

因果推論與效應評估： 區段識別法及其於「選制效應」之應用

黃紀*

《本文摘要》

社會科學中涉及效應評估的問題，都無法迴避因果推論。以觀察研究進行因果推論之所以棘手，癥結在於比較研究的組別，往往取決於因和果之間的內部因素，也就是所謂「內因性」(endogeneity)，造成平均因果效應的識別問題。一般分析因果效應的參數模型 (parametric models)，雖有考慮內因問題，但多建立在很特定的函數形式及變數分佈等假定之上。如果研究的主题及資料的確符合這些假定，自可充分運用；但社會科學研究也常常會碰到與假定不符的情況，此時 Manski 的無母數局部識別法 (nonparametric partial identification) 最為適合，因為這個方法從無假定出發，逐步帶入不同強度的假定，檢視其對於參數區段的影響，將假定與推論之間的關係完全透明化，避免為了達到「定點識別」而強加或暗藏與實際不符的假定，導致過當的推論。

本文從「反事實因果模型」(counterfactual model of causality) 的角度，以最基礎的邏輯與機率論，探討 Manski 的區段識別法，及各種學理假定與「平均因果效應」之上下限的關係，並以 2008 年立委選舉台聯提名區域立委對其政黨票得票率之影響為例，將區段識別法應用於分析混合選制中所謂之「污染效應」(contamination effect)。

關鍵詞：因果推論、內因性問題、區段識別、混合選制、污染效應

* 國立政治大學講座教授、政治學系教授暨選舉研究中心合聘研究員。E-mail: chihuang@nccu.edu.tw。

壹、前言

因果關係是個既有趣卻又動輒引發爭議的問題，然而社會科學研究中所有涉及效應評估的問題，都無法迴避因果推論。早在 1975 年 Deming 便指出：「評估就是研究因果 (Evaluation is a study of causes)」(Deming 1975)。道理很簡單：若想嚴謹地評估某種制度或某項政策方案的成效，必先確立該效應確實是由該制度或政策所造成，而非其他的干擾因素 (confounders) 所造成；也就是說，先要確立因和果之間的關連並非假像 (spurious)，以免歸功或究責不當。

在因果推論的方法上，「比較法」扮演舉足輕重的角色，但運用時卻需謹慎，除非拿來比較的各組是隨機 (random) 產生，確保各組在各方面都旗鼓相當，否則光憑組間之異同驟下論斷，往往容易誤導。例如在評估政府某個職業訓練方案的成效時，若發現受過訓練者的平均薪資，比起沒受過訓練者還低，不宜驟下論斷認定該方案成效適得其反，因為很可能當初自己選擇 (self-select) 去受訓者本來就有工作能力上的困難，因此薪資較低也許只是反映原先能力上的差異，並非職訓造成反效果。換言之，逕行比較有受訓和沒受訓這兩群人，太過於直觀，尚不足以斷定職訓方案的效果；唯有確保這兩組人在職訓開始之前的工作能力等諸多因素的分佈都相當 (equivalent)，之後受完訓者薪資仍較低，我們才能排除其他可能的干擾原因，推論係職訓方案本身成效不彰。同理，制度的效應評估，也不宜逕行比較採行兩種制度的國家或地區在依變數上的差異，就驟然歸因於該制度，因為可能有其他歷史及社會經濟因素形塑了不同的制度，而這些因素也造就了依變數的差異，由於因與果之間的關係受到其他因素干擾，不易清楚識別。

社會科學的研究基於法律、道德、成本等的考量，往往無法隨機分派研究對象到各組，而必須仰賴「觀察研究」(observational study) 推論因果。以觀察研究進行因果推論之所以棘手，癥結就在於拿來比較的各組，組別的形成往往取決於因和果之間的內部因素，也就是所謂「內因性」(endogeneity)，造成平均因果效應的識別問題。

本文旨在闡述觀察研究中，以無母數局部識別法 (nonparametric partial identification) 進行因果推論的原則與方法。全文安排如下：第貳節先釐清因果推論的關鍵與觀察研究的推論瓶頸——內因問題，接著第參節以最基礎的邏輯與機率論，解析 Manski 提出的區段識別法如何抽絲剝繭，逐步釐清實證資料與假定對識別「平均因果效應」上下限區段的作用；第肆節則舉 2008 年立委選舉為例，檢視台聯提名區域立委對其政黨票得票率有無因果效應；最後第伍節為小結。

貳、因果推論的關鍵：選組機制與識別問題

「同中求異法」(method of difference, 或譯為「差異法」)的邏輯並不複雜：如果比較的對象原先處處相同，唯有我們感興趣的「因」不同，此時「果」的差異方能歸因給這個「僅此一家，別無分號」的因。推論的關鍵既然在「控制」原因變數 (causal variable) 以外其他的自變數 (covariates) 都相同，那麼最理想的方式就是同一個人¹自己與自己相比，我們可以想像同一時空內其他條件都相同，唯有原因變數不同，接下來再比較其「果」的差異，因果效應就一清二楚了。可惜實際上，同一個人要不然就受到因的影響、要不然就是沒受到因的影響，無法同時「一人分飾兩角」來自己相比，造成 Holland(1986, 947) 所謂之「因果推論的根本問題」(The fundamental problem of causal inference)。不過這個「自己(本尊)與另一個可能的自己(分身)相比」的想像，卻正是近年盛行的「反事實因果模型」(counterfactual model of causality, 參見 Rubin(1974; 1990; 1991; 2005) 等)的思維根源。

為了說明的簡潔起見，本文聚焦於二分類之原因變數 (binary causal variable) D ，亦即在研究者感興趣的母群中分成兩組：1 代表有施測「因」的處理組 (treatment group)，0 代表未施測的控制組 (control group)。²「反事實因果模型」設想，母群中的每一份子 i 在施測前，其連續依變數 Y 都有兩種可能的結果 (potential outcomes)：若有施測則結果為 Y^1 、若未施測則為 Y^0 。至於施測後實際發生並觀察到的 Y ，其實是這兩種可能的值 Y^1 及 Y^0 的事後體現 (realization)，端視該份子選入 (或被歸入) D 的哪一組而定：

$$\begin{aligned} Y_i &= \begin{cases} Y_i^1 & \text{if } D = 1 \\ Y_i^0 & \text{if } D = 0 \end{cases} \\ &= DY_i^1 + (1-D)Y_i^0 \\ &= Y_i^0 + (Y_i^1 - Y_i^0)D \end{aligned}$$

由於 D 的作用就好像電閘開關 (switch) 一樣：按 $D=1$ 是開，看到 Y^1 ；按 $D=0$ 則是關，看到 Y^0 ，因此上式也稱為「開闔機制」(switching process) (Manski 1995, 38-39)。

¹ 本文所稱「人」，泛指觀察研究的分析單位 (unit of analysis)，可能是個人，也可能是組織、地理單位等，視研究主題而定。

² 相關文獻使用的符號系統十分多樣，本文基本上是採 Morgan 與 Winship(2007) 及 Angrist 與 Pischke(2009) 較為簡易的符號。將因果推論問題稍加符號化，目的只在求追本溯源的過程中，推理清晰且敘述簡明。

依照比較法「同中求異」的原則，原因變數 D 對母群中某一個份子 i 產生的效應 (individual-level treatment effect or unit causal effect)，就反映在該份子兩種可能結果的差：

$$\delta_i = Y_i^1 - Y_i^0$$

但棘手的是，同一個人在同一時間點，要不是在處理組 $D_i=1$ ，只觀察到 Y_i^1 卻觀察不到 Y_i^0 ；要不是就是在控制組 $D_i=0$ ，只觀察到 Y_i^0 卻觀察不到 Y_i^1 ，無法一人同時分飾二角，因此難以計算個人層次的因果效應，Holland(1986, 947) 將這個挑戰稱為「因果推論的根本問題」，如表 1 所示。於是從事因果分析者，便常將焦點放在母群之中處理組與控制組這兩群人的相互比較上，也就是「平均效應」(Average treatment effect, 簡稱 ATE) $E(\delta) = E(Y^1 - Y^0) = E(Y^1) - E(Y^0)$ 。平均效應推論的標的是：母群之中有人屬於處理組、有人屬於控制組，而研究者卻想知道「若整個母群都在處理組，比起若整個母群都在控制組，其 Y 的平均差別（即平均因果效應）為何」(Manski 1995, 37-38)？本文以下的討論，都以「平均效應」為推論之母群參數。³

表 1 事實 vs. 反事實

Group 組別	Potential Outcomes 可能的結果	
	Y^1	Y^0
Treatment Group 處理組 ($D=1$) $P(D=1) = \pi$	Observable as Y 可觀察到的事實 Y $(Y_i^1 D=1) = (Y_i D=1)$ (處理組的本尊)	Counterfactual 觀察不到的「反事實」 $(Y_i^0 D=1)$ (處理組的分身)
Control Group 控制組 ($D=0$) $P(D=0) = 1-\pi$	Counterfactual 觀察不到的「反事實」 $(Y_i^1 D=0)$ (控制組的分身)	Observable as Y 可觀察到的事實 Y $(Y_i^0 D=0) = (Y_i D=0)$ (控制組的本尊)

資料來源：修改自 Morgan 與 Winship(2007, 35)。

一、平均效應之參數識別問題

一般統計推論，重點在樣本數的大小，如果樣本夠大，多數的參數估計問題就可迎刃而解。但是參數的識別 (identification)，⁴ 處理的則是更根本的邏輯問題：亦即邏輯上，有好幾種理論結構都可以解釋觀察到的變數分佈，⁵ 而我們研究者能否根據實際觀察到的變數資料，從這許多種可能性之中識別 (identify) 或辨認出潛藏的真正結構 (underlying

³ 關於其他因果推論的參數，詳見 Heckman、Tobias 與 Vytlacil(2001; 2003)。

⁴ 或翻譯為「認定」。不過「認定」二字似已執著於傳統的「定點」觀念，故本文不予採用。

⁵ 稱為「數個結構都可產生同一組觀察資料」(observationally equivalent structures)。

structure) (Hsiao 1983, 224-226)？在統計上，這就是解模型參數 (parameters) 的數值，也就是「姑且不論樣本，模型中的參數究竟能否一致的估計」(Greene 2008, 361-370; Wooldridge 2009, 552-560)？如果母群參數無法識別，則樣本數再大也無濟於事。傳統上認為能解出參數唯一的解值（也稱為「定點識別」(point identification)），才算「可識別」(identifiable)；反之，如果可觀察到的資料本身並不足以提供唯一的答案（即參數唯一的解值），就常稱為「無法識別」(unidentified)，此時常見的做法是引進強勢 (strong) 的「未驗假定」(assumptions)⁶來「解決」識別問題。

Manski(2003; 2007) 批評這種傳統上對定點識別的過度執著，常驅使研究者為達目的（定點）而不擇手段（強加假定），但關鍵在於該假定是否可信 (credible)？一個斬釘截鐵的結論，如果是建立在一串強勢卻不大可信的假定之上，就好像沙灘上的城堡，再富麗堂皇也終究禁不起考驗！Manski(2003, 1) 稱之為「可信度遞減律」(The law of decreasing credibility)：「下的假定越重，得到的結論看似越斬釘截鐵，實則推論卻越不可信」。針對假定的強弱度與推論的可信度間無可迴避的抵換 (trade-off) 關係，他提出了「局部識別理論」(theory of partial identification)，強調：識別並非「能或不能定點識別」的二元對立，而是有程度之分的。倘若僅有部分資料可觀察得到 (partially observable)，那麼實證研究者不應視而不見粉飾太平，而應坦然面對。畢竟「假定」都是未驗的訊息 (prior information)，目的在幫我們限縮 (restrict) 識別的範圍，故處理識別問題的正途，不在強加或暗藏不可信的假定來達到「定點識別」的目的，而在正視資料的性質、引用比較輕微卻比較可信的假定 (weak but credible assumptions)，先識別參數值的可能範圍 (region)，寧可坦然承認推論的侷限，也不譁眾取寵，或自以為是地妄下斷語。準此，識別問題可修正為「手邊的數據資料，在某種假定下，能得到什麼程度範圍的結論」？換言之，「局部識別」的起點，就是邏輯與證據本身，能推論到哪裡就先推到哪裡，不增一分、不減一分，以上下限區段確立學界無可爭議的「共識域」(a domain of consensus)；這個區段看似不起眼，其實卻非常有用 (Cho and Manski 2008, 562-563)：

1. 倘若區段很窄，則推論已相當精確。
2. 倘若區段很寬，正反映手邊實證資料的侷限，應予正視。此時，研究者可依序加上爭議程度由低至高的各種假定，逐一檢視這些假定對識別區段的影響。此一由寬至窄的敏感度分析 (sensitivity analysis)，可促使研究者正視假定與結論間的明確關係，例如定點識別究竟是建立在哪些假定之上？

⁶ Assumption 未檢驗且往往無意檢驗，故譯為「未驗假定」。Hypothesis 提出的目的就在進行檢驗，故譯為「待驗假設」。

3. 倘若某個假定下的識別的區段，不落在純資料、無假定的區段內，亦即兩者的交集為空集合 (empty set)，則表示該假定根本不成立。

因果推論之所以棘手，除了實證研究的資料搜集問題外，癥結還是在識別 (Manski 1989; 1990)。正因為「一人無法分飾二角」，部分資料在邏輯上就觀察不到，因此唯有釐清這個問題的根源，才能對症下藥。以下的討論，就是從邏輯和基本機率論，針對因果推論的識別問題，一步一腳印地抽絲剝繭。

設若母群中，選入處理組的機率 (treatment-selection probability)⁷ 為 $P(D=1)=\pi$ ，則依「期望值總和律」(The law of total expectation)， Y^1 的期望值為：

$$E(Y^1) = \pi \times E(Y^1 | D=1) + (1-\pi) \times E(Y^1 | D=0)$$

上式之中，

1. $\pi = P(D=1)$ 指屬於處理組的機率（或比例），可觀察得到。
2. $1-\pi = P(D=0)$ 指屬於控制組的機率（或比例），可觀察得到。
3. $E(Y^1 | D=1) = E(Y | D=1)$ 指處理組（即接受施測者）結果變數之期望值，亦屬可觀察的資料。
4. 唯獨 $E(Y^1 | D=0)$ 為反事實，代表「倘若控制組的人也接受施測，其期望值會是多少？」但實際上控制組並未接受施測， $E(Y^1 | D=0)$ 與事實相反，故成了觀察不到的缺漏值 (missing value)，造成識別 $E(Y^1)$ 的困難。

同理，依「期望值總和律」可知 Y^0 的期望值為：

$$E(Y^0) = \pi \times E(Y^0 | D=1) + (1-\pi) \times E(Y^0 | D=0)$$

其中 π 、 $1-\pi$ 及 $E(Y^0 | D=0) = E(Y | D=0)$ 都可觀察，唯獨 $E(Y^0 | D=1)$ 代表「倘若處理組的人沒有接受施測，其期望值會是多少？」但實際上處理組的人已經接受施測，故亦為無法觀察的「反事實」。

平均因果效應為 $E(Y^1)$ 及 $E(Y^0)$ 兩者之差 (Morgan and Winship 2007, 45)：

$$\begin{aligned} E(\delta) &= E(Y^1 - Y^0) \\ &= E(Y^1) - E(Y^0) \\ &= \left[\pi E(Y^1 | D=1) + (1-\pi) E(Y^1 | D=0) \right] - \left[\pi E(Y^0 | D=1) + (1-\pi) E(Y^0 | D=0) \right] \end{aligned}$$

⁷ 文獻中也稱為 propensity score，或譯為「入選機率」。

上式中，出現了兩個無法觀察的「反事實」： $E(Y^1 | D=0)$ 及 $E(Y^0 | D=1)$ ，這兩項就構成了識別因果效應的障礙。Manski(1989; 1990; 1995; 2003; 2005; 2007; 2008) 明確指出，困難的根源在於研究對象選組 (treatment selection) 時，也決定了研究者只能觀察到實際選組的結果、卻無法觀察到「如果他們做了另類選擇則會如何」的反事實；換言之，「選組問題」(selection problem) 就是參數的識別問題。解決之道必須對症下藥，針對無法觀察的「反事實」進行腦力激盪式的反思。但反思絕非漫無邊際的遐想，而是具體設定清晰的推論條件，因此其最大的挑戰就在引入正確且有用的未驗訊息來限縮參數可能的範圍，或替無法觀察的「反事實」找到可觀察得到的分身。

二、最強勢的假定：外因選組 (Exogenous Treatment Selection, ETS)

如果研究者假定：決定選組的因素都與 Y 無關，也稱為「外因選組」(Exogenous treatment selection, ETS)，那麼事情單純，因為這表示反事實的缺漏為隨機， $E(Y^1)$ 不會因 D 的組別而異，亦即 $E(Y^1 | D=1) = E(Y^1 | D=0)$ ，故 $E(Y^1 | D=0) = E(Y^1 | D=1) = E(Y | D=1)$ ，代入上式後， $E(Y^1) = E(Y | D=1)$ ，也就是等於處理組之結果變數的期望值，這正是 $E(Y^1)$ 之直觀估計式。同理， $E(Y^0) = E(Y | D=0)$ 。因此在隨機缺漏的假定下，平均因果效應就簡化為：

$$E(\delta) = E(Y^1) - E(Y^0) = E(Y | D=1) - E(Y | D=0)$$

這個定點識別的式子，就是直觀的因果效應估計式。換言之，直觀上想當然爾的估計 $\delta_{Naive} = E(Y | D=1) - E(Y | D=0)$ ，其實形同強加了「外因選組」的假定。衡諸實際，除非研究者採隨機分派之實驗設計 (randomized experiment)，否則此一假定恐難以服眾。

三、以觀察研究推論因果的罩門：內因選組 (Endogenous Selection)

社會科學研究中，研究的對象究竟屬於處理組還是控制組，往往非研究者所能掌控，而是取決於其他未觀測到 (unobserved) 的因素之影響，或者根本是由研究對象依照他們對 Y 的預期值而自行做選擇 (self-selection)，故稱為「內因選組」(endogenous selection)。

「內因性」使得選組的機制「不可忽視」(non-ignorable)，這正是「觀察研究」與「隨機分派實驗」最大的差異 (Rosenbaum 2002)。研究者如對此「內因性」視而不見，或不由分說強加外因之假定，則會產生因果效應估計之「選組偏誤」(selection bias 或 selectivity bias) (Barnow, Gain, and Goldberger 1981)，形成了社會科學研究做因果推論的罩門 (黃紀 2008a, 8-9)。

文獻中有幾種因應之道，例如最典型的工具變數 (Instrumental variable, IV) 及其延伸 (參見 Jackson(2008))，以及同時建立選組方程式 (selection equation) 和結構方程式 (structural equation) 的「處理效果模型」(treatment-effects model)(Heckman 1978) 等，都是以定點識別參數為職志，不惜加上線型函數、常數係數、常態分佈等諸多假定，但關鍵就在這些假定與應用之研究情境是否相符？如是，則內因問題迎刃而解；如否，則彷彿飲鴆止渴。另一個途徑，則不貿然求唯一解，寧可沈住氣穩紮穩打，從機率理論的基礎出發，先識別感興趣參數的上下限區段，確立無可爭議的參數範圍後，再逐步帶入不同強弱度的假定，檢視各假定對區段之影響；Manski 的「局部識別理論」即屬於後者。「定點識別」比較能滿足一般「求解」的心理需求，故得到青睞，教科書都會詳加討論；「區段識別」則在區段的集合 (set) 內有多組可能的解，區段越寬，解值也越多，這種不確定性常讓人誤以為「有說等於沒說」，欠缺「定點識別」斬釘截鐵、信心滿滿的光環，因此一般教科書不大討論，或頂多點到為止 (Morgan and Winship 2007)。但如果看似信心滿滿的「定點識別」其實建立在無法自圓其說甚或錯誤的一串假定之上，斬釘截鐵反映的可能只是剛愎自用；而樸拙木訥的「區段識別」卻是反映謹慎縝密、據實以告。因此研究者若無法確定其研究主題及數據資料均與「定點識別」模型的假定相符，此時 Manski 「可信度遞減律」的忠告是：寧可穩建地選擇「區段識別」，以免誤判因果卻仍自以為是。畢竟「權衡適度的假定做出可信的推論，研究者責無旁貸」(黃紀 2008a, 7)。

參、區段識別法

區段識別純粹建立在邏輯和機率論的法則之上。顧名思義，區段 (interval) 是介於上下限之間的實數，寫成 $B=[B^L, B^U]$ ，其寬度 (width) 為 B^U-B^L 。下限 (lower bound, B^L) 是指未知數剛好都「落底」的最差狀況下參數的值，上限 (upper bound, B^U) 則反映未知數剛好都「攻頂」的最佳狀況下參數的值。至於參數的真值，則必落在這個區段內，亦即 $B^L \leq E(\delta) \leq B^U$ 。理想上，如果上限等於下限 $B^U-B^L=0$ ，那參數只有唯一的解值，就是「定點識別」；區段越窄，可能的解值越少；區段越寬，可能的解值越多。研究者無疑都希望區段越窄越好。必須強調的是，本文的區段是確定的 (deterministic)，參數真值必落於這個區段內；有別於一般統計推論中，在 $(1-\alpha)\%$ 信心水準下的信心區間 (confidence interval)。

一、純資料、尚無假定之區段 (No-Assumptions Bounds)

平均效應的區段，當然和 Y 的值域 (range) 息息相關。設若 Y^0 及 Y^1 的值域已知為有限 (bounded)，必介於其最小及最大值 $[y^l, y^m]$ 之間，那麼邏輯上，平均

效應 $E(\delta) = E(Y^1 - Y^0) = E(Y^1) - E(Y^0)$ ，再小也小不過「 $E(Y^1)$ 」的下限減去 $E(Y^0)$ 的上限，即「 $(y^l - y^u)$ 」，再大也大不過「 $E(Y^1)$ 」的上限減去 $E(Y^0)$ 的下限，即「 $(y^u - y^l)$ 」，也就是在沒有任何觀察資料佐證下，平均因果效應的上下限區段為 $(y^l - y^u) \leq E(\delta) \leq (y^u - y^l)$ (Manski 1989, 344-346; 1990, 320) (其推導參見附錄)。⁸

上述之邏輯，若再加上經驗證據，可幫我們識別出更窄的上下限區段。Manski 提出的識別原則，就是把已知之 Y 值域 $[y^l, y^u]$ 代入上述未知之「反事實項」中，可得參數之上下限區段。在 $E(Y^1)$ 式子中， $E(Y^1 | D=0)$ 為無法觀察之反事實，但已知 $y^l \leq E(Y^1 | D=0) \leq y^u$ ，故以 $[y^l, y^u]$ 代入後可得其上下限區段 (Manski 1990, 320-321)：

$$\begin{aligned} \therefore E(Y^1) &= \pi \times E(Y^1 | D=1) + (1-\pi) \times E(Y^1 | D=0), \text{ and } y^l \leq E(Y^1 | D=0) \leq y^u \\ \therefore \pi \times E(Y | D=1) + (1-\pi) \times y^l &\leq E(Y^1) \leq \pi \times E(Y | D=1) + (1-\pi) \times y^u \end{aligned}$$

同理， $E(Y^0)$ 式子中， $E(Y^0 | D=1)$ 為無法觀察之反事實，亦以 $[y^l, y^u]$ 代入後可得其上下限區段：

$$\begin{aligned} \therefore E(Y^0) &= \pi \times E(Y^0 | D=1) + (1-\pi) \times E(Y^0 | D=0), \text{ and } y^l \leq E(Y^0 | D=1) \leq y^u \\ \therefore \pi \times y^l + (1-\pi) \times E(Y | D=0) &\leq E(Y^0) \leq \pi \times y^u + (1-\pi) \times E(Y | D=0) \end{aligned}$$

平均效應既為 $E(\delta) = E(Y^1) - E(Y^0)$ ，則其下限 B^l 等於「 $E(Y^1)$ 之下限」減去「 $E(Y^0)$ 之上限」：

$$\begin{aligned} B^l &= [E(Y^1) \text{ 之下限}] - [E(Y^0) \text{ 之上限}] \\ &= [\pi \times E(Y | D=1) + (1-\pi) \times y^l] - [\pi \times y^u + (1-\pi) \times E(Y | D=0)] < 0 \end{aligned}$$

而其上限 B^u 等於「 $E(Y^1)$ 之上限」減去「 $E(Y^0)$ 之下限」：

⁸ 上下限公式之推導，散見於相關之著作，本文將之綜合整理，並改以較簡潔易懂的統一符號表達，其原出處如下：

- (1) 「純邏輯、尚無資料之區段」 (Manski 1989, 344-346; 1990, 320)；
- (2) 「純資料、尚無假定之區段」 (Manski 1990, 320-321)；
- (3) 「Roy 模型」 (Heckman and Vytlačil 2007b, 5084-5086)；
- (4) 「單向組間平均差之區段」 (Manski and Pepper 2000, 1001)；
- (5) 「單向個別效應差之區段」 (Manski 2007, 189-191; Manski and Pepper 2000, 1002-1003)；
- (6) 「單向組間平均差且個別效應差之區段」 (Manski 2007, 191-193; Manski and Pepper 2000, 1004)。

$$\begin{aligned} B^U &= [E(Y^1) \text{ 之上限}] - [E(Y^0) \text{ 之下限}] \\ &= [\pi \times E(Y|D=1) + (1-\pi) \times y^u] - [\pi \times y^l + (1-\pi) \times E(Y|D=0)] > 0 \end{aligned}$$

此一區段 B 的寬度為 $B^U - B^L = y^u - y^l$ 。雖然仍很寬，且必然包含 0 在內，尚不足以識別平均效應的正負號，但已比尚無資料時之區段寬度 $2 \times (y^u - y^l)$ 縮窄一半，有用得多。

若依變數 Y 為比例，必介於 0 與 1 之間，則 $y^l=0$ ， $y^u=1$ ，則 $E(\delta)$ 的上下限簡化為：

$$\begin{aligned} &[\pi \times E(Y|D=1) - (1-\pi) \times E(Y|D=0)] - \pi \leq E(\delta) \\ &\leq [\pi \times E(Y|D=1) - (1-\pi) \times E(Y|D=0)] + (1-\pi) \end{aligned}$$

B 的寬度縮窄為 1（即 100%）。

二、對「選組機制」的幾種輕微假定及其區段

研究者可根據學理，引入未驗的訊息，進一步限縮平均因果效應的區段。局部識別法之重心，就在檢視不同的假定對參數區段寬窄的影響。以下討論的幾種假定，一方面依據學理或對研究主題的實質理解，但另一方面卻不強加諸如線型函數、常數係數、常態分佈等很重的假定，而僅對 Y^1 、 Y^0 值的相對大小來做假定，既未假定任何函數形式、亦未設定機率分佈，因此都比參數模型的假定輕微 (weak)，但仍有程度之分。以下四者，由最輕之 Roy 模型逐步加重，識別區段也逐步縮窄。

(一) Roy 模型 (Roy Model)

若研究對象可自行選組 (self-selection)，則依據學理或對實務的理解，往往可以合理的假定：受測者會「趨利避害」，選擇預期結果 Y 最好的那一組。例如經濟學家 Roy(1951) 就認為：人會選擇預期收入較高的職業，因此這類模型往往稱為 Roy model。用內因之開闢機制 (endogenous switching process) 來表示：

$$\begin{cases} D = 1 \text{ if } Y^1 \geq Y^0 \\ D = 0 \text{ if } Y^1 < Y^0 \end{cases}$$

這個假定可以進一步限縮識別的區段，因為假定一旦成立，兩個「反事實項」的理論最大可能值 y^u ，就可用觀察到的兩組之期望值取代，而該期望值不會大於（通常小於）最大可能值 y^u (Heckman and Vytlacil 2007b, 5084-5086; Manski 1995, 45-46)：

$$E(Y^0 | D=1) = E(Y^0 | Y^1 \geq Y^0) \leq E(Y^1 | Y^1 \geq Y^0) = E(Y^1 | D=1) = E(Y | D=1) \leq y^u$$

同理，

$$E(Y^1 | D=0) = E(Y^1 | Y^1 < Y^0) \leq E(Y^0 | Y^1 < Y^0) = E(Y^0 | D=0) = E(Y | D=0) \leq y''$$

因此 Roy 模型假定下，區段之下限成爲：

$$\begin{aligned} B^L &= [\pi \times E(Y | D=1) + (1-\pi) \times y'] - [\pi \times y'' + (1-\pi) \times E(Y | D=0)] \\ &= [\pi \times E(Y | D=1) + (1-\pi) \times y'] - [\pi \times E(Y | D=1) + (1-\pi) \times E(Y | D=0)] \\ &= (1-\pi) \times [y' - E(Y | D=0)] \end{aligned}$$

而區段之上限成爲：

$$\begin{aligned} B^U &= [\pi \times E(Y | D=1) + (1-\pi) \times y''] - [\pi \times y' + (1-\pi) \times E(Y | D=0)] \\ &= [\pi \times E(Y | D=1) + (1-\pi) \times E(Y | D=0)] - [\pi \times y' + (1-\pi) \times E(Y | D=0)] \\ &= \pi \times [E(Y | D=1) - y'] \end{aligned}$$

其寬度則爲：

$$\begin{aligned} B^U - B^L &= \pi \times E(Y | D=1) + (1-\pi) \times E(Y | D=0) - y' \\ &= E(Y) - y' \\ &\leq y'' - y' \end{aligned}$$

儘管這個最基本的 Roy 模型之區段必包含 0 在內，尚不足以識別平均效應的正負號，⁹但仍十分重要，可做爲後續分析的基準，倘若從 Roy 模型延伸之參數模型，其定點估計值落在該區段之外，均可加以質疑。

若依變數 Y 爲比例，介於 0 與 1 之間，則 $y'=0$ ，Roy 模型之區段更簡化爲：

$$-(1-\pi) \times E(Y | D=0) \leq E(\delta) \leq \pi \times E(Y | D=1)$$

其區段之寬度 $\pi \times E(Y | D=1) - [-(1-\pi) \times E(Y | D=0)] = E(Y) \leq 1.0$ ，小於或頂多等於無假定情況下的區段寬度。

(二) 單向組間平均差 (Monotone Treatment Selection, MTS) 假定

若依據學理可以合理假定：處理組的 Y 平均值，橫豎都不會小於（含大於或等於）控制組的 Y 平均值。Manski 與 Pepper(2000, 1001) 將此稱爲「單向組間平均差」(Monotone treatment selection, MTS) 假定，具體表示爲：

⁹ Heckman(1978) 著名的「因果效應模型」及其延伸 (Heckman and Vytlačil 2007a; Heckman, Tobias, and Vytlačil 2001; 2003) 也是以 Roy 模型爲基礎，而其之所以能定點識別參數，主要是因爲該參數模型又加上了依變數呈多變數常態分佈等假定，但關鍵是該假定是否成立 (Manski 2007, 151)。

$$\begin{cases} y' \leq E(Y^1 | D=0) \leq E(Y^1 | D=1) \leq y'' \\ y' \leq E(Y^0 | D=0) \leq E(Y^0 | D=1) \leq y'' \end{cases}$$

若此一假定成立，下限 B^L 仍然維持「無假設」情況之下限：

$$B^L = [\pi \times E(Y | D=1) + (1-\pi) \times y'] - [\pi \times y'' + (1-\pi) \times E(Y | D=0)];$$

而其上限 B^U 雖仍等於「 $E(Y^1)$ 之上限」減去「 $E(Y^0)$ 之下限」，但「 $E(Y^1)$ 之上限」中未知的反事實項 $E(Y^1 | D=0)$ 依 MTS 的第一項假定必不大於可觀察之 $E(Y^1 | D=1) = E(Y | D=1)$ ，故可以後者取代 y'' ；同理，「 $E(Y^0)$ 之下限」中未知的反事實項 $E(Y^0 | D=1)$ 依 MTS 的第二項假定必不小於可觀察之 $E(Y^0 | D=0) = E(Y | D=0)$ ，故可以後者取代 y' 。因此：

$$\begin{aligned} B^U &= [\pi \times E(Y | D=1) + (1-\pi) \times y''] - [\pi \times y' + (1-\pi) \times E(Y | D=0)] \\ &= [\pi \times E(Y | D=1) + (1-\pi) \times E(Y | D=1)] - [\pi \times E(Y | D=0) + (1-\pi) \times E(Y | D=0)] \\ &= E(Y | D=1) - E(Y | D=0) \end{aligned}$$

換言之，MTS 假定可以進一步將平均效應之上限，限縮為直觀估計式之上限。其區段之寬度為：

$$\begin{aligned} B^U - B^L &= [E(Y | D=1) - E(Y | D=0)] \\ &\quad - \{ [\pi \times E(Y | D=1) + (1-\pi) \times y'] - [\pi \times y'' + (1-\pi) \times E(Y | D=0)] \} \\ &= (1-\pi) \times E(Y | D=1) - \pi \times E(Y | D=0) - (1-\pi) \times y' + \pi \times y'' \\ &= \pi \times [y'' - E(Y | D=0)] - (1-\pi) \times [y' - E(Y | D=1)] \\ &\leq (y'' - y') \end{aligned}$$

故此一寬度小於或等於無假定情況下的區段寬度。此一區段未必包含 0 在內，若不包含 0，則識別出了參數的正負號。

若依變數 Y 為比例，介於 0 與 1 之間，則 $y'=0$ ， $y''=1$ ，MTS 假定下之區段更簡化為：

$$\pi \times [\pi \times E(Y | D=1) - (1-\pi) \times E(Y | D=0)] - \pi \leq E(\delta) \leq E(Y | D=1) - E(Y | D=0)$$

其寬度 $\pi \times [1 - E(Y | D=0)] + (1-\pi) \times E(Y | D=1) \leq 1.0$ ，小於或頂多等於無假定情況下的區段寬度。

(三) 單向個別效應差 (有益無害) (Monotone Treatment Response, MTR) 假定

若依據學理可以假定：原因變數對每個人 i 的 Y 都「有益無害」，亦即每個人的 Y^1 值，橫豎都不會小於（含大於或等於）其 Y^0 值。Manski 與 Pepper(2000, 998) 將此稱為「單向個別效應差」(Monotone treatment response, MTR) 假定，具體表示為： $Y_i^1 \geq Y_i^0$ 或 $\delta_i \geq 0$ ，這意味著兩組 Y 平均值有下列的關係 (Manski 2007, 189-191; Manski and Pepper 2000, 1002-1003)：

$$\begin{cases} y^l \leq E(Y^0 | D=0) \leq E(Y^1 | D=0) \leq y^u \\ y^l \leq E(Y^0 | D=1) \leq E(Y^1 | D=1) \leq y^u \end{cases}$$

MTR 假定若成立，平均效應 $E(\delta) = E(Y^1 - Y^0) \geq 0$ ，故下限 B^L 必為 0。詳言之，因為 B^L 雖仍等於「 $E(Y^1)$ 之下限」減去「 $E(Y^0)$ 之上限」，但「 $E(Y^1)$ 之下限」中未知的反事實項 $E(Y^1 | D=0)$ 依 MTR 的第一項假定必不小於可觀察之 $E(Y^0 | D=0) = E(Y | D=0)$ ，故可以後者取代 y^l ；同理，「 $E(Y^0)$ 之上限」中未知的反事實項 $E(Y^0 | D=1)$ 依 MTR 的第二項假定必不大於可觀察之 $E(Y^1 | D=1) = E(Y | D=1)$ ，故可以後者取代 y^u 。因此：

$$\begin{aligned} B^L &= [\pi \times E(Y | D=1) + (1-\pi) \times y^l] - [\pi \times y^u + (1-\pi) \times E(Y | D=0)] \\ &= [\pi \times E(Y | D=1) + (1-\pi) \times E(Y | D=0)] - [\pi \times E(Y | D=1) + (1-\pi) \times E(Y | D=0)] \\ &= 0 \end{aligned}$$

故若 MTR 假定成立，下限必為 0，形同設定平均效應為正值。至於上限，則仍維持無假設情況之上限：

$$B^U = [\pi \times E(Y | D=1) + (1-\pi) \times y^u] - [\pi \times y^l + (1-\pi) \times E(Y | D=0)]$$

在 MTR 假定下，區段的寬度為：

$$\begin{aligned} B^U - B^L &= \{[\pi \times E(Y | D=1) + (1-\pi) \times y^u] - [\pi \times y^l + (1-\pi) \times E(Y | D=0)]\} - 0 \\ &= (1-\pi) \times [y^u - E(Y | D=0)] - \pi \times [y^l - E(Y | D=1)] \\ &\leq (y^u - y^l) \end{aligned}$$

故此一寬度小於或等於無假定情況下的區段寬度。

若依變數 Y 為比例，介於 0 與 1 之間，則 $y^l=0$ ， $y^u=1$ ，MTR 假定下之區段更簡化為：

$$0 \leq E(\delta) \leq (1-\pi) \times [1 - E(Y|D=0)] + \pi \times E(Y|D=1)$$

其寬度 $(1-\pi) \times [1 - E(Y|D=0)] + \pi \times E(Y|D=1) \leq 1.0$ 。

(四) 單向組間平均差且個別效應差 (MTS-MTR) 假定

假定 MTS 及 MTR 均成立，則平均效應之下限為 0，亦形同設定平均效應為正值；上限為直觀估計式之上限 (Manski 2007, 191-193; Manski and Pepper 2000, 1004)：

$$B^L=0$$

$$B^U = E(Y|D=1) - E(Y|D=0)$$

若 MTS 及 MTR 均成立，其區段之寬度是 $E(Y|D=1) - E(Y|D=0)$ ，比無假定 MTS、或 MTR 假定下之區段都更窄。

若 Y 為介於 y^l 與 y^u 間之連續變數時，以上四種假定下之平均效應上下限，以及外因選組假定下之定點識別，均摘要列於表 2；若 Y 為比例或非 0 即 1 之二分類虛擬變數時，其上下限則列於表 3。

表 2 Y 為介於 y' 與 y'' 間之連續變數時，平均效應 (ATE) 之上下限

假定	下限 B^L	上限 B^U
純邏輯、無資料	$y' - y''$	$y'' - y'$
有資料、無假定	$[\pi \times E(Y D=1) + (1-\pi) \times y'] - [\pi \times y'' + (1-\pi) \times E(Y D=0)]$	$[\pi \times E(Y D=1) + (1-\pi) \times y''] - [\pi \times y' + (1-\pi) \times E(Y D=0)]$
Roy 模型 (Roy Model)	$(1-\pi) \times [y' - E(Y D=0)]$	$\pi \times [E(Y D=1) - y']$
單向組間平均差 (MTS)	$[\pi \times E(Y D=1) + (1-\pi) \times y'] - [\pi \times y'' + (1-\pi) \times E(Y D=0)]$	$E(Y D=1) - E(Y D=0)$
單向個別效應差 (MTR)	0	$[\pi \times E(Y D=1) + (1-\pi) \times y''] - [\pi \times y' + (1-\pi) \times E(Y D=0)]$
單向組間平均差 且 個別 效應差 (MTS-MTR)	0	$E(Y D=1) - E(Y D=0)$
外因選組 (ETS)		$E(Y D=1) - E(Y D=0)$

資料來源：作者自行整理。

表 3 Y 為介於 0 與 1 間之比例，或 Y 為 0 或 1 之二分類變數時，平均效應 (ATE) 之上下限

假定	下限 B^L	上限 B^U
純邏輯、無資料	-1	+1
有資料、無假定	$\pi \times E(Y D=1) - (1-\pi) \times E(Y D=0) - \pi$	$\pi \times E(Y D=1) - (1-\pi) \times E(Y D=0) + (1-\pi)$
Roy 模型 (Roy Model)	$-(1-\pi) \times E(Y D=0)$	$\pi \times E(Y D=1)$
單向組間平均差 (MTS)	$\pi \times E(Y D=1) - (1-\pi) \times E(Y D=0) - \pi$	$E(Y D=1) - E(Y D=0)$
單向個別效應差 (MTR)	0	$\pi \times E(Y D=1) - (1-\pi) \times E(Y D=0) + (1-\pi)$
單向組間平均差 且個別效應差 (MTS-MTR)	0	$E(Y D=1) - E(Y D=0)$
外因選組 (ETS)		$E(Y D=1) - E(Y D=0)$

資料來源：作者自行整理。

肆、區段識別法之應用：混合選制下所謂之「污染效應」

憲政結構與選舉制度，是一個國家政治制度的兩大支柱，而其中國會選舉制度，關乎選票如何轉換為國會席次，其重要性不言可喻（王業立 2008；林繼文 2006；黃紀 2008b；2008c）。近二十餘年，「混合選制」(mixed-member electoral systems) 受許多國家青睞，不但新興民主國家頗多採用，連若干老牌民主國家也改採此一選制 (Farrell 2001)。例如紐西蘭於 1993 年由純單一選區制改採「單一選區兩票聯立制」(Mixed-member proportional system, MMP)，而日本則於 1994 年由原「單記非讓渡制」(Single nontransferable vote, SNTV) 改採「單一選區兩票並立制」(Mixed-member majoritarian system, MMM)。我國亦於 2005 年修憲，立委選舉由 SNTV 制改為類似日本之「單一選區兩票並立制」。

關於選制之影響，早在半個世紀前，Duverger(1959, 217, 239) 提出之「律則」(單一選區制 (Single-member district, SMD) 有利於兩黨) 與「假設」(比例代表制 (Proportional representation, PR) 易促成多黨)，迄今仍被許多學者奉為選舉制度對政黨數影響之皋瑤。至於「單一選區兩票制」，儘管除了德國實施 MMP 歷史悠久之外，其餘各國採行時間並不算長，但也已引起學界的高度興趣，研究方興未艾。由於兩票制兼採「單一選區」及「比例代表」兩者，對鑽研比較選制及想驗證 Duverger 律則及假設的學者，似提供了絕佳的研究機會。以往分析選制之效應，多採跨國比較，惟因各國文化及社會結構殊異，故因果推論上總不免引發質疑與挑戰。如今採混合選制的國家兼具 SMD 與 PR 兩制，同一國的風土民情彷彿是「控制不變」的常數，因此許多學者寄予厚望，認為此乃天然之「控制實驗」(controlled experiments) (如 Moser 與 Scheiner(2004; 2009) 等)，同一個國家兩種選票結果的差異必屬選制的效應，因果推論的效度無可置疑。依照 Duverger 的機械因素及心理因素，SMD 有利於大黨，小黨的支持者為了不浪費選票，往往採策略投票 (strategic voting)，轉而支持較有希望當選的候選人；至於政黨票，則仍依其偏好支持原屬意的政黨 (sincere voting)。在宏觀的層次，這意味著在選區參選的政黨會減少；在微觀的層次，這意味著小黨支持者傾向分裂投票。

但也有部分學者對此提出嚴峻之挑戰 (如 Cox 與 Schoppa(2002)、Ferrara、Herron 與 Nishikawa(2005)、Herron 與 Nishikawa(2001)、Huang(2007) 等)，認為混合選制的 SMD 及 PR 兩票不能拆開 (separated) 個別檢視，因為選民既然可同時投兩票，其考量就會有連動效果 (interaction effect)，與採純 SMD 制或純 PR 制下的投票抉擇不同；而政黨亦深諳此理，在提名及競選策略上都竭盡所能操弄兩票之互動關係，以收席次極大化之效，尤其是小黨雖明知在單一選區中難敵大黨，其候選人當選無望，但仍提名以壯聲勢，藉此鞏固票

源、並吸引獨立選民在政黨票上給予支援，如候選人的魅力夠，甚至還可能吸引他黨選民的選票。換言之，SMD 對 PR 的影響，破壞了兩票各自的純淨效應，有如「控制實驗」遭到污染一般，故亦稱為「污染效應」(contamination effect)。影響所及，在宏觀的層次，這意味著在選區參選的政黨並不會減少，導致非 Duverger 均衡狀態；在選民的層次，這意味著各黨有提名候選人之選區，不但其支持者傾向一致投票，且還會吸引獨立甚或他黨選民的政黨票，因而提升該黨政黨票的得票率。

從方法論的角度看，如要比較兩種選制的效應，「處理組」和「控制組」不能是同一群受試者，也就是兩者必須互斥，就此而言，「污染效應說」對「控制實驗說」的批評確實言之成理。不過令人訝異的是，文獻中驗證污染效應的方法，卻又常想當然爾地以「處理組的政黨票得票率」減去「控制組政黨票得票率」，自陷於隨機分派控制實驗的「外因選組」強勢假定，自相矛盾。正確的做法應該要跳脫「SMD 及 PR 兩種選制之間」的比較，而將研究問題界定為「混合選制之內」政黨在單一選區提名對其政黨票可能產生之效應；亦即思考政治人物在兩票制之下，以超越政黨票法定門檻並極大化總席次為目標，如何選擇提名策略？在此前提下，才能正視政黨佈局時考慮在選區是否提名對整體得票與席次的考量，這種思維下的提名決策，是典型的「內因選組」。

準此，「污染效應說」認為即使小黨也有強烈動機要在單一選區提名候選人，其立論似乎太偏重拉抬政黨票的效益面，而忽略了此一策略的成本面，故其學理的考量還不夠縝密。也就是說，政黨在選區提名候選人及其競選活動，都需要投入經費資源及人力，政黨固然希望衝高政黨票，但也要考慮提名候選人參選所需付出之代價是否符合成本效益，因此，政黨應該會選擇性地僅在較有把握或至少可「雖敗猶榮」的選區參選，而非有選必提。選擇的考量因素，包括制度面與實務面：

一、制度上，例如日本的 MMM 選制允許在 SMD 及 PR 票重複提名 (dual candidacy)，即使未在單一選區當選，也可能因惜敗率而在其比例代表選區（全國共分 11 個 PR 選區）中「敗部復活」，取得席次，提高政黨提名候選人參選之動機（林繼文 2008）。但我國新選制並不允許重複提名，故無此一制度誘因，惟選罷法第 24 條規定，除非過去三屆立委選舉得票率曾達 2% 或現有五位立委，否則列入政黨票的條件是「該次區域及原住民立法委員選舉推薦候選人達十人以上」，使新的政黨不得不至少提名 10 位選區候選人。2008 年立委選舉，在「全國不分區」選票（即政黨票）上共有 12 個政黨，其提名區域立委之選區數，如表 4 所示。

表 4 2008 年立委選舉政黨票上之 12 個政黨有提名區域立委之選區數

政黨	有提名區域立委之選區數
國民黨	69 ^a
民進黨	69
新黨	0
台聯	13
紅黨	11
第三社會黨	10
綠黨	10
台灣農民黨	10
公民黨	10
客家黨	10
制憲聯盟	10 ^b
無盟	4

資料來源：作者自行整理。

說明：^a 含國民黨苗栗縣第 2 選區開放參選。未提名之 4 個選區係「禮讓」無盟。

^b 制憲聯盟在臺北市第 3 及第 6 選區各提名兩人。

二、務實面，MMM 選制下，兩票之應選席次為固定，而立委選舉新制又不允許雙重提名，SMD 票對 PR 票的影響理應甚小。不過主張兩票連動的「污染效應」學派，均認為小黨在 SMD 提名，會對該黨之 PR 票有正面影響，因為政黨傾向挑選強棒候選人，或在地盤較穩、支持者較多、或估計在政黨票上策略投票者較眾的選區提名候選人。此一觀點有幾種不同的解讀：

1. 擇優候選人提名說。在不訴諸黨內初選的機制下，政黨菁英提名的標準，是預期區域立委候選人的形象或能力會提升該選區政黨票者，才刻意挑選提名，否則寧可不提。此一觀點，其實就是 Roy 模型的假定。
2. 反映選區原本的實力說。被提名候選人的單一選區，本來就是該黨較有實力及較有把握的選區，或是政黨較能操弄「以輸為贏」策略（林繼文 2008）的選區，因此平均而言有被提名之選區政黨票得票率，高於未被提名候選人之選區。此一觀點，正是 MTS 假定。
3. 有益無害之拉抬效應說。一個選區有提名候選人的話，其政黨票得票率只會比沒提名好、不會比沒提名差，亦即該選區之政黨票得票率有增無減，有百利而無一害。原因是以候選人的人氣壯黨氣，除了鞏固該區既有票源外，還可能贏得中立選民的青睞，甚至說服他黨支持者轉向 (Ferrara, Herron, and Nishikawa 2005)。此一純效益

的觀點，正是 MTR 假定，也是三個假定中最重的一個，因其設定了平均效應必為正值或 0。

4. 如虎添翼說。在已有實力的選區提名候選人，進一步拉抬其政黨票。此一觀點，是結合了 MTS 及 MTR 之假定。

總之，政黨是否在選區提名區域候選人，是一種刻意的選擇 (selection)，並非隨機，依前述關於因果推論的學理，若要評估 SMD 提名對 PR 票的影響，其因果推論必須顧及「SMD 提名與否」之內因性，否則估計將產生偏誤。Maeda(2008) 指出「污染效應說」在實證研究中忽略內因性問題，故改採「處理效果模型」(Heckman 1976; 1978; 1979; Maddala 1983) 重新分析日本 1996 至 2005 的四次眾議員選舉之選區資料，發覺並無顯著之「污染效果」。Maeda(2008) 的分析正視原因變數之內因性，大方向雖然正確，但仍有其盲點。由於「處理效果模型」對函數形式（直線型）及依變數之機率分佈（多變數常態分佈）做了很強的假定，可能如前述 Manski(2003) 之「可信度遞減律」所言，假定下得太重損及因果推論之效度。

一、第七屆立委選舉台聯之案例

本文以 2008 年立委選舉中，台灣團結聯盟（以下簡稱台聯）為例，評估台聯在單一選區中提名候選人，對其政黨票有無影響？如有，其影響多大？之所以分析台聯，是因為：

1. 第七屆立委選舉前，泛藍整合成功，泛綠卻未達成共識，台聯與當時執政之民進黨頗多爭執，台聯面臨新選制下很難與兩大黨在單一選區抗衡、而民進黨似無協調禮讓提名之誠意，頗有背水一戰、衝刺不分區選票 5% 門檻的態勢。依選罷法 24 條，台聯即使不提名區域立委亦可列名政黨票之中，但仍選擇提名 13 位區域立委候選人（見表 4），與新黨選擇不提名區域立委，形成有趣的對比。
2. 國民黨、民進黨有提名區域立委的單一選區，在 73 個中佔了 69 個 (94.5%)，未提名的選區比例甚低，尤其國民黨未提名的 4 個選區其實是禮讓給結盟之無黨團結聯盟，因此很難比較有提名和未提名這兩類選區。而小黨之中，新黨選擇完全不提名區域立委，亦無從比較有無提名的差異。其餘新登記的政黨之所以在 10 個左右的選區提名區域立委，只是為了符合選罷法第 24 條規定可列入政黨票的基本條件，且 PR 選票上的紅黨、綠黨、台灣農民黨、第三社會黨、公民黨、客家黨、制憲聯盟七者全國合計的政黨票只佔了 3.68%，而台聯則在全國還得到 3.53% 的政黨

票，雖未達 5% 的門檻，但僅次於國民黨的 51.23%、民進黨的 36.91%、及新黨的 3.95%。

二、台聯選前之提名過程

台聯的區域立委提名過程，因為與民進黨的角力與協商，可謂一波三折，然在在顯示最後提名之選區均為刻意之選擇及與現實之妥協。台聯提名無須黨內初選，黨中央之提名小組搶在 2007 年 2 月 26 日民進黨尚未開始初選前即公佈首波 17 個參選名單（自由時報電子報，2007 年 2 月 27 日），¹⁰ 以當時台聯之現任立委為主、任黨職者為輔，卻引起民進黨不快。台聯提出泛綠以 2004 年第六屆立委選舉得票率為依據，推估民進黨至少應禮讓台聯 13 個選區，而民進黨則直言不可能（自由時報電子報，2007 年 6 月 20 日）。¹¹ 在與民進黨多次協商無具體結論下，台聯於 8 月 20 日公佈第二波之 3 人名單。9 月 19 日民進黨中執會決議禮讓台聯「三加一」選區，卻反被台聯斥為毫無誠意，10 月 4 日台聯發言人周美里表示，台聯接下來提名以民進黨未提名選區、雷雨奇兵選區以及台灣主體意識被質疑的民進黨候選人選區為主（大紀元新聞網，2007 年 10 月 4 日），¹² 接著 10 月 5 日公佈第三波 3 人名單，都是李登輝學校學員；再於 11 月 19 日參選登記前補提 2 人，並公佈不分區之 15 人名單，力拼 10% 政黨票（自由時報電子報，2007 年 11 月 21 日）。¹³ 然而台聯前後共提名的 25 個區域立委中，6 月時為了和民進黨協商而勸退 3 人、10 月底及 11 月為「清黨」而出走 4 人（其中 2 人被民進黨徵召參選）、退黨改入民進黨 1 人、自行退選 4 人，故最後台聯正式提名之區域立委共 13 人（見表 5）。

¹⁰ 黃忠榮，2007，〈區域立委提名 台聯首波名單出爐〉，自由時報電子報，2 月 27 日，<http://www.libertytimes.com.tw/2007/new/feb/27/today-p2.htm>，檢索日期：2009 年 9 月 20 日。

¹¹ 李欣芳，2007，〈初選舞弊清查緩慢 綠委提名 沒把握如期公告〉，自由時報電子報，6 月 20 日，<http://www.libertytimes.com.tw/2007/new/jun/20/today-p2.htm>，檢索日期：2009 年 9 月 20 日。

¹² 謝佳珍，2007，〈台聯擬再提名區域立委 李登輝學校學員為主〉，大紀元新聞網，10 月 4 日，<http://www.epochtimes.com/b5/7/10/4/n1856312.htm>，檢索日期：2009 年 9 月 20 日。

¹³ 彭顯鈞，2007，〈台聯搶奪 28 席 三社黨打青年牌〉，自由時報電子報，11 月 21 日，<http://www.libertytimes.com.tw/2007/new/nov/21/today-fo1-3.htm>，檢索日期：2009 年 9 月 20 日。

表 5 2008 年立委選舉台聯有提名區域立委之選區及其得票率

選區名稱	台聯候選人	區域得票率 %	政黨票得票率 %
臺北市第三選區	謝馥米	1.118	3.124
臺北市第七選區	張國慶	0.934	2.697
臺北縣第一選區	莊孟學	1.125	3.680
臺北縣第二選區	林志嘉	6.008	4.024
臺北縣第三選區	劉一德	1.083	3.695
臺北縣第四選區	鄭餘豪	0.650	3.111
臺北縣第七選區	林育任	1.344	3.287
臺北縣第九選區	周倪安	11.059	2.418
桃園縣第六選區	姚吉鴻	2.571	2.530
臺中縣第四選區	高基讚	35.998	3.955
彰化縣第二選區	林招膨	3.120	3.795
高雄縣第四選區	吳隆傑	0.743	3.366
嘉義市	凌子楚	12.631	4.501

資料來源：作者自行整理。

從上述台聯提名區域立委的曲折過程，可以看出：

1. 台聯最後提名之選區均為刻意之選擇及與現實之妥協，並非隨機。
2. 台聯一開始之提名以現任立委為主、任黨職者為輔，應是考量其地方之經營，偏向「反映選區實力」；但第二波之後的提名，則較接近「擇優提名」與寄望其「拉抬效應」。

三、台聯提名區域立委之平均因果效應估算

儘管台聯提名區域立委頗多波折，畢竟在 73 個選區中提名了 13 個（佔 17.81%），其餘 60 個選區未提名，應是 12 個列名政黨票中最適合評估提名區域立委對政黨票得票率影響的案例。不過以單一選區為單位，總共才 73 個觀察值，且為母群資料而非樣本，不但總觀察數少，宣稱其政黨票得票率近乎常態分佈亦很牽強，故不適合採用 Heckman(1978; 1979; 2005) 的「處理效果模型」或其延伸。倘若我們想一窺台灣首度採用之 MMM 選制，對背水一戰、慎重提名區域候選人的小黨，選擇提名候選人對其政黨票之影響如何，則採 Manski 的無母數 (nonparametric) 區段推論法顯然最切合實際，無需顧慮違反常態分佈等牽強的假定。

首先，表 6 顯示，以選區為分析單位，台聯有提名的 13 個選區佔 73 個選區的比例為 $\pi = P(D=1) = 0.17808$ ，而台聯在這 13 個選區中政黨票之平均得票率為

$E(Y|D=1)=0.03399$ ；台聯未提名的選區佔 $1-\pi=P(D=0)=0.82192$ ，其政黨票之平均得票率為 $E(Y|D=0)=0.03501$ 。一般直覺上想當然爾的推論，就是拿台聯有提名區域候選人的 13 個單一選區的政黨票平均得票率，與其他未提名候選人的 60 個選區相比，兩者之差就視為是提名區域立委對政黨票的因果效應。也就是：

$$E(Y|D=1) - E(Y|D=0) = 0.03399 - 0.03501 = -0.00102 \text{ (或 } -0.102\%)$$

平均效應為負值，表示台聯有提名之選區，政黨票平均得票率不增反減，比未提名者還低了約千分之一。這與「互動效應說」的預期相反。

表 6 2008 年台聯提名區域候選人之效應：基本數據資料

組別	Y^1	Y^0
有提名之選區 $\pi=P(D=1)$ $=0.17808$	$E(Y D=1)=0.03399$	觀察不到的「反事實」 $0 \leq E(Y^0 D=1) \leq 1$
未提名之選區 $1-\pi=P(D=0)$ $=0.82192$	觀察不到的「反事實」 $0 \leq E(Y^1 D=0) \leq 1$	$E(Y D=0)=0.03501$

資料來源：作者自行整理。

不過如前節所述，這個直觀的估計方式，其實無形中強加了外因選組 (ETS) 的強勢假定，但這個假定很難自圓其說，因為無法置信台聯的領袖會是以隨機的方式決定哪些選區提名、哪些不提名，因此 ETS 假定雖得到了「定點識別」，但立基過於牽強，故該數值恐禁不起挑戰。

純從邏輯來看，提名候選人對政黨票的效應，再拉抬也大不過 1.0（即 100%，從無人支持變成全民擁戴），再折損也低不過 -1.0（即 -100%，從全民擁戴變成眾叛親離），其理甚明。接著，區段識別法從純資料、無假定開始，逐步引入輕微但比較可信的假定，在最寬的區段 [-1, +1] 中，識別提名候選人之平均因果效應的上下限。由於結果變數 Y 為政黨票之得票率，其值必介於 0 與 1 之間，故區段識別可適用表 3 之上下限計算式。其識別之區段上下限值，列於表 7。

表 7 台聯提名區域立委對政黨票得票率之影響：區段識別

假定	區段 [B ^t , B ^u]
純資料、無假定	[-0.20081, 0.79919]
擇優候選人提名 (Roy 模型)	[-0.02878, 0.00605]
反映選區實力 (MTS 假定)	[-0.20081, -0.00102]
有益無害之拉抬效應 (MTR 假定)	[0, 0.79919]
在有實力的選區拉抬 (MTS-MTR 假定)	[0, -0.00102]

資料來源：作者自行整理。

- (一) 純資料、不加任何假定的情況下，平均因果效應的區段為 [-0.20081, 0.79919]，亦即介於 -20.081% 與 79.919% 之間。
- (二) 結合數據資料與 Roy 模型（假定台聯擇優候選人提名），平均因果效應的區段縮窄為 [-0.02878, 0.00605]，亦即介於 -2.878% 與 0.605% 之間。如前所述，此一區段包含 0 在內，未識別正負號，無法排除負面效應和無效應的可能，但亦顯示即使提名之區域立委對政黨票得票率有正面幫助，在最理想的情況下效果也非常之低，僅約百分之 0.6。
- (三) 結合數據資料與 MTS（假定台聯之提名反映選區原有之實力），平均因果效應的區段為 [-0.20081, -0.00102]，亦即介於 -20.081% 與 -0.102% 之間。換言之，平均因果效應竟均為負值，與「污染效應說」預期的正面效應相反。
- (四) 結合數據資料與 MTR（假定台聯提名區域立委對該選區之政黨票有利無害），平均因果效應的區段為 [0, 0.79919]，亦即介於 0% 與 79.919% 之間。不過如前節所述，此一正面效應是來自於 MTR「有利無害」的純效益假定。
- (五) 結合數據資料與 MTS 暨 MTR（假定台聯在已有實力的選區提名候選人並發揮純效益），平均因果效應的區段為 [0, -0.00102]，上限竟低於下限，是個不合理的區段，顯示 MTS-MTR 假定不成立。因 0 之下限，係來自頗強之 MTR 假定，因此連帶對 MTR 假定也產生質疑。

值得注意的是，上述（二）至（五）的區段識別，是在具體考量各種學理與實務上可能的「內因性」後，針對「污染效應說」提出的解讀，對內因所做的假定由弱漸強（但均未加上線型函數或常態分佈等過度強勢的假定），所推算之平均因果效應值的範圍，四者確實都比（一）純數據資料之區段更窄。在這四種解讀之中，首先，假定下得最重的「如虎添翼說」（MTS 暨 MTR 假定）確定不成立，故予以拒斥。其次，「有益無害拉抬說」（MTR）與「反映選區實力說」（MTS）兩者僅能其中之一成立，或兩者皆不成立。乍看之

下，純就區段的正負號而言，「有益無害拉抬說」的區段均落於正值，似印證正向之污染效應，但如 Manski「可信度遞減律」之警告，即使一開始就假定個別選區之效應均為正值，其區段仍甚寬，介於 0 至近乎八成之間，足見 MTR 假定雖強，識別力卻很弱，可信度甚低。接著，如果再稍微放鬆，僅假定組間之平均差為單向，區段值卻呈現與正向污染效應的預期完全相反之負號，亦即平均而言有提名候選人的選區，政黨得票率反而低於未提名候選人之選區。換言之，台聯反而可能是考量在部分原實力較弱的選區提名候選人，藉此拉抬其弱勢選區之政黨票。最後，假定最寬鬆之「擇優提名說」(Roy 模型)的區段也以負面居多，即使在最佳的狀況下，正向假定平均因果效應也微乎其微。

因此在排除太重的 MTS 暨 MTR 假定、及 MTR 假定後，MTS 假定又否證了「污染效應說」的正面效應，而即使最寬鬆之 Roy 模型在最佳的狀況下，正向平均因果效應也微乎其微。因此我們可以說，2008 年立委選舉台聯提名區域立委，對其政黨票得票率並無「污染效應說」學理預期之正面影響。誠然，2008 年是台灣首次在 MMM 選制下進行之立委選舉，未來政黨能否運用提名選區候選人來影響政黨得票，還值得持續觀察分析。

伍、結語

以觀察研究進行因果推論之所以棘手，癥結在於比較研究的組別，往往取決於因和果之間的內部因素，也就是所謂「內因性」，造成平均因果效應的識別問題。一般典型的「因果效應參數模型」，雖有考慮內因問題，但多建立在很強的函數形式及變數分佈等假定之上。如果研究的主題及資料確實符合這些假定，則可充分運用；但社會科學研究也常常碰到與假定不符的情況，此時 Manski 的區段識別法更為適用，因為這個方法從純邏輯出發，逐步帶入資料及不同強度的假定，檢視其對於參數區段的影響，將假定與推論之間的關係完全透明化，避免為了達到「定點識別」而強加與實際不符的假定，導致過當的推論。本文依據文獻中「污染效應說」提出四個與之相容的假定，檢視 2008 年立委選舉台聯提名區域立委及政黨票得票的資料後，發覺除非強行假定個別效應必為正值，否則在該次選舉並無「污染效應說」宣稱之正面影響。

* * *

投稿日期：99.03.22；修改日期：99.05.27；接受日期：99.07.25

附錄：純邏輯、尚無資料之區段 (Logical Bounds)

平均效應的區段和 Y 的值域 (range) 息息相關。設若 Y^0 及 Y^1 的值域已知為有限 (bounded)，必介於其最小及最大值 $[y^l, y^u]$ 之間，那麼邏輯上，平均效應 $E(\delta) = E(Y^1) - E(Y^0)$ ，再小也小不過「 $E(Y^1)$ 的下限減去 $E(Y^0)$ 的上限，即 $(y^l - y^u)$ 」、再大也大不過「 $E(Y^1)$ 的上限減去 $E(Y^0)$ 的下限，即 $(y^u - y^l)$ 」，也就是在沒有任何觀察資料佐證下，平均因果效應的上、下限各為 (Manski 1989, 344-346; 1990, 320)：

$$B^L = [E(Y^1) \text{ 之下限 } - [E(Y^0) \text{ 之上限 }] = (y^l - y^u)$$

$$B^U = [E(Y^1) \text{ 之上限 } - [E(Y^0) \text{ 之下限 }] = (y^u - y^l)$$

換言之，平均效應的最窄 (sharp) 區段為：

$$(y^l - y^u) \leq E(\delta) \leq (y^u - y^l)$$

而此一區段的寬度為 $B^U - B^L = 2 \times (y^u - y^l)$ 。若依變數 Y 為比例 (proportion)，必介於 0 與 1 之間，則邏輯上平均效應必介於 -1 與 +1 之間，該區段 $[-1, +1]$ 的寬度為 2。

參考文獻

I. 中文部分

王業立，2008，《比較選舉制度》（五版），台北：五南。

(Wang, Yeh-li. 2008. *Comparative Electoral Systems* (5th ed.). Taipei: Wunan.)

林繼文，2006，〈政府體制、選舉制度與政黨體系：一個配套論的分析〉，《選舉研究》，13(2): 1-35。

(Lin, Jih-wen. 2006. "Power Division, Voting Rule, and Party System: What Do We See when They Are Considered Together?" *Journal of Electoral Studies* 13(2): 1-35.)

-----，2008，〈以輸為贏：小黨在單一選區兩票制下的參選策略〉，《選舉研究》，15(2): 37-66。

(-----, 2008. "To Lose Is to Win: The Candidate-Placement Strategy of Minor Parties under Japan's Mixed-Member Majoritarian System." *Journal of Electoral Studies* 15(2): 37-66.)

黃紀，2008a，〈因果推論與觀察研究：「反事實模型」之思考〉，《社會科學論叢》，2(1): 1-21。

(Huang, Chi. 2008a. "Causal Inference and Observational Study: On the Counterfactual Model of Causality." *Journal of Social Sciences* 2(1): 1-22.)

-----，2008b，〈緒論：選舉制度的脈絡與效應〉，載於《如何評估選制變遷：方法論的探討》，黃紀、游清鑫主編，台北：五南。

(-----, 2008b. "Contexts and Consequences of Electoral Systems: An Introduction." In *Consequences of Electoral System Change: Methodological Perspectives*, eds. Chi Huang and Ching-hsin Yu. Taipei: Wunan.)

-----，2008c，〈單一選區兩票並立制下選民之投票抉擇：分析方法之探討〉，載於《如何評估選制變遷：方法論的探討》，黃紀、游清鑫主編，台北：五南。

(-----, 2008c. "Voter Choice under Mixed-Member Majoritarian System: A Multilevel Mixed Logit Model." In *Consequences of Electoral System Change: Methodological Perspectives*, eds. Chi Huang and Ching-hsin Yu. Taipei: Wunan.)

II. 外文部分

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

- Barnow, Burt S., Glen G. Gain, and Arthur S. Goldberger. 1981. "Issues in the Analysis of Selectivity Bias." In Vol.5 of *Evaluation Studies Review Annual*, eds. Ernst Stromsdorfer and George Farkas. Beverly Hills: Sage.
- Cho, Wendy K. Tam, and Charles F. Manski. 2008. "Cross-Level/Ecological Inference." In *The Oxford Handbook of Political Methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford: Oxford University Press.
- Cox, Karen, and Leonard J. Schoppa. 2002. "Interaction Effects in Mixed-Member Electoral Systems: Theory and Evidence from Germany, Japan, and Italy." *Comparative Political Studies* 35(9): 1027-1053.
- Deming, W. Edwards. 1975. "The Logic of Evaluation." In Vol.1 of *Handbook of Evaluation Research*, eds. Elmer L. Struening and Marcia Guttentag. Beverly Hills: Sage.
- Duverger, Maurice. 1959. *Political Parties* (English edition, revised). New York: Wiley.
- Farrell, David M. 2001. *Electoral Systems: A Comparative Introduction*. New York: Palgrave.
- Ferrara, Federico, Erik S. Herron, and Misa Nishikawa. 2005. *Mixed Electoral Systems: Contamination and Its Consequences*. New York: Palgrave Macmillan.
- Greene, William H. 2008. *Econometric Analysis* (6th ed.). Upper Saddle River: Prentice Hall.
- Heckman, James J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5(4): 475-492.
- 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46(6): 931-959.
- 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 547(1): 153-161.
- 2005. "The Scientific Model of Causality." In Vol. 35 of *Sociological Methodology*, ed. Ross M. Stolzenberg. New York: American Sociological Association.
- Heckman, James J., and Edward Vytlacil. 2007a. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." In Vol. 6B of *Handbook of Econometrics*, eds. James J. Heckman and Edward E. Leamer. Amsterdam: North Holland.
- 2007b. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments." In Vol. 6B of *Handbook of Econometrics*, eds. James J. Heckman and Edward E. Leamer. Amsterdam: North Holland.

- Heckman, James J., Justin L. Tobias, and Edward Vytlacil. 2001. "Four Parameters of Interest in the Evaluation of Social Programs." *Southern Economic Journal* 68(2): 210-223.
- , 2003. "Simple Estimators for Treatment Parameters in a Latent-Variable Framework." *The Review of Economics and Statistics* 85(3): 748-755.
- Herron, Erik S., and Misa Nishikawa. 2001. "Contamination Effects and the Number of Parties in Mixed-Superposition Electoral Systems." *Electoral Studies* 20(1): 63-86.
- Holland, Paul W. 1986. "Statistics and Casual Inference." *Journal of the American Statistical Association* 81(396): 945-960.
- Hsiao, Cheng. 1983. "Identification." In Vol. I of *Handbook of Econometrics*, eds. Zvi Griliches and Micael D. Intriligator. Amsterdam: North-Holland.
- Huang, Chi. 2007. "Assessing the Impact of Mixed Electoral System in Taiwan: Methodological Challenges of Testing Interaction Effects." Presented at the International Symposium on Mixed Electoral Systems in East Asia, Taipei.
- Jackson, John E. 2008. "Endogeneity and Structural Equation Estimation in Political Science." In *The Oxford Handbook of Political Methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford: Oxford University Press.
- Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- Maeda, Ko. 2008. "Re-Examining the Contamination Effect of Japan's Mixed Electoral System Using the Treatment-Effects Model." *Electoral Studies* 27(4): 723-731.
- Manski, Charles F. 1989. "Anatomy of the Selection Problem." *The Journal of Human Resources* 24(3): 343-360.
- , 1990. "Nonparametric Bounds on Treatment Effects." *American Economic Review Papers and Proceedings* 80(2): 319-323.
- , 1995. *Identification Problems in the Social Science*. Cambridge: Harvard University Press.
- , 2003. *Partial Identification of Probability Distributions*. New York: Springer.
- , 2005. *Social Choice with Partial Knowledge of Treatment Response*. Princeton: Princeton University Press.
- , 2007. *Identification for Prediction and Decision*. Cambridge: Harvard University Press.
- , 2008. "Partial Identification in Econometrics." In Vol.6 of *The New Palgrave Dictionary of Econometrics*, eds. Steven N. Durlauf and Lawrence E. Blume. New York: Palgrave Macmillan.

- Manski, Charles F., and John V. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68(4): 997-1010.
- Morgan, Stephen L., and Christopher Winship. 2007. *Counterfactual and Casual Inference: Method and Principles for Social Science*. Cambridge: Cambridge University Press.
- Moser, Robert G., and Ethan Scheiner. 2004. "Mixed Electoral Systems and Electoral System Effects: Controlled Comparison and Cross-National Analysis." *Electoral Studies* 23(4): 575-599.
- , 2009. "Strategic Voting in Established and New Democracies: Ticket Splitting in Mixed-Member Electoral Systems." *Electoral Studies* 28(1): 51-61.
- Rosenbaum, Paul R. 2002. *Observational Studies* (2nd ed.). New York: Springer.
- Roy, A.D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers, New Series* 3(2): 135-146.
- Rubin, Donald B. 1974. "Estimating Casual Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5): 688-701.
- , 1990. "Formal Modes of Statistical Inference for Causal Effects." *Journal of Statistical Planning and Inference* 25(3): 279-292.
- , 1991. "Practical Implications of Modes of Statistical Inference for Causal Effects and the Critical Role of the Assignment Mechanism." *Biometrics* 47(4): 1213-1234.
- , 2005. "Casual Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association* 100(469): 322-331.
- Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach* (4th ed.). Canada: South-Western.

Causal Inference and Treatment Effect Evaluation: Partial Identification Approach and Its Application to Electoral System Effect

Chi Huang^{*}

Abstract

In social science we routinely ask questions of the form: What is the effect of X on Y? Attempts to answer these questions unavoidably involve causal inference. However, social scientists relying on observational studies are often plagued by the endogeneity problem. That is, the treatment and control groups are not randomly assigned by researchers but formed spontaneously by some factors related to the causal variable of interest. Some existing parametric models, such as the popular Heckman's treatment-effects model, do take account endogeneity problem but are built upon quite stringent functional and distributional assumptions such as linearity and bivariate Normal distribution. Powerful as they are in point identifying causal parameters, their assumptions are not always met in reality. When these assumptions are violated, a better alternative is to adopt Charles F. Manski's nonparametric partial identification approach. This uncommon approach promotes forthright acknowledge of ambiguity in social science research and discredits misplaced certainty of point identification at the cost of imposing strong and yet incredible assumptions. Relying on available data and weak but credible assumptions, partial identification theory reveals the causal effect parameter that lies in a set that is smaller than the logical range

^{*} University Chair Professor, Department of Political Science, and Research Fellow, Election Study Center, National Chengchi University.

of the parameter but larger than a single point. Yet it makes transparent the relationship between maintained assumptions and causal inference.

Starting from the counterfactual model of causality, this article introduces Manski's partial identification theory and examines its implications on the upper and lower bounds of the average treatment effect (ATE). We then illustrate the approach by applying it to the case of Taiwan's 2008 Legislative Yuan election and examining whether Taiwan Solidarity Union's nomination in 13 single-member districts had any "contamination effect" on its party list vote shares.

Keywords: causal inference, endogeneity problem, partial identification, mixed-member electoral systems, contamination effect