

因果推論在經濟學門的理論發展與應用

賴宗志[†]

國立中正大學經濟學系

摘要

方案評估 (program evaluation) 一向是經濟實證研究的重要議題，而循證決策 (evidence-based policy making) 更是近年國際公共政策的主流趨勢。在過去 30 多年間，計量經濟學家發展出許多不僅能回答相關性問題，還能提供因果證據的研究方法。而這一系列「因果革命」更使三位長期探索因果關係的學者榮獲 2021 年諾貝爾經濟學獎。因此，我們希望藉由本文介紹因果推論在經濟學門的理論發展，並以諾貝爾獎得主等人的經典實證研究作為應用範例。

關鍵詞：處理效應、迴歸調整、工具變數、差異中差異、合成控制法。

JEL classification: C21, C31.

[†]通訊作者：賴宗志
E-mail: tclai@ccu.edu.tw

1. 前言

方案評估 (program evaluation) 一向是經濟實證研究的重要議題，而循證決策 (evidence-based policy making) 更是近年國際公共政策的主流趨勢。在過去 30 多年間，計量經濟學家發展出許多能克服實證侷限以提高評估可信度的研究方法。這些方法不僅可用於回答相關性問題，還能提供因果證據作為施政參考。而這一系列「因果革命」更使三位長期探索因果關係的學者榮獲 2021 年諾貝爾經濟學獎。¹他們提出的研究方法經常被收錄在各回顧文章或書籍中 (例如 [Imbens and Wooldridge 2009](#) ; [Angrist and Pischke 2009, 2015](#) ; [Imbens and Rubin 2015](#) ; [Hernán and Robins 2020](#)) , 然而中文文獻的相關討論卻相對缺乏。²因此，我們希望藉由本文介紹因果推論在經濟學門的理論發展，並以諾貝爾獎得主等人的經典實證研究作為應用範例。

我們首先第 2 節介紹潛在結果模型 (potential outcome model) 與常見的處理 (treatment effect) 指標，並說明消除選擇偏誤對於因果推論和政策評估的重要性。接著，第 3 節探討在外生性假設成立下如何利用迴歸模型控制可觀察特徵以估計同質或異質處理效應。我們也利用 [Dale and Krueger \(2002\)](#) 針對美國公私立大學畢業生薪資差異的研究作為實證範例。第 4 節則討論當有內生性問題時該如何透過工具變數 (instrumental variable) 進行分析。本節將依序說明兩階段最小平方法 (two-stage least squares) 及局部平均處理效應 (local average treatment effect) , 並分別介紹 [Angrist and Krueger \(1991\)](#) 和 [Angrist, Dynarski, Kane, Pathak, and Walters \(2010\)](#) 對於不同教育階段成效的研究。

在追蹤資料框架中，第 5 節首先回顧差異中差異 (difference-in-differences) 與 [Card and Krueger \(1994\)](#) 有關最低工資對就業影響的經典文章。我們接著說明在交錯採用 (staggered adoption) 設計中傳統雙因子固定效果模型 (two-way fixed effects model) 不適用於因果推論的原因及解決方法。最後，第 6 節介紹合成控制法 (synthetic control method) 。該方法被 [Athey and Imbens \(2017\)](#) 評價為近年來在政策評估文獻中最重要的創新。我們也利用 [Abadie, Diamond, and Hainmueller \(2010\)](#) 評估加州菸草控制法案的成效作為應用實例。第 7 節則為結論。

¹分別為 David Card、Joshua Angrist 以及 Guido Imbens。

²許育進與賴宗志 (2018) 為少數例外。

2. 潛在結果與處理效應

本節介紹 [Rubin \(1974\)](#) 的潛在結果模型 (potential outcome model) 以進行因果推論。對第 i 個人來說，令 D_i 為區分其為處理組 ($D_i = 1$) 或對照組 ($D_i = 0$) 的處理變數 (treatment variable)，而 Y_i 則為實際觀察結果 (observed outcome)。定義 Y_i^1 和 Y_i^0 為潛在結果 (potential outcome)，其中 Y_i^1 為第 i 人若接受處理時的結果，而 Y_i^0 則為第 i 人若不接受處理時的結果。由於對同一人而言 Y_i^1 及 Y_i^0 的唯一差別在於是否接受處理，因此其差異 $Y_i^1 - Y_i^0$ 即可代表個人處理效應 (individual treatment effect; ITE)。文獻中常見的平均處理效應 (average treatment effect; ATE) 及處理組平均處理效應 (average treatment effect on the treated; ATT) 亦可定義為：

$$\text{ATE} \equiv \mathbb{E}(Y_i^1 - Y_i^0)$$

$$\text{ATT} \equiv \mathbb{E}(Y_i^1 - Y_i^0 | D_i = 1)$$

顧名思義，ATE 反映的是 D_i 對 Y_i 的全體 (處理組與對照組) 平均因果效應，而 ATT 僅衡量 D_i 對 Y_i 在處理組的平均因果效應。但由於許多對照組成員無論如何並不會接受處理，估計其效應意義不大，因此經濟文獻中普遍認為 ATT 比 ATE 更具政策意涵 ([Imbens and Wooldridge 2009](#))³。

儘管潛在結果可用於定義處理效應，但由於同一人無法同時身兼處理組與對照組，代表我們只能觀察 Y_i^1 及 Y_i^0 兩者之一，且其與 Y_i 具有以下關係⁴：

$$Y_i = \begin{cases} Y_i^1 & \text{若 } D_i = 1 \\ Y_i^0 & \text{若 } D_i = 0 \end{cases} \quad \text{或} \quad Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0 \quad (1)$$

換句話說，ATE 或 ATT 無法直接從資料得知，需要其他認定條件 (identification conditions) 才能分析。值得一提的是，由於我們可以觀察處理組與對照組的平均差異

³文獻中主要探討 ATE 和 ATT 而非 ITE 的主因是 ITE 無法被估計。事實上，由於 ITE 為隨機，[Hernán and Robins \(2020\)](#) 證明即便 (D_i, Y_i^0, Y_i^1) 的聯合分配已知，ITE 仍無法被認定。

⁴[Hernán and Robins \(2020\)](#) 稱呼該關係為一致性 (consistency) 假設。

$\mathbb{E}(Y_i|D_i = 1) - \mathbb{E}(Y_i|D_i = 0)$ ，且由 (1) 可知：

$$\begin{aligned} & \mathbb{E}(Y_i|D_i = 1) - \mathbb{E}(Y_i|D_i = 0) \\ &= \mathbb{E}(Y_i^1|D_i = 1) - \mathbb{E}(Y_i^0|D_i = 0) \\ &= \underbrace{\mathbb{E}(Y_i^1 - Y_i^0|D_i = 1)}_{\text{ATT}} + \underbrace{\mathbb{E}(Y_i^0|D_i = 1) - \mathbb{E}(Y_i^0|D_i = 0)}_{\text{選擇偏誤}} \end{aligned} \quad (2)$$

代表可觀察的組間差異與 ATT 之間存在選擇偏誤 (selection bias)，也就是處理組和對照組在皆未接受處理時的固有差異 $\mathbb{E}(Y_i^0|D_i = 1) - \mathbb{E}(Y_i^0|D_i = 0)$ 。從 (2) 亦可看出在隨機對照試驗 (randomized controlled trial) 或隨機實驗 (randomized experiment) 中，由於處理變數 D_i 為隨機指派且獨立於其他變數 $D_i \perp\!\!\!\perp (Y_i^1, Y_i^0)$ ，選擇偏誤將不存在，且觀察差異即為 ATT (或 ATE)。

然而，大多數經濟實證研究使用的是非實驗性資料 (non-experimental data) 或觀察性資料 (observational data)，無法操弄處理變數，意味著研究者必須透過假設或資料特性來消除選擇偏誤。我們在第 3 節說明當外生性假設成立時如何利用迴歸模型進行調整，並在第 4 節介紹允許內生處理的工具變數法第 5 節和第 6 節則利用追蹤資料特性，透過不同方式建立具有比較基準的對照組以進行因果推論。

3. 迴歸調整

由 (2) 可知，欲認定 ATT，關鍵在於消除處理組和對照組的固有差異 (選擇偏誤)。我們首先假設選擇偏誤僅受到可觀察個人特徵 $\mathbf{X}_i = (X_{i1}, \dots, X_{ik})'$ 的影響，不被其他不可觀察因素干擾。換句話說，在給定相同個人特徵下，處理組和對照組成員並無系統性差別，也表示此時的處理變數與隨機指派無異 (as good as randomly assigned)。該假設的定義如下：

假設 1 (selection on observables). $D_i \perp\!\!\!\perp (Y_i^1, Y_i^0) | \mathbf{X}_i$ 。

假設 1 又稱為無混淆性 (unconfoundedness)、可忽略性 (ignorability)、外生性 (exogeneity) 或條件獨立假設 (conditional independence assumption)。在該假設成立下，估計處理效應的常見作法是利用迴歸模型控制可觀察特徵，故此方法亦稱為迴歸調整 (regression adjustment)：

$$Y_i = \alpha + \beta D_i + \gamma' \mathbf{X}_i + \varepsilon_i \quad (3)$$

我們可用上式認定 ATE 及 ATT。首先，將 (1) 改寫為：

$$Y_i = Y_i^0 + (Y_i^1 - Y_i^0)D_i \quad (4)$$

比較 (3) 和 (4) 後可得：

$$\begin{aligned} Y_i^0 &= \alpha + \gamma' \mathbf{X}_i + \varepsilon_i \\ Y_i^1 &= \alpha + \beta + \gamma' \mathbf{X}_i + \varepsilon_i \end{aligned}$$

上述潛在結果隱含處理效應具有同質性 (homogeneity)，亦即對所有人來說，ITE $Y_i^1 - Y_i^0$ 皆為迴歸係數 β ，而 ATE 及 ATT 自然也相同。值得一提的是 (3) 中的 γ 並不重要，也代表該迴歸模型中各解釋變數扮演的角色有所不同： D_i 為我們感興趣的處理變數，而 \mathbf{X}_i 則是用於消除選擇偏誤的控制變數 (control variables)。最後，從假設 1 和 Y_i^0 及 Y_i^1 的線性迴歸模型可知 $\mathbb{E}(\varepsilon_i | D_i, \mathbf{X}_i) = \mathbb{E}(\varepsilon_i | \mathbf{X}_i)$ ，由此可證明 (3) 中 β 的 OLS 估計式具有不偏性與一致性，但 γ 的 OLS 估計式則未必^{5,6}。

縱然 (3) 相當直觀，但其不允許異質 (heterogeneous) 處理效應的限制卻大幅降低實用性。一個簡單的解決方法是在迴歸模型中加入交乘項：

$$Y_i = \alpha + \beta D_i + \gamma' \mathbf{X}_i + \delta' \mathbf{X}_i D_i + \varepsilon_i \quad (5)$$

同理，比較 (4) 和 (5) 可得：

$$\begin{aligned} Y_i^0 &= \alpha + \gamma' \mathbf{X}_i + \varepsilon_i \\ Y_i^1 &= \alpha + \beta + (\gamma + \delta)' \mathbf{X}_i + \varepsilon_i \end{aligned}$$

此時的 ITE 將隨著 \mathbf{X}_i 變動，亦即 $Y_i^1 - Y_i^0 = \beta + \delta' \mathbf{X}_i$ ，代表 ATE 和 ATT 不再全然由迴歸係數 β 決定。事實上，在 (5) 成立下：

$$\begin{aligned} \text{ATE} &= \beta + \delta' \mathbb{E}(\mathbf{X}_i) \\ \text{ATT} &= \beta + \delta' \mathbb{E}(\mathbf{X}_i | D_i = 1) \end{aligned}$$

⁵由於 Y_i^0 及 Y_i^1 服從線性迴歸模型，其支集 (support) 相同，代表 Hernán and Robins (2020) 要求的正值性/重疊性 (positivity/overlap) 假設在迴歸調整中自動成立。然而，該假設在機率倒數加權法 (inverse probability weighting) 中則未必成立，詳情可參考 許育進與賴宗志 (2018)。

⁶文獻中存在許多估計 β 的方法，舉例來說，Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018) 提出雙重機器學習法 (double machine learning) 以放寬 (3) 的線性迴歸假設並允許控制高維特徵向量 \mathbf{X}_i 。

表 1: Angrist and Pischke (2015), Table 2.1。

The college matching matrix

Applicant group	Student	Private			Public			Altrerd earnings
		Ivy	Leafy	Smart	All State	Tall State	State	
A	1		Reject	Admit			Admit	110,000
	2		Reject	Admit			Admit	100,000
	3		Reject	Admit			Admit	110,000
B	4	Admit			Admit		Admit	60,000
	5	Admit			Admit		Admit	30,000
C	6		Admit					115,000
	7		Admit					75,000
D	8	Reject			Admit	Admit		90,000
	9	Reject			Admit	Admit		60,000

Note: Enrollment decisions are highlighted in gray

意味著只要能估計 β 和 δ ，ATE 與 ATT 則唾手可得。與同質處理效應情況類似，當假設 1 和 (5) 成立時可知 $\mathbb{E}(\varepsilon_i|D_i, \mathbf{X}_i) = \mathbb{E}(\varepsilon_i|\mathbf{X}_i)$ ，且可證明 β 和 δ 的 OLS 估計式具有不偏性與一致性。此外，(5) 亦可延伸至廣義線性模型並得到類似結論。

利用迴歸調整，Dale and Krueger (2002) 估計在美國就讀私立大學相較於公立大學對收入的影響。⁷單純比較公私立大學畢業生薪資，可發現私校生平均薪資較公校生高出 13.5%，但其入學成績與家庭背景等可能影響未來薪資的個人特質與公校生也有著明顯差異。控制上述可觀察特徵後，公私立學生的薪資差距縮小為 8.6%，惟此差距仍可能來自無法觀察的非認知能力，而非就讀私校的薪資加成。為解決此問題，Dale and Krueger (2002) 進一步將公私立學生申請入學時錄取和拒絕的學校名單進行匹配（見表 1），以消除無法觀察的個人特質差異。在控制所有變因後，他們發現就讀私立大學並不比就讀公立大學的學生薪資更高，代表個人能力或動機才是決定收入的關鍵因素。

以上介紹的迴歸調整僅是在假設 1 成立下的眾多認定及估計方法其中之一，有興趣的讀者可參考 許育進與賴宗志 (2018) 所討論的機率倒數加權法 (inverse probability weightin) 及其變化，例如同時利用迴歸調整及機率倒數加權的雙重穩健法 (doubly

⁷以下內容為 Angrist and Pischke (2015), Chapter 2 的重製版本。

robust methods)。

4. 工具變數

4.1 兩階段最小平方法

本節介紹當**假設 1** 不成立，亦即處理變數為內生（endogenous）時傳統計量經濟學如何利用工具變數（instrumental variable）進行因果推論。為了方便說明，我們首先忽略 (3) 中的可觀察個人特徵 \mathbf{X}_i 並假設選擇偏誤來自研究者無法觀察的個人特徵 U_i ：

$$Y_i = \alpha + \beta D_i + \gamma U_i + \varepsilon_i \quad (6)$$

其中 $\text{Cov}(D_i, U_i) \neq 0$ 。由於 U_i 無法觀察，直接估計 (6) 中代表同質處理效應的 β 會產生遺漏變數偏誤（omitted variable bias）。然而，若存在一個影響 D_i 但卻不影響 U_i 的工具變數 Z_i ，則 β 依然可被認定及估計。理想的工具變數必須滿足以下條件：

假設 2 (工具變數)。 Z_i 與 D_i 有關且僅透過 D_i 與 Y_i 有關。

文獻中一般將**假設 2** 的前半部分稱為 first stage 而後半部分稱為排他限制（exclusion restriction）。在該假設成立下， Z_i 與 Y_i 的相關性只會經由 D_i 傳遞而非其他管道，意味著 Z_i 與無法觀察的 U_i （和 ε_i ）無關，即 $\text{Cov}(Z_i, U_i) = \text{Cov}(Z_i, \varepsilon_i) = 0$ 。**圖 1** 描繪了工具變數法的基本概念。由圖中可看出 Z_i 與 Y_i 的相關可拆解為 Z_i 與 D_i 的相關乘上 D_i 與 Y_i 在控制 U_i 之後的偏相關（亦即 (6) 中的 β ），也代表 β 可由 $\text{Cov}(Z_i, Y_i)$ 和 $\text{Cov}(Z_i, D_i)$ 的比值來認定。事實上，(6) 和**假設 2** 即蘊含以下等式成

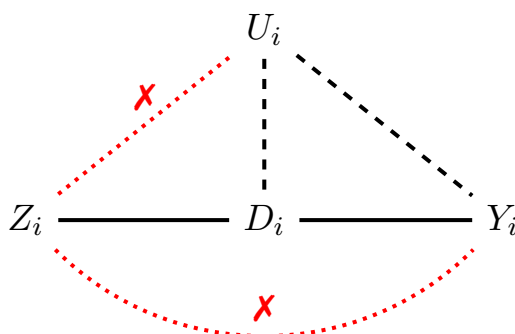


圖 1: 工具變數法。

立：

$$\frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, D_i)} = \frac{\text{Cov}(Z_i, \alpha + \beta D_i + \gamma U_i + \varepsilon_i)}{\text{Cov}(Z_i, D_i)} = \beta \quad (7)$$

此外，從 (6) 和假設 2 亦可看出 $\mathbb{E}(Y_i|Z_i) = \alpha + \beta\mathbb{E}(D_i|Z_i)$ 。若進一步假設工具變數為二元變數 $Z_i \in \{0, 1\}$ ，則 β 也可表示為 Wald (1940) 的形式：

$$\beta = \frac{\mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 0)}{\mathbb{E}(D_i|Z_i = 1) - \mathbb{E}(D_i|Z_i = 0)} \quad (8)$$

儘管 (7) 和 (8) 提供了兩種方式估計 β ，但文獻中更常見的作法是透過兩階段最小平方法 (two-stage least squares; 2SLS) 納入外生控制變數 \mathbf{X}_i 或多個工具變數 $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iq})'$ 。舉例而言，考慮以下迴歸模型：

$$Y_i = \alpha + \beta D_i + \gamma' \mathbf{X}_i + \mu_i \quad (9)$$

$$D_i = \zeta + \eta' \mathbf{Z}_i + \theta' \mathbf{X}_i + \nu_i \quad (10)$$

其中 $(\mathbf{Z}_i', \mathbf{X}_i')$ 與 $(\mu_i, \nu_i)'$ 皆無關但 $\text{Cov}(D_i, \mu_i) \neq 0$ 。由於 (10) 無內生性問題，我們可用 OLS 區分 D_i 當中被外生變數 \mathbf{Z}_i 和 \mathbf{X}_i 解釋的配適值 \hat{D}_i 以及無法被模型解釋的殘差 $\hat{\nu}_i$ 。將 $D_i = \hat{D}_i + \hat{\nu}_i$ 代入 (9) 後可得：

$$Y_i = \alpha + \beta \hat{D}_i + \gamma' \mathbf{X}_i + \varepsilon_i \quad (11)$$

其中 $\varepsilon_i \equiv \mu_i + \beta \hat{\nu}_i$ 。由於 μ_i 與 $(\hat{D}_i, \mathbf{X}_i)'$ 無關，而根據 OLS 性質可知殘差 $\hat{\nu}_i$ 與 $(\hat{D}_i, \mathbf{X}_i)'$ 亦無關，由此可知 (11) 無內生性問題，可再一次用 OLS 得到 β 的一致估計式，故該方法被稱為 2SLS。值得一提的是，由於 (11) 中的 \hat{D}_i 為生成自變數 (generated regressor) 存在估計誤差，這將導致 β 的 2SLS 估計式僅為漸近不偏，且進行統計推論時需調整標準誤才能得到正確結果。⁸

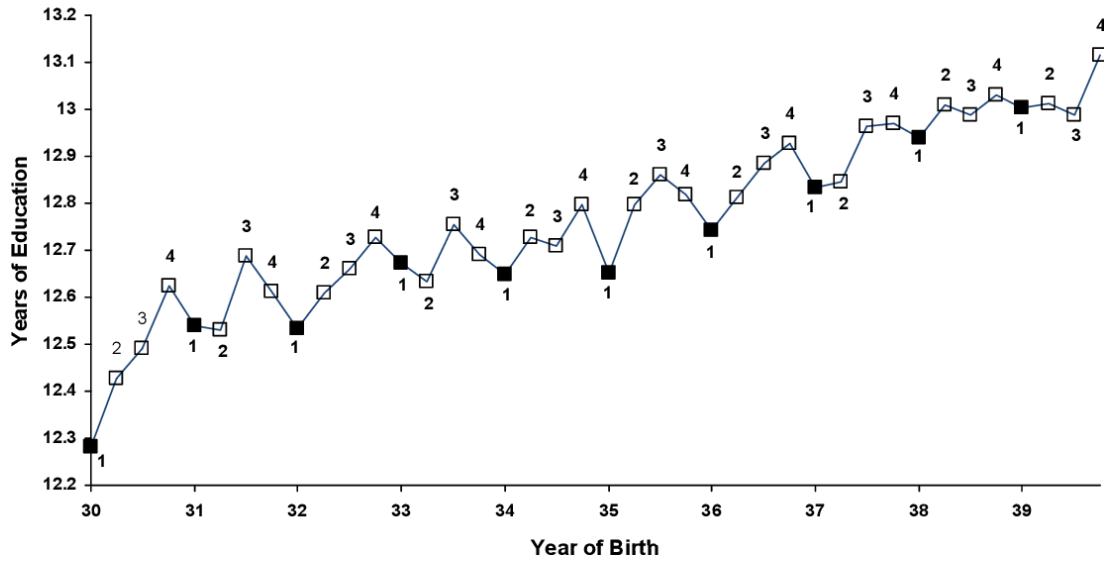
此外，透過 Frisch-Waugh-Lovell 定理可知 (11) 中的 β 亦具備 (7) 的形式：

$$\beta = \frac{\text{Cov}(\hat{\nu}_i, Y_i)}{\text{Var}(\hat{\nu}_i)} = \frac{\text{Cov}(\hat{\nu}_i, Y_i)}{\text{Cov}(\hat{\nu}_i, D_i)}$$

其中 $\hat{\nu}_i$ 是 \hat{D}_i 對 \mathbf{X}_i 進行迴歸的 OLS 殘差，而上式第二個等號來自 $D_i = \hat{D}_i + \hat{\nu}_i$ 且可證明 $\text{Cov}(\hat{\nu}_i, \hat{\nu}_i) = 0$ 。換句話說，2SLS 實際上是把多個外生變數 \mathbf{X}_i 和 \mathbf{Z}_i 濃縮成單一「工具變數」 $\hat{\nu}_i$ 並利用 (7) 來進行估計。

⁸就作者所知大部分統計套裝軟體皆已調整。

A. Average Education by Quarter of Birth (first stage)



B. Average Weekly Wage by Quarter of Birth (reduced form)

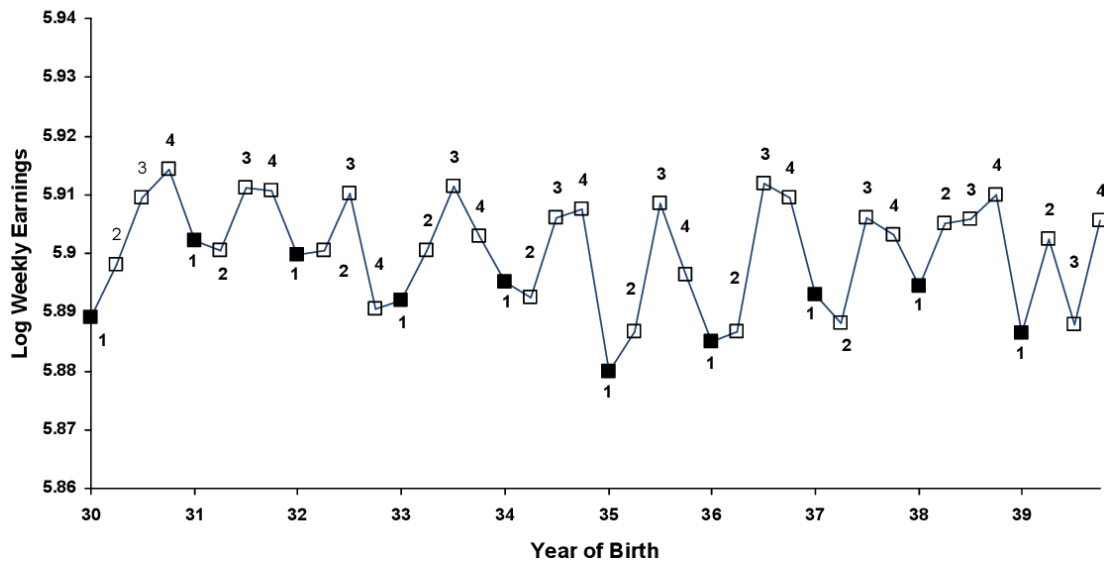


圖 2: Angrist and Pischke (2009), Figure 4.1.1。

利用 2SLS，Angrist and Krueger (1991) 將出生季節作為工具變數來估計多接受一年教育對薪資的影響。⁹支持該工具變數滿足假設 2 的論述如下：在同一年（無論年初或年底）滿法定年齡即可入學的規定下，年初出生的人入學年齡較大，因此與年底出生的人相比更早成年，不受未成年人需接受義務教育的限制。由此可知年初出生的人提前輟學機率較高，接受的教育年數普遍較低，更可能導致未來薪資受此影響而偏低。圖 2 利用美國 1980 年普查資料驗證上述說法。在出生季節不透過其他管道影響薪資的前提下，Angrist and Krueger (1991) 分別估計 2SLS 和 OLS 結果。有趣的是，他們發現兩者並無明顯差異，代表內生性問題在此議題中影響不大，也間接支持 Mincer (1974) 薪資方程式的正確性。

4.2 局部平均處理效

縱然上節介紹的 2SLS 在文獻中廣為使用，但其隱含的同質處理效應卻過於嚴苛，不見得符合實際情況。有趣的是，利用潛在結果模型和二元工具變數，Imbens and Angrist (1994)、Angrist, Imbens, and Rubin (1996) 發現 (8) 中的 Wald (1940) 形式實際上可表示為允許異質性存在的局部平均處理效應 (local average treatment effect; LATE)。本節將介紹其概念及相關應用。首先，由於除了處理變數 D_i 之外還有工具變數 Z_i 會影響結果 Y_i ，我們應重新定義潛在結果為 $Y_i^{d,z}$ ，亦即第 i 個人在 $D_i = d$ 且 $Z_i = z$ 之下的結果。然而在假設 2 成立下 Z_i 只會透過 D_i 影響 Y_i ，意味著一但固定 $D_i = d$ 後 Z_i 並不會對 Y_i 造成影響，也代表潛在結果可維持第 2 節的定義：

$$Y_i^{d,z} = Y_i^d$$

與潛在結果類似，當工具變數為二元變數 $Z \in \{0, 1\}$ 時可定義 D_i^1 和 D_i^0 為第 i 個人的潛在處理狀態 (potential treatment status)，其中 D_i^1 為若 $Z_i = 1$ 時第 i 人的處理狀態，而 D_i^0 則為若 $Z_i = 0$ 時第 i 人的處理狀態。值得注意的是，潛在處理狀態 D_i^z 除了受到工具變數 $Z_i = z$ 影響之外也與個人行為有關。表 2 列出四種可能的類型 (Angrist, Imbens, and Rubin 1996)：

- (i) 永遠接受者 (always-taker)：無論工具變數為何，此類型的人永遠接受處理，即 $D_i^1 = D_i^0 = 1$ 。
- (ii) 永不接受者 (never-taker)：無論工具變數為何，此類型的人永不接受處理，即

⁹以下內容為 Angrist and Pischke (2009), Chapter 4 的重製版本。

表 2: 以潛在處理狀態區分的四種類型。

$D_i^0 \backslash D_i^1$	1	0
1	永遠接受者 (always-taker)	違逆者 (defier)
0	依從者 (complier)	永不接受者 (never-taker)

$$D_i^1 = D_i^0 = 0。$$

(iii) 依從者 (complier)：此類型的人處理狀態完全依從工具變數，即 $D_i^1 = 1$ 且 $D_i^0 = 0$ (或 $D_i^1 > D_i^0$)。

(iv) 違逆者 (defier)：此類型的人處理狀態完全違逆工具變數，即 $D_i^1 = 0$ 且 $D_i^0 = 1$ (或 $D_i^1 < D_i^0$)。

然而，由於我們只能觀察 D_i^1 及 D_i^0 兩者之一，且其與 D_i 具有以下關係：

$$D_i = \begin{cases} D_i^1 & \text{若 } Z_i = 1 \\ D_i^0 & \text{若 } Z_i = 0 \end{cases} \quad \text{或} \quad D_i = Z_i D_i^1 + (1 - Z_i) D_i^0 \quad (12)$$

意味著 Z_i 對 D_i 的「平均處理效應」 $\mathbb{E}(D_i^1 - D_i^0)$ 無法由資料認定。Imbens and Angrist (1994) 因而引入工具變數獨立性 (independence) 作為認定條件之一：

假設 3 (獨立性). $Z_i \perp (Y_i^1, Y_i^0, D_i^1, D_i^0)$ 。¹⁰

由假設 3 和 (12) 可知 $\mathbb{E}(D_i^1 - D_i^0)$ 等於 (8) 的分母，而 (8) 的分子在假設 3 成立下亦可證明與 $\mathbb{E}[(D_i^1 - D_i^0)(Y_i^1 - Y_i^0)]$ 相等。換言之，(8) 可改寫為：

$$\frac{\mathbb{E}(Y_i | Z_i = 1) - \mathbb{E}(Y_i | Z_i = 0)}{\mathbb{E}(D_i | Z_i = 1) - \mathbb{E}(D_i | Z_i = 0)} = \frac{\mathbb{E}[(D_i^1 - D_i^0)(Y_i^1 - Y_i^0)]}{\mathbb{E}(D_i^1 - D_i^0)} \quad (13)$$

進一步觀察可知 (13) 的分子即為利用 $D_i^1 - D_i^0$ 加權的 ATE，其中永遠接受者和永不接受者的權重皆為 0，而依從者和違逆者的權重正負號卻剛好相反。為了確保 (13) 能反映單一類型的異質處理效應，Imbens and Angrist (1994) 假設潛在處理狀態具有單調性 (monotonicity)：

¹⁰ 此假設可放寬至條件獨立假設： $Z_i \perp (Y_i^1, Y_i^0, D_i^1, D_i^0) | \mathbf{X}_i$ ，見 Abadie (2003)。

假設 4 (單調性). $D_i^1 \geq D_i^0$ 。

假設 4 要求潛在處理狀態為工具變數的單調遞增函數，其意味著工具變數會「鼓勵」人們接受處理，以排除違逆者 ($D_i^1 < D_i^0$) 的存在。在該假設成立下，可知 (13) 的分母 $\mathbb{E}(D_i^1 - D_i^0) = \mathbb{P}(D_i^1 > D_i^0)$ 為依從者的比例，且 (13) 的分子 $\mathbb{E}[(D_i^1 - D_i^0)(Y_i^1 - Y_i^0)] = \mathbb{E}(Y_i^1 - Y_i^0 | D_i^1 > D_i^0) \mathbb{P}(D_i^1 > D_i^0)$ 為依從者的平均處理效應乘上其比例。綜上所述，Imbens and Angrist (1994) 定義依從者的平均處理效應為 LATE，並證明在假設 2 至 4 成立下，LATE 可由 (13) 認定：

$$\text{LATE} \equiv \mathbb{E}(Y_i^1 - Y_i^0 | D_i^1 > D_i^0) = \frac{\mathbb{E}(Y_i | Z_i = 1) - \mathbb{E}(Y_i | Z_i = 0)}{\mathbb{E}(D_i | Z_i = 1) - \mathbb{E}(D_i | Z_i = 0)} \quad (14)$$

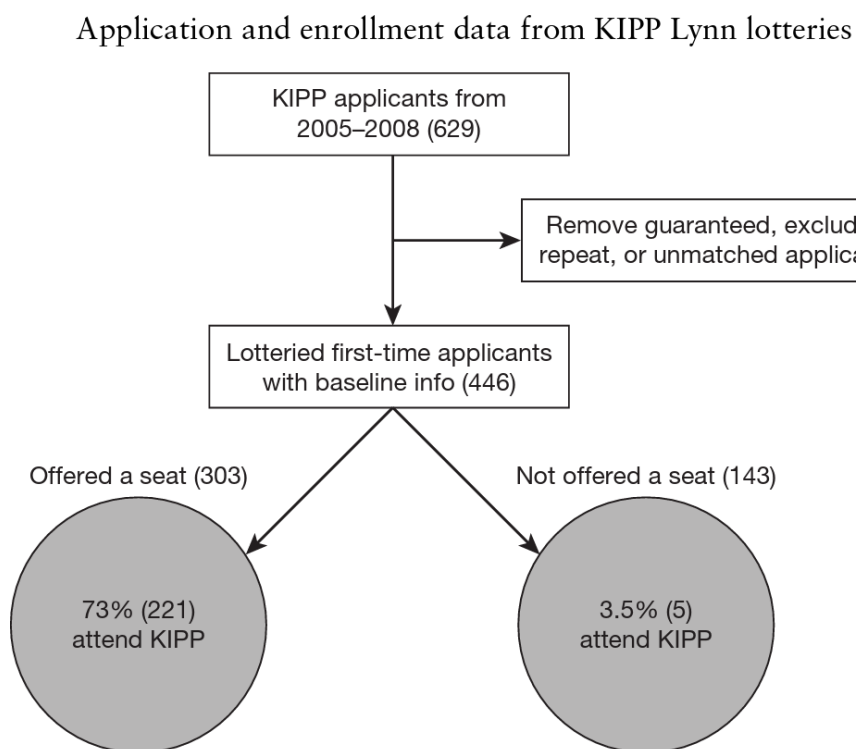
上述發現開拓了新的研究領域，也讓人們了解在允許異質性之下工具變數實際衡量的對象為何。然而，由於 D_i^1 和 D_i^0 無法同時被觀察，我們並不能分辨樣本中誰是依從者，誰又是永遠接受者或永不接受者。此外，依從者的定義會隨著不同工具變數而改變，因此以政策評估的角度來看，LATE 提供的資訊並不如 ATE 或 ATT 來得明確。有趣的是，在引入 Bloom (1984) 提出的單邊不依從 (one-sided non-compliance) 假設下可證明 LATE 即為 ATT。該假設的定義如下：

假設 5 (單邊不依從). $D_i^0 = 0$ 。

假設 5 要求若工具變數 $Z_i = 0$ 時所有人皆不接受處理 (亦即 $D_i^0 = 0$)，但 D_i^1 則不設限。換句話說，該假設比起單調性假設排除違逆者外還多排除了永遠接受者，因此必然隱含假設 4 成立。透過假設 5、假設 3 和 (12) 可知：

$$\begin{aligned} \text{LATE} &= \mathbb{E}(Y_i^1 - Y_i^0 | D_i^1 > D_i^0) \\ &= \mathbb{E}(Y_i^1 - Y_i^0 | D_i^1 = 1) \\ &= \mathbb{E}(Y_i^1 - Y_i^0 | Z_i = 1, D_i^1 = 1) \\ &= \mathbb{E}(Y_i^1 - Y_i^0 | D_i = 1) = \text{ATT} \end{aligned}$$

另一方面，由於假設 5 要求處理組只會有依從者，故在放寬假設 3 下亦可證明處理組局部平均處理效應 (local average treatment effect on the treated; LATT) 等於 ATT。由於估計 ATT 一般而言需要假設 1 成立但 LATT 則否，因此 Donald, Hsu, and Lieli (2014) 便利用此關係提出 Durbin-Wu-Hausman 類型的檢定來驗證假設 1 是否成立。此外，Swanson, Hernán, Miller, Robins, and Richardson (2018) 也提出即使在假設 4



Note: Numbers of Knowledge Is Power Program (KIPP) applicants are shown in parentheses.

圖 3: Angrist and Pischke (2015), Figure 3.1。

不成立的情況下，仍可用二元工具變數對 ATE 進行部分認定（partial identification）。¹¹

利用 LATE，Angrist, Dynarski, Kane, Pathak, and Walters (2010) 評估就讀特許學校（charter school）是否能提升美國中小學生的學習表現。¹²特許學校是由政府資助的私立學校，雖然其教師相對而言較無經驗或未經認證，但由於擁有免學費的優勢，熱門特許學校的申請人數往往超過招生名額。為了公平起見，位於麻州的特許學校 KIPP Lynn 通過抽籤來決定誰具有入學資格。由圖 3 可看出在 303 位具有資格的人中有 221 位選擇就讀，而 143 位無資格的人僅有 5 位透過其他管道入學（可能謊稱兄弟姐妹中有人具有資格）。若將有無資格當作工具變數且將是否就讀當作處理變數，則 (i) 入學資格只會透過影響就讀狀態來間接影響學習表現；(ii) 入學資格為隨機指派；(iii) 取得入學資格應提高而非降低就讀意願。以上三點代表假設 2 至 4 皆應成立，可用 (14) 估計就讀特許學校對學習表現的 LATE。此外，若能完全排除其他入學管

¹¹ 作者感謝匿名評審的補充。

¹² 以下內容為 Angrist and Pischke (2015), Chapter 3 的重製版本。

道，**假設 5** 也應成立，代表該 LATE 即為 ATT。利用 KIPP Lynn 的資料，Angrist, Dynarski, Kane, Pathak, and Walters (2010) 發現就讀特許學校可提高學生數學成績 0.48 個標準差，大約從成績分佈的底部三分之一上升至中間，成效可說相當卓越。

5. 差異中差異

上述方法均建立在研究者僅有橫斷面資料 (cross-sectional data) 的前提，但為了克服外生性假設或工具變數限制，本節及第 6 節介紹兩個利用追蹤資料 (panel data) 進行因果推論的方法。我們首先考慮有兩個組別 (處理組與對照組) 及兩個時間點 (事前與事後) 的 2×2 差異中差異法 (difference-in-differences; DID)。由於存在時間維度 $t \in \{0, 1\}$ ，潛在結果和實際觀察結果應重新定義為 $(Y_{it}^1, Y_{it}^0)'$ 和 Y_{it} ，其中事前 ($t = 0$) 所有人皆未接受處理，亦即 $Y_{i0} = Y_{i0}^0$ ，而事後 ($t = 1$) 僅處理組接受處理而對照組則否，亦即 $Y_{i1} = D_i Y_{i1}^1 + (1 - D_i) Y_{i1}^0$ 。同樣地，我們感興趣的 ATT 亦需重新定義為：

$$ATT \equiv \mathbb{E}(Y_{i1}^1 - Y_{i1}^0 | D_i = 1) = \mathbb{E}(Y_{i1} | D_i = 1) - \mathbb{E}(Y_{i1}^0 | D_i = 1) \quad (15)$$

由 (15) 可知，認定 ATT 的關鍵在於 $\mathbb{E}(Y_{i1}^0 | D_i = 1)$ ，也就是處理組在事後若未接受處理的平均結果。儘管我們可以觀察處理組在事前未接受處理的平均結果 $\mathbb{E}(Y_{i0}^0 | D_i = 1) = \mathbb{E}(Y_{i0} | D_i = 1)$ ，但這並無法反映處理組 (即便在未接受處理下) 的結果會隨時間變化的事實。基於該變化與是否接受處理無關，我們可假設處理組與對照組在未接受處理下的變化趨勢相同，也就是所謂的平行趨勢假設 (parallel trends assumption)：

假設 6 (平行趨勢). $\mathbb{E}(Y_{i1}^0 - Y_{i0}^0 | D_i = 1) = \mathbb{E}(Y_{i1}^0 - Y_{i0}^0 | D_i = 0)$ 。

利用 (15)、**假設 6** 以及對照組 $Y_{it}^0 = Y_{it}$ 恆成立的關係，可知 ATT 為處理組和對照組前後期差異之間的差異：

$$\begin{aligned} ATT &= \mathbb{E}(Y_{i1} | D_i = 1) - \mathbb{E}(Y_{i1}^0 | D_i = 1) \\ &= \mathbb{E}(Y_{i1} - Y_{i0}^0 | D_i = 1) - \mathbb{E}(Y_{i1}^0 - Y_{i0}^0 | D_i = 1) \\ &= \mathbb{E}(Y_{i1} - Y_{i0} | D_i = 1) - \mathbb{E}(Y_{i1}^0 - Y_{i0}^0 | D_i = 0) \\ &= \mathbb{E}(Y_{i1} - Y_{i0} | D_i = 1) - \mathbb{E}(Y_{i1} - Y_{i0} | D_i = 0) \end{aligned} \quad (16)$$

這也是 DID 名稱的由來。事實上，我們也可先對組別差分再對時間差分來認定 ATT，亦即圖 4 所示的垂直差異。然而若要估計 ATT，最常見的作法反而是利用帶有交乘項的迴歸模型：

$$Y_{it} = \alpha + \beta D_i + \gamma T_t + \delta D_i T_t + \varepsilon_{it} \quad (17)$$

其中 $T_t = t \in \{0, 1\}$ 為時間二元變數且 $\mathbb{E}(\varepsilon_{it}|D_i, T_t) = 0$ 。表 3 列出在不同組別及時間下用迴歸係數表示的平均結果。綜合表 3 和 (16)，可看出交乘項係數 δ 即為 2×2 DID 架構下的 ATT。

利用 DID，Card and Krueger (1994) 研究提高最低工資對基層勞工就業的影響。在完全競爭市場理論中，當最低工資高於均衡工資時，廠商會減少對勞動者的需求量，進而導致就業人口下降。為了驗證該理論是否正確，Card and Krueger (1994) 將 1992 年 4 月曾調漲近 20% 最低工資的紐澤西州作為處理組，並將鄰近但未調整最低工資的賓州視為對照組。他們比較 410 間位於邊界附近的速食餐廳（見圖 5）在政策實施前後的聘用人數，以提高假設 6 在相似就業環境下的可信度。令人意外的是，DID 結果顯示紐澤西州在工資調漲後相較於賓州並未減少聘用員工，與理論預測不符。此結論開啟了後續許多相關研究（支持與反對皆有），直到現在最低工資對勞動市場的影響仍然是熱門研究議題之一。

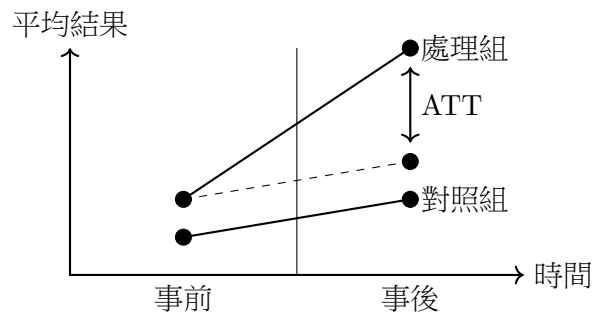


圖 4: 差異中差異法。

表 3: DID 與 (17) 之迴歸係數。

平均結果	事前 ($T_t = 0$)	事後 ($T_t = 1$)	差異
處理組 ($D_i = 1$)	$\alpha + \beta$	$\alpha + \beta + \gamma + \delta$	$\gamma + \delta$
對照組 ($D_i = 0$)	α	$\alpha + \gamma$	γ
差異	β	$\beta + \delta$	δ

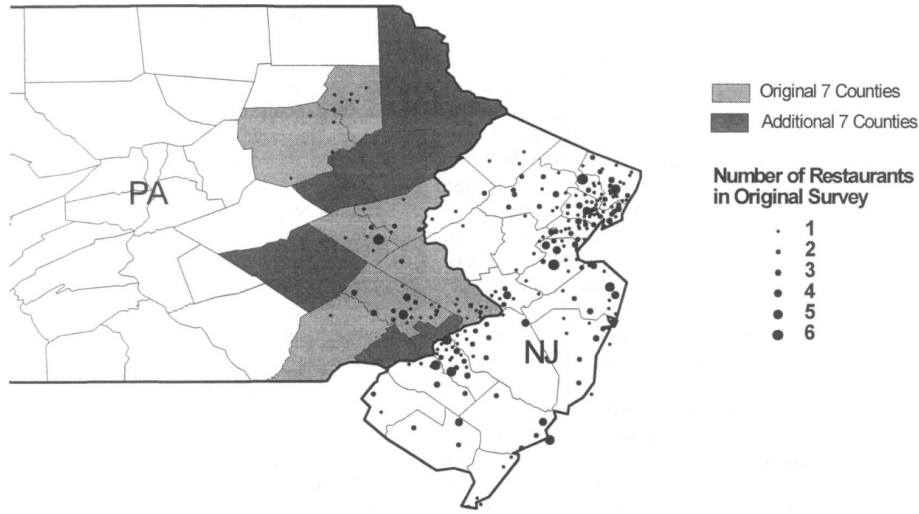


圖 5: Card and Krueger (2000), Figure 1。

我們接著考慮比 2×2 DID 更一般化的設定。令時間長度為 T 且 $t = 1, \dots, T$ ，則 (17) 可改寫為雙因子固定效果模型 (two-way fixed effects model)：

$$Y_{it} = \beta_i + \gamma_t + \delta D_{it} + \varepsilon_{it} \quad (18)$$

其中 β_i 和 γ_t 分別代表個人和時間固定效果，而 $D_{it} \in \{0, 1\}$ 為第 i 人在第 t 期是否接受處理的二元變數。在 $T = 2$ 且平行趨勢假設成立時可驗證 (18) 中的 δ 即為 ATT。然而，在 $T > 2$ 且允許接受處理時間不同的交錯採用 (staggered adoption) 設定下，上述關係並不成立，甚至可能連正負號都相反。以下我們略加修改 Callaway and Sant'Anna (2021) 的符號來說明其原因。令 $G_i = g \in \{2, \dots, T\}$ 代表第 i 人在第 g 期開始接受處理的群組變數，而 $G_i = 0$ 代表第 i 人為從未接受處理的對照組。同理，定義 Y_{it}^g 為第 i 人若在第 g 期開始接受處理時其在第 t 期 ($t \geq g$) 的潛在結果，而 Y_{it}^0 為第 i 人若從未接受處理時其在第 t 期的潛在結果。我們接著修改假設 6 為以群組為單位的平行趨勢假設：

假設 7 (群組平行趨勢). 對 $t \geq g \in \{2, \dots, T\}$ 而言， $\mathbb{E}(Y_{i,t}^0 - Y_{i,t-1}^0 | G_i = g) = \mathbb{E}(Y_{i,t}^0 - Y_{i,t-1}^0 | G_i = 0)$ 。

與 (16) 類似，在假設 7 成立下我們可認定處理組中第 g 組在第 t 期 ($t \geq g$) 的

ATT 為：

$$\begin{aligned} \text{ATT}(g, t) &\equiv \mathbb{E}(Y_{it}^g - Y_{it}^0 | G_i = g) \\ &= \mathbb{E}(Y_{it} - Y_{i,g-1} | G_i = g) - \mathbb{E}(Y_{it} - Y_{i,g-1} | G_i = 0) \end{aligned} \quad (19)$$

根據 (19)，我們可用處理組中第 g 組和對照組之間的 DID 來正確估計 $\text{ATT}(g, t)$ 。然而，值得注意的是 de Chaisemartin and d'Haultfoeuille (2020) 證明了 (18) 中的 δ 可表示為 $\text{ATT}(g, t)$ 的加權平均：

$$\delta = \sum_{(g,t): D_{it}=1} w(g, t) \text{ATT}(g, t) \quad (20)$$

其中 $w(g, t)$ 為總和為 1 但可能為負的權重。換言之，在某些情況下即便 $\text{ATT}(g, t)$ 皆為正，加權後反而變成負值，也導致 δ 和 $\text{ATT}(g, t)$ 的正負號可能不同，無法用於估計處理效應。

為什麼 (20) 中的權重可能為負？Goodman-Bacon (2021) 給出一個相當直覺的解釋：(18) 中的 δ 可視為處理狀態改變及處理狀態不變群組之間 DID 的加權平均，而這些 DID 包括 (i) 使用從未接受處理的群組作為比較對象（如圖 6 上半部分），(ii) 使用尚未接受處理的群組作為比較對象（如圖 6 左下部分）以及 (iii) 使用已接受處理的群組作為比較對象（如圖 6 右下部分）。其中情況 (iii) 的比較對象亦接受處理，代表其 DID 並不等於 $\text{ATT}(g, t)$ 。換句話說，即便 δ 為不同情況下 DID 的加權平均，但這些權重並不能轉換為 (20) 中跟隨著 $\text{ATT}(g, t)$ 的權重 $w(g, t)$ ，甚至在極端情況下會導致 $w(g, t)$ 為負。

6. 合成控制法

相較於 DID，Abadie and Gardeazabal (2003) 提出的合成控制法（synthetic control method；SCM）僅針對單一處理組成員進行因果推論，因此精神上更接近傳統比較研究（comparative study）。然而不同於比較研究逕行挑選與處理組最相似的比較對象，也不像 DID 利用控制組平均變化與處理組平均變化相比，SCM 的作法是將潛在控制組成員（donor pool）加權平均後得到一個「合成控制」，再將此「合成控制」的結果當作處理組成員在未接受處理下的潛在結果。在良好媒合下，處理組成員與「合成控制」除了是否接受處理外應高度相似，也意味著兩者結果差異即為處理組個人處理效應（individual treatment effect on the treated；ITT）。

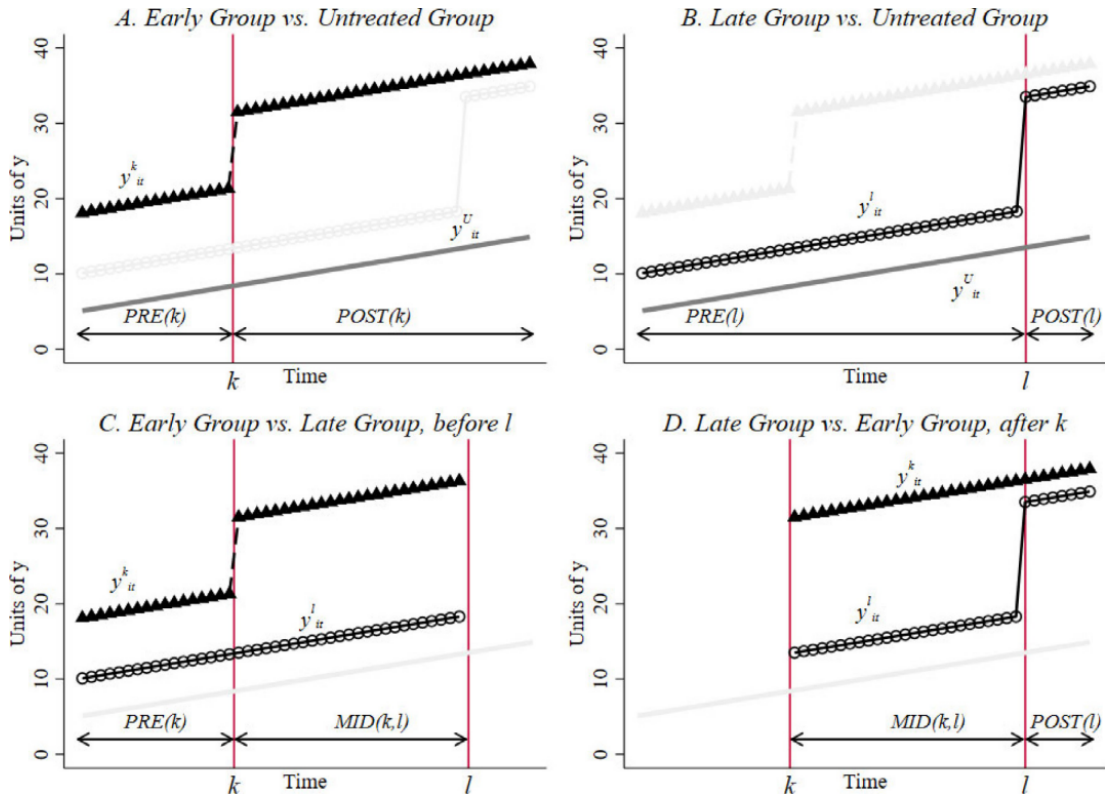


圖 6: Goodman-Bacon (2021), Figure 2。

我們利用 Abadie, Diamond, and Hainmueller (2010) 的符號來介紹 SCM。首先，令所有成員為 $i = 1, \dots, J + 1$ 且時間為 $t = 1, \dots, T$ 。其中 $i = 1$ 代表處理組成員而 $i = 2, \dots, J + 1$ 為其他 J 個控制組成員，而 $t = 1, \dots, T_0$ 代表事前而 $t = T_0 + 1, \dots, T$ 為事後。定義 $(Y_{1t}^1, Y_{1t}^0)'$ 為處理組成員在第 t 期若接受處理或若不接受處理的潛在結果，而 Y_{1t} 為處理組成員在第 t 期的實際觀察結果，則事後第 t 期 ($t > T_0$) 的 ITT 即為：

$$ITT(t) \equiv Y_{1t}^1 - Y_{1t}^0 = Y_{1t} - Y_{1t}^0$$

與 (15) 類似，只要能認定事後的 Y_{1t}^0 則可認定 $ITT(t)$ 。SCM 的認定條件建立 (但不限於) 以下因子模型 (factor model)。對 $i = 1, \dots, J + 1$ 和 $t = 1, \dots, T$ 而言，令

$$Y_{it}^0 = \delta_t + \theta_t' \mathbf{Z}_i + \lambda_t' \boldsymbol{\mu}_i + \varepsilon_{it} \quad (21)$$

其中 δ_t 為時間趨勢， \mathbf{Z}_i 和 $\boldsymbol{\mu}_i$ 分別為可觀察和不可觀察的預測變數向量，而 θ_t 和 λ_t 則為隨時間變動的參數向量。Abadie, Diamond, and Hainmueller (2010) 證明了若控

制組權重 $\mathbf{W}^* \equiv (w_2^*, \dots, w_{J+1}^*)'$ 滿足下列**假設 8**，則隨著 T_0 增加，該權重亦可用於合成事後 ($t > T_0$) 的 Y_{1t}^0 ，亦即：

$$Y_{1t}^0 \approx \sum_{i=2}^{J+1} w_i^* Y_{it}$$

假設 8 (凸性). 對 $t = 1, \dots, T_0$ 而言， $Y_{1t} = \sum_{i=2}^{J+1} w_i^* Y_{it}$ 且 $\mathbf{Z}_1 = \sum_{i=2}^{J+1} w_i^* \mathbf{Z}_i$ 。

我們接著說明如何用**假設 8** 得到認定結果。從潛在結果的定義和 (21) 可得：

$$\begin{aligned} Y_{1t}^0 - \sum_{i=2}^{J+1} w_i^* Y_{it} &= \boldsymbol{\theta}_t \left(\mathbf{Z}_1 - \sum_{i=2}^{J+1} w_i^* \mathbf{Z}_i \right) \\ &+ \boldsymbol{\lambda}_t \left(\boldsymbol{\mu}_1 - \sum_{i=2}^{J+1} w_i^* \boldsymbol{\mu}_i \right) + \left(\varepsilon_{1t} - \sum_{i=2}^{J+1} w_i^* \varepsilon_{it} \right) \end{aligned} \quad (22)$$

而上式當 $t = 1, \dots, T_0$ 時 $Y_{1t}^0 = Y_{1t}$ 可被觀察。由此可知，若在**假設 8** 成立下控制組權重 \mathbf{W}^* 能有效擬合事前結果變數 Y_{11}, \dots, Y_{1,T_0} 與可觀察預測變數 \mathbf{Z}_1 ，則該權重擬合不可觀察預測變數 $\boldsymbol{\mu}_1$ 的表現應同樣優異，畢竟 (22) 最後一項的期望值為零，隨著 T_0 增加可利用不等式限制其大小。綜上所述，該控制組權重可用於擬合不受時間影響的 \mathbf{Z}_1 和 $\boldsymbol{\mu}_1$ ，也可進一步透過 (22) 擬合事後 ($t > T_0$) 的 Y_{1t}^0 。

基於上述結果， Y_{1t}^0 的 SCM 估計式為：

$$\hat{Y}_{1t}^0 = \sum_{i=2}^{J+1} \hat{w}_i Y_{it}$$

其中控制組權重 $\hat{\mathbf{W}} = (\hat{w}_2, \dots, \hat{w}_{J+1})'$ 為下列受限制最小化問題的解：

$$\begin{aligned} \hat{\mathbf{W}} &= \operatorname{argmin}_{\mathbf{W} \in \mathbb{W}} \|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}\|_{\mathbf{V}} \\ \text{s.t. } \mathbb{W} &= \{\mathbf{W} \in [0, 1]^J : \boldsymbol{\iota}' \mathbf{W} = 1\} \end{aligned}$$

其中 $\mathbf{X}_1 \equiv (Y_{11}, \dots, Y_{1,T_0}, \mathbf{Z}_1)'$ 為處理組擬合變數向量， $\mathbf{X}_0 \equiv (\mathbf{X}_2, \dots, \mathbf{X}_{J+1})$ 為 J 個控制組成員擬合變數合併後的矩陣， \mathbf{V} 為衡量擬合變數重要程度的對角矩陣，而 $\|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}\|_{\mathbf{V}} \equiv \sqrt{(\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W})^\top \mathbf{V} (\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W})}$ 且 $\boldsymbol{\iota} = (1, \dots, 1)'$ 。文獻中有許多估計 \mathbf{V} 的方法，除了由研究者主觀定義其重要程度之外，Abadie and Gardeazabal (2003) 亦提出基於交叉驗證 (cross-validation) 的方法，有興趣的讀者可參閱 Abadie (2021)。

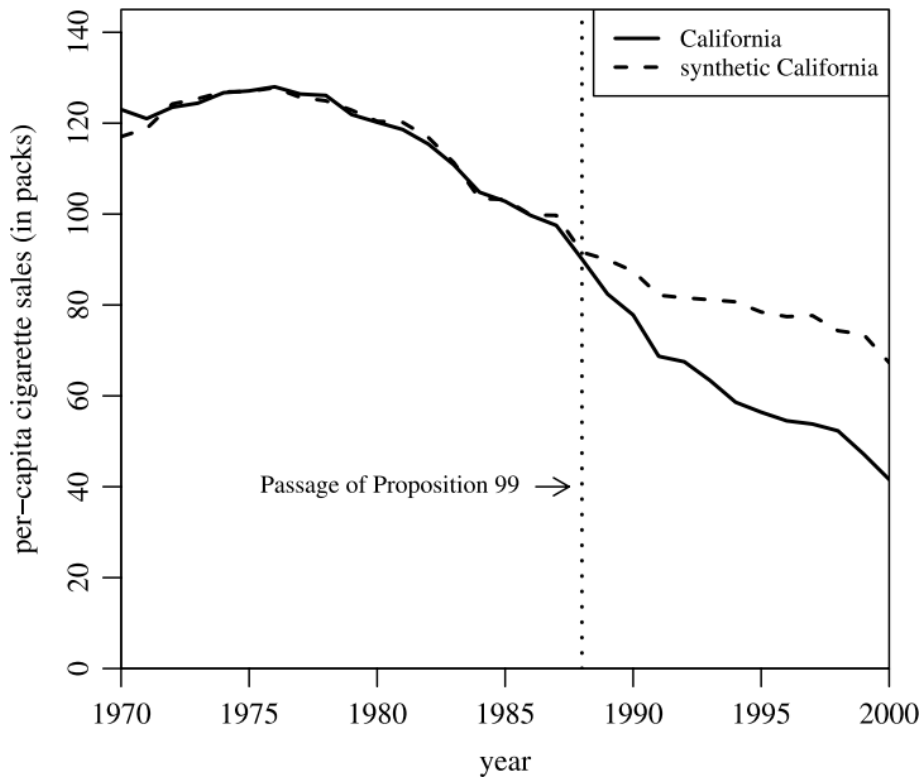


圖 7: [Abadie, Diamond, and Hainmueller \(2010\)](#), Figure 2。

利用 SCM 與 1970-2000 年美國各州資料，[Abadie, Diamond, and Hainmueller \(2010\)](#) 探討加州於 1988 年實施的菸草控制法案是否能有效遏止香菸消費。以其餘 38 個在樣本期間未推動相關法案或大幅調整菸稅的州作為潛在控制組，該研究擬合了香菸價格、人均所得、青少年人口比例、人均啤酒消費量、以及事前（1975、1980、1988 年）的香菸消費量來得到未實施菸草控制法案的「合成加州」：

$$\text{合成加州} = 0.33 \text{ 猶他} + 0.23 \text{ 內華達} + 0.20 \text{ 蒙大拿} + 0.16 \text{ 科羅拉多} + 0.07 \text{ 康乃狄克}$$

由圖 7 可看出真實加州與「合成加州」的香菸消費量軌跡在 1988 年前幾乎重疊，證實了 SCM 的有效性。從圖中亦可看出真實加州的香菸消費量在 1988 年後比「合成加州」下降不少，可視為該法案能有效降低香菸消費的支持性證據。

7. 結論

本文回顧計量經濟學在過去 30 多年間發展的因果推論方法及應用。這些方法不僅帶來學術上的「因果革命」，更提供許多可信實證給施政者作為參考依據。表 4 提

表 4: 本文所述方法與假設。

方法	同質／異質效應	估計目標	資料	假設和模型
迴歸調整	同質	ATE = ATT	橫斷面資料	假設 1 和 (3)
迴歸調整	異質	ATE/ATT	橫斷面資料	假設 1 和 (5)
工具變數	同質	ATE = ATT	橫斷面資料	假設 2 和 (9) 和 (10)
工具變數	異質	LATE	橫斷面資料	假設 2 至 4
工具變數	異質	LATE = ATT	橫斷面資料	假設 2 至 5
差異中差異	同質	ATT	追蹤資料	假設 6 和 (17)
差異中差異	異質	ATT(g, t)	追蹤資料	假設 7
合成控制法	異質	ITT(t)	追蹤資料	假設 8 和 (21)

供了本文所述方法的統整比較，以幫助讀者更好地理解每種方法的適用範圍和限制。然而礙於篇幅所限，文獻中還有許多相關研究議題無法詳列，例如分位數處理效應 (quantile treatment effect)、中介處理效應 (mediated treatment effect)、條件處理效應 (conditional treatment effect) 以及最適處理指派 (optimal treatment assignment)。有興趣的讀者可進一步搜尋查閱。

8. 致謝

作者感謝期刊主編及匿名評審的寶貴意見與建議，以及科技部專題研究計畫 (MOST 110-2628-H-194-001-MY2) 的部分經費補助。文中如有任何疏誤，悉由作者負責。

參考文獻

- [1] Abadie, A. (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models. *Journal of Econometrics*, 113, pages 231-263.
- [2] Abadie, A. (2021). Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature*, 59(2), pages 391-425.
- [3] Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco

- Control Program. *Journal of the American statistical Association*, 105(490), pages 493-505.
- [4] Abadie, A., and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, 93(1), pages 113-132.
- [5] Angrist, J. D., Dynarski, S. M., Kane, T. J., Pathak, P. A., and Walters, C. R. (2010). Inputs and Impacts in Charter Schools: KIPP Lynn. *American Economic Review*, 100(2), pages 239-243.
- [6] Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434), pages 444-455.
- [7] Angrist, J. D., and Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, 106(4), pages 979-1014.
- [8] Angrist, J. D., and Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, United States.
- [9] Angrist, J. D., and Pischke, J. S. (2015). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.
- [10] Athey, S., and Imbens, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives*, 31(2), pages 3-32.
- [11] Bloom, H. S. (1984). Accounting for No-Shows in Experimental Evaluation Designs. *Evaluation Review*, 8(2), pages 225-246.
- [12] Callaway, B., and Sant'Anna, P. H. C. (2021). Difference-in-Differences With Multiple Time Periods. *Journal of Econometrics*, 225(2), pages 200-230.
- [13] Card, D., and Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4), pages 772-793.

-
- [14] Card, D., and Krueger, A. B. (2000). Minimum Wages and Employment: A case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply. *American Economic Review*, 90(5), pages 1397-1420.
- [15] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased Machine Learning for Treatment and Structural Parameters. *Econometrics Journal*, 21(1), pages C1-C68.
- [16] Dale, S. B., and Krueger, A. B. (2002). Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables. *Quarterly Journal of Economics*, 117(4), pages 1491-1527.
- [17] De Chaisemartin, C., and d'Haultfoeuille, X. (2020). Two-Way Fixed Effects Estimators With Heterogeneous Treatment Effects. *American Economic Review*, 110(9), pages 2964-2996.
- [18] Donald, S. G., Hsu, Y. C., and Lieli, R. P. (2014). Testing the Unconfoundedness Assumption via Inverse Probability Weighted Estimators of (L)ATT. *Journal of Business and Economic Statistics*, 32(3), pages 395-415.
- [19] Goodman-Bacon, A. (2021). Difference-in-Differences With Variation in Treatment Timing. *Journal of Econometrics*, 225(2), pages 254-277.
- [20] Hernán M. A., and Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman and Hall/CRC.
- [21] Imbens, G. W., and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), pages 467-475.
- [22] Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.
- [23] Imbens, G. W., and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1), pages 5-86.
- [24] Mincer, J. (1974). *Schooling, Experience, and Earnings*. Columbia University Press.

- [25] Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5), pages 688-701.
- [26] Swanson, S. A., Hernán, M. A., Miller, M., Robins, J. M., and Richardson, T. S. (2018). Partial Identification of the Average Treatment Effect Using Instrumental Variables: Review of Methods for Binary Instruments, Treatments, and Outcomes. *Journal of the American Statistical Association*, 113(522), pages 933-947.
- [27] Wald, A. (1940). The Fitting of Straight Lines if Both Variables are Subject to Error. *Annals of Mathematical Statistics*, 11(3), pages 284-300.
- [28] 許育進與賴宗志 (2018)。處理效果文獻回顧。《經濟論文叢刊》，46 卷 4 期，頁 501-521。

[Received October 2023; accepted February 2024.]

Causal Inference in Economics: Theoretical Development and Applications

Tsung-Chih Lai[†]

Department of Economics, National Chung Cheng University

ABSTRACT

Program evaluation has occupied a pivotal role in empirical economics, and evidence-based policy making has ascended to a prevailing norm in public policy worldwide. Over the past 30 years, econometricians have developed methods capable of elucidating not only correlations but also causality. This "causal revolution" has led to three distinguished scholars being awarded the Nobel Prize in Economics in 2021 for their seminal contributions to understanding causal relationships. This article aims to introduce the theoretical development of causal inference in economics and use the canonical empirical applications of Nobel laureates and others as illustrative examples.

Key words and phrases: treatment effect, regression adjustment, instrumental variable, difference-in-differences, synthetic control method.

JEL classification: C21, C31.

[†]Corresponding to: Tsung-Chih Lai
E-mail: tc lai@ccu.edu.tw