

# 基于市场的减排制度

黎睿 2020201623

## 1 研究动机与研究问题

为了减少污染物排放，政府政策制定者往往有各种各样的工具可以选择。传统情况下，监管机构采取包括规定排放限值或污染控制技术标准的“命令和控制”（CAC）方法。然而，后来排放市场项目逐渐受到政策制定者们的喜爱。许多国家逐渐制定了各种各样的基于市场的排放制度，比如排放交易。这类总量管制与交易计划在应对气候变化的地区和联邦提案中占据了重要地位，吸引了很多学者的注意力。研究排放交易项目在实际情况中的作用自然而然成为了一个重要话题。而关于这一话题的研究无外乎两个角度，本文的研究也主要围绕这两个问题展开：一是这种基于市场的减排项目是否能相比于传统 CAC 制度达到更好的减排效果；二是这种依赖市场许可证的制度是否会导致环境不公正。

从直觉上来说，由于基于市场的方法花费的遵从成本更低，并且更灵活，因此可以实现更严格的减排目标，从而带来更大的公共健康和环境利益。至于环境公正性，有些人认为排放交易计划没有考虑到污染损害的分配，而根据 CAAAs 的许可可以明确考虑环境正义问题。如果污染设施能够购买许可证而不是减少排放，就有可能使被允许的污染流入穷人或少数民族人口居住的地区，从而加剧先前存在的环境风险分配方面的不平等。因此人们对这类项目的评价相差甚远。然而这些直觉都很难通过实证分析进行验证。

Meredith Fowlie 等人 (2012) 基于区域清洁空气激励市场 (RECLAIM)，使用双重差分计量模型和半参数匹配估计方法，识别了采取排放交易计划或传统的 CAC 政策对设备水平上的排放的因果效应，并通过检验不同社会经济特征导致的异质性探究 RECLAIM 是否会导致环境不公正。本文研究问题与原文基本一致，是对原文话题和数据集的机器学习方法尝试，即使用因果森林估计 RECLAIM 对于设备减排量的条件平均政策效应 (CATE) 并通过根据 1990 年社会人口经济特征划分样本来检验异质性。本文主要由两大部分组成，一是使用 R 语言对 Meredith Fowlie 等人的文章中部分数据描述和结果内容进行复现，二是用因果森林方法估计条件样本平均政策效应并检验异质性，两部分的具体 Rmarkdown 内容均存于附件中。

## 2 数据描述

本文所使用的数据主要包括污染性设备每年的排放量及其自身信息，和各设备所处街区水平上的人口普查数据。其中前者主要从加州空气资源委员会（ARB 1987-2005）统计的数据库中获取，其包括大约 10000 个污染性设备的年度排放量，及一些设备自身的信息，如所处地址、邮政编码、工业分类等。后者的获取方式为在每个设备周围构造不同长度的半径，确定街区组地理区域的百分比，结合该百分比和每个街区的人口普查数据，算出受到该设施影响的人口数。这些数据包括 1989 年和 1999 年的家庭收入中位数以及按种族和民族划分的人口，其中，非西班牙裔、黑人或西班牙裔人口的百分比用来衡量少数民族百分比。我们在后面的研究中强调 1990 年的社区人口特征数据而非 2000 年的数据，因为后者可能因为分类（加入 RECLAIM 计划或者依旧受 CAC 制度管制）而对排放存在内生性。

图 1 展示了在加利福尼亚的采取 RECLAIM（实验组）和持续采用 CAC 制度（对照组）的设备在 1990 年到 2005 年的总  $NO_x$  排放量变化趋势，显然，在实施 RECLAIM 计划之前（1990-1993 年），两组设施的排放量以相近的比例下降。在 RECLAIM 计划实施早期（1995-2000 年），实验组的设备采取 RECLAIM 计划， $NO_x$  排放量相对于对照组的设施略有上升。2000 年之后，实验组的排放量相对于对照组下降迅速。总体而言，与 1993 年之前水平相比，实验组的  $NO_x$  排放量下降了 72%，而对照组只下降了 62%。所以简单来看，在长期中，RECLAIM 计划能导致更好的减排效果。

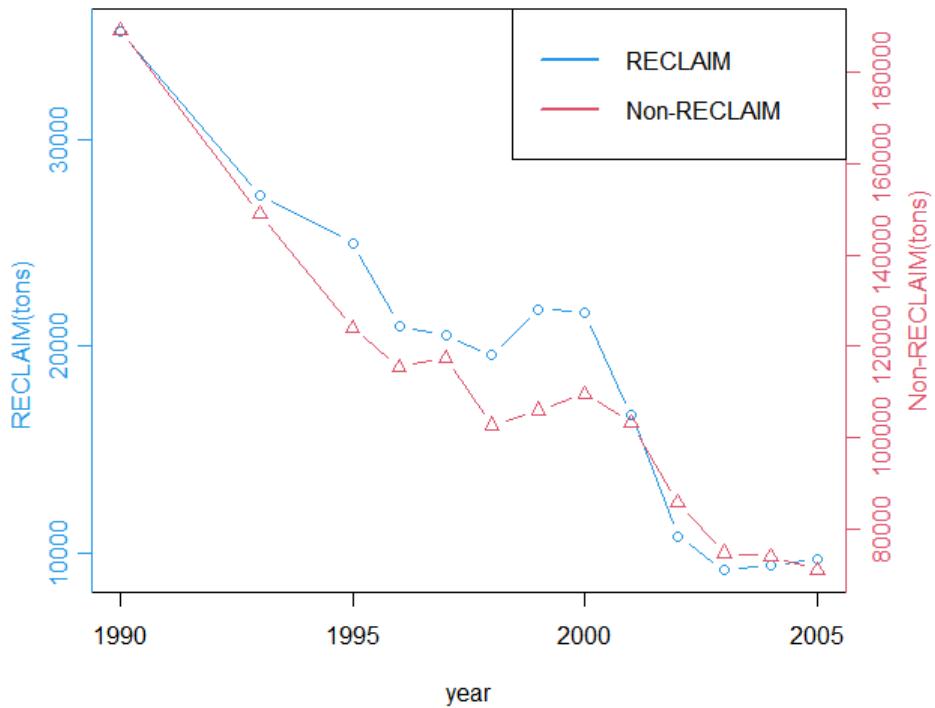


Figure. 1: RECLAIM 与非 RECLAIM 设备的  $NO_x$  排放总量

与 Meredith Fowlie 等人的文章一致，本文将所获取的面板数据切割为四个阶段，分别为第一阶段 1990–1993 年，第二阶段 1997–1998 年，第三阶段 2001–2002 年，第四阶段 2004–2005 年。表 1 则简要描述了按是否采取 RECLAIM 计划的两组设备在不同阶段的排放数据的平衡面板<sup>1</sup>。显然，在第一阶段还未引入 RECLAIM 计划时，实验组和对照组的平均排放量相似，而在 RECLAIM 计划引进之后，采取 RECLAIM 计划的实验组  $NO_x$  排放量相对于继续采用 CAC 制度的对照组的设备们下降得更快。具体地，参与 RECLAIM 的设备平均排放下降大约 90%，而对照组仅下降了大约 50%。

阶段	实验组 (RECLAIM)			对照组 (Non-RECLAIM)			总和		
	样本数	均值	标准差	样本数	均值	标准差	样本数	均值	标准差
阶段一 (1990-1993)	213	101.8	304.4	1052	102.8	430.5	1265	102.6	411.9
阶段二 (1997-1998)	213	62.7	179.8	1052	80	371	1265	77.1	346.3
阶段三 (2001-2002)	213	43.8	125.4	1052	67.9	339.6	1265	63.8	314
阶段四 (2004-2005)	213	30.8	117.1	1052	53	290.8	1265	49.2	269.6

Table. 1:  $NO_x$  排放数据的描述性统计

### 3 传统计量模型

之前对 RECLAIM 计划的排放影响的估计是有条件的，并且高度敏感于一个存在很大争议的假设，即在没有 RECLAIM 的情况下排放会是多少。在 Meredith Fowlie 等人的研究中，他们构建了更合理和透明的“反事实”（即假设未加入 RECLAIM 这些设备会排放多少）排放估计。他们没有依赖于事先预期还未加入 RECLAIM 设备的总排放轨迹，而是使用计量经济学中的时候观察调整后，在同一时期受 CAC 监管的设施的排放，接下来，我将简单介绍一下他们的实证框架和识别策略。

#### 3.1 实证框架

基于目前在项目评估文献中已成为标准的潜在结果框架，Meredith Fowlie 等人假设有两个监管州可以将加利福尼亚州的工业  $NO_x$  排放设备分配到这两个州，即实施以市场为基础的 RECLAIM 计划的州或在 SCAQMD 以外的未达标地区<sup>2</sup> 普遍存在的 CAC 制度的州。其中前者为实验组，后者为对照组（控制组），并且假设所有的设备在加入 RECLAIM 计划

<sup>1</sup>该数据包括了所有满足平衡面板要求的设备样本

<sup>2</sup>实际情况中，California 内只有未达标县的部分工业设备采取 RECLAIM 计划

之前都从属于 CAC。根据潜在产出模型，作者设置了如下变量：

政策变量 (treat) $D_i$   $i \in 1, 2, \dots, n$ :

$$D_i = \begin{cases} 1 & \text{设备 } i \text{ 属于实验组} \\ 0 & \text{设备 } i \text{ 属于对照组} \end{cases} \quad (3.1)$$

潜在排放变量 (potential outcomes) $Y_{it}(D_i)$ ,  $D_i = 1, 0$ :

$$\begin{aligned} Y_{it}(1) &: \text{设备 } i \text{ 属于实验组的潜在年度排放} \\ Y_{it}(0) &: \text{设备 } i \text{ 属于对照组的潜在年度排放} \end{aligned} \quad (3.2)$$

Meredith Fowlie 等人主要关注的是估计实验组的样本平均政策效应，即如下条件平均政策效应 (treatment effect)(SATT):

$$\alpha_{TT} = E[Y_{it'}(1) - Y_{it'}(0)|D_i = 1] \quad (3.3)$$

其中， $E[Y_{it'}(0)|D_i = 1]$  无法利用实验组的观测值直接估计，作者提议使用控制组在同一时间段内的观测值间接估计，再通过条件 DID 模型识别政策效应。为了避免个体选择性偏误，该文使用半参数匹配估计 (semiparametric matching estimators )DID 模型，即使用广义 DID 匹配估计量估计政策效应，具体公式如下：

$$Y_{it'} - Y_{it^0} = \beta' X_i + \alpha D_i + \varepsilon_i \quad (3.4)$$

$$\widehat{\alpha_{DID}} = \frac{1}{N_1} \sum_{j \in I_1} \{(Y_{jt'}(1) - Y_{jt^0}(1)) - \sum_{k \in I_0} w_{jk} (Y_{kt'}(0) - Y_{kt^0}(0))\} \quad (3.5)$$

其中， $t'$  表示引入 RECLAIM 计划后一年的时刻， $t^0$  表示引入 RECLAIM 前的时刻，假设  $\varepsilon$  与设备特征向量  $X$  和政策变量  $D$  都独立； $I_1$  和  $I_0$  分别表示加入 RECLAIM 的设备集和留在 CAC 的设备集， $N_1$  是实验组的设备个数，实验组的个体用下标  $j$  表示，对照组的个体用下标  $k$  表示  $Y_{jt'}(1) - Y_{jt^0}(1)$  为实验组的排放量改变量， $\sum_{k \in I_0} w_{jk} (Y_{kt'}(0) - Y_{kt^0}(0))$  为基于所有对照组数据估计的设备  $j$ (属于实验组) 若继续保持 CAC 的同期潜在  $NO_x$  排放量改变量，其中  $w_{jk}$  是基于设备  $j$  与设备  $k$  的相似度（可由设备特征变量  $X$  用最近邻匹配算法求出）算得的权重。

## 3.2 识别假设

为了说明如上政策效应估计量的合理性，Meredith Fowlie 等人做出了如下一系列识别假设。

1. 条件不混淆 (conditional unconfoundedness):  $Y_{it}(0) \perp D_i | X_i$ ，即在给定设备信息和周围环境信息时，控制组设备和实验组设备采取 CAC 政策遵循一样的排放轨迹。如果这个假设得以满足，只要对可观测的差异进行调整，就可以将观测结果的差异解释为采取 RECLAIM 相对于采取 CAC 的政策效应。

2. 实验组和对照组的设备的特征变量  $X$  的分布的支集存在重叠。
3. 稳定单位政策假设 (SUTVA):  $(Y_{it}(0), Y_{it}(1)) \perp D_j \quad i \neq j$ 。该假设用于排除溢出效应和一般均衡效应。

其中假设 2 可以通过直接观察实验组和对照组中协变量的联合分布来验证，而假设 1 和假设 3 则在原则上无法通过实证进行检验，但 Meredith Fowlie 等人还是通过间接测试评估了这些假设的合理性。由于文章篇幅有限，本文并不详细复现原文对识别假设的检验过程，只展示主要问题的估计结果。

### 3.3 政策效应异质性

除了 RECLAIM 制度对设施排放的平均政策效应，Meredith Fowlie 等人还关注在具有不同社会经济特征的社区中，不同设施的政策效应是否存在系统性的差异，即 RECLAIM 对设备排放的效应是否关于设备所处社区的社会经济特征存在异质性。因此他们估计了如下回归模型：

$$Y_{it'} - Y_{it^0} = \delta_j + \beta' X_i + \theta' X_i D_i + \alpha D_i + \varepsilon_i \quad (3.6)$$

其中  $\delta_j$  是组固定效应，组  $j$  包含属于实验组的设备  $j$  以及它的  $m_j$  个最近的对照组中的“匹配”。该模型与更标准的基于回归的策略的不同之处在于，观察结果在匹配时是加权的。为了调查 RECLAIM 相对于 CAC 法规在何种程度上加剧（或减轻）了环境不公现象，协变量  $X$  中包含了社会经济和人口统计变量。

## 4 传统模型结论

Meredith Fowlie 等人用前文所述计量模型估计了样本平均政策效应，并进行了一系列的稳健性检查和安慰剂检验。然后她们在政策效应评估中就其对社会人口特征的异质性进行了检验，并讨论了选择性问题。与他们一致，本文感兴趣的主要是一些不同时期设施水平的年度  $NO_x$  排放量的变化。主要报告使用水平值和取对数转换数据生成的结果，取对数的情况下 SATT 可以理解为对平均政策效果的百分比估计。

注意到有 14 家生产者在 2001 年（第三阶段）离开了 RECLAIM 市场，但是之后又重新加入了该市场。因此在长期政策效应角度，作者分析了第一阶段和第四阶段之间设备水平的排放变化，而在评估在许可证价格飙升周围的关键窗口期间的排放变化（即第二阶段和第三阶段之间的排放变化）时，比较了不包含这 14 家生产者和包含这 14 家生产者后政策效应的差异。

### 4.1 政策效应估计

非参数最近邻 (NN) 匹配估计器使用最接近实验个体的控制组中的观测值构造每个实验个体的“反事实”估计，传统非参数 NN 为每个实验个体选择  $m$  个最近的“邻居”，并对每

个被选择的邻居设置  $w_{jk} = 1/m$  的权重。Meredith Fowlie 等人加强了重叠条件，即只有与实验个体属于同一行业的控制组的设施才有机会参与与该实验个体的匹配，并且要求所有设施位于臭氧不达标地区。并且在将实验个体与  $m$  个最近的邻居匹配后，使用协变量  $X$  上的控制结果的参数回归调整对内差异。Meredith Fowlie 等人主要采取了两种最近邻匹配具体规格：第一个是基础规格，只考虑臭氧是否达标，SIC 编码，历史  $NO_x$  排放（阶段一）三个协变量；第二个替代规格则除了这三个基础协变量，还包括种族和社会人口特征等协变量。均采取标准的匹配到三个最近邻居的方法。

具体估计结果如表 2 所示，其中面板 A 估计了 RECLAIM 对阶段一到阶段四的  $NO_x$  排放量变化的政策效应，即排放的长期总体变化，忽略了有 14 个设备在 2001 年退出了 RECLAIM 这一现象；而面板 B 则估计了阶段二到阶段三的  $NO_x$  排放量变化的政策效应，此时作者将这 14 个设备视为对照组中个体。并且所有的观测样本都来自历史上的臭氧排放非达标县。

	水平值	取对数	实验组设备	对照组设备
<b>面板 A：阶段 1 到阶段 4 的 <math>NO_x</math> 排放变化量</b>				
<b>OLS</b>	-32.58** (13.77)	-0.30*** (0.10)	212	1222
<b>最近邻估计 (基础规格)</b>	-20.59*** (7.63)	-0.25*** (0.09)	212	1222
<b>最近邻估计 (替代规格)</b>	-18.12 (11.51)	-0.11 (0.08)	211	1191
<b>最近邻估计 (受限制的样本集)</b>	-14.16** (6.86)	-0.20** (0.09)	199	1222
<b>面板 B：阶段 2 到阶段 3 的 <math>NO_x</math> 排放变化量</b>				
<b>OLS</b>	-6.84 (6.65)	-0.22*** (0.04)	255	1577
<b>最近邻估计 (基础规格)</b>	-8.29** (3.85)	-0.26*** (0.06)	255	1577
<b>最近邻估计 (替代规格)</b>	-6.18 (5.06)	-0.16*** (0.06)	252	1493
<b>最近邻估计 (受限制的样本集)</b>	-6.37 (4.57)	-0.23*** (0.06)	268	1577

<sup>1</sup> \*\*\* 在 1% 水平上显著； \*\* 在 5% 水平上显著； \* 在 10% 水平上显著。

<sup>2</sup> OLS 估计量控制了第一阶段的  $NO_x$  排放量和四字节的 SIC 码哑变量，并且估计聚类到大气限域的标准误。

<sup>3</sup> 基础规格的 NNM 估计基于历史排放和四字节的 SIC 码，替代规格还包括了特定行业的排放四分位数指标和预先确定的人口特征（种族和收入）。

Table. 2: 平均政策效应的最近邻估计

对于排放的长期总体变化，广义 DID 匹配估计结果为每年减少 20.59 吨  $NO_X$  排放，这一结果在 5% 的水平上显著。这相当于实验组的设备在第 1 阶段平均年排放量的 20%。使用对数转换过的排放数据重新估计，估计系数为 -0.25，这意味着实验组中的设备与匹配的控制组的设备相比，多减排了约 25%。对于许可证价格变化剧烈的第二阶段与第三阶段的排放差的研究显示，NN 估计值为 -8.29，并且在 5% 的显著性水平上显著，这意味着实验组的设备在第二阶段的平均排放量减少了 12%。使用对数转换过的排放数据重新估计，估计系数为 -0.26，即 RECLAIM 使设备多减排 25%。

总之，无论是在长期的 15 年时间里（第一阶段到第四阶段的 1990 年至 2005 年），还是在实验组的减排速度超过对照组的减排速度的交叉点周围的时段时，RECLAIM 的政策效应都显著为负。并且在对排放数据进行对数变换依然不改变估计结果的显著性。

## 4.2 政策效应异质性估计

Meredith Fowlie 等人通过估计式 (3.6) 的  $\theta$  来检验政策效应是否对历史排放、所处社区人口平均收入、少数民族聚集比例等协变量存在异质性，来分析 RECLAIM 计划是否会导致环境不公正，即 RECLAIM 计划是否会导致 SCAQMD 的传统弱势社区和其他社区遭受不同的排放变化。表 3 即为估计结果，其中面板 A 分析所有样本第一阶段到第四阶段的  $NO_X$  排放差，面板 B 将那 14 个特殊的设备视为对照组中个体，研究了第二阶段到第三阶段的  $NO_X$  排放差。

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<b>面板 A：阶段 1 到阶段 4 的 NOX 排放变化量</b>							
RECLAIM	-20.64** (7.81)	-20.38* (8.85)	-17.49** (6.17)	-20.46** (7.41)	-18.52** (7.04)	-15.26*** (4.36)	-17.71** (5.29)
RECLAIM*	-0.19			-0.19	-0.19		-0.18
阶段 1NOX 排放	(0.11)			(0.11)	(0.11)		(0.11)
RECLAIM* 收入		-1.27 (0.96)		-0.65 (1.09)		0.42 (1.95)	-0.02 (1.53)
RECLAIM*			0.94		0.43	1.04	0.41
少数民族比例			(0.60)		(0.36)	(0.96)	(0.51)
阶段 1NOX 排放	-0.48*** (0.11)	-0.49** (0.15)	-0.49** (0.15)	-0.48*** (0.11)	-0.48*** (0.11)	-0.49** (0.14)	-0.48*** (0.11)
收入		0.10 (0.80)		0.16 (0.74)		-0.66 (1.47)	-0.24 (1.04)
少数民族比例			0.35 (0.31)		-0.22 (0.26)	-0.52 (0.56)	-0.28 (0.37)
R2	0.87	0.85	0.85	0.87	0.87	0.85	0.87

面板 B: 阶段 2 到阶段 3 的 NOX 排放变化量							
<b>RECLAIM</b>	-6.70*** (1.43)	-7.19** (2.22)	-6.29*** (1.35)	-7.16*** (1.45)	-6.62*** (1.25)	-6.45*** (1.85)	-7.05*** (1.23)
<b>RECLAIM*</b>	-0.06***			-0.07***	-0.07***		-0.07***
阶段 1NOX 排放	(0.02)			(0.02)	(0.02)		(0.02)
<b>RECLAIM*</b>		-0.16		-0.09		-0.12	-0.22
收入		(0.24)		(0.17)		(0.36)	(0.35)
<b>RECLAIM*</b>			0.09* (0.04)		-0.004 (0.045)	0.05 (0.11)	-0.07 (0.14)
阶段 1NOX 排放	-0.35*** (0.08)	-0.34*** (0.05)	-0.34*** (0.05)	-0.34*** (0.08)	-0.34*** (0.08)	-0.34*** (0.06)	-0.34*** (0.08)
收入		0.19 (0.36)		0.16 (0.33)		0.05 (0.47)	0.15 (0.46)
少数民族比例			-0.11 (0.07)		-0.05 (0.06)	-0.10 (0.11)	-0.02 (0.11)
<b>R2</b>	0.52	0.47	0.47	0.49	0.49	0.47	0.49

<sup>1</sup> \*\*\* 在 1% 水平上显著; \*\* 在 5% 水平上显著; \* 在 10% 水平上显著; ' 不显著

<sup>2</sup> 本结果由基础规格 NNM 估计得到, 对于 1990 年人口数据的回归分析, 面板 A 和面板 B 分别有 875 个和 1043 个观察值。

<sup>3</sup> 没有显示群体固定效应。

<sup>4</sup> 处理过的观察值的权重为  $1m_j$ , 其中  $m_j$  是实验组中设备  $j$  匹配得到的控制组的规模。

<sup>5</sup> 少数民族比例是黑人或西班牙裔人口的百分比。

Table. 3: 环境公正性评估 (NNM 估计)

结果显示, RECLAIM 组中设施在阶段一排放的更多, 则减排效果更好, 但表中所有的  $\theta$  估计均不显著, 即 Meredith Fowlie 等人并没有找到 1990 年人口统计数据中存在影响设施减排的重要决定性因素。因此未检测出任何社区人口特征导致的异质性, 没有找到环境不公正的证据。

## 5 因果森林估计结果

无论是经济学领域还是统计学领域, 对观察性研究中异质性治疗方法的估计都是相当热门的一个话题, 前文简要展示了 Meredith Fowlie 等人结合 DID 与 NNM 识别 RECLAIM 对加州污染性设备减排效果的政策效应及其异质性的过程, 接下来本文将进一步对这一话题进行机器学习的尝试。本文采用的是 Athey 等人在 2014 年 JASA 的文章《Estimation and inference of heterogeneous treatment effects using random forests》一文中提出的, 基于决策树和随机森林的因果森林 (Causal forests) 方法。

本文选择这一机器学习方法进行因果分析主要出于两点考虑，一是该方法本质上是一个非参数的方法，与传统非参数方法，比如前文所用的临近匹配、基于 kernel 的方法等相比，该方法具有一个很好的优势，即该方法没有“维数诅咒”的问题，而传统方法当变量的维数增加时效果会大打折扣。而本文所要考虑的变量维度较高，因此使用因果森林可以避免对巨大数据量的要求。二是已有许多学者通过模拟实验发现因果森林实质上比基于最近邻匹配的经典方法更强大（Susan Athey 等，2014），而 Meredith Fowlie 等人使用的正是 NNM 方法。

显然，本文使用的数据集为观测性数据而非随机试验获得的数据，因此在这组数据中存在一定的选择性。如图 2 所示，其中左边是研究阶段一到阶段四中 RECLAIM 政策效应的数据集在不同 1993 年（即还未加入 RECLAIM 之前） $NO_X$  排放程度下在未来受 RECLAIM 计划约束的倾向分数。显然，1993 年的  $NO_X$  排放越高的设备，越可能在未来“选择”受 RECLAIM 计划约束。右边则是研究阶段一到阶段四中 RECLAIM 政策效应的数据集，情况与左图类似。因此，为了识别因果效应，本文假设条件不混淆（conditional unconfoundedness），即在给定协变量条件时，潜在产出联合分布与是否受 RECLAIM 计划约束相互独立，该识别假设 Meredith Fowlie 等人也有提出。

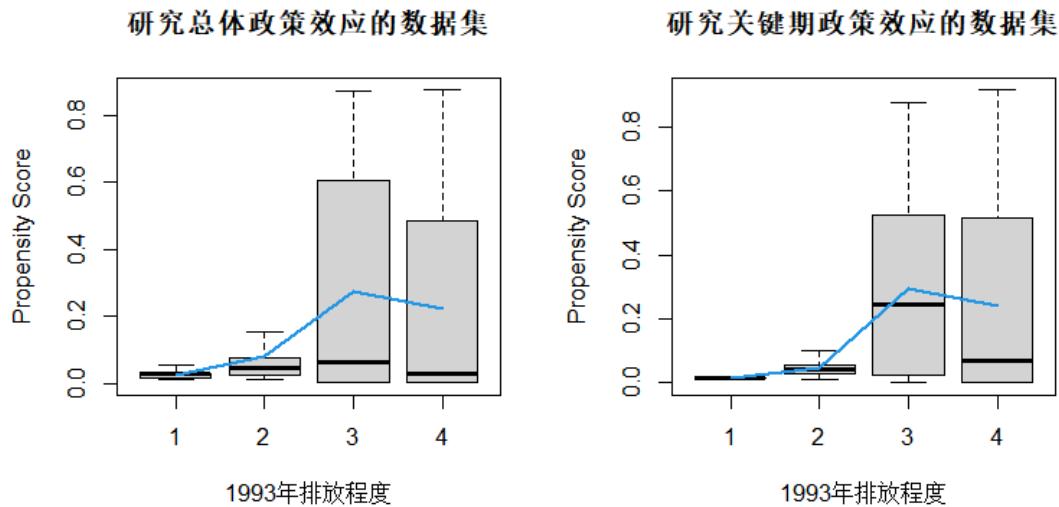


Figure. 2: 关于原始  $NO_X$  排放量的选择性问题

具体估计结果如表 4 所示，其中面板 A 估计了在不同社会人口特征约束条件下 RECLAIM 对阶段一到阶段四的  $NO_X$  排放量变化的政策效应，ATE 即所有样本的平均政策效应，其估计值显著为负，为  $-31.444$ 。而在施加其他社会人口特征约束后，除了原始排放规模较少，即 1993 年  $NO_X$  排放量少的子样本的 RECLAIM 政策效应显著为正外，其余子样本的政策效应估计量均在 10% 显著性水平下为负。对于许可证价格变化剧烈的第二阶段与第三阶段的排放差的研究显示，所有的子样本的平均政策效应均在 10% 显著性水平上显著为负。因此可以认为受 RECLAIM 计划约束的设施的减排效果超过了受 CAC 监管的非常相似的设施在同一时期内实现的减排效果。注意到按减排效果高低和原始排放规模分组的政

<sup>2</sup>1993 年  $NO_X$  排放程度 1 ~ 4 即为 1993 年  $NO_X$  排放量数据分别在 25%，50%，75%，100% 的分位数。

策效应估计数值极端，并且十分显著，说明原始排放规模对未来排放规模具有很强的解释作用，这与 Meredith Fowlie 等人的估计一致，但由于其不属于社区人口特征，因此不对其展开分析。

针对环境公正性问题，由于本文采取的因果森林估计方法相比于 NNM 方法可以避免维度灾难，所以相对于 Meredith Fowlie 等人的研究，我将更多的社会人口特征纳入异质性的考虑范畴，包括家庭收入中位数、少数民族占比、平均就业率、平均工资率、基础建设率。注意到 1990 年的社会人口特征数据中的少数民族占比以及就业率对阶段一到阶段四的总体政策效应和对阶段二到阶段三的关键时期政策效应都会导致异质性。即少数民族占比越高，就业率越低，RECLAIM 计划的政策效应可能越差。但这并不能证明前文所述的环境不公正性的存在，因为这只能说 RECLAIM 计划相对于 CAC 监管的效果在贫穷落后的地区较差，但子样本估计结果显示即使在贫穷落后的地区，RECLAIM 计划相对于 CAC 监管的政策效应都显著为负，并不存在所谓的基于市场的制度会让污染流入贫困落后的地区，使之环境更加恶化。事实上，表 4 中所有按社区人口特征（收入、民族、就业、工资、建设）划分的子样本的政策效应估计值都在 10% 的显著性水平上为负，即无论是高收入还是低收入地区，无论是少数民族聚集还是非聚集...，RECLAIM 项目都对这些地区有相对 CAC 制度更高效的减排作用。此外，注意到在面板 B 的关键时期政策效应估计中，家庭收入中位数较低地区的政策效应估计量为 -12.453，而家庭收入中位数较高地区仅有 -2.168，即低收入地区在许可证价格迅速变化的关键时期受 RECLAIM 约束导致的减排效果更好，这更与环境不公正的观点相悖。

图 3 则通过箱形图可视化了不同家庭收入中位数和不同少数民族占比这两个典型特征，也是 Meredith Fowlie 等人主要关注的社会人口特征变量导致的 RECLAIM 政策效应异质性。显然少数民族占比越高，说明地区越落后，条件政策效应估计值越大，其绝对值越小，即政策效应越弱。而条件政策效应虽平均收入的变化与直觉相反，收入越高条件政策效应更弱。但不难发现，箱形图主体部分均位于 0 值以下，即几乎所有子样本的条件政策效应的因果森林估计均为负，即 RECLAIM 计划可以使绝大部分的地区的污染排放下降。

		估计值	异质性检验 p 值	样本数
面板 A：阶段 1 到阶段 4 的 $NO_x$ 排放变化量				
ATE		-31.444***	8.72e-3**	1428
CATE-减排效果	高	-62.403***	2.2e-16***	714
	低	-0.485***		714
CATE-原始排放规模	大	-63.544***	2.2e-16***	714
	小	0.656***		714
CATE-家庭收入	高	-28.346***	0.266	713
	低	-34.533***		715
CATE-少数民族占比	低	-55.341***	2.135e-7***	714
	高	-7.547***		714

CATE-就业率	高	-44.386***	6.776e-3**	674
	低	-19.876.		754
CATE-工资	高	-22.243.		647
	低	-39.066.	0.892	781
CATE-基础建设	好	-31.558.		491
	差	-31.384.	0.4412	937
CATE-高收入非少数民族聚集地		-36.744.		489
CATE-低收入少数民族聚集地		-6.419.	0.0104.	490
CATE-极度富裕地区		-65.575.		204
CATE-极度落后地区		-8.272.	1.124e-4***	386
面板 B: 阶段 2 到阶段 3 的 $NO_X$ 排放变化量				
ATE		-7.311**	0.0570.	1752
CATE-减排效果	高	-13.115**		876
	低	-1.507**	2.2e-16***	876
CATE-原始排放规模	大	-13.495**		875
	小	-1.141**	2.2e-16***	877
CATE-家庭收入	高	-2.168**		876
	低	-12.453**	2.711e-5***	876
CATE-少数民族占比	低	-9.57**		876
	高	-5.051**	1.817e-6***	876
CATE-就业率	高	-10.23*		823
	低	-4.725.	0.0494*	929
CATE-工资	高	-10.486.		858
	低	-4.263.	0.416	894
CATE-基础建设	好	-15.183.		643
	差	-2.746.	0.146	1109
CATE-高收入非少数民族聚集地		-1.718.		615
CATE-低收入少数民族聚集地		-5.824.	0.705	615
CATE-极度富裕地区		-21.429.		390
CATE-极度落后地区		-3.987.	2.901e-5***	548

<sup>1</sup> \*\*\* 在 0.1% 水平上显著; \*\* 在 1% 水平上显著; \* 在 5% 水平上显著; . 在 10% 水平上显著; . 不显著

<sup>2</sup> ATE 的异质性检验使用 Omnibus test; 其余均使用 t 检验。

<sup>3</sup> 极度富裕地区指高工资率、高就业率、基础建设较好、少数民族占比低的地区; 反之, 极度贫困地区指低工资率、低就业率、基础建设较差、少数民族占比高的地区。

<sup>4</sup> 面板 A 使用了所有样本, 面板 B 则剔除了那 14 个在 2001 年退出 RECLAIM 的设备。

<sup>5</sup> 原始排放规模指设备在 1993 年的排放量。

Table. 4: 环境公正性评估 (因果森林估计)

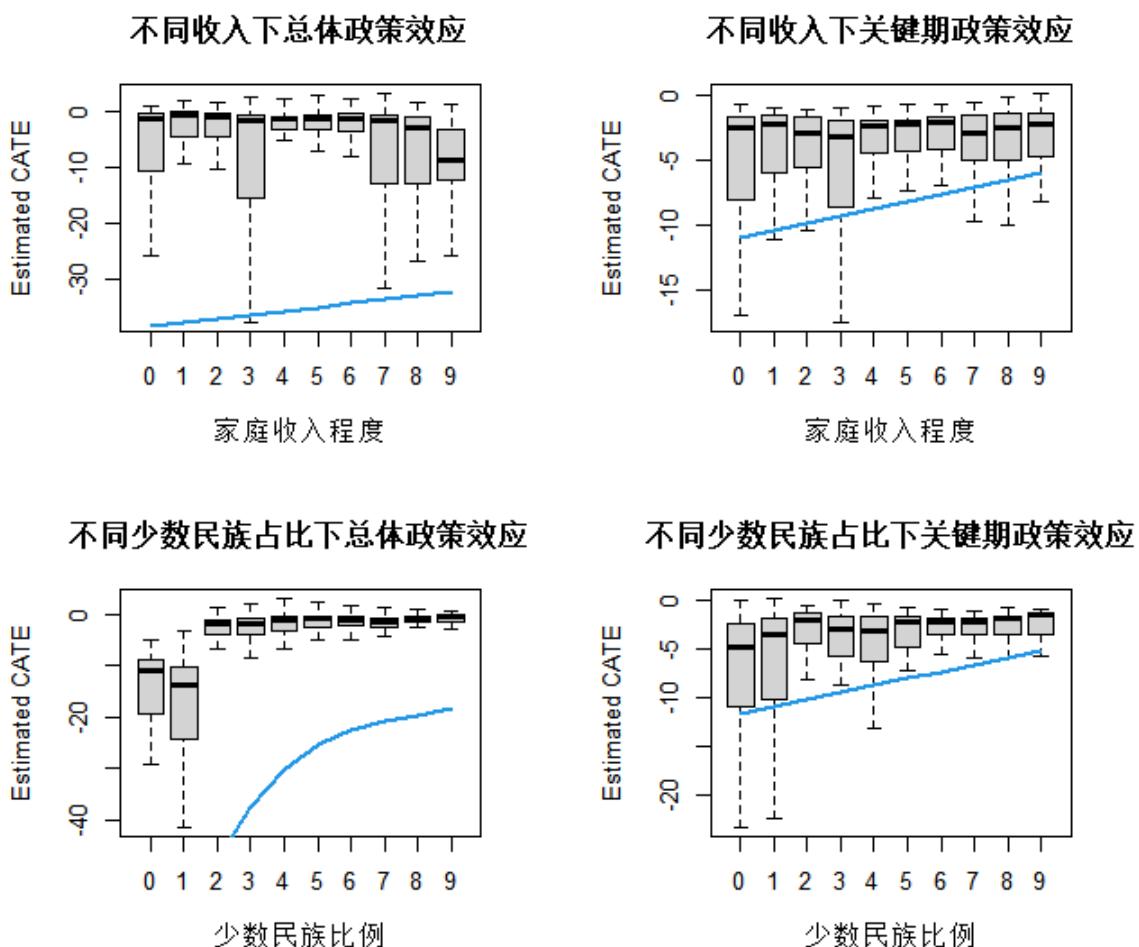


Figure. 3: 社会人口特征异质性可视化 - 节选

为了验证结论的稳健性，本文在表 5 和图 4 中给出对主要变量进行对数转换后的因果森林估计结果。结果显示受 RECLAIM 计划约束相较于受 CAC 制度管制在开始的十年内能减少大约 27%  $NO_x$  排放。所有的估计结果方向与显著性均与表 4 结果相似。

<sup>2</sup>横坐标的家庭收入程度和少数民族比例 1 ~ 9 分别为 1990 年家庭收入中位数数据和少数民族比例数据在 10%, 20%, 30%...90% 的分位数。

		估计值	异质性检验 p 值	样本数
面板 A: 阶段 1 到阶段 4 的 $NO_X$ 排放变化率				
ATE		-0.27***	6.105e-6**	1428
CATE-减排效果	高	-0.173***		714
	低	-0.118***	2.2e-16***	714
CATE-原始排放规模	大	-0.164***		714
	小	0.109***	2.2e-16***	714
CATE-家庭收入对数	高	-0.043***		713
	低	-0.011***	0.72	715
CATE-少数民族占比/100	低	-0.101***		714
	高	-0.047***	0.090.	714
CATE-就业率/100	高	-0.088***		674
	低	-0.13.	6.205e-3***	754
CATE-工资率/100	高	-0.014.		647
	低	-0.061.	2.187e-10***	781
CATE-基础建设率/100	好	-0.185.		491
	差	-0.139.	2.2e-16***	937
CATE-高收入非少数民族聚集地		-0.117.		489
CATE-低收入少数民族聚集地		-0.092.	0.2024	490
CATE-极度富裕地区		-0.12.		204
CATE-极度落后地区		-0.001.	1.502e-6***	386
面板 B: 阶段 2 到阶段 3 的 $NO_X$ 排放变化率				
ATE		-0.153***	2.2e-16***	1752
CATE-减排效果	高	-0.12***		876
	低	-0.187***	2.2e-16***	876
CATE-原始排放规模	大	-0.175***		875
	小	-0.131***	2.2e-16***	877
CATE-家庭收入对数	高	-0.178***		876
	低	-0.178***	2.2e-16***	876
CATE-少数民族占比/100	低	-0.136***		876
	高	-0.171***	2.2e-16***	876

CATE-就业率/100	高	-0.095.	2.2e-16***	823
	低	-0.205**		929
CATE-工资率/100	高	-0.095**	2.2e-16***	858
	低	-0.21**		894
CATE-基础建设率/100	好	-0.079**	2.2e-16***	643
	差	-0.196**		1109
CATE-高收入非少数民族聚集地		-0.131**		615
CATE-低收入少数民族聚集地		-0.12**	2.2e-16***	615
CATE-极度富裕地区		-0.11**		390
CATE-极度落后地区		-0.238**	2.2e-16***	548

<sup>1</sup> \*\*\*\*, 在 0.1% 水平上显著; \*\*, 在 1% 水平上显著; \*, 在 5% 水平上显著; , 在 10% 水平上显著; , 不显著

<sup>2</sup> ATE 的异质性检验使用 Omnibus test; 其余均使用 t 检验。

<sup>3</sup> 极度富裕地区指高工资率、高就业率、基础建设较好、少数民族占比低的地区; 反之, 极度贫困地区指低工资率、低就业率、基础建设较差、少数民族占比高的地区。

<sup>4</sup> 面板 A 使用了所有样本, 面板 B 则剔除了那 14 个在 2001 年退出 RECLAIM 的设备。

<sup>5</sup> 原始排放规模指设备在 1993 年的排放量对数。

Table. 5: 对数变换后环境公正性评估 (因果森林估计)

## 6 结果比较与分析

本文同 Meredith Fowlie 等人一致, 通过挖掘 RECLAIM 计划的一些独特特点, 为两个重要经济环境问题的研究提供新的实证证据。问题 1 即, 受 RECLAIM 计划约束的设施的减排效果是否超过了受 CAC 监管的非常相似的设施在同一时期内实现的减排? 问题 2 即, 以市场为基础的环境法规所提供的合规灵活性, 是否在传统的弱势社区造成了更多(或更少)的污染?

对于第一个问题, Meredith Fowlie 等人的结果表明, 在 RECLAIM 项目的前十年, 相对于控制设施, 实验设施(加入 RECLAIM 计划)的排放量平均下降了大约 20%。这些结果对其他匹配策略是稳健的。在许可证价格大幅波动的时期, 政策效应强度相对较小。在此期间, 14 家发电厂被从该项目中移除。当这些设施被排除在分析之外时, 他们同样发现实验组设施的排放量相对于非常相似的控制组设施减少 7% 左右。而本文采取机器学习方法估计的结果对 RECLAIM 项目的总体估计政策效应要强于 Meredith Fowlie 等人的结果, 因果森林估计量显示相对于控制设施, 实验设施(加入 RECLAIM 计划)的排放量平均下降了大约 30%。对许可证价格大幅波动的时期的平均政策效应估计与 Meredith Fowlie 等人结果相近, 均为 7% 左右。

对于第二个问题, Meredith Fowlie 等人没有找到实证证据说明 RECLAIM 计划对设施

水平排放的估计相对影响随着社区人口特征的变化而有系统地变化。特别的是，他们发现他们估计的效果与社区对收入或少数族裔百分比的衡量没有相关性。因此他们认为 RECLAIM 计划不会使任何种族或收入群体经历显著的排放增加。而本文采取的机器学习方法研究出的结果虽然也未能找到支持环境不公正理论的证据，但却发现少数民族占比和平均就业率等社会人口特征变量会系统性地影响 RECLAIM 计划对设施减排的作用。即少数民族占比越高，就业率越低，RECLAIM 计划的政策效应可能越差。此外还发现家庭收入中位数的增加与 RECLAIM 计划效果成反比，这似乎不可思议，难以用经济学直觉解释，但却存在统计学的意义。

因果森林方法与 Meredith Fowlie 等人的研究在社区人口特征异质性的分析上产生了矛盾，这可能是因为 Meredith Fowlie 等人在进行异质性分析过程中采取的是半参数分析框架，即只通过估计政策选择虚拟变量和社区人口特征的交互项来判断异质性的存在，而因果森林估计量作为无参估计量避免了设置变量具体形式（Meredith Fowlie 等人设置的交互项形式）从而减少了偏误。又因为随机森林很大程度上避免了一般非参数方法面临的最大问题——“维度灾难”，因此基于随机森林的因果森林也较一般的非参方法更优越。因此从方法角度来看，因果森林的估计结果应该比广义 DID 最近邻匹配的估计结果准确。此外，从经济学直觉角度来说，同一政策对不同地区的影响效果存在较大差异更符合直觉，更落后的地区往往因为缺乏理念知识、物质经费等，在改善环境问题上面临的阻碍应远远大于更发达的地区，从而抑制环境政策效果。因此本文认为，因果森林的估计结果更合理。

## 7 结论

本文的最终结论与 Meredith Fowlie 等人结论相似，解决了两个排放交易项目在实际情况中的这一热门话题中的两个重要问题，即基于市场的减排项目是否比传统 CAC 制度更有效，以及这种项目是否会导致环境不公正。本文利用因果森林的机器学习方法分析了这两个问题，结果显示，RECLAIM 项目相对于 CAC 制度在 RECLAIM 项目的前十年的总体政策效应和在许可证价格迅速变化的 2001 ~ 2005 年内的政策效应都显著为负，后者估计结果绝对值要小于前者估计结果，即基于市场的减排项目确实比传统 CAC 制度更有效，并且长期稳定地起着作用。并且，本文使用因果森林估计 CATE 从而分析异质性发现 RECLAIM 计划对设施水平排放的估计相对影响随着社区人口特征的变化而有系统地变化，但这种异质性并不能说明环境不公正。因为虽然 RECLAIM 项目相对于 CAC 制度在低就业率的少数民族聚集地的减排促进效果要低于在高就业率的非少数民族聚集地。但在无论是高就业率地区还是低就业率地区，无论是少数民族聚集地还是非少数民族聚集地，用因果森林估计的 RECLAIM 项目的 CATE 均显著为负，即在弱势社区受 RECLAIM 计划约束仍比受 CAC 管制减少更多的  $NO_X$  排放。

## 参考文献

- [1] Alberto Abadie and Guldo W. Imbens. Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235 – 267, 2006.
- [2] Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7353–7360, 2016.
- [3] Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7353 – 7360, 2016.
- [4] Carol Joyce Blumberg. Causal inference for statistics, social, and biomedical sciences: An introduction. *International Statistical Review*, 84(1):159, 2016.
- [5] Dallas Burtraw, David A. Evans, Alan Krupnick, Karen Palmer, and Russell Toth. Economics of pollution trading for so<sub>2</sub> and nox. *Annual Review of Environment & Resources*, 30(1):253 – 289, 2005.
- [6] Péter Elek and Anikó Bíró. Regional differences in diabetes across europe – regression and causal forest analyses. *Economics & Human Biology*, 40:100948, 2021.
- [7] Meredith Fowlie, Stephen P. Holland, and Erin T. Mansur. What do emissions markets deliver and to whom? evidence from southern california’s nox trading program. *American Economic Review*, 102(2):965–93, April 2012.
- [8] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228 – 1242, 2018.