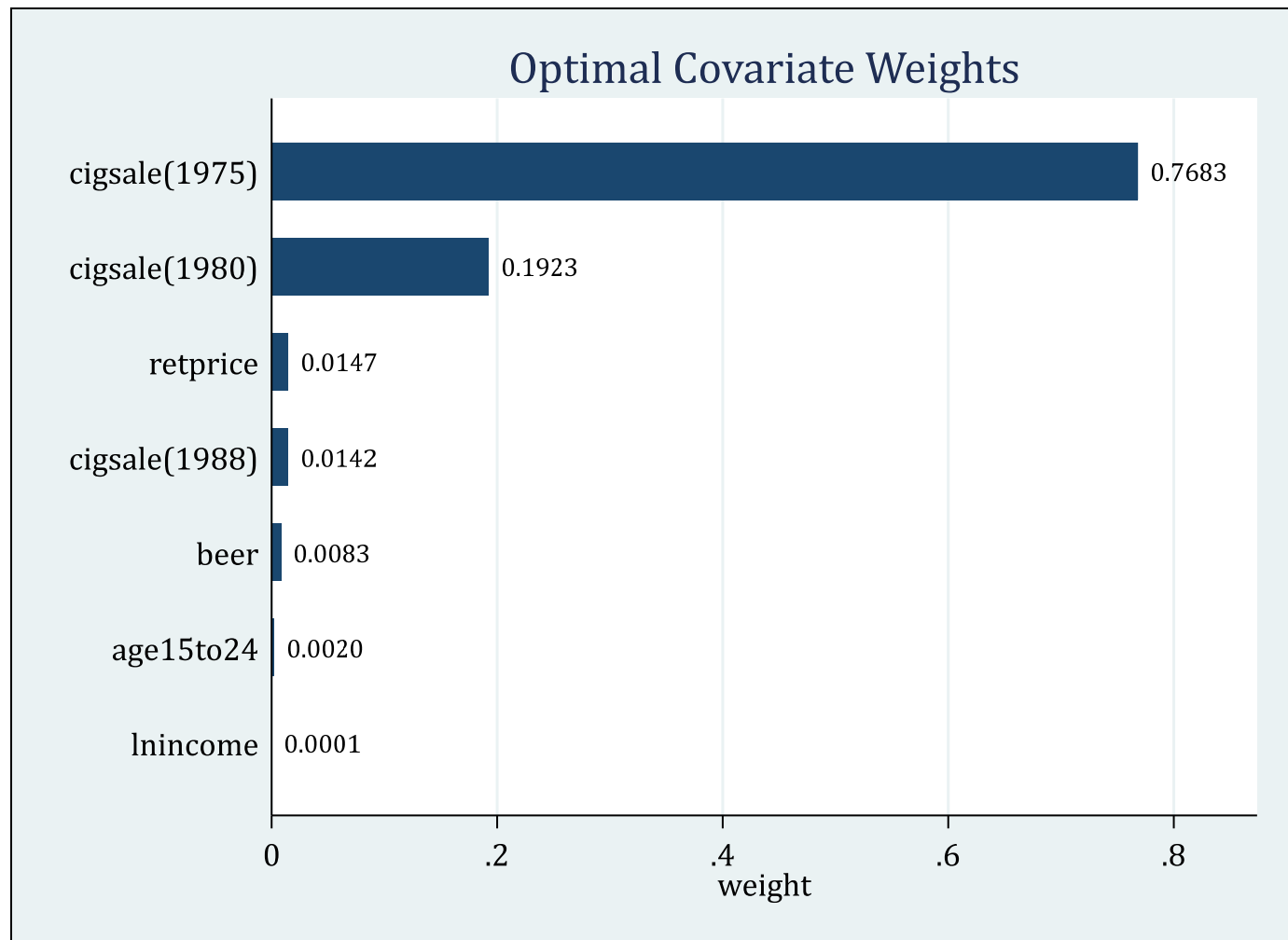


图18.5 最优变量权重的条形图

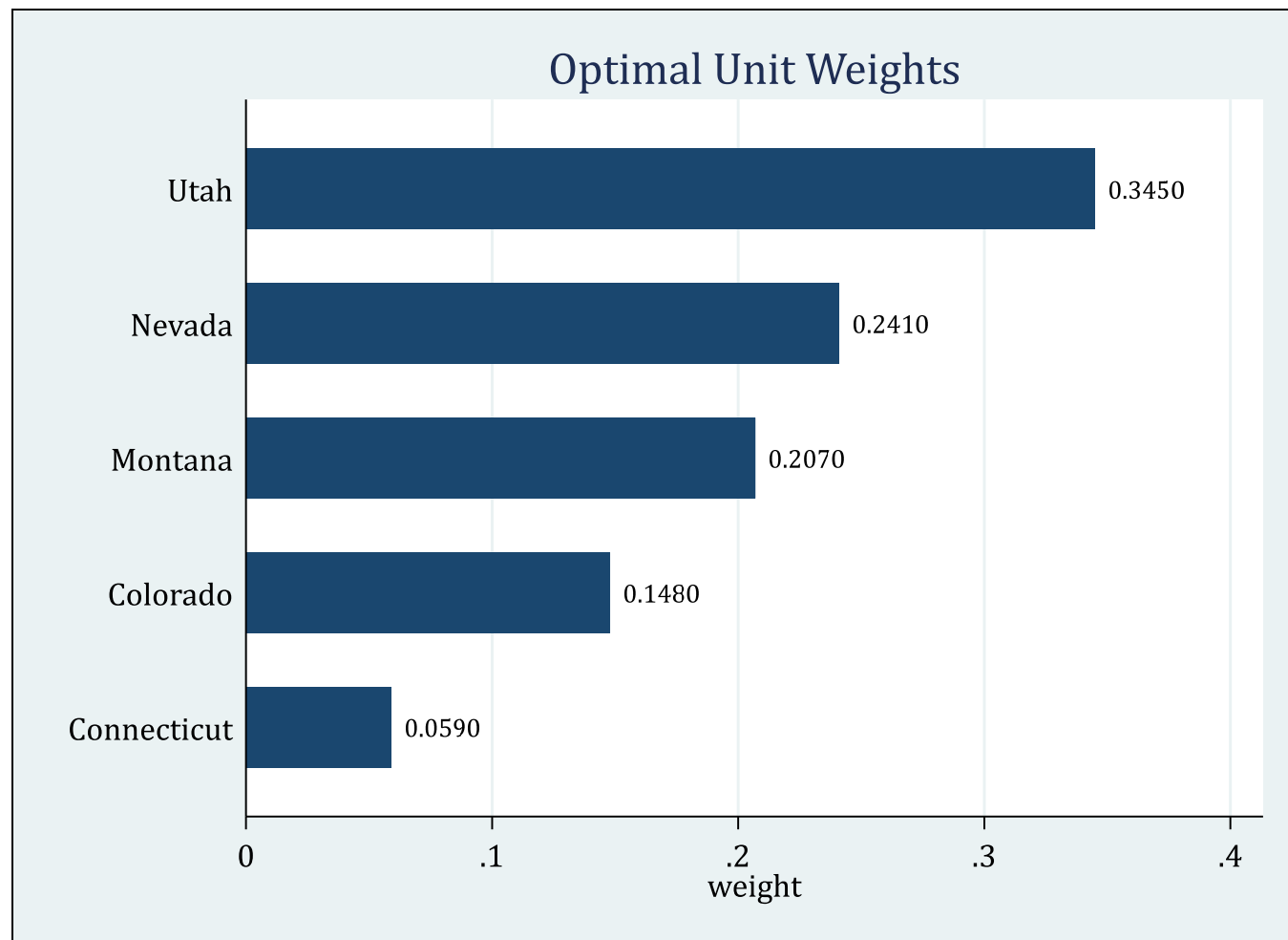


图18.6 最优个体权重的条形图

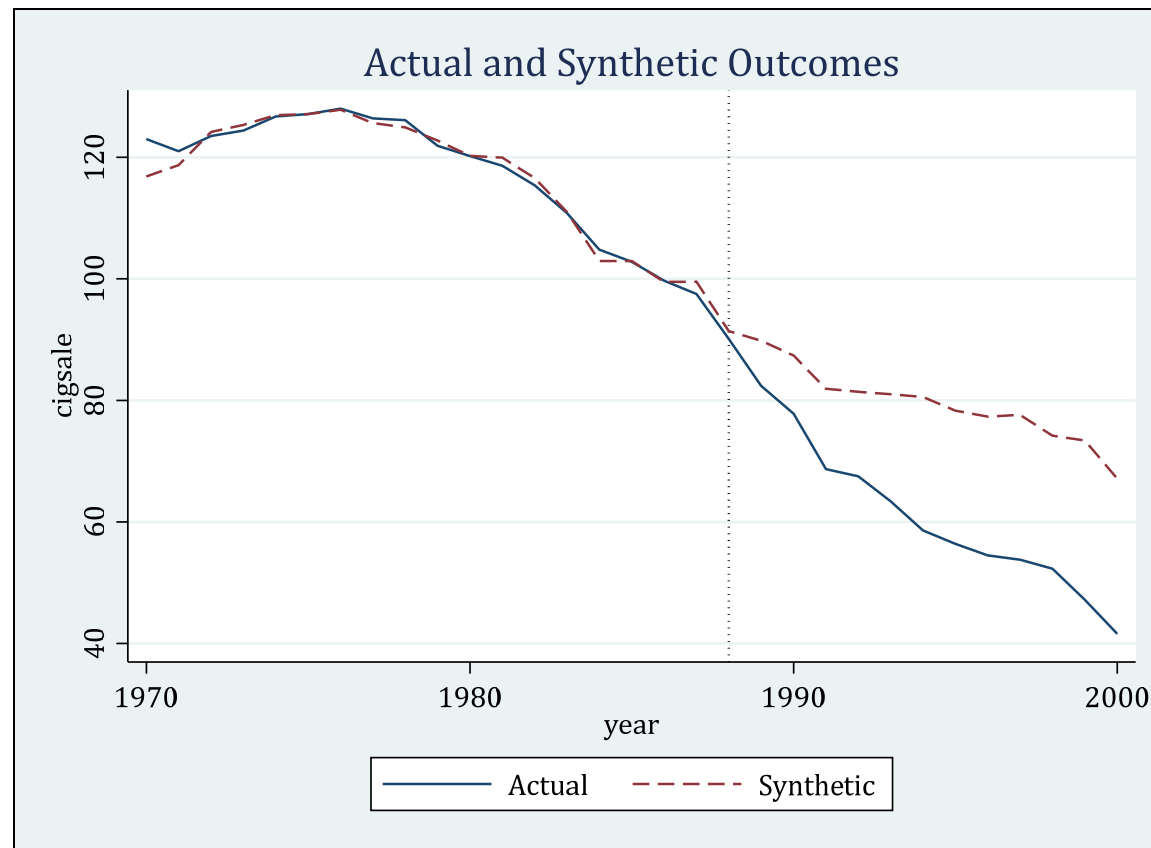


图18.7 结果变量的缺口图

此缺口图与命令synth所画的图18.3几乎一样, 但将垂直虚线画在处理前的最后一期(即1988年), 以达到更好的视觉效果。

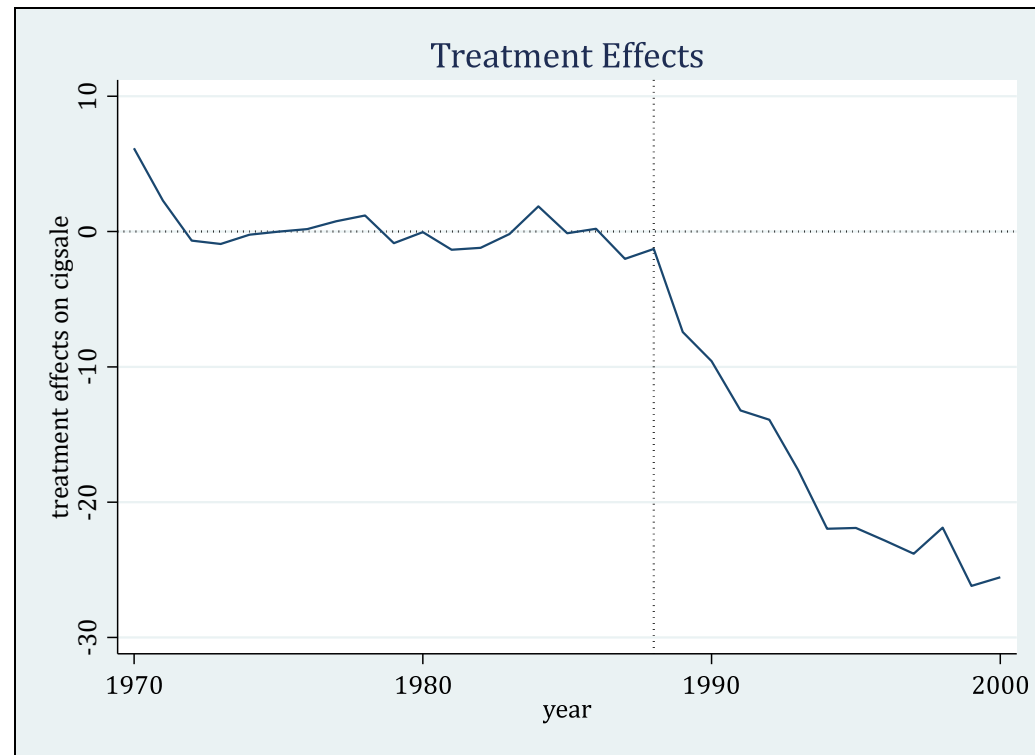


图18.8 处理效应图

加州控烟法对于人均香烟消费量有很大的负效应，而且此效应随着时间推移而变大。在1989-2000年间，加州的人均年香烟消费减少了20多包，大约下降25%，故经济效应十分显著。



In-space placebo test results using fake treatment units:

Unit	Pre MSPE	Post MSPE	Post/Pre MSPE	Pre MSPE of Fake Unit/ Pre MSPE of Treated Unit
California	3.1668	391.2533	123.5490	1.0000
Alabama	5.4170	8.2108	1.5157	1.7106
Arkansas	4.5587	28.8239	6.3228	1.4395
Colorado	17.6103	68.6736	3.8996	5.5609
Connecticut	20.6396	118.9189	5.7617	6.5175
Delaware	30.3949	499.0694	16.4195	9.5980
Georgia	1.4610	116.8893	80.0074	0.4613
Idaho	5.8142	39.1830	6.7392	1.8360
Illinois	4.3146	89.0552	20.6406	1.3624
Indiana	14.4145	469.4150	32.5654	4.5518
Iowa	14.6527	31.5816	2.1553	4.6270
Kansas	14.1121	9.0349	0.6402	4.4563
Kentucky	431.7229	1475.7975	3.4184	136.3284
Louisiana	2.0183	94.1070	46.6279	0.6373
Maine	8.6989	115.5412	13.2822	2.7469
Minnesota	14.1736	52.5217	3.7056	4.4757
Mississippi	4.0894	37.2754	9.1151	1.2913
Missouri	1.2009	85.1794	70.9308	0.3792
Montana	5.2861	54.8978	10.3853	1.6692
Nebraska	4.8287	36.5597	7.5713	1.5248
Nevada	40.6500	83.4186	2.0521	12.8364
NewHampshire	3436.5980	134.9018	0.0393	1085.2007
NewMexico	5.0577	63.7459	12.6036	1.5971
NorthCarolina	90.2241	67.3144	0.7461	28.4907
NorthDakota	8.0725	72.3200	8.9588	2.5491
Ohio	2.9585	12.3535	4.1757	0.9342

Oklahoma	5.7128	267.8078	46.8786	1.8040
Pennsylvania	2.7029	6.7073	2.4815	0.8535
RhodeIsland	87.9904	242.6697	2.7579	27.7854
SouthCarolina	2.1997	41.2941	18.7727	0.6946
SouthDakota	7.5688	32.2367	4.2592	2.3901
Tennessee	5.2043	123.3097	23.6938	1.6434
Texas	4.6691	239.8559	51.3707	1.4744
Utah	593.7643	223.2758	0.3760	187.4975
Vermont	15.3860	116.8473	7.5944	4.8585
Virginia	2.7825	219.8136	78.9994	0.8786
WestVirginia	8.1492	242.1734	29.7175	2.5733
Wisconsin	2.8950	75.2425	25.9901	0.9142
Wyoming	83.7717	31.8266	0.3799	26.4532

Note: (1) Using all control units, the probability of obtaining a post/pretreatment MSPE ratio as large as California's is 0.0256.

(2) Excluding control units with pretreatment MSPE 2 times larger than the treated unit, the probability of obtaining a post/pretreatment MSPE ratio as large as California's is 0.0500.

(3) The pointwise p-values below are computed by excluding control units with pretreatment MSPE 2 times larger than the treated unit.

(4) There are total 19 units with pretreatment MSPE 2 times larger than the treated unit, including Colorado Connecticut Delaware Indiana Iowa Kansas Kentucky Maine Minnesota Nevada NewHampshire NorthCarolina NorthDakota RhodeIsland SouthDakota Utah Vermont WestVirginia Wyoming.

第五步“空间安慰剂检验”结果显示，加州的处理后与处理前的MSPE比值(Post/Pre MSPE)遥遥领先，在所有39个地区中达到最大，故相应的 $p$ 值为 $1/39 = 0.0256$ 。

如果加州控烟法无处理效应，则看到此极端结果的概率仅为2.56%，故在5%的水平上显著。

共有19个控制地区的处理前MSPE超过加州的两倍。

若排除这19个处理前拟合不佳的控制地区，则安慰剂检验的 $p$ 值为 $1/(39-19) = 0.05$ ，仍在5%的水平上显著。

图18.9展示了所有地区处理后与处理前的MSPE比值的条形图。

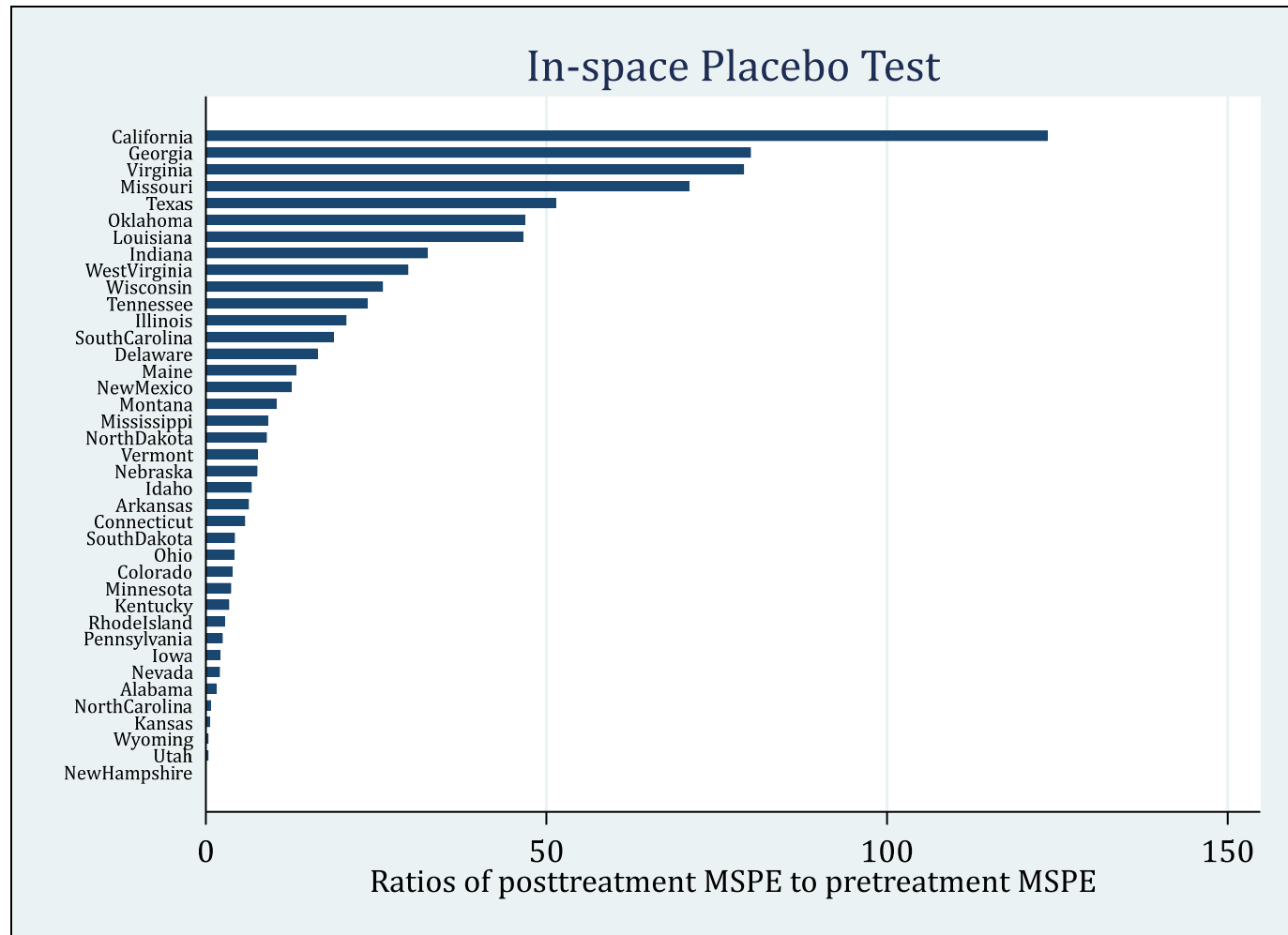


图18.9 处理后与处理前的MSPE比值的条形图

安慰剂效应的分布图参见图18.10, 也称为意面图(spaghetti plot), 因为形似意大利面。

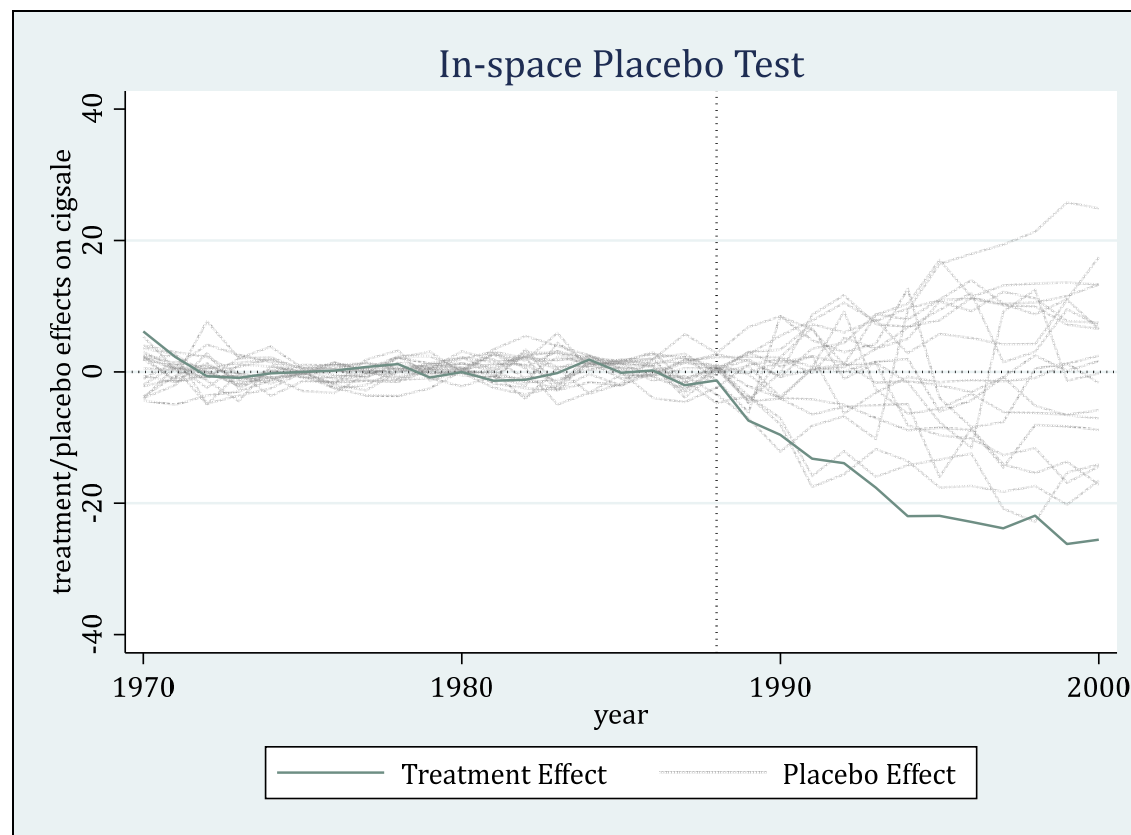


图18.10 安慰剂效应的分布图

根据图18.10的安慰剂效应分布，此命令还提供了每期处理效应的三种不同 $p$ 值。

在计算这些 $p$ 值时，均排除了处理前MSPE超过加州两倍的那19个控制地区。

针对原假设 $H_0: \tau_{1t} = 0$ ，若替代假设为 $H_1: \tau_{1t} \neq 0$ ，则为双边检验，拒绝域在分布的两侧，应使用**双边 $p$ 值**(two-sided  $p$ -value)。

双边 $p$ 值为“安慰剂效应的绝对值”大于或等于“处理效应绝对值”的比重。

若替代假设为 $H_1: \tau_{1t} > 0$ , 则为右边检验, 拒绝域在分布的右侧, 应使用**右边 $p$ 值(right-sided  $p$ -value)**, 即安慰剂效应大于或等于处理效应的比重。

反之, 若替代假设为 $H_1: \tau_{1t} < 0$ , 则为左边检验, 拒绝域在分布的左侧, 应使用**左边 $p$ 值(left-sided  $p$ -value)**, 即安慰剂效应小于或等于处理效应的比重。

一般地, 单边检验的功效(power)高于双边检验。

In-space placebo test results using fake treatment units (continued, cutoff = 2):

Time	Treatment Effect	p-value of Treatment Effect		
		Two-sided	Right-sided	Left-sided
1989	-7.4201	0.0500	1.0000	0.0500
1990	-9.5789	0.1000	0.9500	0.1000
1991	-13.2182	0.1500	0.9000	0.1500
1992	-13.9061	0.1000	0.9500	0.1000
1993	-17.6228	0.0500	1.0000	0.0500
1994	-21.9678	0.0500	1.0000	0.0500
1995	-21.9083	0.0500	1.0000	0.0500
1996	-22.8429	0.0500	1.0000	0.0500
1997	-23.8174	0.0500	1.0000	0.0500
1998	-21.8877	0.1000	0.9500	0.1000
1999	-26.1950	0.0500	1.0000	0.0500
2000	-25.5478	0.0500	1.0000	0.0500

Note: (1) The two-sided p-value of the treatment effect for a particular period is defined as the frequency that the absolute values of the placebo effects are greater than or equal to the absolute value of treatment effect.  
(2) The right-sided (left-sided) p-value of the treatment effect for a particular period is defined as the frequency that the placebo effects are greater (smaller) than or equal to the treatment effect.  
(3) If the estimated treatment effect is positive, then the right-sided p-value is recommended; whereas the left-sided p-value is recommended if the estimated treatment effect is negative.



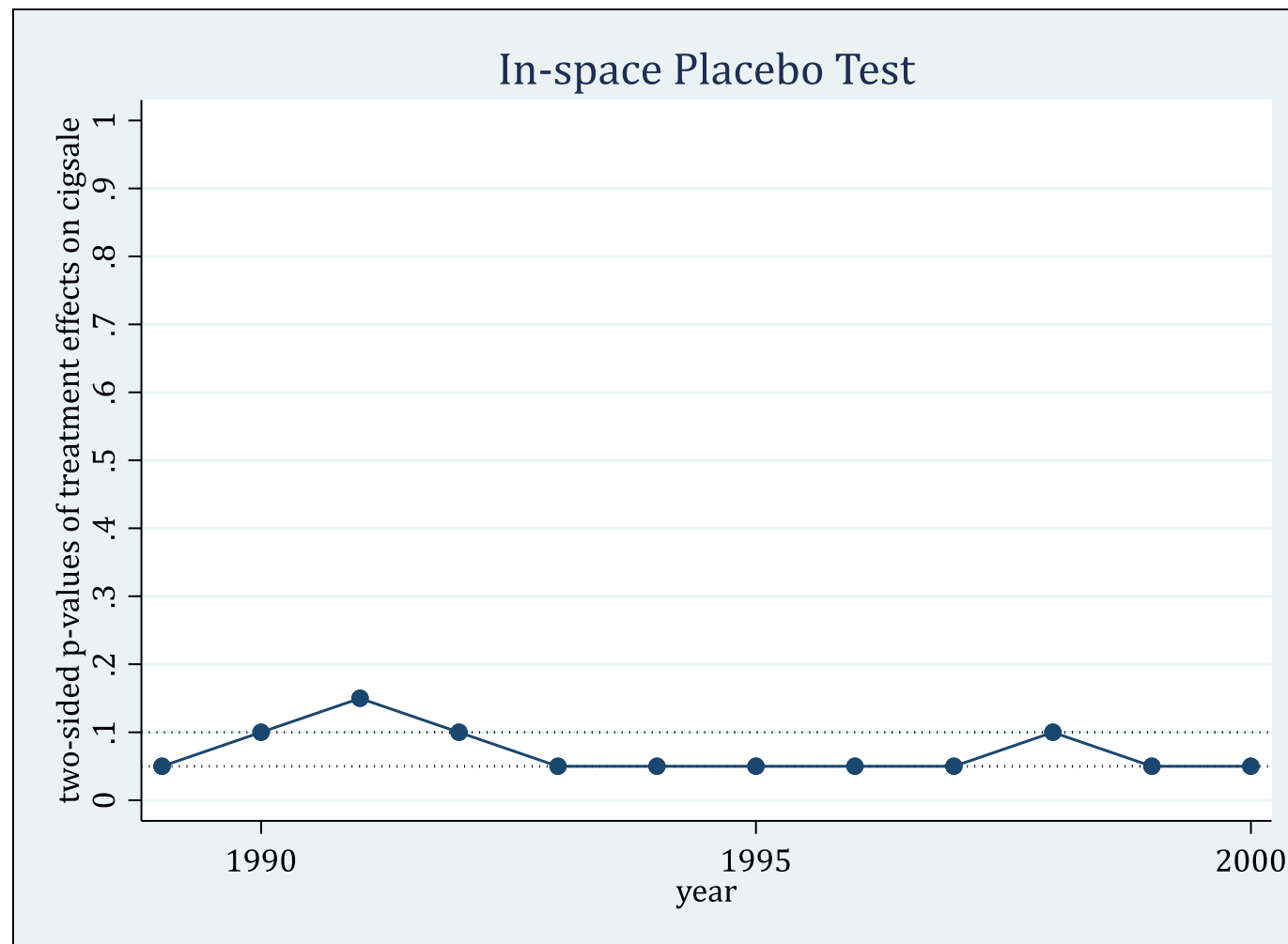


图18.11 各期处理效应的双边 $p$ 值

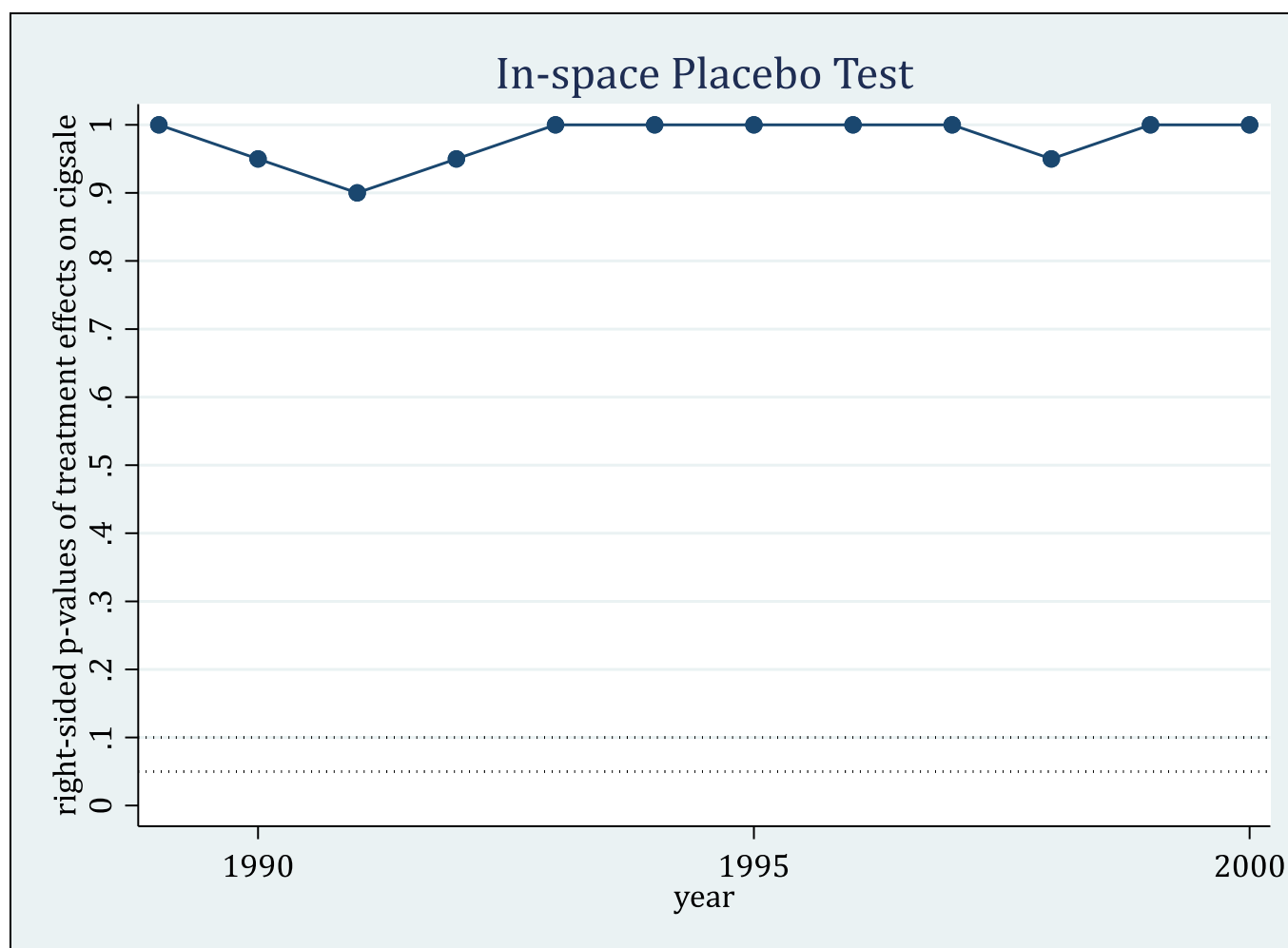


图18.12 各期处理效应的右边 $p$ 值

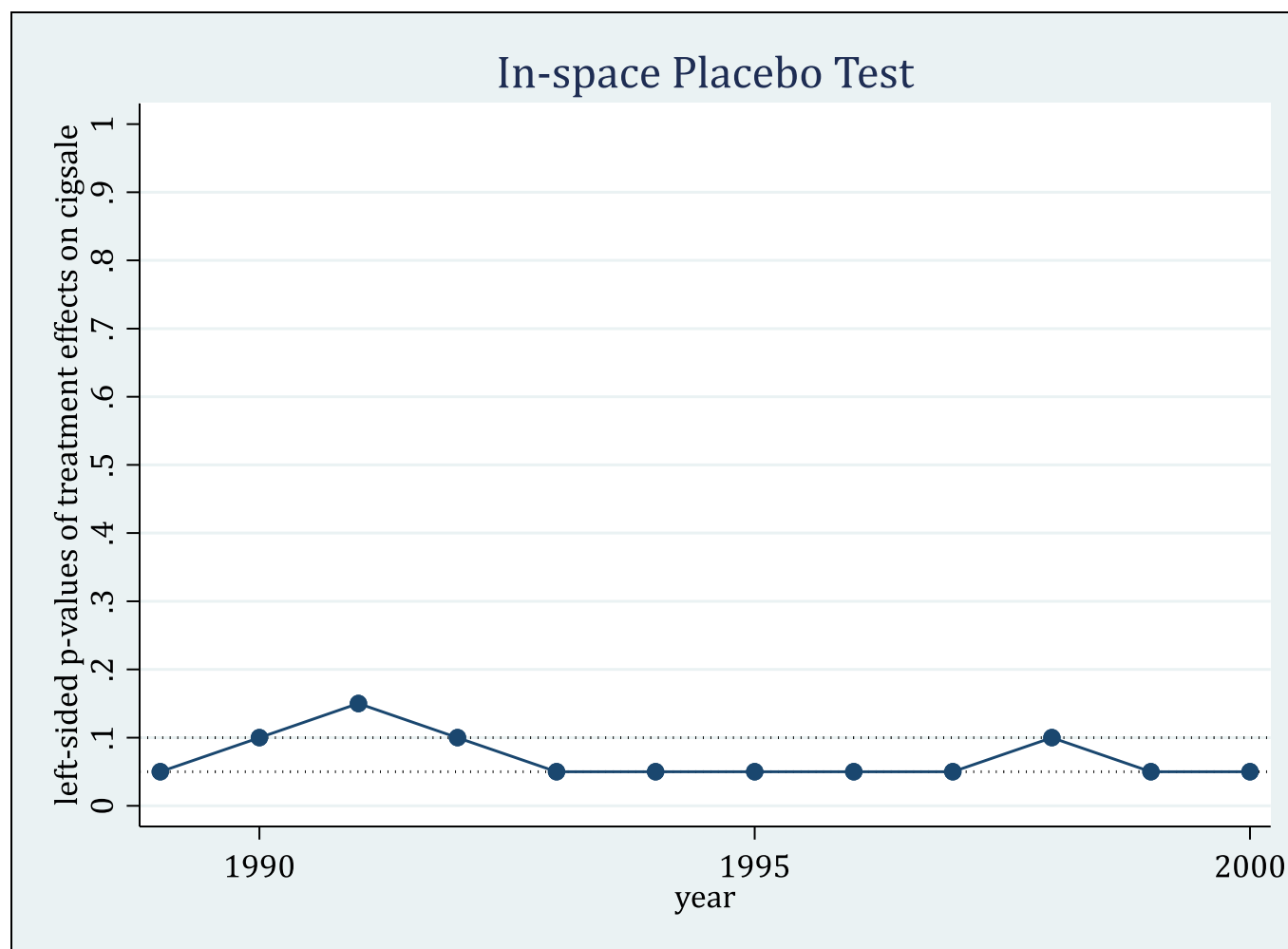


图18.13 各期处理效应的左边 $p$ 值

第六步“留一稳健性检验”汇报了依次去掉一个权重为正的制地区的SCM估计结果。限于篇幅，仅汇报在所有留一样本中，结果变量预测值与处理效应的最小值与最大值。

Leave-one-out robustness test results in the posttreatment period:				
Time	Outcome		Synthetic Outcome (LOO)	
	Actual	Synthetic	Min	Max
1989	82.4000	89.8201	87.9898	91.9821
1990	77.8000	87.3789	83.2283	88.8714
1991	68.7000	81.9182	80.5846	82.5416
1992	67.5000	81.4061	80.6038	81.9059
1993	63.4000	81.0228	79.7538	81.2011
1994	58.6000	80.5678	78.5506	82.2610
1995	56.4000	78.3083	76.0281	79.6610
1996	54.5000	77.3429	75.0686	78.8747
1997	53.8000	77.6174	71.7959	81.7168
1998	52.3000	74.1877	71.1697	77.4632
1999	47.2000	73.3950	71.5312	76.1493
2000	41.6000	67.1478	65.1462	68.3202

Note: The last two columns report the minimum and maximum synthetic outcomes when one control unit with a nonzero weight is excluded at a time.

Time	Treatment Effect	Treatment Effect (LOO)	
		Min	Max
1989	-7.4201	-9.5821	-5.5898
1990	-9.5789	-11.0714	-5.4283
1991	-13.2182	-13.8416	-11.8846
1992	-13.9061	-14.4059	-13.1038
1993	-17.6228	-17.8011	-16.3538
1994	-21.9678	-23.6610	-19.9506
1995	-21.9083	-23.2610	-19.6281
1996	-22.8429	-24.3747	-20.5686
1997	-23.8174	-27.9168	-17.9959
1998	-21.8877	-25.1632	-18.8697
1999	-26.1950	-28.9493	-24.3312
2000	-25.5478	-26.7202	-23.5462

Note: The last two columns report the minimum and maximum treatment effects when one control unit with a nonzero weight is excluded at a time.

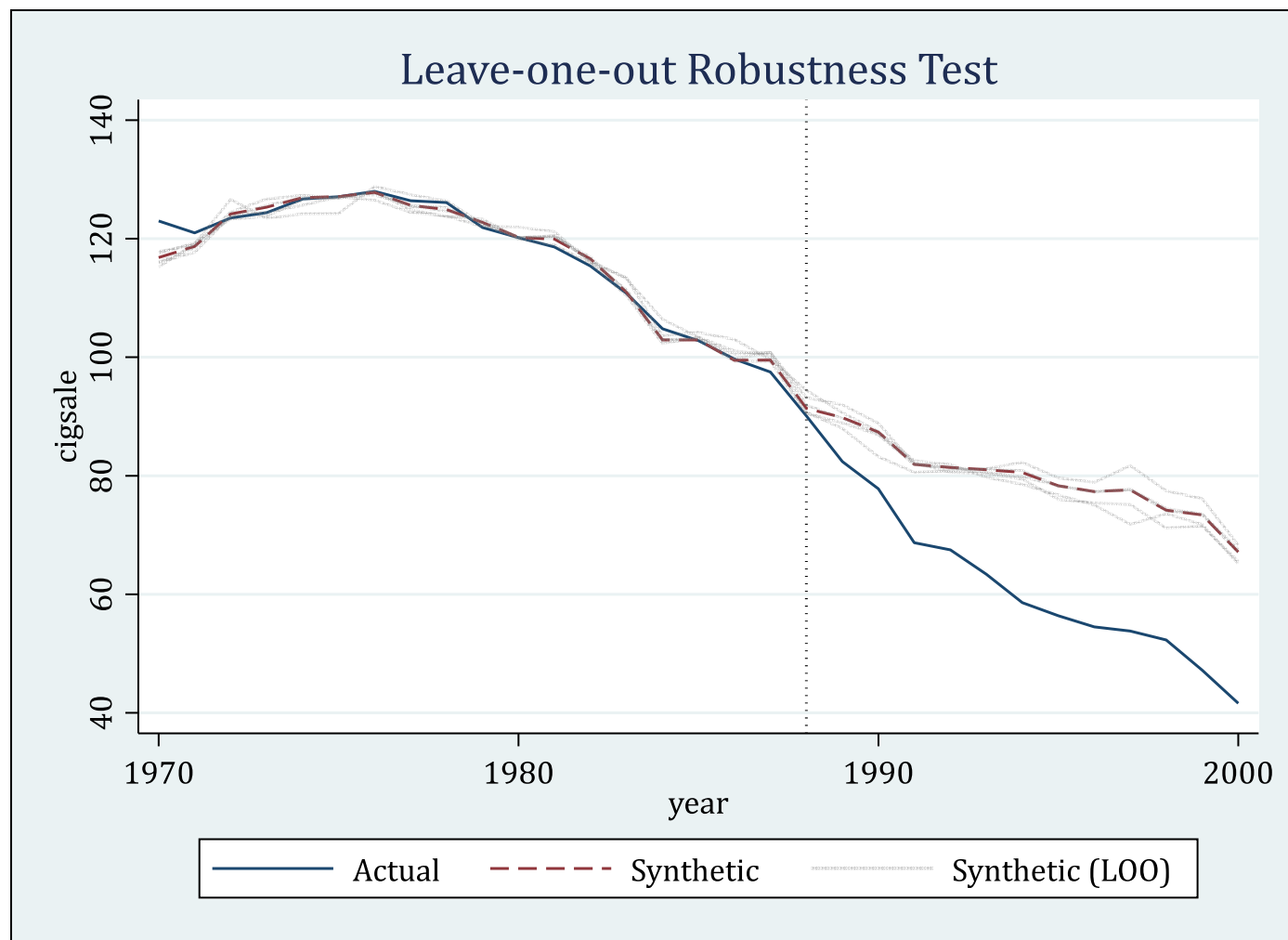


图18.14 结果变量的留一稳健性检验

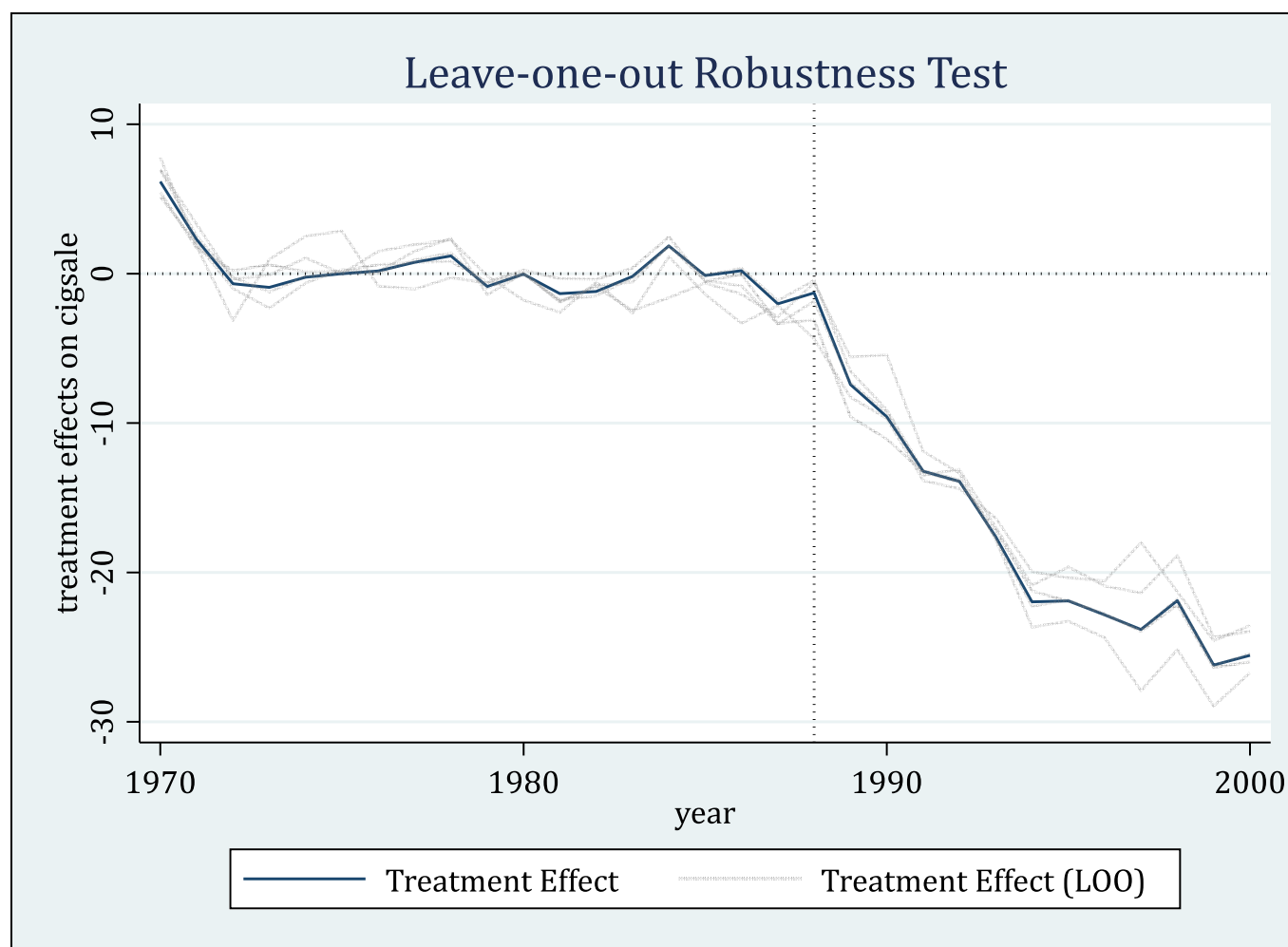


图18.15 处理效应的留一稳健性检验

下面进行时间安慰剂检验，以1985年为伪处理时间，可通过选择项`placebo(period(1985))`来实现。

由于变量`cigsale(1988)`发生于1985年之后，故去掉此变量。

原来的选择项“`xperiod(1980(1)1988)`”也相应地更新为“`xperiod(1980(1)1984)`”：

```
. synth2 cigsale lnincome age15to24 retprice beer  
cigsale(1980) cigsale(1975), trunit(3)  
trperiod(1989) xperiod(1980(1)1984) nested  
sigf(6) placebo(period(1985))
```



In-time placebo test results using fake treatment time 1985:			
Time	Actual Outcome	Synthetic Outcome	Treatment Effect
1985	102.8000	106.1262	-3.3262
1986	99.7000	103.2850	-3.5850
1987	97.5000	106.1524	-8.6524
1988	90.1000	98.4873	-8.3873
1989	82.4000	96.5237	-14.1237
1990	77.8000	91.9127	-14.1127
1991	68.7000	83.7156	-15.0156
1992	67.5000	81.4730	-13.9730
1993	63.4000	79.7911	-16.3911
1994	58.6000	77.9078	-19.3078
1995	56.4000	76.2193	-19.8193
1996	54.5000	75.2010	-20.7010
1997	53.8000	75.1958	-21.3958
1998	52.3000	71.9437	-19.6437
1999	47.2000	72.2260	-25.0260
2000	41.6000	67.1861	-25.5861
Mean	69.6437	85.2092	-15.5654

Note: The average treatment effect over the posttreatment period is -15.5654.

上表汇报了以1985年为假处理时间的SCM估计结果。

图18.16提供了结果变量的缺口图，假定政策冲击发生于1985年(图中左边的垂直虚线)。

在1985-1988年间，结果变量与反事实预测似乎有明显的差距；尽管加州控烟法在此期间并未实施。

图19.17提供了安慰剂效应的趋势图，其中安慰剂效应在1985-1988年间似乎也较明显。一

个可能的解释是，控烟运动或许在几年前就已在加州启动，并最终导致1988年通过加州控烟法。

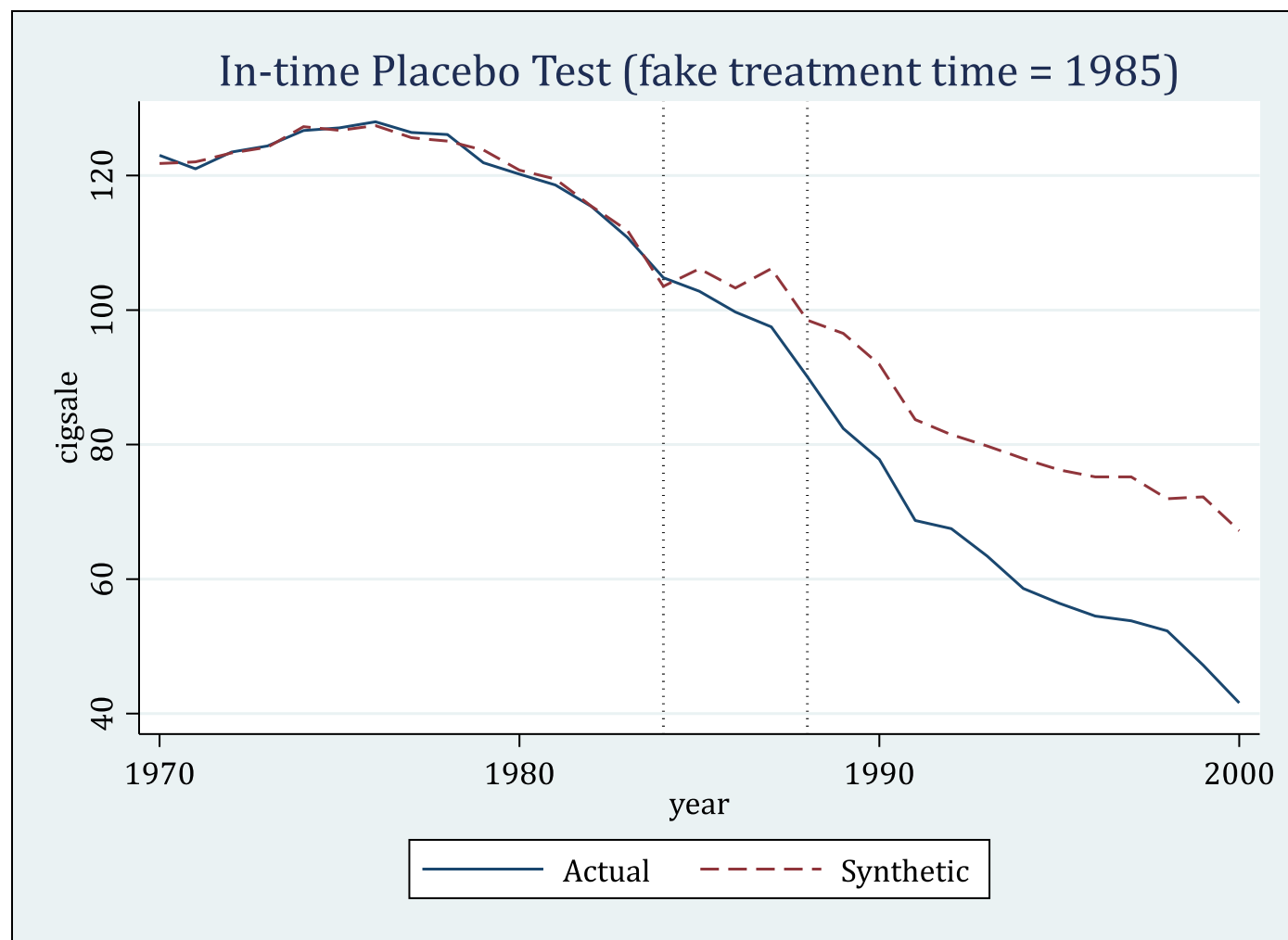


图18.16 结果变量的缺口图(时间安慰剂检验)

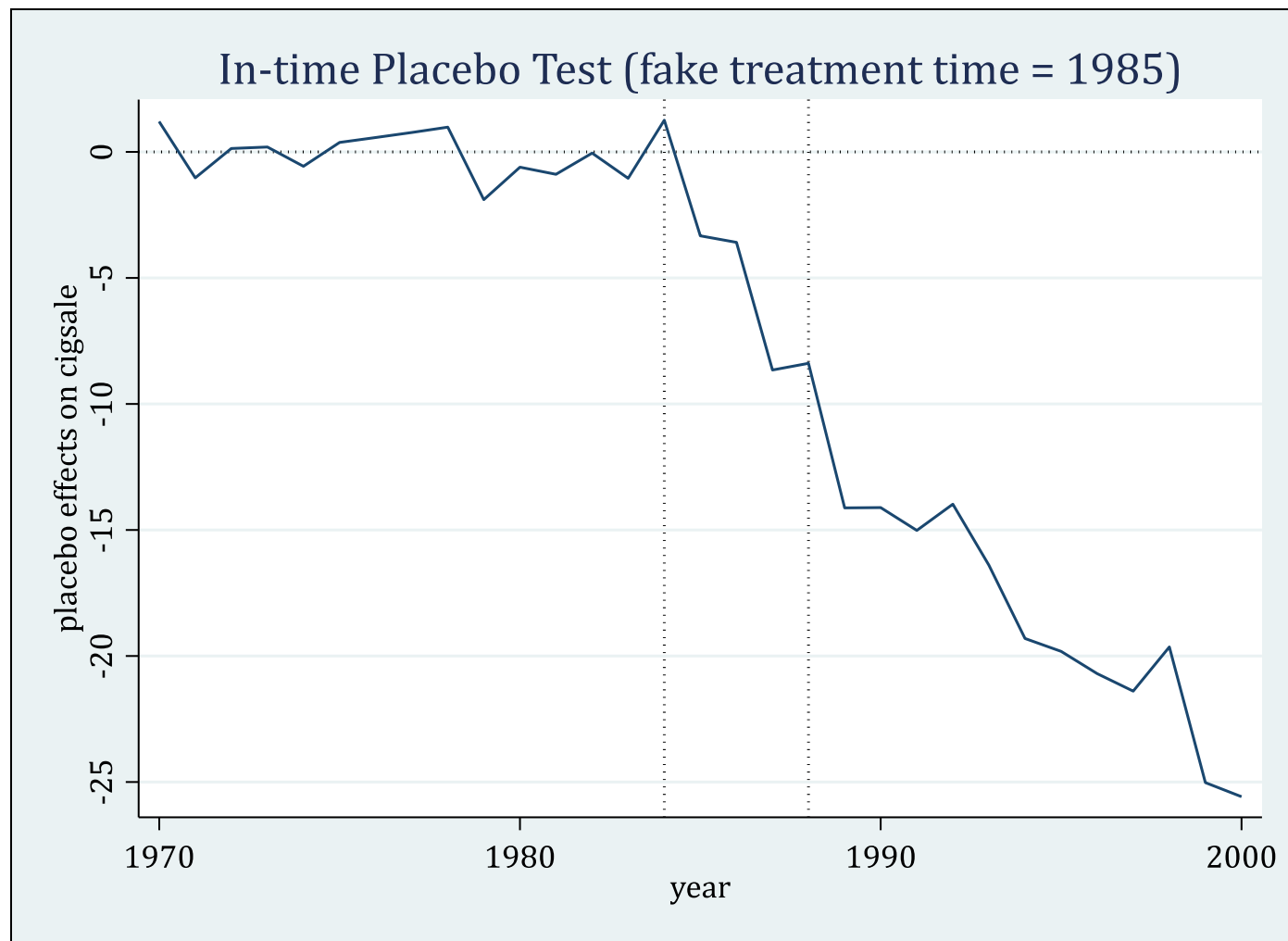


图18.17 处理效应(时间安慰剂检验)

由于传统的时间安慰剂检验并无 $p$ 值，故下面进行混合安慰剂检验(Chen and Yan, 2023)，同时使用伪处理时间与伪假处理地区。

以1985年为伪处理时间，但使用选择项“`trperiod(1985)`”将其视作真实的处理时间，然后进行空间安慰剂检验：

```
. synth2 cigsale lnincome age15to24 retprice beer  
cigsale(1980) cigsale(1975), trunit(3)  
trperiod(1985) xperiod(1980(1)1984)  
postperiod(1985(1)1988) nested placebo(unit  
cut(10)) sigf(6)
```

其中，选择项“`postperiod(1985(1)1988)`”指定处理后时期为1985-1988年，因为1989年后处理效应即开始起作用。

选择项 “`placebo(unit cut(10))`” 进行空间安慰剂检验，并指定控制地区的处理前MSPE不超过加州的10倍。

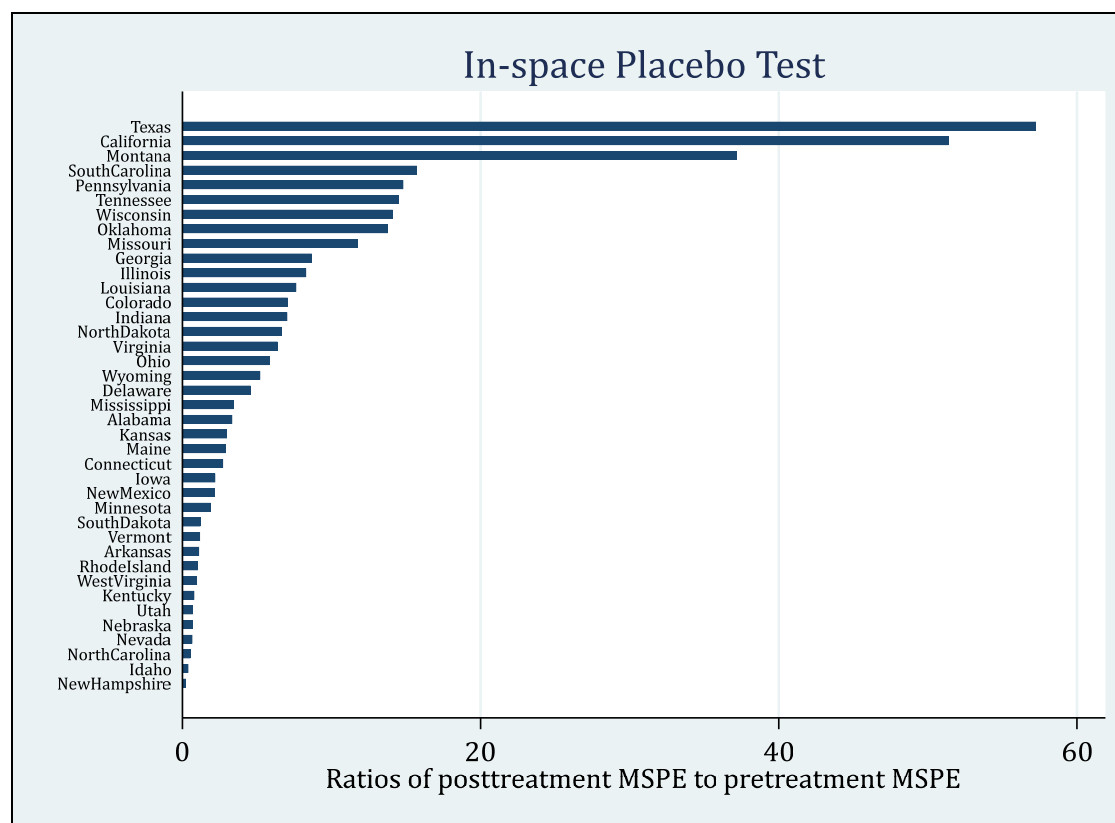


图18.18 处理后与处理前的MSPE比值(混合安慰剂检验)

样本中所有 39 个州中，加州的 MSPE 比值为第二大，这意味着  $p$  值为  $2/39 = 0.0513$ 。

可在 5% 的水平上接受“加州安慰剂效应不显著”的原假设。

如果去掉处理前 MSPE 超过加州 10 倍的 19 个控制地区，则加州的 MSPE 比值依然在剩余的 20 州中排名第二，故  $p$  值为  $2/20 = 0.1$ ，仍可在 5% 的水平上接受原假设。

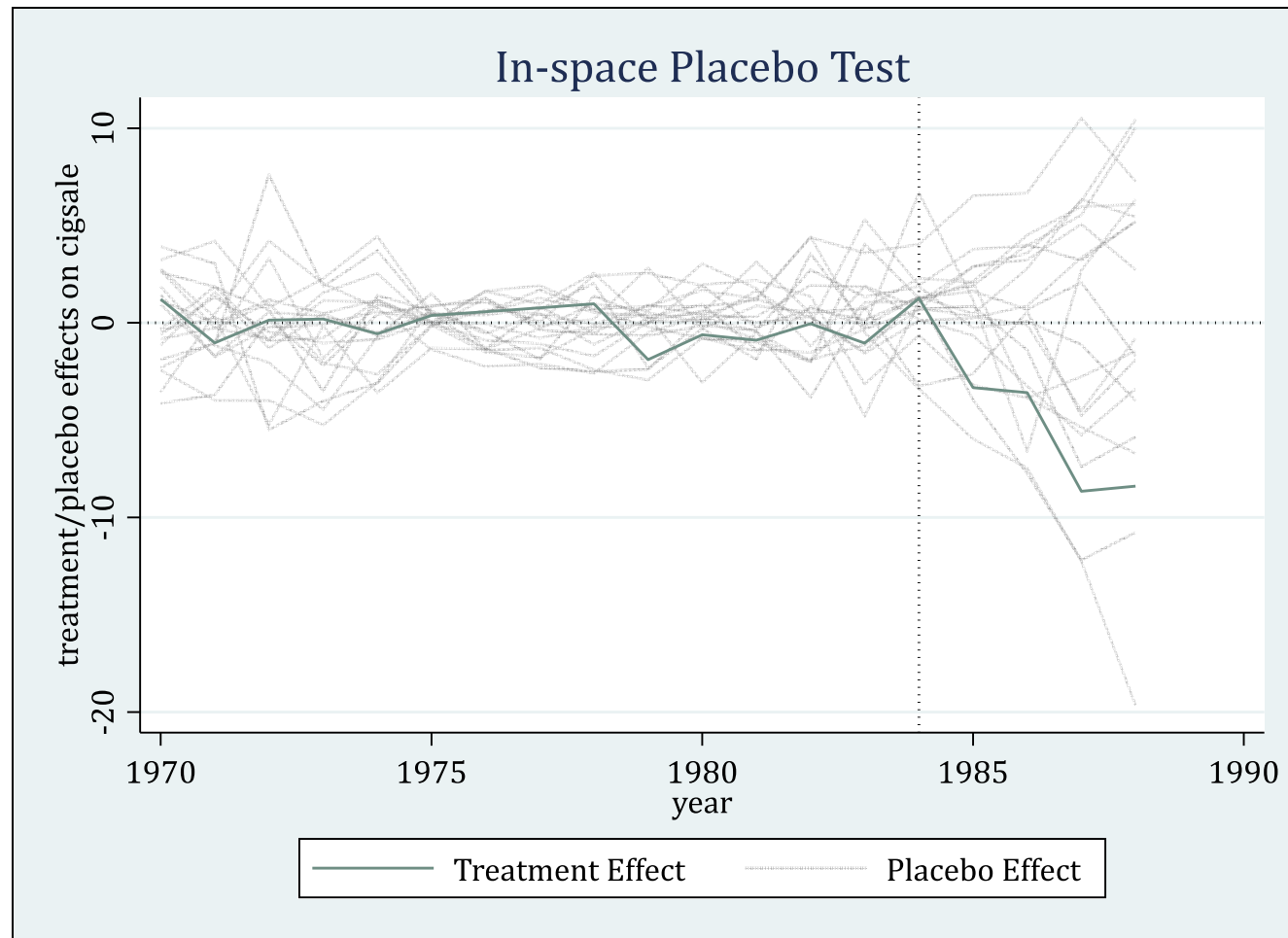


图18.19 安慰剂效应的分布图(混合安慰剂检验)



加州的安慰剂效应虽然位于分布的下方，但并非在最底部。

下表提供了在 1985-1988 年间，各期安慰剂效应的三种  $p$  值。

由于各期安慰剂效应均为负数，故应使用功效最高的左边  $p$  值。

最小的左边  $p$  值为 0.15，故可在 10% 的水平上接受“加州安慰剂效应为 0”的原假设。

In-space placebo test results using fake treatment units (continued, cutoff = 10):				
Time	Treatment Effect	p-value of Treatment Effect		
		Two-sided	Right-sided	Left-sided
1985	-3.3262	0.2500	0.9000	0.1500
1986	-3.5850	0.5000	0.7500	0.3000
1987	-8.6524	0.2000	0.9000	0.1500
1988	-8.3873	0.2500	0.9000	0.1500

为了进一步检验 SCM 估计结果的稳健性，Abadie et al. (2010) 加入了更多的预测变量，比如失业率、收入不平等、贫困率、福利转移、犯罪率、毒品相关的逮捕率、香烟税、人口密度等；发现结果依然稳健。

另一担心是，地区间无互相影响的 SUTVA 假定可能不满足，比如加州的反烟运动可能波及其他州，烟草行业或将其他州的香烟广告预算投入到加州，甚至从其他州走私便宜香烟到加州。

Abadie et al. (2010) 根据史实对此进行了探讨，认为这些效应均不大，至少不可能导致上文图中如此大的处理效应。

## 18.8 使用合成控制法的注意事项

Abadie et al. (2015)探讨了使用合成控制法的前提与局限性。

在构建潜在控制地区的捐赠池(donor pool)时尤其应小心。

首先，也受到此政策冲击的地区应从捐赠池中去掉。

其次，在样本期间受到很大特殊冲击的地区应排除在捐赠池之外。

再次，为了避免“内插偏差”(interpolation bias)，应将捐赠池限定为与处理地区具有相似特征的控制地区。

在应用合成控制法时，还要求干预前的时期数 $T_0$ 达到一定规模(a sizable number of preintervention periods)。

合成控制法的可信度取决于，合成控制能在干预前相当一段时期内很好地追踪处理地区的经济特征与结果变量。

如果干预前的拟合不好，或干预前期数太短，则不建议使用合成控制法。

如果政策冲击的效应需要一段时间才会显现(即滞后效应)，则也要求干预后的期数足够大。