DOI: 10.6277/TER.201812_46(4).0001

處理效果文獻回顧

許育進・賴宗志*

處理效果文獻在過去30年間有不少新的理論發展, 這篇文章將回顧在 Rubin 因果模型中, 當條件獨立假設成立時, 如何以機率倒數加權法得到平均處理效果之認定及估計結果。我們也將說明在此假設不成立的內生性架構之下, 該如何利用工具變數來對局部平均處理效果進行機率倒數加權之認定與估計。最後, 我們介紹一個對條件獨立假設的統計檢定, 並總結近年來處理效果文獻的發展, 以及提出幾個未來可能的研究方向。

關鍵詞: Rubin 因果模型, 處理效果, 傾向分數, 機率倒數加權
JEL 分類代號: C21, C31

1 前言

衡量政策效果一直以來都是經濟、財務及會計領域中的重要議題。在過去的30年裡, 計量經濟學文獻發展出許多有關因果推論的理論文章, 其中也包含了數篇相當知名的文獻綜述, 例如 Imbens (2004) 和 Imbens and Wooldridge (2009)。然而, 中文文獻中仍缺乏對於處理效果模型的系統性整理, 因此, 我們希望能藉由這篇文章說明處理效果的基本模型架構, 並回顧近年來的理論發展。[1]

首先, 我們在第2節介紹處理效果文獻中最常採用的 Rubin 因果模型。此模型由 Rubin (1974) 等一系列文章發展而成, 其中關鍵在於引進潛在

*中央研究院經濟研究所副研究員、國立中央大學財務金融學系合聘副教授與國立政治大學經濟學系合聘副教授與逢甲大學經濟學系助理教授。許育進爲通訊作者。作者感謝編輯委員及匿名評審的寶貴意見, 我們受益良多。本文若有任何謬誤, 當屬作者之責。

[1]這篇文章僅回顧處理效果文獻中部分的結果。

經濟論文叢刊 (*Taiwan Economic Review*), 46:4 (2018), 501–521。
國立台灣大學經濟學系出版

結果 (potential outcomes) 之概念, 使研究者可輕易透過此模型定義各種因果推論或政策參數。我們接著介紹在控制可觀察個體特徵之下, 當處理指標與潛在結果互為條件獨立 (unconfounded 或 conditionally independent) 時, 研究者該如何針對平均處理效果 (average treatment effect, ATE) 及受處理族群的平均處理效果 (average treatment effect on the treated, ATT) 進行機率倒數加權 (inverse probability weighting) 之無母數認定及估計。除了介紹無母數估計方法外, 我們也將簡單討論這些估計式的漸近分配與半參數效率等統計性質。

接著, 第3節討論當條件獨立假設不成立, 而研究者僅能依靠一個二元 (binary) 工具變數時, 該如何在此內生性架構下把母體分為四個子族群, 並針對服從者族群的處理效果進行分析。具體來說, 本節將討論如何利用機率倒數加權法, 來對局部平均處理效果 (local average treatment effect, LATE) 及受處理族群的局部平均處理效果 (local average treatment effect on the treated, LATT) 進行無母數認定及估計。然而, 儘管 LATE 及 LATT 所需之假設較弱, 一般而言研究者感興趣的政策參數還是定義在整個母體的 ATE 及 ATT。換句話說, 條件獨立假設的成立與否仍十分重要, 我們也在第4節中介紹一個利用工具變數來檢驗此假設的統計檢定。

最後, 第5節回顧一些處理效果文獻近年來的發展。其中包括了可用來衡量組內異質性的分位數處理效果 (quantile treatment effect, QTE)、可衡量外部效度的平均虛擬處理效果 (average counterfactual treatment effect, ACTE), 以及可衡量組間異質性的條件平均處理效果 (conditional average treatment effect, CATE)。在第6節裡我們也建議一些處理效果理論未來可能的研究方向。

2 條件獨立假設下的處理效果

2.1 Rubin 因果模型

Rubin 因果模型 (Rubin, 1974) 在處理效果文獻中是最基本, 也是最被廣泛應用的模型之一。使用此模型的好處在於能輕易地定義政策參數, 並釐清其所帶來的因果效應。一個傳統的 Rubin 因果模型包含了以下變數: $D \in \{0, 1\}$ 是個二元變數, 通常稱為處理指標 (treatment indicator) 或處理

指派 (treatment assignment)。當 D 等於1時, 代表某個體接受處理, 而 D 等於0則代表此個體並無接受處理。$Y(1)$ 及 $Y(0)$ 爲一組潛在結果 (potential outcomes)。$Y(1)$ 代表某個體若接受處理時 (也就是當 $D = 1$ 時), 該個體最終會得到的結果。而 $Y(0)$ 則代表某個體若不接受處理 ($D = 0$) 其最終會出現的結果。很明顯地, 由於實際上同一觀察值的處理指標只可能爲0或1, 代表這一組潛在結果只會有一個存在於現實世界中。換句話說, 若將實際觀察到的結果表示爲 Y, 則實際結果與潛在結果會有以下關係: $Y = DY(1) + (1 - D)Y(0)$。除了 D 和 Y 之外, 我們也可觀察到每個個體的一組特徵向量 X。總而言之, 對每個個體我們都能觀察到 (Y, D, X), 然而, 同一時間下我們只能觀察到潛在結果 $Y(1)$ 或 $Y(0)$ 其中之一。

2.2 平均處理效果

在評估某政策是否有效時, 政策制定者與研究人員最爲關心的政策參數可能是平均處理效果 (average treatment effect, ATE)。ATE 衡量了整個母體的平均效應, 也就是平均而言, 接受處理會比不接受處理帶來多少影響。其數學定義爲:

$$\beta = E[Y(1) - Y(0)]。 \tag{1}$$

除了 ATE 以外, 另一政策制定者感興趣的參數爲受處理族群的平均處理效果 (average treatment effect on the treated, ATT)。顧名思義, ATT 衡量的是對於實際上接受處理的這群個體而言, 其平均處理效果爲何。ATT 數學上的定義爲:

$$\beta_t = E[Y(1) - Y(0)|D = 1]。 \tag{2}$$

相對於 ATT, 我們也能定義未受處理族群的平均處理效果 (average treatment effect on the non-treated, ATNT)。ATNT 的定義爲 $\beta_{nt} = E[Y(1) - Y(0)|D = 0]$, 其所提供的資訊爲實際上未接受處理的族群若接受處理時, 其平均效果應當爲何。然而, ATT 跟 ATNT 的認定跟估計結果相似, 因此接下來的討論將著重於 ATE 與 ATT。

2.3 平均處理效果的認定與估計

在 Rubin 因果模型的框架中, 由於每個個體的潛在結果並不能同時被觀測, 代表我們無法根據式 (1) 及 (2) 來直接估計 ATE 及 ATT。也就是說, 因爲有缺失變量 (missing variable) 的問題, 我們必須引入額外的假設才能得到認定結果。

第一個假設可追溯至 Rosenbaum and Rubin (1983), 它要求在控制可觀察的個體特徵後, 處理指標與潛在結果必須互爲條件獨立 (unconfounded 或 conditionally independent)。換句話說, 這個假設允許存在樣本選擇偏誤, 只是此選擇偏誤必須被可觀察之個體特徵所刻劃。若兩個體具有相同的可觀察特徵, 則接受處理與否應與其他 (同時會影響潛在結果的) 特徵無系統性關連。此假設的正式定義如下:

假設 2.1. 條件獨立: $D \perp (Y(1), Y(0))|X$。

接著, 第二個假設跟傾向分數函數 (propensity score function) 有關。我們首先定義傾向分數函數爲 $p(x) = P(D = 1|X = x)$, 也就是某群特徵爲 $X = x$ 的個體會接受處理的條件機率。

假設 2.2. 傾向分數函數: 對於 $0 < \delta < 1/2, \delta \leq p(x) \leq 1 - \delta$。

假設2.2要求傾向分數函數不能太接近0或1, 代表著擁有相同特徵的一群個體中, 必須有些個體接受處理, 也有些個體不接受處理。此假設一般被稱爲重疊假設 (overlap assumption), 主要原因來自於它要求接受及不接受處理這兩個族群中的特徵支撐集合 (support) 必須互相重疊。若非重疊, 則表示一定有某群個體全都選擇接受或都不接受處理, 無法判定此群個體在與事實相反狀態下的潛在結果應該爲何。

給定假設2.1及2.2, 可應用 Horvitz and Thompson (1952) 來證明 ATE 及 ATT 的認定式爲:

$$\beta = E\left[\frac{DY}{p(X)} - \frac{(1-D)Y}{1-p(X)}\right], \tag{3}$$

$$\beta_t = E\left[p(X)\left(\frac{DY}{p(X)} - \frac{(1-D)Y}{1-p(X)}\right)\right] \Big/ P(D=1)。 \tag{4}$$

我們簡單說明式 (4) 的認定過程。首先, 由迭代期望值定律 (law of iterated expectations)、假設2.1, 與 $p(x) = f_{X,D}(x,1)/f_X(x)$ 可得

$$\begin{aligned}
\beta_t &= E[Y(1) - Y(0)|D = 1] \\
&= E[E[Y(1) - Y(0)|X, D = 1]|D = 1] \\
&= E[E[Y(1) - Y(0)|X]|D = 1] \\
&= \int E[Y(1) - Y(0)|X = x] f_{X|D}(x|1) dx \\
&= \int E[Y(1) - Y(0)|X = x] \frac{f_{X,D}(x,1)}{f_X(x)} \frac{f_X(x)}{P(D = 1)} dx \\
&= \frac{E[p(X)E[Y(1) - Y(0)|X]]}{P(D = 1)}。
\end{aligned}$$

接著, 由 $p(X) = E[D|X]$、假設2.1, 以及 $Y = DY(1) + (1 - D)Y(0)$ 可進一步推導

$$\begin{aligned}
E[Y(1) - Y(0)|X] &= \frac{E[D|X]E[Y(1)|X]}{p(X)} - \frac{E[1 - D|X]E[Y(0)|X]}{1 - p(X)} \\
&= \frac{E[DY(1)|X]}{p(X)} - \frac{E[(1 - D)Y(0)|X]}{1 - p(X)} \\
&= E\left[\frac{DY(1)}{p(X)} - \frac{(1 - D)Y(0)}{1 - p(X)} \middle| X \right] \\
&= E\left[\frac{DY}{p(X)} - \frac{(1 - D)Y}{1 - p(X)} \middle| X \right]。
\end{aligned}$$

其中, 假設2.2必須成立才能使上式分母不爲0。最後, 再一次利用迭代期望值定律即可得到式 (4) 之認定結果。

給定一組隨機抽樣資料 $\{(Y_i, D_i, X_i)\}_{i=1}^n$, 根據認定式 (3) 及 (4), 可以利用機率倒數加權 (inverse probability weighting, IPW) 方法來估計 ATE 及 ATT:

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n \frac{D_i Y_i}{\hat{p}(X_i)} - \frac{(1 - D_i) Y_i}{1 - \hat{p}(X_i)}, \tag{5}$$

$$\hat{\beta}_t = \frac{1}{n} \sum_{i=1}^n \hat{p}(X_i) \left(\frac{D_i Y_i}{\hat{p}(X_i)} - \frac{(1 - D_i) Y_i}{1 - \hat{p}(X_i)} \right) \Big/ \frac{1}{n} \sum_{i=1}^n D_i, \tag{6}$$

其中 $\hat{p}(x)$ 是傾向分數函數的無母數估計式。

傾向分數函數的無母數估計式可以有很多種, 例如級數羅吉特法 (series logit method, Hirano, Imbens, and Ridder (2003))、局部多項式法 (local polynomial method, Ichimura and Linton (2005)), 以及高階內核法 (higher-order kernel method, Li, Racine, and Wooldridge (2009))。我們在此簡單說明如何操作級數羅吉特法來估計傾向分數函數。令 $R^K(x) = (x^{\lambda(1)}, x^{\lambda(2)}, \cdots, x^{\lambda(K)})'$ 爲一個由 x 的冪函數 (power function) 所組成之升冪級數向量, 其中對於 $k = 1, \cdots, K$, $\lambda(k)$ 爲非負整數, 且爲 k 的單調遞增函數。令 $L(a) = \exp(a)/(1 + \exp(a))$ 爲羅吉斯函數 (logistic function)。傾向分數函數的估計式即爲

$$\hat{p}(x) = L\left(R^K(x)'\hat{\pi}_K\right),$$

其中

$$\begin{aligned}\hat{\pi}_K = \arg\min_{\pi_K} \frac{1}{n}\sum_{i=1}^{n}\Big(D_i \cdot \ln L\left(R^K(X_i)'\pi_k\right)\\ + (1 - D_i)\cdot \ln\left(1 - L\left(R^K(X_i)'\pi_k\right)\right)\Big)。\end{aligned}$$

由上式可發現, 使用級數羅吉特法的好處在於其估計值必然落在0與1之間, 符合 (條件) 機率之性質。

更值得一提的是, 在不同的正規條件 (regularity conditions) 之下, 無論用上述何種無母數方法所得的 ATE 及 ATT 估計式都具有漸進常態分配的性質。也就是說,

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}), \tag{7}$$

$$\sqrt{n}\left(\hat{\beta}_t - \beta_t\right) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_t), \tag{8}$$

且

$$\begin{aligned}\mathcal{V} = E\Bigg[&\frac{Var(Y(1)|X)}{p(X)} + \frac{Var(Y(0)|X)}{1 - p(X)}\\ &+ (E[Y(1) - Y(0)|X] - \beta)^2\Bigg],\end{aligned}$$

$$\mathcal{V}_t = E\left[\frac{p(X)^2}{E[p(X)]^2}\left(\frac{Var(Y(1)|X)}{p(X)} + \frac{Var(Y(0)|X)}{1-p(X)} + (E[Y(1)-Y(0)|X]-\gamma)^2\right)\right]。$$

其中, Hahn (1998) 證明了無論傾向分數函數已知或未知, ATE 之半參數效率性界限 (semiparametric efficiency bound) 與式 (7) 的漸近變異數 $\mathcal{V}$ 皆相同, 代表式 (5) 的 ATE 估計式具有半參數效率 (semiparametric efficient) 性質。然而, Hahn (1998) 也指出當傾向分數函數未知時, ATT 的半參數效率性界限會大於 $\mathcal{V}_t$, 代表式 (6) 之無母數 ATT 估計式比起用真實 (母體) 傾向分數函數所得的 ATT 估計式還來得有效率。對 ATE 及 ATT 之漸進分配及半參數效率性界限有興趣的讀者, 請參見 Hahn (1998) 及 Hirano, Imbens, and Ridder (2003)。統計推論的部分也請參考 Hirano, Imbens, and Ridder (2003)。

3 內生處理指派下的處理效果

第2節討論在條件獨立假設下處理效果的認定及估計, 在這一節中, 我們將討論當處理指派爲內生 (endogenous) 時, 也就是假設 2.1 不成立的情況下該如何進行分析。

我們再定義一些變數。首先, 假設存在一個二元的工具變數 (instrumental variable) $Z \in \{0, 1\}$。透過此工具變數, 我們可以進一步定義潛在處理狀態 (potential treatment status) 爲 $D(1)$ 及 $D(0)$。與潛在結果的概念類似, $D(1)$ 代表當某個體的工具變數爲1時, 此個體會被觀察到的處理狀態。$D(0)$ 則代表此個體的工具變數爲0時相對應的處理狀態。很明顯地, 潛在結果與潛在處理狀態最大不同處在於潛在結果可以是任意數值, 但潛在處理狀態只會是1或0。無論 $z = 1$ 或0, 當 $D(z) = 1$ 時代表了接受處理, 而 $D(z) = 0$ 則代表沒有接受處理。同樣地, 實際觀察到的處理狀態與潛在處理狀態也存在以下的關係: $D = ZD(1) + (1 - Z)D(0)$。

由於 $D(1)$ 及 $D(0)$ 皆爲二元變數且同一時間下只能觀察到其中之一, Angrist, Imbens, and Rubin (1996) 根據四個可能的組合把母體分成四個

族群:

$$\begin{cases} \text{總是接受者 (Always-Taker)} & \text{當} D(1) = D(0) = 1; \\ \text{永不接受者 (Never-Taker)} & \text{當} D(1) = D(0) = 0; \\ \text{服從者 (Complier)} & \text{當} D(1) = 1 \text{ 及 } D(0) = 0; \\ \text{不服從者 (Defier)} & \text{當} D(1) = 0 \text{ 及 } D(0) = 1\text{。} \end{cases}$$

顧名思義, 無論當工具變數的值爲何, 總是接受者一定會接受處理。相反地, 永不接受者無論在任何情況下都不會接受處理。至於服從者是否接受處理則全然受到工具變數的影響: 當工具變數爲1時, 服從者會接受處理; 當工具變數爲0時, 服從者則不接受處理。不服從者最爲叛逆, 接受處理與否跟工具變數之值完全相反: 當工具變數爲1時, 不服從者不接受處理; 當工具變數爲0時, 不服從者則選擇接受處理。

Imbens and Angrist (1994) 在此模型架構中定義了局部平均處理效果 (local average treatment effect, LATE) 以衡量服從者中的平均處理效果。其數學定義爲:

$$\tau = E[Y(1) - Y(0)|D(1) = 1, D(0) = 0]\text{。} \tag{9}$$

類似於 ATT, Donald and Hsu (2014) 針對受處理族群的局部平均處理效果 (local average treatment effect on the treated, LATT) 進行分析。LATT 數學上的定義爲:

$$\tau_t = E[Y(1) - Y(0)|D(1) = 1, D(0) = 0, D = 1]\text{。} \tag{10}$$

由於假設2.1不成立, 我們介紹一組新的假設來幫助認定 LATE 及 LATT。首先, 先定義工具變數傾向分數函數 (instrument propensity score function) 爲 $q(x) = P(Z = 1|X = x)$。

假設 3.1. 工具變數條件獨立: $(Y(0), Y(1), D(1), D(0)) \perp Z|X$。

假設 3.2. 工具變數傾向分數函數: 對於 $0 < \delta < 1/2, \delta \leq q(x) \leq 1 - \delta$。

假設 3.3. 單調性: $P(D(1) \geq D(0)) = 1$。

假設 3.4. 工具變數: $P(D(1) = 1) > P(D(0) = 1)$。

假設3.1的無條件 (unconditional) 版本由 Imbens and Angrist (1994) 所提出, 而 Abadie (2003) 及 Frölich (2007) 則放寬限制至目前的版本, 也就是在控制個體特徵 X 後, 工具變數 Z 和所有潛在變數互相獨立即可。假設3.2則跟假設2.2相似, 只是換成要求 $q(x)$ 不能太過接近0或1。假設3.3透過要求潛在處理狀態爲工具變數的單調遞增函數, 來排除掉母體中存在不服從者 ($D(1)=0$ 及 $D(0)=1$) 的可能性。假設3.4則藉由要求工具變數與接受處理的機率爲正相關, 以確保母體中不僅有總是接受者與永不接受者, 同時也存在我們感興趣的服從者族群。

在假設3.1、3.2、3.3及3.4成立之下, Donald and Hsu (2014) 提出可將 LATE/LATT 表示成 Z 對 Y 的 ATE/ATT 除上 Z 對 D 的 ATE/ATT, 並且進一步證明 LATE 及 LATT 的認定式爲:

$$\tau = \frac{E\left[\frac{ZY}{q(X)} - \frac{(1-Z)Y}{1-q(X)}\right]}{E\left[\frac{ZD}{q(X)} - \frac{(1-Z)D}{1-q(X)}\right]}, \tag{11}$$

$$\tau_t = \frac{E\left[q(X)\left(\frac{ZY}{q(X)} - \frac{(1-Z)Y}{1-q(X)}\right)\right] \Big/ P(Z=1)}{E\left[q(X)\left(\frac{ZD}{q(X)} - \frac{(1-Z)D}{1-q(X)}\right)\right] \Big/ P(Z=1)}。 \tag{12}$$

LATE 及 LATT 之無母數機率倒數加權估計式則爲:

$$\hat{\tau} = \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i Y_i}{\hat{q}(X_i)} - \frac{(1-Z_i)Y_i}{1-\hat{q}(X_i)}}{\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i D_i}{\hat{q}(X_i)} - \frac{(1-Z_i)D_i}{1-\hat{q}(X_i)}}, \tag{13}$$

$$\hat{\tau}_t = \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{q}\,(X_i)\left(\frac{Z_i Y_i}{\hat{q}(X_i)} - \frac{(1-Z_i)Y_i}{1-\hat{q}(X_i)}\right) \Big/ \frac{1}{n}\sum_{i=1}^{n} Z_i}{\frac{1}{n}\sum_{i=1}^{n} \hat{q}\,(X_i)\left(\frac{Z_i D_i}{\hat{q}(X_i)} - \frac{(1-Z_i)D_i}{1-\hat{q}(X_i)}\right) \Big/ \frac{1}{n}\sum_{i=1}^{n} Z_i}, \tag{14}$$

其中, $\hat{q}(x)$ 與 $\hat{p}(x)$ 類似, 都可以透過前述之無母數估計方法來進行估計。

由於 LATE/LATT 的估計式是由兩個 ATE/ATT 的估計式相除, 所以漸進常態的結果可以直接應用 delta 法 (delta method) 求得。也就是說,

$$\sqrt{n}\left(\hat{\tau} - \tau\right) \xrightarrow{d} \mathcal{N}\left(0, E\left[\psi(Y, D, Z, X)^2\right]\right), \tag{15}$$

$$\sqrt{n}\left(\hat{\tau}_t - \tau_t\right) \xrightarrow{d} \mathcal{N}\left(0, E\left[\psi_t(Y, D, Z, X)^2\right]\right), \tag{16}$$

其中, 影響函數 (influence function) ψ 及 ψ_t 分別爲

$$\begin{aligned}
\psi(Y, D, Z, X) = \frac{1}{\Gamma}\Bigg\{ & \frac{Z\left[Y - m_1(X) - \tau\left(D - \mu_1(X)\right)\right]}{q(X)} \\
& - \frac{(1-Z)\left[Y - m_0(X) - \tau\left(D - \mu_0(X)\right)\right]}{1 - q(X)} \\
& + m_1(X) - m_0(X) - \tau\left(\mu_1(X) - \mu_0(X)\right)\Bigg\}, \\
\psi_t(Y, D, Z, X) = \frac{q(X)}{P(Z=1)\Gamma_t}\Bigg\{ & \frac{Z\left[Y - m_1(X) - \tau_t\left(D - \mu_1(X)\right)\right]}{q(X)} \\
& - \frac{(1-Z)\left[Y - m_0(X) - \tau_t\left(D - \mu_0(X)\right)\right]}{1 - q(X)} \\
& + \frac{Z\left[m_1(X) - m_0(X) - \tau_t\left(\mu_1(X) - \mu_0(X)\right)\right]}{q(X)}\Bigg\},
\end{aligned}$$

對於 $z = 1$ 或 0, $m_z(x) = E[Y|X = x, Z = z]$, $\mu_z(x) = E[D|X = x, Z = z]$ 且

$$\begin{aligned}
\Gamma &= E\left[\frac{ZD}{q(X)} - \frac{(1-Z)D}{1-q(X)}\right], \\
\Gamma_t &= E\left[q(X)\left(\frac{ZD}{q(X)} - \frac{(1-Z)D}{1-q(X)}\right)\right] \Big/ P(Z=1).
\end{aligned}$$

對半參數效率性質有興趣的讀者可進一步參閱 Frölich (2007), Hong and Nekipelov (2010) 及 Donald and Hsu (2014)。

4 條件獨立假設的檢定

即便在內生性架構下可以估計出服從者的平均處理效果, LATE 及 LATT 仍然遭受到一些批評聲浪。首先, 四個子族群的分組與所使用的工具變數有關, 當實證問題中同時存在兩個或多個合理的二元工具變數時, 使用不同工具變數所定義的服從者族群也會跟著改變, 代表 LATE 及 LATT 會隨之變化爲不同族群的平均處理效果。另一方面來說, 由於四個子族群是由潛在處理狀態 $(D(1), D(0))$ 所定義而成, 但潛在處理狀態與潛在結果相似, 在同一時間點只能觀察到其中一個變數, 意味著研究者實際上無法

分辨出哪些個體為服從者, 哪些又是總是接受者或永不接受者。換句話說, 即使我們能得到服從者的平均處理效果, 但我們卻不能確定這會是哪些個體的平均效果。

正因如此, 比起 LATE 及 LATT, ATE 與 ATT 這兩個定義明確的平均處理效果還是受到更多實證研究的青睞。然而如同上一節所述, 在內生性框架中我們只能得到 LATE 跟 LATT 的認定及估計, 無法更進一步分析 ATE 跟 ATT。因此, 檢驗條件獨立假設的合理性便顯得十分重要。雖然這個假設在沒有額外條件下是不能被檢定的 (untestable), 但 Donald and Hsu (2014) 提出一個可在工具變數架構下檢驗假設2.1的統計檢定。為了介紹這項檢定, 我們先引進以下的假設:

假設 4.1. 單邊不服從: $P(D(0)=0)=1$。

單邊不服從假設指當某個體的工具變數 $Z=0$ 時, 此個體一定不會接受處理, 也就是 $D=D(0)=0$。然而當 $Z=1$ 時, 我們允許個體接受或不接受處理, 代表在此情況下 D 跟 Z 的值可以不相同。這個假設可適用於許多實證案例, 舉例來說, 為了檢驗就業培訓計畫的成效, 許多隨機試驗會將申請者隨機分為實驗組與對照組, 實驗組成員具有得到參加訓練的資格 (但不強迫一定要受訓), 然而控制組成員卻被禁止接受培訓。若將隨機分組視為工具變數, 而接受訓練與否視為處理指標, 那麼控制組成員 $(Z=0)$ 一律不能接受訓練, 代表其 $D=0$。而實驗組成員 $(Z=1)$ 雖然擁有受訓的資格, 但仍可能選擇不接受, 導致 $D=0$ 或 1, 單邊不服從假設成立。

在工具變數的架構下, 加上單邊不服從假設, 我們可以證明 LATT 等於 ATT:

$$
\begin{aligned}
\tau_t &= E[Y(1)-Y(0)|D(1)=1, D(0)=0, D=1] \\
&= E[Y(1)-Y(0)|D(1)=1, D=1] \\
&= E[Y(1)-Y(0)|D=1]=\beta_t。 \qquad (17)
\end{aligned}
$$

首先, $D(1)=1$ 及 $D(0)=0$ 代表服從者族群。接著, 第二個等號來自於單邊不服從假設。而第三個等號則是因為在 $D=1$ 的情況下, 無論總是

接受者或服從者都一定符合 $D(1) = 1$。[2]

根據式 (17), Donald and Hsu (2014) 提出一個對於條件獨立假設的 Durbin-Wu-Hausman 型檢定, 其基本概念如下: 由於式 (14) 中 LATT 的估計式無論條件獨立假設成立與否都會是一致估計式, 然而一般來說, 式 (6) 中 ATT 的估計式只有在條件獨立假設成立下才會是一致估計式。因此, 我們可以藉由比較 LATT 跟 ATT 估計式的差距來衡量條件獨立假設的合理性。有關更多檢定細節或統計推論的部分請參考 Donald and Hsu (2014)。

5 其他相關議題

在這一節中, 我們將討論處理效果文獻中近年來興起的議題。

5.1 分位數處理效果

前面的討論主要聚焦於某些族群的平均處理效果, 然而, 同一族群內的不同個體並不一定受到相同影響, 有些可能效果顯著, 而另一些的效果卻又不如預期, 此現象通常被稱爲組內異質性 (intra-group heterogeneity)。爲了評估處理效果的組內異質, Firpo (2007) 針對分位數處理效果 (quantile treatment effect, QTE) 進行分析, 並提出無母數估計式來描繪不同分位數上之處理效果。令 $\tau \in (0, 1)$, QTE 的數學式定義如下:

$$\beta(\tau) = Q_{Y(1)}(\tau) - Q_{Y(0)}(\tau), \tag{18}$$

對於 $d = 0$ 或1, $Q_{Y(d)}(\tau) = \inf\{y : F_{Y(d)}(y) \geq \tau\}$ 爲 $Y(d)$ 的分位數函數 (quantile function)。由式 (18) 可知, QTE 乃是兩個潛在結果分位數函數的差異, 而非潛在結果差異的分位數函數。也就是說,

$$\beta(\tau) = Q_{Y(1)}(\tau) - Q_{Y(0)}(\tau) \neq Q_{Y(1)-Y(0)}(\tau)。$$

一般而言, 在沒有更強的保序假設 (rank preservation) 下此二者並不會相等。而在處理效果文獻中, 前者比後者吸引了更多的關注, 也被認爲更能

[2]由於假設3.3已先將不服從者剔除, 因此 $D = 1$ 受處理族群裡只會有總是接受者及服從者。

捕捉組內異質性, 詳細原因請參考 Imbens and Wooldridge (2009)。[3] 在假設2.1及2.2成立之下, Firpo (2007) 證明 QTE 可被認定爲

$$\beta(\tau) = \arg\min_q E\left[\frac{D\rho_\tau(Y-q)}{p(X)}\right] - \arg\min_q E\left[\frac{(1-D)\rho_\tau(Y-q)}{1-p(X)}\right], \tag{19}$$

其中 $\rho_\tau(a) = a(\tau - 1\{a \le 0\})$ 爲核對函數 (check function) 而 $1\{\cdot\}$ 則爲指示函數 (indicator function)。很自然地, 我們可針對上式進行二階段估計:

$$\hat{\beta}(\tau) = \arg\min_q \frac{1}{n}\sum_{i=1}^{n} \frac{D_i \rho_\tau(Y_i - q)}{\hat{p}(X_i)} - \arg\min_q \frac{1}{n}\sum_{i=1}^{n} \frac{(1-D_i)\rho_\tau(Y_i - q)}{1-\hat{p}(X_i)}, \tag{20}$$

其中 $\hat{p}(x)$ 代表第一階段傾向分數函數的無母數估計式。在適當的正規條件之下, Firpo (2007) 推導出 $\hat{\beta}(\tau)$ 的逐點 (pointwise) 漸近理論, 並用其來檢定在某特定分位數上之處理效果是否顯著。然而, 當研究者欲檢定之假說牽涉到某一段分位數之間的處理效果時 (例如, 所得低於中位數者參加就業培訓計畫的效果是否爲正), 使用逐點漸近理論會有資料窺視偏差 (data snooping bias) 的疑慮, 並不適用於檢定此假說。爲解決這個問題, Donald and Hsu (2014)。推導出均匀 (uniform) 漸近理論來進行統計推論。礙於篇幅限制, 更多均匀推論的細節請參考 Donald and Hsu (2014)。此外, 有關內生性分位數處理效果也請參見 Frölich and Melly (2013) 及 Hsu, Lieli, and Lai (2018)。

5.2 虛擬處理效果

第2及第3節分別介紹在給定條件獨立假設或在工具變數架構之下, 研究者該如何對不同的處理效果進行認定及估計。然而這些方法只能對最初定

[3] 由於 $\beta = E[Y(1)] - E[Y(0)] = E[Y(1) - Y(0)]$, 平均處理效果之分析無須擔心此問題。

義的母體 (或某些子族群) 處理效果加以分析, 無法將所估計到的處理效果推廣至其他母體。也就是說, 即便上述方法具備了良好的內部效度 (internal validity), 但卻缺乏提供外部效度 (external validity) 的能力, 無法在政策實施前執行事前政策評估。有鑑於此, 近年來處理效果文獻開始重視外部效度, 有興趣的讀者可參閱 Athey and Imbens (2017) 之回顧性文章。其中, Hsu, Lai, and Lieli (2018) 提出可透過平均虛擬處理效果 (average counterfactual treatment effect, ACTE) 來將已估計的處理效果外推至其他母體, 以下我們將簡單介紹此方法。

我們首先推廣 Rubin 因果模型至現狀 (status quo) 與虛擬 (counterfactual) 母體。除了原本定義在現狀母體中的潛在結果 $Y(1)$ 及 $Y(0)$ 與可觀察變數 (Y, D, X) 之外, 我們在虛擬母體也定義類似的變數: 令 $Y^*(1)$ 與 $Y^*(0)$ 爲虛擬母體之潛在結果, 而 (Y^*, D^*, X^*) 分別代表虛擬母體的結果、處理指標, 以及個體特徵。同樣地, 以下的關係式也一定成立: $Y^* = D^*Y^*(1) + (1 - D^*)Y^*(0)$。然而, 由於虛擬母體尚未實施政策, 我們僅能觀察到 (Y, D, X, X^*) 而非 Y^* 及 D^*, 代表著無論是 $Y^*(1)$ 或 $Y^*(0)$ 都無法被觀察。值得一提的是, 此虛擬母體可以是真實存在的另一母體, 也可以是研究者自行定義的反事實母體。舉例來說, 假設政策制定者想將台北市某福利政策推廣至台中市, 那麼 (Y, D, X) 分別代表了台北市民衆的所得、接受補貼與否, 以及個體特徵 (例如, 民衆年齡), 而 X^* 則爲台中市民的年齡。另一方面來說, 若政策制定者想知道當台北市民年齡結構改變時此政策效果會受到什麼影響, 可令 X^* 之分配服從新的 (虛擬) 年齡結構, 來估計此時的虛擬處理效果。ACTE 的定義式爲:

$$\beta^* = E\left[Y^*(1) - Y^*(0)\right]。 \tag{21}$$

與式 (1) 相比, 式 (21) 中的 $Y^*(1)$ 及 $Y^*(0)$ 都無法被觀察, 代表我們必須引進更強的假設來得到認定結果。首先, 令 $F_{Y|X}(y|x)$ 爲條件累積分布函數, 而 $\mathcal{X}$ 爲 X 的支撐集合。

假設 5.1. 條件分佈不變性: 對於 $d = 0$ 或 1, $F_{Y^*(d)|X^*}(y|x) = F_{Y(d)|X}(y|x)$。

假設 5.2. 支撐集合: $\mathcal{X}^* \subseteq \mathcal{X}$。

假設5.1來自於 Fortin, Lemieux, and Firpo (2011), 其要求對於 $d = 0$ 或1, $Y^*(d)$ 與 $Y(d)$ 的差異只會來自於不同的個體特徵 X 與 X^*, 而非其他總體效果。[4] 假設5.2則要求虛擬母體的特徵支撐集合必須爲現狀母體的特徵支撐集合之子集, 若違反此假設, 則需重新定義虛擬母體。在假設2.1、2.2、5.1、5.2 成立下, Hsu, Lai, and Lieli (2018) 證明了 ACTE 的認定式爲

$$\beta^* = \int_{\mathcal{X}} (E[Y|X = x, D = 1] - E[Y|X = x, D = 0]) dF_{X^*}(x)。 \quad (22)$$

給定隨機抽樣資料 $\{(Y_i, D_i, X_i)\}_{i=1}^n$ 及 $\{X_j^*\}_{j=1}^{n^*}$, Hsu, Lai, and Lieli (2018) 提出類似於 Heckman and Todd (1998) 的核迴歸 (kernel regression) 估計式來進行無母數估計:

$$\hat{\beta}^* = \frac{1}{n^*} \sum_{j=1}^{n^*} \left\{ \frac{\sum_{i=1}^{n} D_i Y_i \mathcal{K}_{X_j^*,h}\left(X_i - X_j^*\right)}{\sum_{i=1}^{n} D_i \mathcal{K}_{X_j^*,h}\left(X_i - X_j^*\right)} - \frac{\sum_{i=1}^{n} (1 - D_i) Y_i \mathcal{K}_{X_j^*,h}\left(X_i - X_j^*\right)}{\sum_{i=1}^{n} (1 - D_i) \mathcal{K}_{X_j^*,h}\left(X_i - X_j^*\right)} \right\}, \quad (23)$$

其中 $\mathcal{K}_{x,h}(\cdot)$ 爲高階邊界核 (higher-order boundary kernel) 且 h 爲相對應的帶寬 (bandwidth)。給定正規條件之下, 此估計式之漸近分配爲:

$$\sqrt{n}\left(\hat{\beta}^* - \beta^*\right) \xrightarrow{d} \mathcal{N}\left(0, E\left[\varrho(Y, D, X)^2\right] + E\left[\varphi(X^*)^2\right]\right), \quad (24)$$

其中, 令樣本數的極限比率爲 $\lambda = \lim_{n,n^* \to \infty} n/n^*$, ϱ 及 φ 分別爲

$$\varrho(Y, D, X) = \left\{ \frac{D(Y - E[Y(1)|X])}{p(X)} - \frac{(1 - D)(Y - E[Y(0)|X])}{1 - p(X)} \right\} \frac{f_{X^*}(X)}{f_X(X)},$$

$$\varphi(X^*) = \sqrt{\lambda}\left(E\left[Y(1) - Y(0)|X^*\right] - \beta^*\right)。$$

[4]Hotz, Imbens, and Mortimer (2005) 則直稱此假設爲無總體效果 (no-macro effects) 假設。

由上式可發現, 若感興趣的估計目標爲 ATE, 可令 $X^* = X$ 且 $n^* = n$ (代表 $\lambda = 1$), 則式 (24) 之漸近變異數將與式 (7) 之漸近變異數 $\mathcal{V}$ 相同, 代表式 (23) 與式 (5) 同時具有半參數效率性質。除此之外, Hsu, Lai, and Lieli (2018) 也討論了受處理族群的平均虛擬處理效果 (average counterfactual treatment effect on the treated, ACTT)、分位數虛擬處理效果 (quantile counterfactual treatment effect, QCTE) 及受處理族群的分位數虛擬處理效果 (quantile counterfactual treatment effect on the treated, QCTT)。有興趣的讀者可自行參閱。

5.3 條件處理效果

除了組內異質性外, 組間異質性 (inter-group heterogeneity) 也是造成處理效果不齊一的另一原因。在這裡, 我們將組間異質性定義爲以個體特徵來分群之子族群平均處理效果。Abrevaya, Hsu, and Lieli (2015) 在條件獨立假設下提出可用條件平均處理效果 (conditional average treatment effect, CATE) 衡量此組間異質性, 我們簡單說明其內容。假設 X_1 是 X 的一個子向量 (sub-vector), CATE 的定義式爲:

$$\beta(x_1) = E\left[Y(1) - Y(0) | X_1 = x_1\right]。 \tag{25}$$

式 (25) 跟式 (1) 最大的差異在於式 (25) 爲條件期望值, 而式 (1) 爲無條件期望值。除此之外, 我們仍然可用類似方式來得到 CATE 的認定結果:

$$\beta(x_1) = E\left[\frac{DY}{p(X)} - \frac{(1-D)Y}{1-p(X)}\middle| X_1 = x_1\right]。 \tag{26}$$

估計方法則跟 X_1 是離散或連續變數有關。首先假設 X_1 是個二元的變數, 例如 $X_1 = 1$ 代表女性而 $X_1 = 0$ 代表男性。在計算 $\beta(1)$ 時, 我們可以先把全樣本分爲女性和男性兩個子樣本, 並套用式 (5) 的估計式在女性子樣本上, 如此一來即可直接估計 $\beta(1)$。然而當 X_1 是一維連續變數時, 不能使用此樣本分割法, 而必須改用核迴歸估計式 (kernel regression estimator):

$$\hat{\beta}(x_1) = \frac{\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{D_iY_i}{\hat{p}(X_i)} - \frac{(1-D_i)Y_i}{1-\hat{p}(X_i)}\right)\mathcal{K}_1\left(\frac{X_{1i}-x_1}{h_1}\right)}{\frac{1}{nh}\sum_{i=1}^{n}\mathcal{K}_1\left(\frac{X_{1i}-x_1}{h_1}\right)}, \tag{27}$$

其中 $\hat{p}(x)$ 可以爲參數或非參數的傾向分數函數估計式, h 是核迴歸的帶寬 (bandwidth), $\mathcal{K}_1$ 則是核函數 (kernel function)。在適當的正規條件之下, Abrevaya, Hsu, and Lieli (2015) 分別推導出參數化傾向分數估計式以及非參數傾向分數估計式的漸進常態性質。

值得一提的是, Abrevaya, Hsu, and Lieli (2015) 的理論中 X_1 可以推廣至多維度, 而感興趣的參數也包含了受處理族群的條件平均處理效果 (conditional average treatment effect on the treated, CATT)、條件局部平均處理效果 (conditional local average treatment effect, CLATE), 以及受處理族群的條件局部平均處理效果 (conditional local average treatment effect on the treated, CLATT)。另一方面, Delgado and Escanciano (2013)、Lee, Song, and Whang (2013)、Chang, Lee, and Whang (2015) 以及 Hsu (2017) 分別提供了不同的檢定方法來檢驗條件平均處理效果是否對所有不同的子群體都顯著爲正。Lee, Okui, and Whang (2017) 則提供了條件平均處理效果均勻的信任區間帶 (uniform confidence band)。更多的細節請參見這些論文。

6 未來方向

最後, 我們提出幾個未來可能的研究方向。首先, 第2節中提到要估計 ATE 及 ATT, 條件獨立假設扮演著關鍵角色。然而, 即便第 4 節裡我們介紹了 Donald and Hsu (2014) 的條件獨立假設檢定, 研究者仍須先找到可靠的工具變數才能應用, 代表這項檢定的適用性還不夠廣。如何在其他合理假設下對條件獨立假設進行檢定會是未來研究的一個重要方向。其次, 由於在機率倒數加權法中必須先針對傾向分數函數進行無母數估計, 而使用不同無母數方法在小樣本下會有不同的估計表現, 該如何根據資料特性來選擇估計方法 (data dependent method) 也是文獻中未決的問題。最後, 對於如何在級數羅吉特法選取冪級數的數量, 或是在核迴歸估計式中選擇帶寬大小等最適調校參數 (tuning parameter) 問題目前也尚無定論, 還需等待未來處理效果文獻的理論發展。對未來發展方向有興趣的讀者也可參見 Athey et al. (2017)。

參考文獻

Abadie, Alberto (2003), “Semiparametric Instrumental Variable Estimation of Treatment Response Models,” *Journal of Econometrics*, 113, 231–263.

Abrevaya, Jason, Yu-Chin Hsu, and Robert P. Lieli (2015), “Estimating Conditional Average Treatment Effects,” *Journal of Business and Economic Statistics*, 33, 485–505.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin (1996), “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, 91, 444–455.

Athey, Susan and Guido W. Imbens (2017), “The State of Applied Econometrics: Causality and Policy Evaluation,” *Journal of Economic Perspectives*, 31, 3–32.

Athey, Susan, Guido W. Imbens, Thai Pham, and Stefan Wager (2017), “Estimating Average Treatment Effects: Supplementary Analyses and Remaining Challenges,” *American Economic Review*, 107, 278–281.

Chang, Minsu, Sokbae Lee, and Yoon-Jae Whang (2015), “Nonparametric Tests of Conditional Treatment Effects with an Application to Single-Sex Schooling on Academic Achievements,” *Econometrics Journal*, 18, 307–346.

Delgado, Miguel A. and J. Carlos Escanciano (2013), “Conditional Stochastic Dominance Testing,” *Journal of Business and Economic Statistics*, 31, 16–28.

Donald, Stephen G. and Yu-Chin Hsu (2014), “Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models,” *Journal of Econometrics*, 178, 383–397.

Firpo, Sergio (2007), “Efficient Semiparametric Estimation of Quantile Treatment Effects,” *Econometrica*, 75, 259–276.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo (2011), “Decomposition Methods in Economics,” in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, vol. 4, Amsterdam: Elsevier North Holland, 1–102.

Frölich, Markus (2007), “Nonparametric IV Estimation of Local Average Treatment Effects with Covariates,” *Journal of Econometrics*, 139, 35–75.

Frölich, Markus and Blaise Melly (2013), “Unconditional Quantile Treatment Effects under Endogeneity,” *Journal of Business and Economic Statistics*, 31, 346–357.

Hahn, Jinyong (1998), "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331.

Heckman, James J. adn Hidehiko Ichimura and Petra Todd (1998), "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261–294.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189.

Hong, Han and Denis Nekipelov (2010), "Semiparametric Efficiency in Nonlinear LATE Models," *Quantitative Economics*, 1, 279–304.

Horvitz, Daniel G. and Donovan J. Thompson (1952), "A Generalization of Sampling without Replacement from a Finite Universe," *Journal of the American Statistical Association*, 47, 663–685.

Hotz, V. Joseph, Guido W. Imbens, and Julie H. Mortimer (2005), "Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations," *Journal of Econometrics*, 125, 241–270.

Hsu, Yu-Chin (2017), "Consistent Tests for Conditional Treatment Effects," *Econometrics Journal*, 20, 1–22.

Hsu, Yu-Chin, Tsung-Chih Lai, and Robert P. Lieli (2018), "Estimation and Inference for Counterfactual Treatment Effects," Working Paper.

Hsu, Yu-Chin, Robert P. Lieli, and Tsung-Chih Lai (2018), "Estimation and Inference for Distribution Functions and Quantile Functions in Endogenous Treatment Effect Models," Working Paper.

Ichimura, Hidehiko and Oliver Linton (2005), "Asymptotic Expansions for Some Semiparametric Program Evaluation Estimators," in Donald W. Andrews and James H. Stock (eds.), *Identification and Inference for Econometric Models*, Chap. 8, Cambridge: Cambridge University Press, 149–170.

Imbens, Guido W. (2004), "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and statistics*, 86, 4–29.

Imbens, Guido W. and Joshua D. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Imbens, Guido W. and Jeffrey M. Wooldridge (2009), "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47, 5–86.

Lee, Sokbae, Ryo Okui, and Yoon-Jae Whang (2017), "Doubly Robust Uniform Confidence Band for the Conditional Average Treatment Effect Function," *Journal of Applied Econometrics*, 32, 1207–1225.

Lee, Sokbae, Kyungchul Song, and Yoon-Jae Whang (2013), "Testing Functional Inequalities," *Journal of Econometrics*, 172, 14–32.

Li, Qi, Jeffrey S. Racine, and Jeffrey M. Wooldridge (2009), "Efficient Estimation of Average Treatment Effects with Mixed Categorical and Continuous Data," *Journal of Business and Economic Statistics*, 27, 206–223.

Rosenbaum, Paul R. and Donald B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

Rubin, Donald B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

投稿日期: 2018年2月21日, 接受日期: 2018年10月2日

Treatment Effect Models: A Brief Review

Yu-Chin Hsu
Institute of Economics, Academia Sinica
Department of Finance, National Central University and
Department of Economics, National Chengchi University

Tsung-Chih Lai
Department of Economics, Feng Chia University

Over the last three decades, much progress has been made in the treatment effect literature. In this article, we review the identification and estimation results of average treatment effects based on inverse probability weighting under the unconfoundedness assumption. We also discuss similar identification and estimation results of local average treatment effects when treatment assignment is endogenous but a binary instrumental variable is available. We then introduce a test for the unconfoundedness assumption and summarize some recent developments in the treatment effect literature. Finally, we point out some future research directions.

Keywords: rubin causal model, treatment effect, propensity score, inverse probability weighting
JEL classification: C21, C31