

微观计量经济学 合成控制法

阮 睿

中央财经大学
中国财政发展协同创新中心

May 15, 2025

Contents

1	比较案例研究	3
2	合成控制方法	9
2.1	加利福利亚州《99 号提案》	14
2.2	证伪	23

1 比较案例研究

在定性案例研究中，如托克维尔的经典著作《论美国的民主》，其目的是通过逻辑和历史分析，归纳出某个事件或单个个体的特征在结果上的因果效应。但由于定性比较案例研究有时缺乏明确的反事实，所以对这些因果问题可能无法给出非常令人满意的答案。

定量比较案例研究是更明确的因果设计。通常是自然实验，只适用于单个个体，如单个学校、公司、地区或国家。这类定量比较案例研究将总体结果的演变与其他个体的结果进行比较，或者更常见的情况是，选择出一组类似的个体作为对照组。

Athey 和 Imbens (2017) 认为，对定量比较案例研究最重要的贡献之一是合成控制模型。

合成控制模型最优地选择一组权重，当应用到一组相应的个体时，对接受处理的个体产生最优的反事实估计。这个反事实个体就被称为“合成个体”(**synthetic unit**)，用于概述如果不发生处理，总体处理个体将会发生什么。这是对双重差分策略的一个强大而简单的概括。

劳动经济学家多年来一直在辩论移民对当地劳动市场条件的影响（Card 和 Peri 2016）。移民的涌入是否会压低工资并增加本地人在本地劳动市场的就业？

Card（1990）使用了自然实验方法来研究。

1980 年，菲德尔·卡斯特罗宣布任何希望离开古巴的人都可以离开。古巴裔美国人帮助安排了马列尔船运，这是一场从 1980 年 4 月到 10 月从古巴马列尔港到美国（主要是迈阿密）的大规模外流。大约有 125,000 名古巴人移民到佛罗里达州。之后古巴和美国同意停止这次移民活动。这一事件使迈阿密的劳动力增加了 7%，主要是将创纪录数量的低技能工人投入到一个相对较小的地理区域。

- 一个理想的自然实验。这可以说是劳动供给曲线的一个外生性变化
- 使用了来自当前人口调查的关于迈阿密的个人失业数据，并选择了四个比较城市（亚特兰大、洛杉矶、休斯顿和坦帕-圣彼得堡）。Card 认为这几个城市在人口统计学和经济状况上比较相似。
- Card 估计了一个 DiD 模型，但是没有发现此次外生冲击的影响。

Card 这个研究的两个特征：

1. 政策干预发生在总体层面，总体数据是可用的。但研究的问题是控制组的选择是特定情况下的和主观的。
2. 标准误差反映了抽样方差，而不是关于控制组能否再现感兴趣的反事实的不确定性。

Abadie 和 Gardeazabal (2003) 以及 Abadie、Diamond 和 Hainmueller (2010) 引入了合成控制估计器，作为同时解决这两个问题的一种方式。

Abadie 和 Gardeazabal (2003) 的方法使用捐赠池 (donor pool, 意思是用来合成伪处理组的样本) 中单位的加权平均值来模拟反事实。该方法基于这样一个观察：当分析单位是少数总体单位时，比较单位的组合 (“合成控制”) 通常比单独使用单个比较单位更好地再现受处理个体的特征。因此，这里的比较单位被选为加权平均值，这些比较单位在预处理期间最类似于受处理个体 (s) 的特征。

Abadie、Diamond 和 Hainmueller (2010) 认为这种方法比基于回归的方法有许多明显的优势。首先，该方法避免外推。它使用插值，因为估计的因果效应始终基于给定年份中某个结果与同一年份的反事实之间的比较。也就是说，它使用控制组个体的凸包作为反事实，因此反事实基于实际数据的位置，而不是在极端情况下回归可能发生的超出数据支持范围的外推 (King 和 Zeng 2006)。

Card 认为合成控制方法的几个优势：

1. 避免外推 (Extrapolation Avoidance): Abadie, Diamond, 和 Hainmueller (2010) 认为这种方法有许多独特的优势，首先，这种方法避免了外推。它使用的是插值，因为估计的因果效应始终基于同一年份中某个结果和一个反事实之间的比较。也就是说，它使用的反事实是基于控制组个体数据的实际位置，而不是像极端情况下的回归分析那样，进行数据支持范围之外的外推。
2. 数据处理优势 (Advantage in Data Processing): 第二个优势与数据处理有关。构建反事实不需要在研究的设计阶段访问后处理结果，与回归分析不同。这里的优势在于，它帮助研究者在指定模型时避免“偷看”结果。尽管仍需小心和诚实，因为在设计阶段查看结果与不查看结果一样容易，但关键在于，使用这种方法理论上可能只关注设计，而不是估计。

3. 权重的明确性 (**Explicitness of Weights**): 另一个优势, 这通常是人们反对研究的一个理由, 是所选择的权重明确了每个个体对反事实的贡献。现在这在很多方面都是一个明显的优势, 除非是在研讨会上为这些权重辩护时。因为有人可以看到爱达荷州在你的佛罗里达州模型中贡献了 0.3, 他们现在可以争辩说, 认为爱达荷州与佛罗里达州有任何相似之处是荒谬的。但与回归分析对比, 回归分析也对数据进行加权, 但是以一种盲目的方式。没有人反对回归分析产生的权重, 唯一的原因是他们看不到权重。它们是隐含的, 而不是明确的。因此, 我认为这种权重的明确产生是一个明显的优势, 因为它使合成控制比基于回归的设计更加透明 (即使这可能会需要与观众和读者进行你本来不会有的争论)。
4. 定性与定量研究的桥梁 (**Bridge Between Qualitative and Quantitative Types**): 第四个优势, 我认为通常被忽视, 是它架起了定性和定量研究者之间的桥梁。定性研究者经常是那些专注于详细描述单个个体, 如一个国家或一个监狱的人。他们通常是那些机构周围历史的专家。他们通常是首先进行比较案例研究的人。合成控制为他们提供了一个有价值的工具, 使他们能够选择反事实——原则上, 如果他们对评估某个特定干预措施感兴趣, 这个过程可以改善他们的工作。

2 合成控制方法

设 Y_{jt} 为 t 时刻 $J+1$ 个总体中我们感兴趣的 j 个个体的结果，处理组是 $j = 1$ 。合成控制估计模型使用选择出的最佳个体的线性组合作为合成控制。测量了在处理组上 T_0 时间干预产生的影响。对于干预发生后的时期，合成控制估计量测度的因果效应为 $Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$ ，其中 w_j^* 是最优选择权重向量。

选择匹配向量 X_1 和 X_0 作为干预后结果的预测变量，我们必须保证匹配变量不受干预影响。权重选择的目的是最小化受权重约束的形式， $\|X_1 - X_0 W\|$ 。这里有两个权重约束。首先，设 $W = (w_2, \dots, w_{J+1})'$ ，当 $j = 2, \dots, J+1$ 时， $w_j \geq 0$ 。其次，设 $w_2 + \dots + w_{J+1} = 1$ 。换句话说，没有任何个体可以得到负权重。但它们可以得到零权重。所有权重之和必须等于 1。

Abadie(2010) 考虑了：

$$\|X_1 - X_0W\| = \sqrt{(X_1 - X_0W)'V(X_1 - X_0W)}$$

式中， V 为某个 $(k \times k)$ 对称的半正定矩阵。设 X_{jm} 为个体 j 的第 m 个协变量的值。通常， V 是一个主对角线为 v_1, \dots, v_k 的对角矩阵。那么合成控制权重最小化了

$$\sum_{m=1}^k v_m \left(X_{1m} - \sum_{j=2}^{J+1} w_j X_{jm} \right)^2$$

式中， v_m 为一个权重，反映了当我们测量处理个体和合成个体之间的差异时，我们赋予给第 m 个变量的相对重要性。

可以看到， V 的选择是重要的，因为 W^* 依赖于人们对 V 的选择。合成控制 $W^*(V)$ 意味着在没有处理的情况下，为受处理个体重现其结果的行为。因此，权重 v_1, \dots, v_k 应反映协变量的预测值。

Abadie, Diamond 和 Hainmueller (2010 年) 建议了不同的 V 的选择，但最终从实践中看来，大多数人选择一个最小化均方预测误差最小的 V ：

$$\sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^*(V) Y_{jt} \right)^2$$

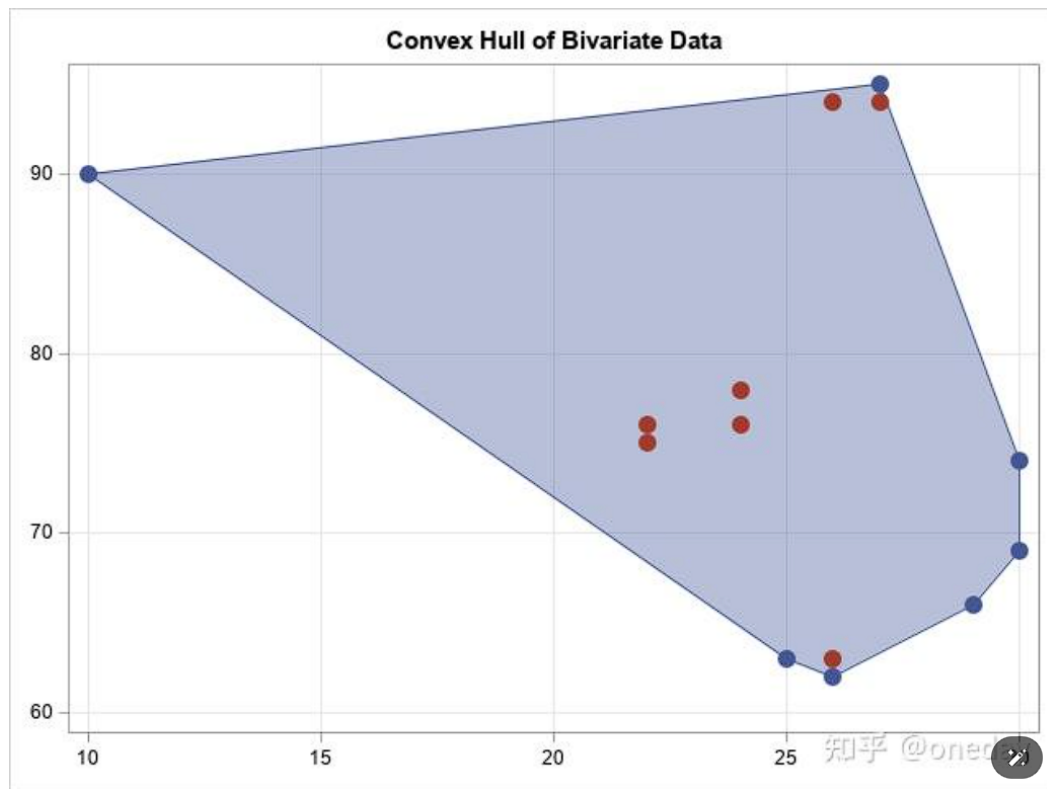
此外，还应对权重的取值范围施加约束。

这就是我们基于线性回归权重生成合成控制的原因，我们的权重结果有很多是属于外插值 (Extrapolation)。从技术角度而言，使用外插值并没有问题。但对于这里的因果推断或策略评估的场景，使用外插值就很危险，会使得我们的结果不可靠。也就是说，我们更希望我们的结果是内插值 **interpolation**，在我们能观察到的数据分布内部)。

这就是我们基于线性回归权重生成合成控制的原因，我们的权重结果有很多是属于外插值 (Extrapolation)。在社会科学里，使用外插值进行预测是件比较危险的事，一个经典的例子是人口增长预测，我们用人口较小时的增长率来预测多年以后的人口基数。同样，对这里的因果推断或策略评估的场景，会使得我们的结果不可靠。也就是说，我们更希望我们的结果是内插值（在我们能观察到的数据分布内部) 预测的结果。

我们可以通过如下方式来保证我们权重结果是内插的。具体而言，我们使用最优化中的受约束优化方法，我们限制 1) 我们捐赠池中的测试个体权重都为正，2) 其权重和为 1，在该约束下我们最小化试验组测试个体的结果变量与捐赠池内测试个体线性加权预测结果。相应地，我们的合成控制就变成优化中基于捐赠池 (donor pool) 中测试个体的凸组合问题 (convex combination)。

我们接下来用最优化的术语进行描述，简单而言，凸包 (convex hull) 是包含观测值的最小凸集 (convex set)，如下图是在给定双变量数据点生成的 convex hull。



$$\begin{aligned} \text{Minimize } \|X_1 - X_0W\| &= \left(\sum_{h=1}^k v_h \left(X_{h1} - \sum_{j=2}^{J+1} w_j Y_{jt} \right) \right) \\ \text{s.t. } \sum w_j &= 1 \\ w_i &\geq 0, \text{ for } j = 2, \dots, J \end{aligned}$$

那么未观察到的因素呢？比较案例研究因未测量的因素影响感兴趣的结果以及观察到的和未观察到的因素效应的异质性而变得复杂。**Abadie, Diamond** 和 **Hainmueller** (2010 年) 指出，如果数据中干预前的时期数量“很大”，那么匹配干预前的结果可以让我们控制对多个未观察到因素的异质性反应。这里的直觉是，只有在不可观察和可观察方面相似的单位才会在处理前遵循类似的轨迹。

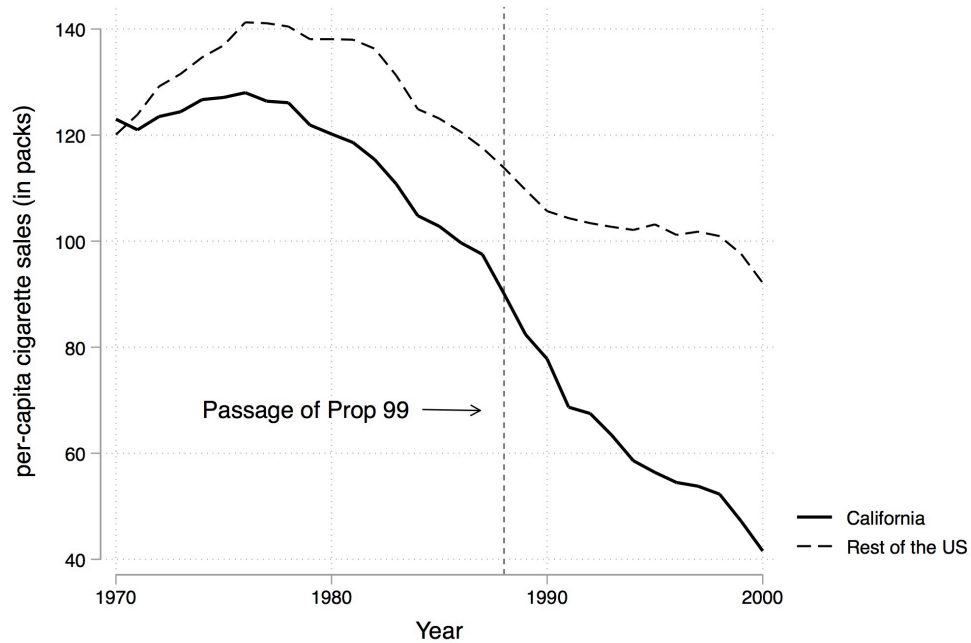
2.1 加利福利亚州《99 号提案》

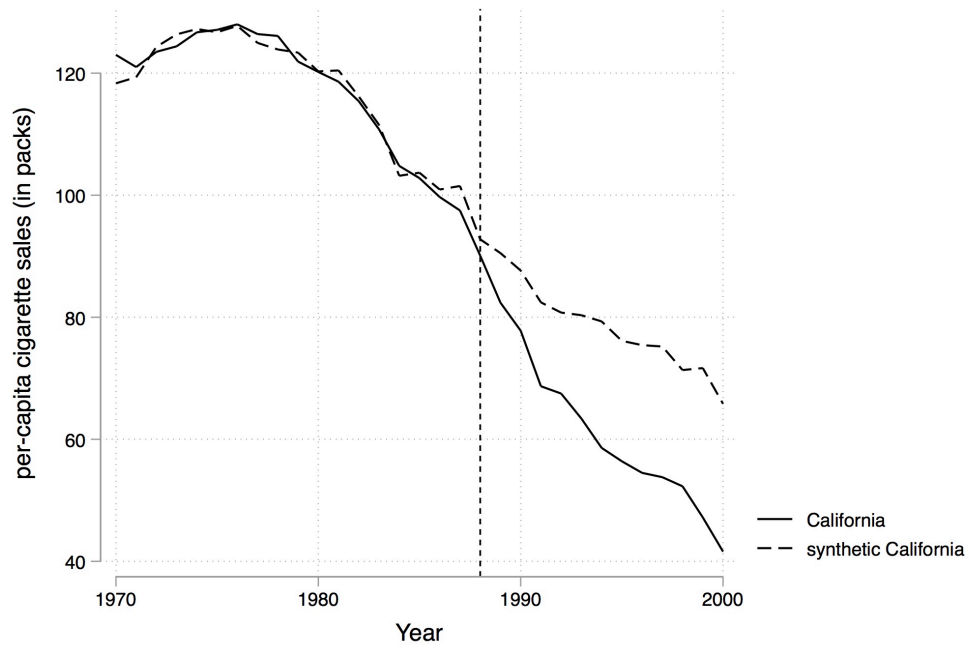
Abadie 和 Gardeazabal (2003) 开发了合成控制估计器，以评估恐怖主义对西班牙巴斯克地区的影响。但是，Abadie, Diamond, 和 Hainmueller (2010) 通过使用加利福利亚州的一项名为《99 号提案》的香烟税来进一步阐述这种方法。他们的例子使用基于安慰剂的方法进行推断，让我们更仔细地看看他们的论文。

1988 年，加利福利亚州通过了一项名为《99 号提案》的全面烟草控制立法。《99 号提案》将香烟税每包增加了 0.25 美元，推动了全州的清洁空气条例，资助了反吸烟媒体活动，将税收收入用于健康和反吸烟预算，并为反烟草项目产生了每年超过 1 亿美元的资金。其他州也有类似的控制计划，它们在分析中被排除了。

图 10.1 显示了从 1970 年到 2000 年，加利福利亚州和美国其他地区每年香烟销售的变化。可以看到，《99 号提案》通过后香烟销售下降，但它们已经在下降，所以不清楚是否有明显的影响——特别是因为它们在美国其他地区同时下降。

然而，使用他们的方法，他们选择了一组最优权重，当应用于美国其他地区时，产生了图 10.2 所示的图形。注意，在预处理期间，这组权重产生的加利福利亚州的时间路径与真实的加利福利亚州几乎相同，但在后处理期间两个系列出现了分歧。乍一看，该计划对香烟销售似乎有影响。





他们用于距离最小化的变量列在表 10.1 中。注意，这个分析是检验处理组和对照组之间的平衡性。用于匹配的变量在两组中非常相似，尤其是滞后值。

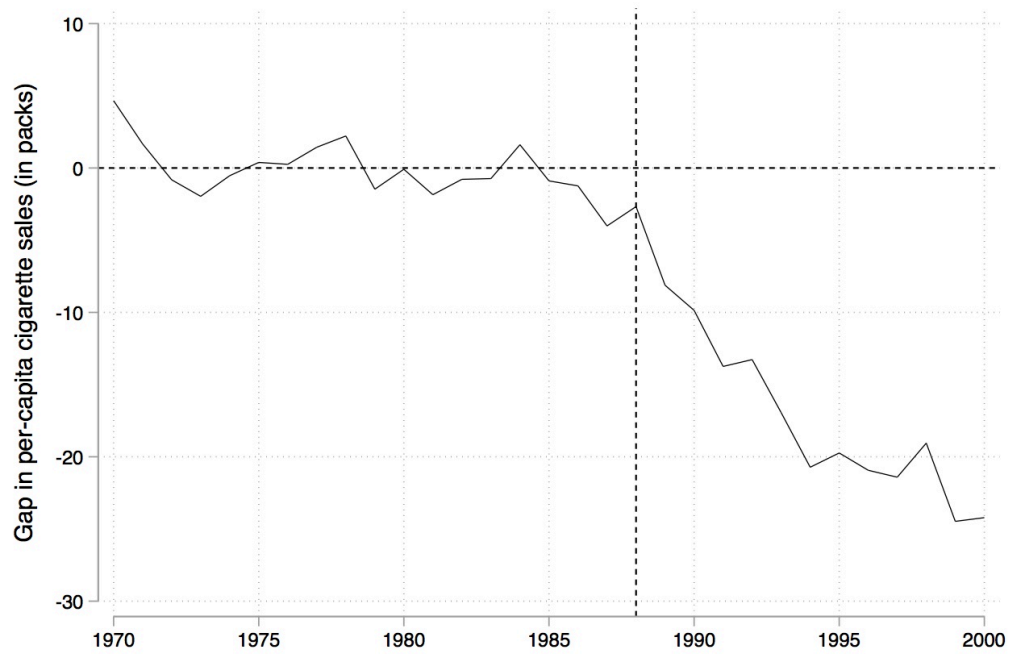
Table 10.1: Balance table



Variables	Real California	Synthetic Calif.	Avg. of 38 Control States
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15–24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

All variables except lagged cigarette sales are averaged for the 1980–1988 period. Beer consumption is averaged 1984–1988.

在使用合成控制方法的时候，你需要大量使用图片的形式展示你的估计结果。如果有因果效应，它们会在后处理期分叉，但在预处理期相似。因此，通常会看到一张图表，仅显示这两个系列之间的差异。



以上是使用合成控制方法进行估计的过程。接下来的问题是，如何确定观测到的两条线之间的差异是否具有统计上的显著性？

难点在于，每年只有两个观测值，无法使用常用的统计检验方法，如 T 检验。

Abadie, Diamond, 和 Hainmueller (2010) 建议我们使用一个经典的方法：构建基于 Fisher (1935) 的确切 P 值。随机化推断将处理分配给每个未处理的个体，重新计算模型的关键系数，并将它们收集到一个分布中，然后用于推断。Abadie, Diamond, 和 Hainmueller (2010) 建议我们计算一套预处理和后处理期间的均方根预测误差 (RMSPE) 值，作为用于推断的测试统计量。我们按以下步骤进行：

1. 将合成控制方法迭代应用于样本池中的每个个体，得到安慰剂效应的分布。
2. 计算每个安慰剂处理前阶段的 RMSPE:

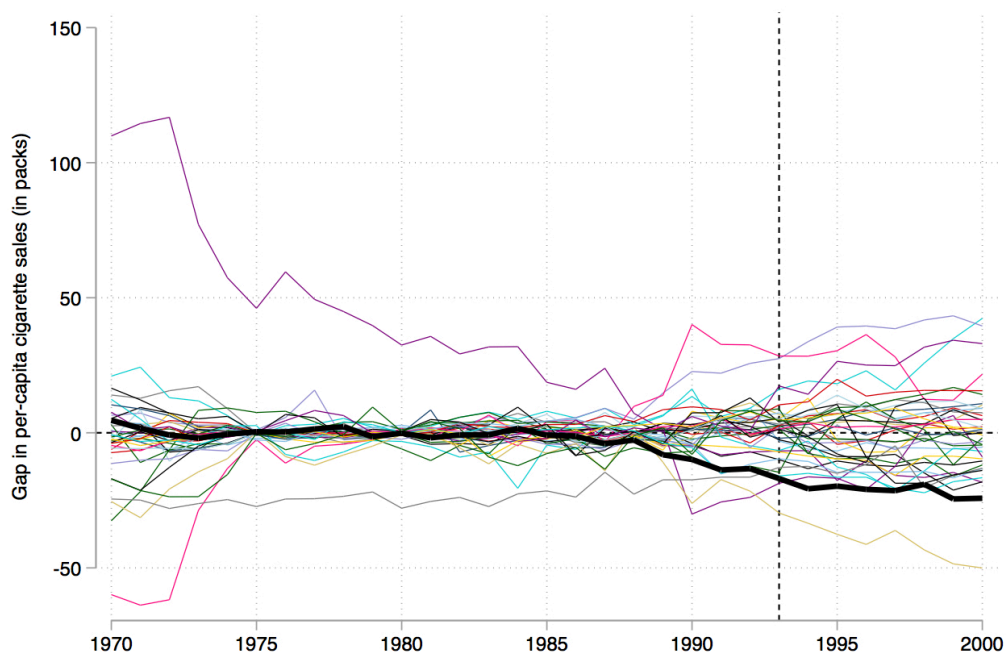
$$RMSPE = \left(\frac{1}{T - T_0} \sum_{t=T_0+t}^T \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \right)^2 \right)^{\frac{1}{2}}$$

3. 计算每个安慰剂的后处理期间的 RMSPE（类似公式，但用于后处理期）。
4. 计算后处理与预处理 RMSPE 的比率。
5. 按从大到小的顺序对这个比率进行排序。
6. 计算处理个体在分布中的比率。

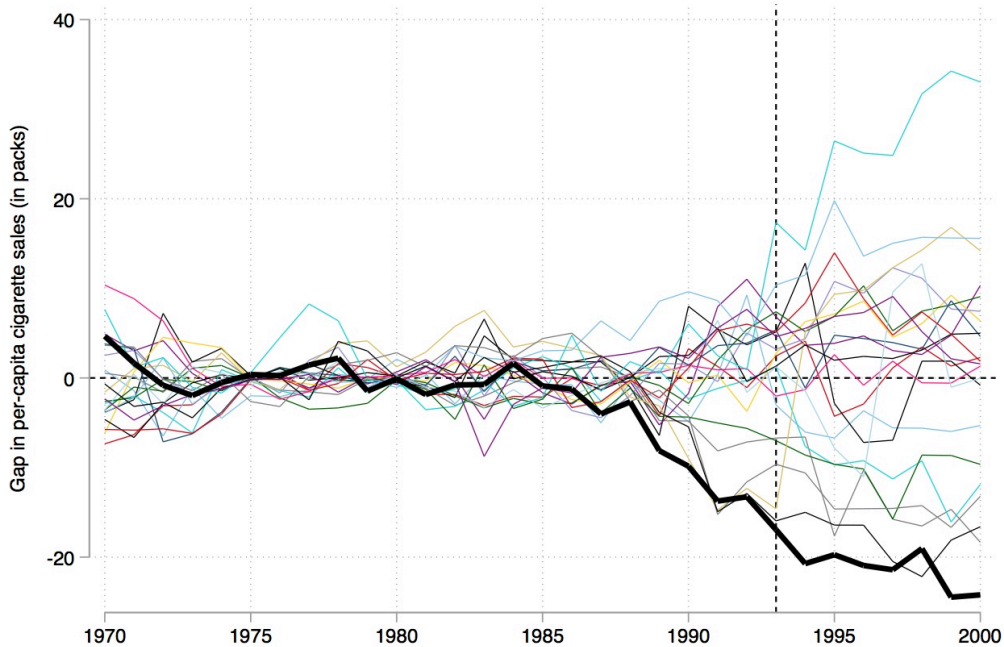
我们想知道的是，与捐赠池自己的安慰剂比率相比，加利福利亚州的处理效果是否极端。

展示结果：

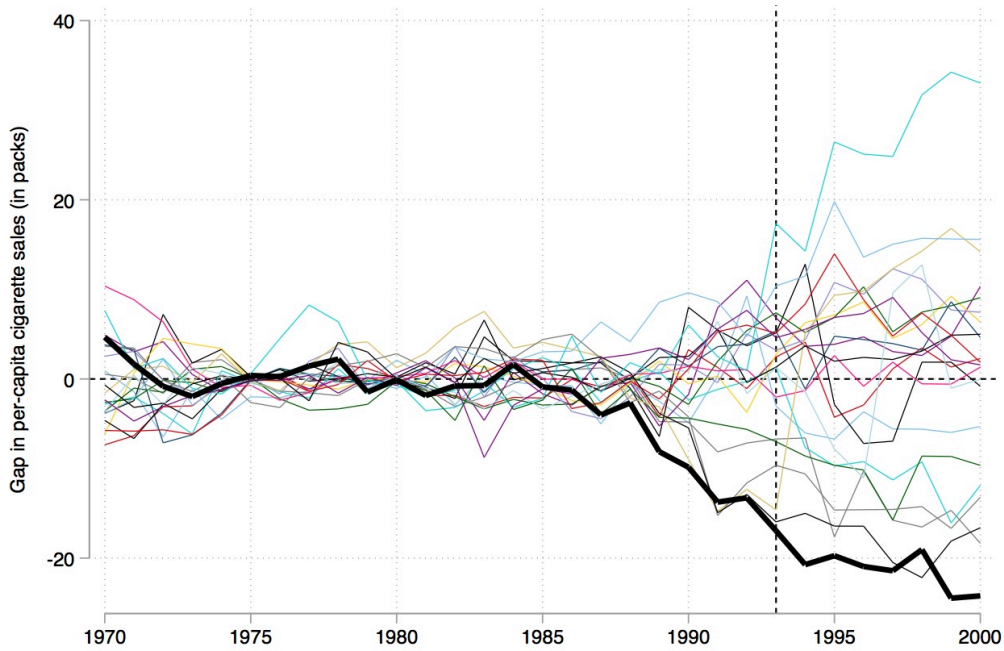
第一种方法是使用 STATA 的 `twoway` 命令将加利福尼亚结果与所有安慰剂结果画在一起，



Abadie, Diamond, 和 Hainmueller (2010) 建议反复排除那些预处理 RMSPE 与加利福尼亚相比明显不同的州。在这篇文章中，作者排除了任何预处理 RMSPE 超过加利福尼亚两倍的州单位，最后得到这个结果：



最终的推断基于这些确切的 P 值。所以我们这样做的方式是，我们简单地创建一个比率的直方图，或多或少地在分布中标记处理组，以便读者可以看到与模型相关联的确切 P 值。从图中可以看出，加利福尼亚在三十八个州个体中排名第一。确切的 P 值为 0.026，低于大多数期刊希望看到的 5% 的常规水平，以确定统计显著性。



2.2 证伪

在 Abadie, Diamond, 和 Hainmueller (2015) 的研究中，作者们研究了德国统一对国内生产总值的影响。然而，这篇论文的一个贡献引入证伪检验来论证合成控制估计方法的有效性。

1990 年德国的统一将东德和西德结合在一起，经过多年的分离，它们发展出了截然不同的文化、经济和政治体系。作者们对评估统一对经济产出的影响感兴趣，但正如在吸烟研究中一样，他们认为这些国家与任何一个单一国家都太不同，无法构成一个有说服力的比较组，因此他们使用合成控制来创建一个基于最优选择国家的复合比较组。

作者实施的做法是，将时间倒回到处理本身的日期，并在更早的（安慰剂）日期上估计他们的模型。由于安慰剂日期对产出应该没有任何影响，这提供了一些保证，即在 1990 年发现的任何偏差可能是由于统一本身引起的结构性断裂。结果表明，当使用 1975 年作为安慰剂日期时，他们没有发现任何影响，这表明他们的模型具有良好的样本内和样本外预测属性。

FIGURE 4 Placebo Reunification 1975–Trends in per Capita GDP: West Germany versus Synthetic West Germany

