

新北市垃圾費隨袋徵收效果-

以 Synthetic Control Method 檢視

410673120 經濟三 邱奕勳

摘要

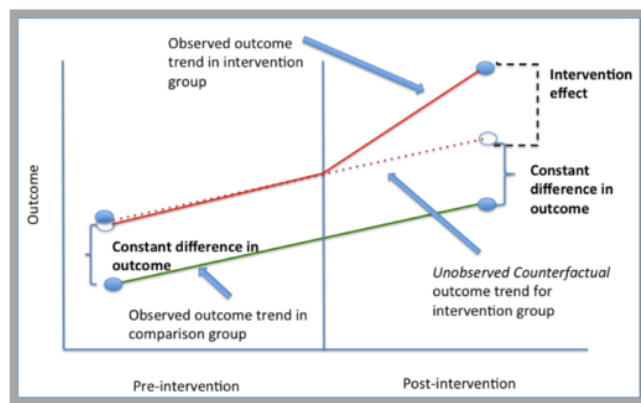
新北市政府自 2008 年開始部分區域試辦、2010 年 12 月全面實施垃圾費隨袋徵收，此報告將以「平均每人每日一般廢棄物產生量 (公斤)」為依變數，探討此政策是否有效降低垃圾產生量。研究方法先採用 Difference in Difference(DID)模型設計，從台灣縣市中找到控制組，估計新北市於政策實施後若未接受實驗的結果，並由此推估政策實施後之效果。然而，DID 模型設計需要符合 Parallel Trend Assumption，亦即實驗前須有相近的依變數走勢，但是怎樣才算符合 Parallel Trend Assumption 並不明確，因此本報告使用 Synthetic Control Method(SCM)，透過特徵合成一個近似於實驗前新北市之合成新北市，其實驗前的走勢與新北市更相符，以之作為控制組，作為最終政策效果之探討。

前言

為了探討政策實施之單獨效果，需要去除其他可能變數對依變數之影響，以本報告主題為例，除了垃圾費隨袋徵收政策會影響依變數外，其他變數如：環保政策宣導、商品普遍減少包裝...等變數都有可能影響依變數，因此需要採行模型設計去除其他變數對依變數的影響，以下分別採 DID 模型以及 SCM，分別探討政策效果，並比較兩研究方法之差異。

研究方法- Difference in Difference

DID 模型有一個強烈的假設: Parallel Trend Assumption-在政策實施前，控制組與實驗組需有相同的走勢，也就是說政策前需要有相同的依變數差距(constant difference in outcome)，因為若此假設成立，控制組與實驗組的依變數受其他變數的影響應相同，只是水準不同而已，因此實驗前的依變數差距需相同。



圖（一） Parallel Trend Assumption

PC: <https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation>

在 Parallel Trend Assumption 成立下，控制組則可以做為預估的反事實論證實驗組，以控制組估計若實驗組沒有政策干擾下的依變數結果，將實驗組事前事後的依變數差距，扣除控制組事前事後的依變數差距，藉此排除其他變數對依變數的影響，即為所謂的差距中差距。

控制組的事前事後差距來自於政策效果外的其他變數效果，因此藉由合理的控制組，透過差異中差異排除其他變數的效果，藉此估計政策的單一效果。

		Pre	Post	
Treatment		20	25	5
Control		18	22	4
				1

圖（二） Difference in Difference

DID- 探討新北市垃圾隨袋徵收效果

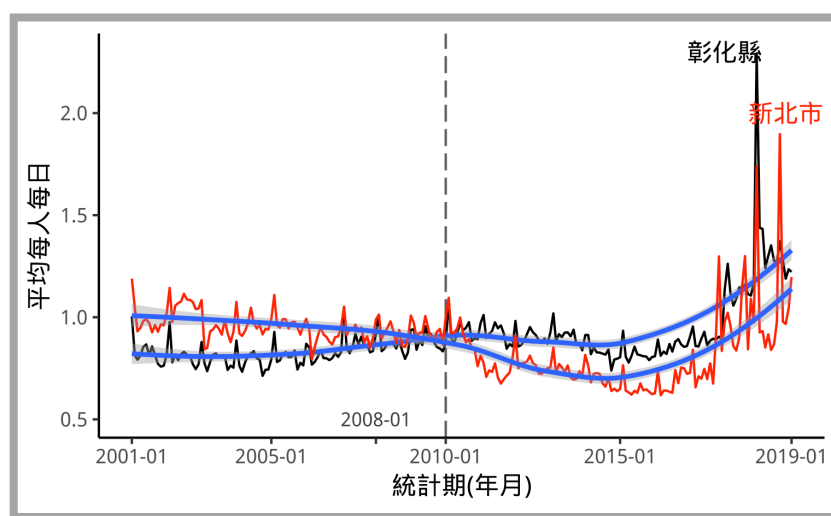
新北市於 2010 年 12 月全面實施垃圾費隨袋徵收效果，因此需要從台灣其他縣市找到一控制組，並且此控制組的政策前依變數的走勢與新北市相近，即最符合 Parallel Trend Assumption，本報告定義最符合 Constant Differences in Outcome 的縣市為控制組，因此透過計算政策前新北市與台灣其他縣市的依變數差額之標準差，並取標準差最小，也就是最符合 Constant Differences in Outcome 的縣市最為控制組。

如圖（三）所示，為新北市與台灣縣市依變數差額之標準差，排序第一為標準差最小，可發現彰化縣與新北市的依變數差距標準差最小，也就是彰化縣與新北市在政策效果前的平均每人每日一般廢棄物產生量（公斤）的差距在台灣縣市中變化幅度最小，最符合 Constant Differences in Outcome。因此依據此標準，選定彰化縣為 DID 模型下的控制組。

新北市 排序					
臺北市	0.2413775	18	南投縣	0.1757619	6
桃園市	0.1866363	11	雲林縣	0.1904631	13
臺中市	0.1638206	2	嘉義縣	0.1909153	14
臺南市	0.1848710	10	屏東縣	0.1807588	9
高雄市	0.1678436	5	臺東縣	0.2248391	16
宜蘭縣	0.1662487	4	花蓮縣	0.2703260	19
新竹縣	0.1805937	8	澎湖縣	0.2097177	15
苗栗縣	0.1891941	12	基隆市	0.1777304	7
彰化縣	0.1592426	1	新竹市	0.2368473	17
			嘉義市	0.1656918	3

圖（三） 臺灣縣市與新北市依變數差距標準差排序

如圖（四）所示，在政策效果(2010-01)前，彰化縣與新北市的走勢相近，唯自 2008 年起，新北市與彰化縣差距已逐漸縮小，其可能原因為新北市於 2008 年已經部分實施垃圾費隨袋徵收。



圖（四） 新北市與彰化縣走勢比較

DID-簡單回歸（實驗組：新北市；控制組：彰化縣）

Y = `平均每人每日一般廢棄物產生量（公斤）`

$treat$ = dummy variable; 如果統計區為新北市為 1, 否則 0

$post$ = dummy variable; 如果統計期大於 2010-12-01 為 1, 否則 0

$effect$ = dummy variable; 如果統計區為新北市且統計期大於 2010-12-01 為 1, 否則 0

$$Y = \beta_0 + \beta_1 treat + \beta_2 post + \beta_3 effect$$

effect 即為本報告所捕捉的政策效果，如圖(五)所示，effect 效果大約為 -0.280，可解讀為新北市垃圾費隨袋徵收政策可降低`平均每人每日一般廢棄物產生量（公斤）`0.28 公斤（約為 10.370 個 600cc 寶特瓶空瓶隻重量），且 p 值小於 0.01，為顯著效果。

Dependent variable:	
`平均每人每日一般廢棄物產生量（公斤）`	
treat	0.114*** (0.019)
post	0.124*** (0.020)
effect	-0.280*** (0.028)
Constant	0.839*** (0.013)

Observations	456
R2	0.190
Adjusted R2	0.185
Residual Std. Error	0.147 (df = 452)
F Statistic	35.354*** (df = 3; 452)
=====	
Note:	*p<0.1; **p<0.05; ***p<0.01

圖（五） DID-簡單回歸結果

DID-思考

此 DID 模型以彰化縣為控制組，然而，彰化縣並非完美的符合 Parallel Trend Assumption，彰化縣與新北市在政策效果前的依變數的差距標準差大約為 0.159，偏離於完美 Parallel Trend Assumption 下的差距標準差 0，因此彰化縣無法完整的預期「新北市如果未接受政策效果下的依變數」，而導致 DID 模型下的政策效果估計有誤，因此本報告進而思考如何找到更適合的控制組。

研究方法-Synthetic Control Method

起源：

西班牙邊境的一個地區-Basque，在 1970-1990 年代不斷受到恐怖組織 ETA 的武裝恐怖攻擊，使得 Basque 的經濟發展受到負面衝擊，在 1970 年代初期，恐怖攻擊尚緩時，Basque 地區為西班牙境內人均 GDP 排名第三富有的國家，而在經過將近三十年的恐怖組織與政治爭鬥後，在 1990 年後期，Basque 地區的人均 GDP 跌落自第六名，而兩名 Basque 出生的經濟學家 Alberto Abadie 以及 Javier Gardeazabal 想要探究政治以及恐怖組織對於 Basque 地區的影響，然而西班牙其他地區的特徵與 Basque 相差極大使得 Basque 無法與其他單一地區比較，因此兩位經濟學家在“The Economic Costs of Conflict: A Case Study of the Basque Country.” American Economic Review, 93(1), 112–132 研究中從西班牙其他區域組合特徵與 Basque 相近的合成控制組，模擬 Basque 如果沒有受到干擾-政治與恐怖組織迫害時的反事實論政 Basque，以此估計干擾的效果，結果發現 Basque 的人均 GDP 因為干擾而下降了 10%。

SCM-模型解釋

觀察單位 $j=1, \dots, J+1$ for 期間 $t=1, \dots, T$

實驗組： $j=1$ ；控制組(Possible donor)： $j=2, \dots, J+1$

干擾時間： $t=T_0+1$ ；干擾前： $t=1, \dots, T_0$ ；干擾後： $t=T_0+1, \dots, T$

Y_{it}^N ：單位 i 在在 t 時刻沒有接收實驗，觀察到的結果

Y_{it}^I ：單位 i 在在 t 時刻有接收實驗，觀察到的結果

探討之效果： $\alpha_{1t} = Y_{1t}^I - Y_{1t}^N$ for $t = T_0 + 1, \dots, T$

效果即為實驗組在政策效果實施後有接收到實驗的結果扣除實驗組在政策效果實施後沒有接收到實驗的結果，然而 Y_{1t}^N 為一反事實結果，無法被觀察，因此需要找到一個估計量，而如同前述起源提及，SCM 透過特徵值形成一合成控制組，因此在此先定義特徵。

$U_i: (\gamma \times 1)$ 觀察之共變量; 為一特徵

$(T_0 \times 1)$ 向量 $K = (k_1, \dots, k_{T_0})'$ 以 K 權重建立干擾前 ($t < T_0$) 的線性組合結果

$$Y_1^K = \sum_{s=1}^{T_0} k_s Y_{is}; \text{也作為一特徵,}$$

進而透過特徵找到給予所有可能控制組 (donor pool) 的權重 W ,

$$W = (w_2, \dots, w_{J+1}) \text{ such that } w_j \geq 0 \text{ for } j = 2, \dots, J+1 \text{ and } \sum_{j=2}^{J+1} w_j = 1$$

$$W^* \text{ such that } 1) \sum_{j=2}^{J+1} W_j^* Y_j^{K_1} = Y_1^{K_1} \dots \sum_{j=2}^{J+1} W_j^* Y_j^{K_M} = Y_1^{K_M}; 2) \sum_{j=2}^{J+1} W^* U_j = U_1$$

最佳化的權重需要符合

1) 以 W^* 組成之合成控制組的干擾前 K 權重線性組合之結果等於實驗組干擾前 K 權重線性組合之結果的結果。

2) 以 W^* 組成之合成控制組的特徵等於實驗組之特徵。

因此以 W^* 組成合成控制組作為實驗組如果沒有接受干擾下的估計;

而得出估計干擾效果量: $\widehat{\alpha}_{1t}$

$$\widehat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} W^* Y_{jt} \text{ for } t = T_0 + 1, \dots, T$$

而選定的權重需要極小化實驗組與合成控制組特徵值的差異,

選定 W^* that solves $\|X_1 - X_0 W\|_v = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$

$(k \times 1)$ 向量 $X_1 = (U_1', Y_1^{K_1}, \dots, Y_1^{K_M})'$;

$(k \times J)$ 矩陣 X_0 = 每個控制組的特徵項;

V 是 $(k \times k)$ 對稱的正半定矩陣,

給予 X_0 與 X_1 不同的權重以極小化 $(Y_1 - Y_0 W^*)' (Y_1 - Y_0 W^*)$,

以上述方法找到最適的權重 W^* 形成合成控制組, 便可以合成控制組作為實驗組

如果沒有接受干擾下的估計, 進而估計實驗組的干擾效果量。

SCM-R synth 套件實作

依變數:

平均每人每日一般廢棄物產生量 (公斤)

Predictor (U):

平均每戶消費支出,

失業率,

平均每戶人數,

平均每人可支配所得,

平均每人每日一般廢棄物產生量 (公斤)

特徵值參考[“A Review on Factors affecting Municipal Solid Waste Generation”](#)以及[“Prediction of municipal solid waste generation using artificial neural network approach enhanced by structural break analysis”](#)，選定與依變數相關之變數作為特徵項。

1) 使用 `dataprep()` 產生準備資料型態以符合 synth 套件

Argument 解釋

`foo`: 資料，本報告使用年資料(2001-2018)，變數為前述提及之依變數及特徵項以及統計期、統計區。

`predictors.op`: 用來計算特徵值的方法。

`times.predictors.prior`: 政策效果前的時間區間。

`unit.variable`: 每個單位(unit)的編號，如：高雄市為 1，花蓮縣為 2。

`unit.names.variable`: 單位的變數名稱

`time.variable`: 時間變數名稱

`treatment.identifier`: 實驗組編號

`controls.identifier`: 控制組編號

`time.optimize.ssr`: 極小化特徵值差異所計算的區間

`time.plot`: 畫圖所包含的區間

```
```{r}
trashprep.out <- dataprep(
 foo= trash_by_year,
 predictors = c("平均每戶消費支出","失業率","平均每戶人數","平均每人可支配所得","平均每人每日一般廢棄物產生量(公斤)"),
 predictors.op = "mean",
 time.predictors.prior = 2001:2010,
 dependent = "平均每人每日一般廢棄物產生量(公斤)",
 unit.variable = "regionno",
 unit.names.variable = "統計區",
 time.variable = "統計期",
 treatment.identifier = 15,
 controls.identifier = c(1:14,16:20),
 time.optimize.ssr = 2001:2010,
 time.plot = 2001:2018
)
```

圖（六） `dataprep()`-資料整理

2) 使用 Synth() 進行最佳化計算，

$$\text{即最小化} \|X_1 - X_0 W\|_v = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$$

```

{r}
synth.out = synth(data.prep.obj = trashprep.out)

```

圖（七） Synth()-最佳化計算

## SCM 結果探討

### 權重

得出最佳化的  $W^*$ ，如圖（八），可得知部分縣市-嘉義縣、臺北市未納入合成控制組。

w.weights	unit.names	unit.numbers
0.008	高雄市	1
0.010	花蓮縣	2
0.297	基隆市	3
0.007	嘉義市	4
0.000	嘉義縣	5
0.015	苗栗縣	6
0.019	南投縣	7
0.007	澎湖縣	8
0.009	屏東縣	9

0.000	臺北市	10
0.010	臺東縣	11
0.013	臺南市	12
0.177	臺中市	13
0.009	桃園市	14
0.275	新竹市	16
0.102	新竹縣	17
0.015	宜蘭縣	18
0.015	雲林縣	19
0.013	彰化縣	20

圖（八） 合成控制組權重

### 特徵比較

由圖（九）可得知合成控制組的特徵值與新北市極度相近

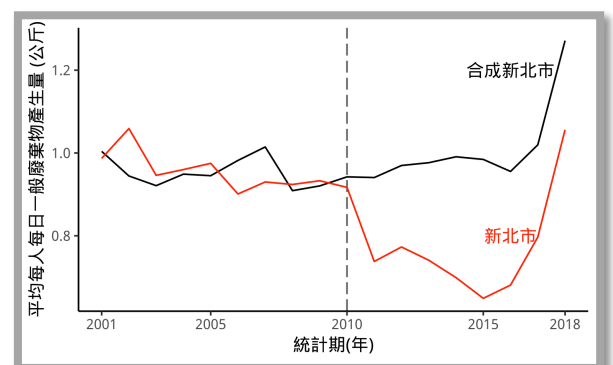
```

{r}
synth.tables$tab.pred

```

	Treated	Synthetic	Sample Mean
平均每戶消費支出	734629.400	734371.533	628338.563
失業率	4.710	4.691	4.616
平均每戶人數	3.485	3.485	3.376
平均每人可支配所得	509251.900	509423.440	470422.116
平均每人每日一般廢棄物產生量 (公斤)	0.953	0.953	0.937

圖（九） 特徵值比較



圖（十） 合成新北市與新北市走勢比較



### SCM-簡單回歸（實驗組：合成控制組；控制組：彰化縣）

$Y$  = `平均每人每日一般廢棄物產生量（公斤）`

$treat$  = dummy variable; 如果統計區為新北市為 1, 否則 0

$post$  = dummy variable; 如果統計期大於 2010 為 1, 否則 0

$effect$  = dummy variable; 如果統計區為新北市且統計期大於 2010 為 1, 否則 0

$$Y = \beta_0 + \beta_1 treat + \beta_2 post + \beta_3 effect$$

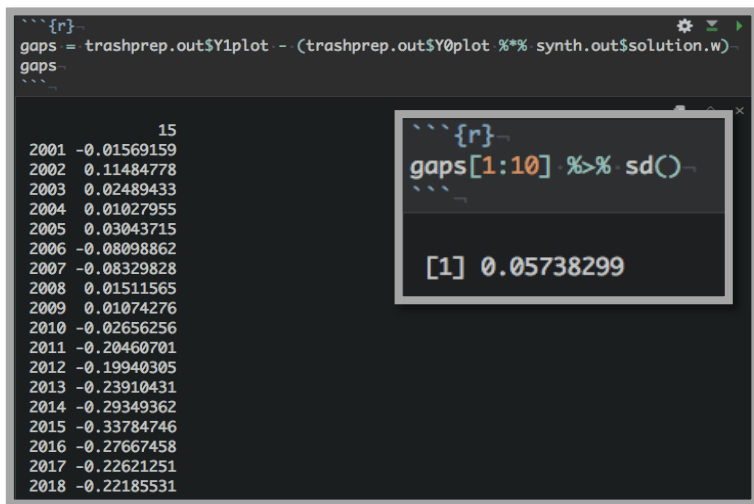
如下圖所示，效果量大約為 -0.247，可解讀為新北市垃圾費隨袋徵收政策可降低`平均每人每日一般廢棄物產生量（公斤）` 0.247 公斤（約為 9.148 個 600cc 寶特瓶空瓶隻重量），且  $p$  值小於 0.01，為顯著效果。

Dependent variable:	
`平均每人每日一般廢棄物產生量（公斤）`	
treat	-0.00001 (0.037)
post	0.060 (0.040)
effect	-0.247*** (0.056)
Constant	0.953*** (0.026)
-----	
Observations	36
R2	0.557
Adjusted R2	0.515
Residual Std. Error	0.083 (df = 32)
F Statistic	13.408*** (df = 3; 32)
=====	
Note:	* $p < 0.1$ ; ** $p < 0.05$ ; *** $p < 0.01$

圖（十一） SCM-簡單迴歸結果

## 結論

DID 模型與 SCM 模型比較，計算合成控制組與新北市的依變數差額標準差，可得標準差為 0.057，此標準差大幅低於先前 DID 模型選定之控制組-彰化縣與新北市的依變數差額標準差:0.159，因此合成新北市比起彰化縣更符合 constant differences in outcome，因此更能準確預測新北市若接收政策效果下的結果。DID 模型下（左方欄位）的簡單迴歸效果量為-0.280( $p < 0.01$ )，SCM 模型下（左方欄位）的簡單迴歸效果量為-0.247( $p < 0.01$ )，兩個效果皆為顯著，然而 SCM 模型下的  $R^2: 0.557 > \text{DID 模型下的 } R^2: 0.190$ ，因此 SCM 模型具有較高的預測能力，所以 SCM 模型下估計的效果量較為準確，因此本報告的結論為新北市垃圾費隨袋徵收效果以 SCM 估計之效果量為降低平均每人每日一般廢棄物產生量 0.280 公斤。



圖（十二） SCM-合成新北市與新北市干擾前依變數差距標準差

Dependent variable:		
	平均每人每日一般廢棄物產生量 (公斤) (1)	平均每人每日一般廢棄物產生量 (公斤) (2)
treat	0.114*** (0.019)	-0.00001 (0.037)
post	0.124*** (0.020)	0.060 (0.040)
effect	-0.280*** (0.028)	-0.247*** (0.056)
Constant	0.839*** (0.013)	0.953*** (0.026)
Observations	456	36
R2	0.190	0.557
Adjusted R2	0.185	0.515
Residual Std. Error	0.147 (df = 452)	0.083 (df = 32)
F Statistic	35.354*** (df = 3; 452)	13.408*** (df = 3; 32)
Note:	*p<0.1; **p<0.05; ***p<0.01	

圖（十二） 模型結果比較

## 問題探討

### Overfitting

當特徵值過多，或控制組(donor pool)過多時將會導致 overfitting 的問題，進而使得模型的預測能力下降，解決方法為：

- 1) 正規化處理 (Regularization)
- 2) 交叉驗證 (Cross-validation)，將干擾前的期間分為兩期間：訓練期間(training period)以及驗證期間(validation period)

## 參考文獻

Abadie A, Gardeazabal J (2003). “The Economic Costs of Conflict: A Case Study of the Basque Country.” *American Economic Review*, 93(1), 112–132

Alberto A., Alexis D., and Jens H., (2011), “Synth: An R Package for Synthetic Control Methods in Comparative Case Studies” *Journal of Statistical Software*

[Causal Inference Using Synthetic Control: The Ultimate Guide](#)

“A Review on Factors affecting Municipal Solid Waste Generation”, 2nd International Engineering Conference. Federal University of Technology, Minna, Niger State, Nigeria. Minna, Nigeria, October 17-19, 2017

Prediction of municipal solid waste generation using artificial neural network approach enhanced by structural break analysis

Robert McClelland, Sarah Gault (2017). “The Synthetic Control Method as a Tool to Understand State Policy” Urban Institute

Alberto Abadie, Alexis Diamond, Jens Hainmueller (2014). “Comparative Politics and the Synthetic Control Method”, Midwest Political Science Association

白經濟 跟著經濟學家玩實驗：合成一個經濟體

利用合成對照組 (Synthetic Control) 了解政策與事件的影響