

廢票哪裡來？ ——無效票定義範圍擴大 對 2004 年總統選舉的影響

駱明慶*

台灣大學經濟系教授

本文使用 2000 年和 2004 年總統選舉各鄉鎮的投票資料，配合 2000 年戶口普查的人口資料，探討 2004 年總統選舉廢票大量增加的原因。首先，廢票比例增加較多的是都市化程度較低的縣分；民進黨主政與否，廢票比例增加的幅度並沒有明顯的差異。其次，不識字比例較高的鄉鎮，在無效票範圍擴大之後，廢票比例顯著地多增加了 1.392%；迴歸分析的結果顯示，不識字比例增加 1%，會使得廢票比例多增加 0.21%。區位推論模型的估計結果亦發現，不識字選民投出廢票的比例在 2004 年大幅增加，且明顯較識字選民高出許多。此外，我們也檢驗「做票說」對於「陳呂得票率越高，廢票比例越高」這個現象的詮釋，說明兩者並沒有因果關係。控制不識字比例之後，陳呂得票率越高的鄉鎮，其廢票比例反而是減少的，但不具統計顯著性。綜合而言，「無效票範圍擴大」是唯一與資料一致的假說。

關鍵詞：廢票、不識字比例、總統選舉、做票

* 本文曾於 2005 年 12 月 23 日在中央研究院人文社會科學研究中心舉辦之《台灣經驗政治研究的現況與展望》學術研討會宣讀，感謝主持人黃秀端教授、評論人莊文忠教授，以及三位匿名審查委員的意見與建議。

收稿日期：94 年 8 月 29 日；接受刊登日期：95 年 5 月 15 日

1 前言

2004年3月20日第11屆總統副總統選舉的開票結果，陳水扁和呂秀蓮（陳呂）的得票數為6,471,970票，連戰和宋楚瑜（連宋）則得到6,442,452票，兩組候選人的差距只有29,518票，得票率的差距則只有0.2286%，如此接近的得票數，引發選後極為嚴重的爭議。¹其中一項開票當晚最明顯的爭議是，該次選舉的廢票數大為增加。根據中央選舉委員會的統計，廢票總數共337,297票，占總投票數13,521,719票的2.55%。相較而言，2000年總統大選時，廢票數只有122,278票，占總投票數12,786,671票的0.96%。換言，2004年的廢票數增加了215,019票，廢票比例也增加了1.59%，廢票比例增加大約1.5倍。相較於兩組候選人勝負差距只有0.2286%，2.55%的廢票比例自然成爲爭議的焦點，在野黨也質疑是否有選舉舞弊的問題，並要求重新驗票，「廢票哪裡來」因而成爲選後的主要爭議之一。

關於廢票大幅增加的假說，至少有三種。首先，中央選舉委員會認爲廢票之所以大幅增加，主要是因爲選罷法修訂，使得無效票範圍擴大，對有效票的認定採取更嚴格的標準，導致圈選在圈選欄之外的選票，都被判定爲廢票所致。其次，此次選舉過程中，民間出現「百萬廢票行動聯盟」，²號召選民拒絕爛蘋果的綁架，並呼籲選民主動投下廢票，可能也是原因之一。最後，在野黨在投票當天晚上也提出「做票」的質疑，認爲民進黨政府可能勾結選務人員，將連宋的選票做成廢票，才導致廢票數目大幅增加。

本文使用2000年和2004年兩次總統選舉，各鄉鎮投票結果的資料，配合2000年戶口普查中各鄉鎮人口特性的資料，³嘗試檢驗上述三種假說的可

1 當選無效訴訟過程中，高等法院作了全面的司法驗票，結果陳呂得票數由6,471,970票減爲6,461,177票，票數減少10,793票。連宋得票數也由6,442,452票減爲6,435,614票，票數減少6,838票，連宋減少的票數較陳呂減少的票數少，雙方差距也由29,518票縮小爲25,563票。（聯合報，2004年11月5日，A3版。）

2 該聯盟網址，<http://www.coolcloud.org.tw/user/nobnog/homepage.asp>。

3 使用2000年戶口普查的資料，是著眼於其龐大的樣本，讓我們可以得到各鄉鎮精確的人口資料。但是，由於2004年並沒有對應的資料，我們必須假設這2000年至2004年的四年期

能性，其中最重要的解釋變數是各鄉鎮 20 歲以上人口的教育程度，尤其是各鄉鎮教育程度為不識字或自修的比例。換言之，由於教育程度較低者可能在過去無效票範圍較小的時候，習慣圈選在候選人名、號次或相片上，中選會所提的「無效票範圍擴大」的假說，隱含的是 2004 年各選區的廢票比例的增加應該和低教育程度的比例成正相關。此外，「百萬廢票行動聯盟」所影響的「主動型廢票」，由於主要是透過媒體做理念上的宣傳，此一運動應該在都會地區、選民教育程度較高的選區成效較好。⁴ 換言之，教育程度較高的選民，應該有較高的可能性投出主動型的廢票，因此廢票比例應與教育程度成正相關。至於做票導致廢票激增的假說，雖然乍看之下似乎與教育程度無關，不過本文將論證，「做票說」的論證基礎——陳呂得票率高的地方，廢票比例也較高，此一確實存在的相關性，其實並不是因果關係，其相關性來自於陳呂得票率和廢票比例兩者都同時受到選民教育程度的影響。

本文除了應用經濟學晚近所發展、為了克服解釋變數內生性所造成估計偏誤問題的 differences-in-differences 估計法，直接估計無效票定義範圍擴大對廢票率的影響，探討「廢票哪裡來」此一實務上的爭議之外，也進一步以「廢票說」為例，說明實證研究中，常見由於混淆了相關性和因果關係所造成的問題。另外，我們也討論了社會學與政治學文獻中關於使用加總資料 (aggregate data) 來推論個體行為時，可能存在的區位謬誤 (ecological fallacy) 問題，並使用政治學文獻中由 King (1997: 77-157) 所發展的區位推論模型 (ecological inference model)，重新估計識字選民和不識字選民投出廢票的比例，並與線性迴歸的結果作一對照。

本文除前言之外，第 2 節由各縣市廢票比例之間的差異和兩次總統選舉廢票比例的變化，觀察廢票增加的區域差異，並初步檢驗各種假說和資料是否一致。第 3 節討論教育程度和廢票比例之間的關係，以及無效票範圍擴大對不同族群的影響，並將此一制度變化視為一個自然實驗 (natural experi-

間，人口特性並沒有明顯的變化。感謝評審之一指出這項隱含的假設 (implicit assumption)。

4 「廢票運動」號召到的廢票，應該比較可能是來自都會地區或教育程度較高的選民，是我們的推測。我們並沒有實證資料來驗證此一假說。感謝評審之一指出這個隱含的假設。

ment)，以 differences-in-differences 法估計此一制度變化對廢票比例的影響。迴歸估計之後，我們也使用區位推論模型，估計 2000 年和 2004 年識字選民和不識字選民投出廢票的比例，以避免使用加總資料可能產生的偏誤。第 4 節則進一步探討「做票說」的可能性，並說明廢票比例和陳呂得票率之間的相關性，其實並非因果關係，而是因為兩者同時受到選民教育程度的影響。第 5 節為結論。

2 各縣市廢票比例的變化

首先，我們以縣市為單位，觀察各縣市在 2000 年和 2004 年總統選舉的廢票比例，以及兩次選舉間廢票比例的變化。

表 1 顯示，2000 年全國平均廢票比例只有 0.96%，各縣市廢票比例並沒有明顯的差異。廢票比例最高的桃園縣（1.22%）和最低的連江縣（0.49%），差異只有 0.73%，各縣市廢票比例的標準差也只有 0.19%，各縣市選民投出廢票的機率並沒有明顯的差異。

反觀 2004 年，廢票比例就呈現相當程度的差異性。廢票比例最高的雲林縣，其廢票比例高達 3.99%，是全國平均 2.55% 的 1.56 倍，廢票比例最低的則是高雄市，只有 1.64%。廢票比例最高與最低差距達 2.35%，各縣市廢票比例的標準差也高達 0.85%，約為 2000 年標準差的 4.5 倍。

此外，就兩次總統選舉全國廢票比例的變化而言，2004 年增加了 1.59%，其中增加最多的是 2004 年廢票比例也最高的雲林縣，增加了 3.20%，其次為金門、宜蘭和嘉義縣，分別增加了 2.79%、2.58% 和 2.57%，其他諸如彰化縣、苗栗縣、連江縣、高雄縣和台南縣，廢票比例的增加都超過 2%。另一方面，廢票比例增加最少的是高雄市，僅增加 0.72%，其次為台中市的 0.90%，其他增加幅度較少的縣市包括台南市、嘉義市、台北市、新竹市和基隆市等。

換言之，表 1 顯示，2004 年廢票比例較低，或是廢票比例增加較少的縣市，都是直轄市或省轄市等都會型縣市，2004 年廢票比例較高或者廢票比例增加較多的縣市，反而是雲林、嘉義和彰化等都市化程度較低的縣分，此與「百萬廢票行動聯盟」的影響應在都市地區較顯著的假說並不一致。

表 1：2000 年和 2004 年廢票比例(%)¹

縣市	2000 年	2004 年	2000-2004 年 增加
雲林縣	0.79	3.99	3.20
金門縣	0.87	3.66	2.79
嘉義縣	0.94	3.51	2.57
彰化縣	1.06	3.46	2.40
宜蘭縣	0.77	3.35	2.58
苗栗縣	1.01	3.34	2.33
桃園縣	1.22	2.98	1.76
高雄縣	0.80	2.92	2.12
台南縣	0.85	2.88	2.03
南投縣	0.84	2.84	2.00
台東縣	1.16	2.67	1.51
連江縣	0.49	2.65	2.16
澎湖縣	0.86	2.64	1.78
新竹縣	1.04	2.55	1.51
屏東縣	0.77	2.53	1.76
花蓮縣	1.16	2.46	1.30
台中縣	1.11	2.44	1.34
台北縣	1.11	2.43	1.32
新竹市	1.05	2.33	1.28
基隆市	0.86	2.20	1.33
台北市	0.78	1.90	1.12
嘉義市	0.75	1.87	1.11
台南市	0.79	1.86	1.08
台中市	0.94	1.84	0.90
高雄市	0.92	1.64	0.72
平均 ²	0.96 (0.19)	2.55 (0.85)	1.59 (0.84)

¹ 資料來源：中央選舉委員會。依 2004 年廢票比例排列。² 括弧內為標準差，平均數均經投票人數加權計算。

另一方面，上面依廢票比例高低分類的縣市組成，民進黨主政與否似乎沒有構成系統性的差異。由於縣市選委會是選務機關最基層的層級，「做票說」最可能的管道在於縣市長擔任主任委員的縣市選舉委員會，⁵ 我們也進一步以 2004 年縣市長的黨籍為分類標準，比較民進黨主政與非民進黨主政的縣市，廢票比例變化的情形，表 2 呈現比較的結果。

表 2 顯示，不論是 2000 年或是 2004 年，確實如表 1 的觀察，民進黨主政與否並沒有構成廢票比例的明顯差異。民進黨主政縣市與非民進黨主政縣市，其廢票比例幾乎完全相同，均為全國的平均值。而兩次選舉間廢票比例增加的幅度，也幾乎沒有差別。

表 2：2000 年和 2004 年廢票比例(%)¹

縣市	2000 年	2004 年	增加
民進黨主政縣市 ²	0.96 (0.18)	2.55 (0.87)	1.59 (0.87)
非民進黨主政縣市	0.96 (0.21)	2.54 (0.83)	1.58 (0.81)
平均	0.96 (0.19)	2.55 (0.85)	1.59 (0.84)

¹ 資料來源：中央選舉委員會。括弧內標準差，平均數均經投票人數加權計算。

² 指 2004 年縣市長為民進黨籍者。

3 教育程度對廢票比例的影響

上一節，我們初步分析了各縣市 2004 年廢票比例以及兩次選舉間廢票比例變化的差異，發現廢票比例變化和民進黨主政縣市的關係不大，廢票比例增加的幅度和都市化程度成反向關係。這兩項特性分別與「做票說」及「廢

5 做票應和執政黨是否在選區執政有關，是依據我們對選務運作的瞭解所做的推測，也是在野黨在選後質疑的論點。不過，我們並沒有資料來驗證此一推測。感謝評審之一指出這項隱含的假設。

票運動說」不一致。

這一節，我們探討無效票範圍擴大對廢票比例的影響。有效票認定標準的變化之所以會影響廢票比例，是因為在兩次選舉之間「總統副總統選舉罷免法」對無效票的定義有所修訂。2000年總統選舉適用的是1995年8月9日公佈施行的「總統副總統選舉罷免法」，其中第53條規定：

選舉票有左列情事之一者無效：

- 一、不用選舉委員會製發之選舉票者。
- 二、圈二組以上者。
- 三、所圈地位不能辨別為何組者。
- 四、圈後加以塗改者。
- 五、簽名、蓋章、按指印、加入任何文字或劃寫符號者。
- 六、將選舉票撕破致不完整者。
- 七、將選舉票污染致不能辨別所圈選為何組者。
- 八、不加圈完全空白者。
- 九、不用選舉委員會製備之圈選工具者。

前項無效票，應由開票所主任管理員會同主任監察員認定；認定有爭議時，由全體監察員表決之。表決結果正反意見同數者，該選舉票應為有效。

在這個無效票的定義之下，原則上只要選票足以辨識選舉人的投票意向，即視為有效票。換言之，圈選在候選人號次或相片的選票均屬有效。

此一定義在2004年的總統選舉產生變化，因為立法院在2003年10月8日，第五屆第四會期第四次會議修正通過，並在2003年10月9日公布施行的「總統副總統選舉罷免法」，在第60條規定無效票的範圍，全文如下：

選舉票有下列情事之一者，無效：

- 一、不用選舉委員會製發之選舉票者。
- 二、未依前條第一項規定圈選一組者。

三、所圈位置全部於圈選欄外或不能辨別為何組者。

四、圈後加以塗改者。

五、簽名、蓋章、按指印、加入任何文字或符號者。

六、將選舉票撕破致不完整者。

七、將選舉票污染致不能辨別所圈選為何組者。

八、不用選舉委員會製備之圈選工具者。

前項無效票，應由開票所主任管理員會同主任監察員認定；認定有爭議時，由全體監察員表決之。表決結果正反意見同數者，該選舉票應為有效。

值得注意的是，這項修正案是行政院提出的，行政院在草案說明中並解釋：「由於選舉人均已普遍習慣於選舉票圈選欄內圈選候選人，為免爭議，防止弊端，爰修正第一項第三款。」⁶

比較修法前後的無效票認定標準，最主要的差別在於修正後的第三項加上「所圈位置全部於圈選欄外」為無效票的規定，使得圈選在候選人號次或相片的選票都變成無效票，此一改變造成 2000 年和 2004 年兩次總統選舉在無效票的定義上有明顯的不同。這也讓中選會在經歷 2004 年總統選舉的嚴重爭議之後，在 2004 年年底立法委員選舉前的 12 月 2 日，還特別發佈新聞稿表示：「總統副總統選舉罷免法與公職人員選舉罷免法針對有效票與無效票認定之標準不同，立法委員選舉的法令較寬，此次如果蓋在候選人號次或相片等不正確位置雖然仍為有效票，但還是希望選民能在圈選欄內圈選，才是正確的投票方式。」⁷

除了總統副總統之外，其他諸如立法委員、縣市議員、縣市長和鄉鎮市長等公職人員，均受「公職人員選舉罷免法」之規範，其中關於無效票的定義規定在該法的第 62 條，內容與修正前之「總統副總統選舉罷免法」之無效

6 立法院公報，第 92 卷，第 4 期，院會記錄。

7 2004 年 12 月 2 日，中央選舉委員會新聞稿。

票定義完全相同。⁸ 換言之，我國過去舉辦的所有選舉，除了2004年總統選舉在「總統副總統選舉罷免法」中規定範圍較大的無效票定義之外，圈選在候選人號次或相片等位置均屬有效票。此一僅只在2004年總統選舉發生的制度變化，非常可能造成長期以來已經習慣圈選在候選人號次或相片的選民，在中選會與各政黨宣導不夠的情況之下，依舊圈選在號次或照片上，而成爲無效票範圍擴大之下的廢票。

附圖1是一張選票的範例，⁹ 由此我們可以清楚看出，在「總統副總統選舉罷免法」修正、無效票定義範圍擴大之前，選民只要圈選在選票的左半欄或者右半欄，不論是號次、「總統候選人」與「副總統候選人」字樣、正副總統候選人照片、名字或推薦政黨，均屬有效票。但在無效票定義範圍擴大之後，只有完全圈選在號次上方圈選欄內的選票才是有效票。以選票面積來看，新制之下的有效票面積，大約只有舊制下有效票面積的八分之一，變動的幅度不可謂不大。

那麼，如果我們把這項無效票定義的變化視爲一個自然實驗 (natural experiment)，哪些選民是最可能受到影響的族群？我們認爲，教育程度是一個重要的決定因素。教育程度偏低者應該是最可能在長期的選舉過程中，爲了投票方便起見而習慣圈選在候選人號次或相片的族群。¹⁰ 因此，我們進一步以教育程度在不識字與自修（即國小未畢業）者，占2000年戶口普查時各鄉鎮20歲以上人口的比例，來衡量這個族群占有選民的比例，並觀察它和廢票比例之間的關係，圖1至圖3分別畫出「2004年廢票比例」、「2000年廢

8 全文爲：「選舉票有左列情事之一者無效：一、不用選舉委員會製發之選舉票者。二、圈二人以上者。三、所圈地位不能辨別爲何人者。四、圈後加以塗改者。五、簽名、蓋章、按指印、加入任何文字或劃寫符號者。六、將選舉票撕破致不完整者。七、將選舉票污染致不能辨別所圈選爲何人者。八、不加圈完全空白者。九、不用選舉委員會製備之圈選工具者。

前項無效票，應由開票所主任管理員會同主任監察員認定；認定有爭議時，由全體監察員表決之。表決結果正反意見同數者，該選舉票應爲有效。」

9 感謝東吳大學政治系黃秀端教授提供選票範例的照片。

10 教育程度較低者較可能養成圈選在候選人號次或相片，是我們的推測，並沒有實證資料可以驗證此一隱含的假設。感謝評審之一指出這項隱含的假設。

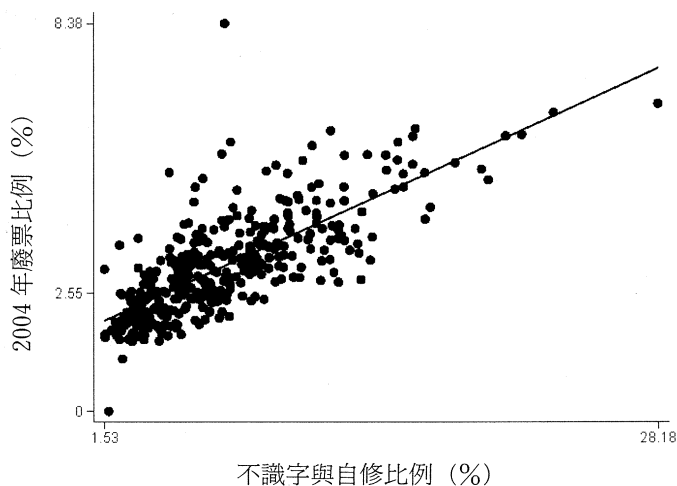


圖 1：不識字比例與 2004 年廢票比例

票比例」以及「2000-2004 廢票比例增加」和各鄉鎮不識字或自修比例（以下簡稱為不識字比例）之間的關係。

圖 1 清楚顯示，2004 年總統選舉中，各鄉鎮的廢票比例和該鄉鎮公民的不識字比例成正相關，相關係數高達 0.8269。不識字比例最高（28.18%）的彰化縣大城鄉，其廢票比例高達 6.65%，僅次於蘭嶼的 8.38%。相反地，全國教育程度最高的台北市大安區，不識字比例僅有 1.55%，其廢票比例也只有 1.66%。

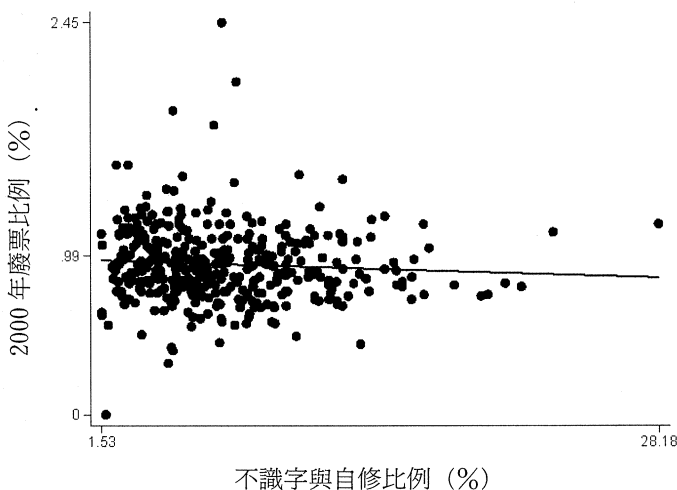


圖 2：不識字比例與 2000 年廢票比例

另一方面，圖 2 顯示，2000 年的廢票比例，除了整體而言較 2004 年為低，各鄉鎮的廢票比例和不識字比例的關係也不明顯，相關係數為 -0.0495 。

當然，各鄉鎮廢票比例的差異，可能來自於各鄉鎮其他獨有的特性，因此我們以 2000 年至 2004 年廢票比例的變化，即各鄉鎮 2004 年廢票比例減 2000 年廢票比例，觀察不識字比例對廢票比例變化的影響。圖 3 顯示，各鄉鎮廢票比例的增加和不識字比例之間，呈現相當清楚的正相關，相關係數更高達 0.8503 。

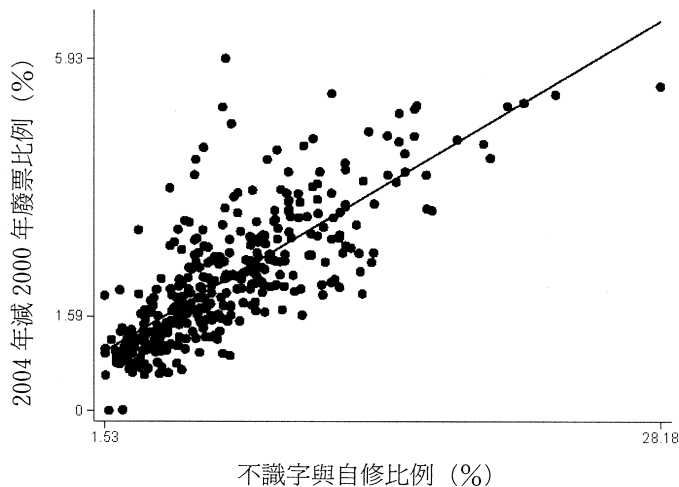


圖 3：不識字比例與廢票比例增加幅度

總的來說，圖 1 至圖 3 顯示，各鄉鎮容易受到無效票定義變化所影響的不識字比例，和 2004 年廢票比例的增加，有相當明顯的關係。

10 餘年來，經濟學的實證研究爲了克服解釋變數與誤差項 (error term) 相關的內生性問題，將研究方法轉向尋找由各種外生的制度變化所產生的自然實驗，試圖由比較被解釋變數在制度變化前後的變動，來獲得具一致性 (consistent) 的估計式 (Meyer, 1995: 151-161)。舉例而言，文獻上關於未來醫療需求的不確定性，對預防性儲蓄性 (precautionary savings) 影響的研究，由於醫療保險的購買和個人儲蓄行爲相關，透過直接比較擁有醫療保險者和沒有醫療保險者的儲蓄差異，來估計醫療保險對儲蓄的影響，其估計結果會產生偏誤。由迴歸分析的角度來看，問題在於作爲解釋變數的醫療保險是內生的。Chou et al. (2003: 1873-1894) 利用我國在 1995 年實施全民健

康保險的制度變化，研究健康保險對儲蓄的影響。全民健保實施前，健康保險的取得有職業屬性上的差異，只有公務員的眷屬擁有健康保險，其他非公務員但參加勞工保險者，其眷屬就沒有被健康保險所涵蓋。全民健保的實施，使得所有眷屬都擁有健康保險。因此，全民健保實施的制度變化，對不同職業屬性的家庭有不同的影響。非公務員家庭所面對的醫療風險，在全民健保實施之後大幅降低了，但對公務員家庭而言，由於原本就有健康保險，其醫療風險並沒有變化。

另外，值得說明的是，上述制度變化對不同群體的界定，並不需要是受到影響與否的「全有」或「全無」的差別，也可以是受到影響「程度」的差別。舉例而言，Duflo (2001: 795-813) 使用 1973-79 年間，印尼政府大量興建小學的過程中，各個地區興建小學密度的差異，對應於個人由於出生地不同所受到小學擴張政策「不同程度」的影響，觀察興建小學對個人教育程度和成人之後薪資的影響，並估計教育的報酬率。由於大量興建小學的政策是全國適用，但各地興建小學的密度並不相同，Duflo (2001: 795-813) 對受影響群體的界定，就是受到政策影響的「程度」，而不是「是否」受到政策的影響。

概念上來看，上述方法是透過區分受到新制度影響的實驗組，和不受新制度影響的對照組，或是區分受到影響較大的實驗組，和受到影響較小的對照組，來估計實驗組與對照組兩者在新舊制度下，被解釋變數變動的差異。因此，這個估計式稱為「差異中的差異」(differences-in-differences) 估計式，可以表示為，

$$\begin{aligned}\hat{\beta} &= \left(\bar{Y}_{after}^{treatment} - \bar{Y}_{before}^{treatment} \right) - \left(\bar{Y}_{after}^{control} - \bar{Y}_{before}^{control} \right) \quad (1) \\ &= \Delta \bar{Y}^{treatment} - \Delta \bar{Y}^{control}\end{aligned}$$

其中 Y 為被解釋變數， \bar{Y} 為 Y 的平均值，上標的 *treatment*、*control* 分別代表受到制度變化影響的實驗組和未受到影響的對照組，下標的 *before*、*after* 分別代表制度變化之前與制度變化之後， $\Delta \bar{Y}$ 則表示制度變化前後 Y

平均值的變動。換言之，differences-in-differences 估計式所計算的是，實驗組的被解釋變數在制度變化前後平均值的變動，減去對照組的被解釋變數在制度變化前後平均值的變動。

雖然不識字比例作為廢票比例的解釋變數，並沒有明顯的內生性問題，我們仍然可以使用 differences-in-differences 估計法的架構，將各鄉鎮依不識字比例的高低分為實驗組和對照組，以 2000 年和 2004 年廢票的比例為目標，觀察 2004 年無效票範圍擴大的制度變化，對廢票比例的影響。¹¹

首先，我們以各鄉鎮不識字比例的中位數 6.615% 為分界，將各鄉鎮依不識字比例高低分為兩組，以高於 6.615% 者為實驗組，低於 6.615% 者為對照組，每組各有 184 個鄉鎮，表 3 為各組在 2000 年和 2004 年的廢票比例，以及不同面向上的差異 (differences)。

表 3：不識字比例對廢票比例的影響¹

鄉鎮	廢票比例(%)		
	2000 年	2004 年	Differences
高不識字比例 (≥6.615%)	0.925 (0.204)	3.575 (0.912)	2.650 (0.069)
低不識字比例 (<6.615%)	0.967 (0.187)	2.225 (0.501)	1.258 (0.040)
Differences	-0.042 (0.023)	1.350 (0.076)	1.392 (0.080)

¹ 資料來源：中央選舉委員會。各組廢票比例均經各鄉鎮投票數加權，括弧中為標準差。

表 3 顯示，就不識字比例較高的 184 個鄉鎮而言，2000 年的廢票比例為 0.925%，而 2004 年無效票範圍擴大之後，廢票比例便增為 3.575%，增加了 2.650%，且其標準差只有 0.069%，兩次選舉的差異相當顯著。

11 這裡，實驗組與對照組的定義和式(1)的定義略有不同。式(1)的定義中，實驗組與對照組的差別是受到制度變化的影響與否，此處實驗組與對照組的差別則是受到制度變化影響程度上的差異。換言之，作為實驗組的不識字比例較高之鄉鎮，受到無效票範圍擴大的影響較大，而作為對照組的不識字比例較低之鄉鎮，受到無效票範圍擴大的影響較小。

另一方面，就不識字比例較低的半數鄉鎮而言，2000 年的廢票比例為 0.967%，2004 年的廢票比例則增加至 2.225%，廢票比例也增加了 1.258%，其標準差為 0.040%，兩次選舉的差異也相當顯著。

再者，就實驗組與對照組的比較而言，前者在無效票範圍擴大之後，廢票比例增加了 2.650%，後者增加了 1.258%，兩者之差距即為 differences-in-differences 方法之下的估計值，其值為 1.392%，標準差則為 0.080%，差異相當顯著。換言之，相較於不識字比例較低的半數鄉鎮，不識字比例較高的鄉鎮，在無效票範圍擴大之後，廢票比例多增加了 1.392%。無效票範圍的擴大，以及教育程度較低選民的投票習慣，造就了廢票比例的提高。

表 3 使用各組的平均廢票比例的計算，可以得到在各鄉鎮依不識字比例高低分組的情況下，以 differences-in-differences 方法估計無效票範圍擴大對廢票比例的影響，但是無法直接估計不識字比例對廢票比例的邊際影響。因此，我們將上面 differences-in-differences 估計方法表示成迴歸式，並進一步控制其他可能影響廢票比例的人口特性，例如 65 歲以上老年人口比例，來估計無效票範圍擴大之後，不識字比例對廢票比例的邊際影響。我們將迴歸式表示為，

$$\begin{aligned} invalid = & \beta_0 + \beta_1 illiterate + \beta_2 year2004 \\ & + \beta_3 illiterate * year2004 + \gamma X + u \end{aligned}$$

其中 *invalid* 表示各鄉鎮廢票比例，*illiterate* 表示不識字比例，*year2004* 是 2004 年的虛擬變數，*X* 為所有其他控制變數，*u* 為誤差項。不識字比例和 2004 年虛擬變數交乘項的係數 $-\beta_3$ 即為無效票範圍擴大對廢票比例影響的 differences-in-differences 估計式，表 4 為迴歸分析的結果。

表 4 的模型(1)中，在放入其他控制變數之前，2004 年的虛擬變數的係數顯示，2004 年的廢票比例顯著地增加了 0.4678%，不識字比例本身的效果則不顯著。最重要地，不識字比例和 2004 年的交乘項顯示，無效票範圍擴大之後，不識字比例每增加 1%，廢票比例增加 0.2133%。換言之，以不識字比例 1.55% 的台北市大安區和不識字比例為 28.18% 的彰化縣大城鄉為例，兩者在

表 4：無效票範圍擴大對廢票比例的影響¹

	(1)	(2)	(3)
不識字比例(%)*2004年	0.2133 (26.81)**	0.2132 (27.10)**	0.2132 (26.54)**
不識字比例(%)	-0.0028 (0.77)	-0.0213 (3.76)**	-0.0476 (6.94)**
2004年	0.4678 (7.72)**	0.4687 (7.80)**	0.4684 (8.03)**
男性比例(%)		0.0369 (6.12)**	0.0252 (4.68)**
65歲以上老人比例(%)		0.0065 (1.21)	0.0144 (2.35)*
就業人口中農業部門比例(%) ²			0.0079 (4.95)**
就業人口中工業部門比例(%)			0.0085 (6.17)**
常數項	0.9717 (30.92)**	-0.8818 (2.92)**	-0.5511 (1.98)*
樣本數	736	736	736
R ²	0.87	0.88	0.89

¹ 括弧內為 t 值，**表示在 1% 的水準之下顯著，*表示在 5% 的水準之下顯著，各迴歸式均以各鄉鎮投票數加權。

² 對照組為就業人口在服務業的比例。

2000 年至 2004 年間的廢票比例增加的差距將為 5.68%。¹²

我們在模型(2)再加入「男性比例」和「65 歲以上老人比例」，其中男性比例顯著地增加廢票的比例，65 歲以上老人的比例則不顯著。理論上，老人也應屬於在無效票範圍擴大之後，較可能投成廢票的族群。不過，高齡和不識字比例應該高度相關，使得老人比例的解釋力不顯著。此外，加入男性比例和老人比例之後，也讓「不識字比例」本身的影響轉為顯著地負向。不過，

¹² $(28.18 - 1.55) * 0.2133 = 5.68$ 。

不識字比例和 2004 年的交乘項，其估計值則沒有改變，仍為顯著的 0.2132。

最後，模型(3)加入各鄉鎮就業人口屬於農業部門和工業部門的比例。相對於服務業，農業人口比例和工業人口比例越高的鄉鎮，廢票比例也越高。此外，加入就業人口比例之後，男性比例的影響略微下降，不識字比例和 2004 年的交乘項則不受影響，仍為 0.2132。

綜合而言，模型(1)-(3)的結果都顯示，無效票範圍的擴大，使得不識字比例每增加 1%，2004 年的廢票比例會多增加 0.21%。

以上的分析中，我們以 differences-in-differences 估計法，分析教育程度（不識字比例）對廢票比例的影響，由於個體投票資料的不可得，我們只能使用加總的資料 (aggregate data)。不過，使用加總資料來推論個人行為，可能存在區位謬誤 (ecological fallacy) 的問題，我們有必要討論此一問題對本文的影響。¹³

自從 Robinson (1950: 351-357) 指出使用加總資料分析個體行為，區位相關 (ecological correlation) 並不一定和個體行為中的相關性一致，甚至可能出現方向完全相反的推論之後，社會科學研究就已經知道此一問題的存在，許多人甚至因此盡量避免使用加總資料來作研究。不過，誠如黃紀 (2001: 541-574) 所言：「該文卻引起了另一種誤解，將 Robinson 的提醒與警告解讀為禁令，認定以集體數據跨層次推估個體層次之事實，若非緣木求魚，徒勞無功，就是煉金詐術，欺世盜名。」由於許多研究問題並不存在個體資料，使用加總資料來推論個體行為，並且盡可能避免偏誤，也是不得不然的研究策略。

Robinson (1950: 351-357) 所舉最明顯的例子是，外來移民和不識字比例之間的相關性。在個體的層次上，外來移民如果來自母語非為英語的國家，其不識字的可能性應該較高，因此「外來移民」與「不識字」之間應該成正相關。不過，使用美國 9 個區域加總資料的分析中，由於外來移民的落腳地，是他們自己選擇工作機會較多、或者都市化程度較高地區的結果，外來移民集中在不識字比例較低的地區，使得外來移民比例較高的地區，其不識字比

13 感謝二位評審對區位謬誤問題的提醒。

例卻較低，造成「外來移民比例」和「不識字比例」成負相關，此一相當偏誤的結果。換言之，當分析對象如外來移民所在的區域是自我選擇（self selection）的內生（endogenous）結果時，直接計算外來移民比例和不識字比例的相關性，將得到相當偏誤的結果。

區位推論（ecological inference）或者區位謬誤（ecological fallacy）的問題，在政治學界和社會學界的文獻上是一重要議題，但在經濟學界卻較少被提及，這並非因為經濟學不使用加總資料來分析問題，而是經濟學界將此一問題放在「解釋變數具內生性問題」的架構來處理。換言之，經濟學的個體計量（microeconometrics）方法，相當強調解釋變數與誤差項如果相關時，迴歸估計的結果將是偏誤的。因此，我們必須使用追蹤調查資料（panel data）、工具變數估計法（instrumental variable estimation），甚至尋找政策外生變動所造成的自然實驗（natural experiment），來處理內生性的問題以獲取具一致性（consistent）的估計係數。本文使用的 differences-in-differences 估計法，即為此類方法之一。

和許多選舉研究所面臨的問題相同，我們無法取得同時包含投出廢票和個人教育程度等資料的個人資料，因此我們必須連結中選會的投票資料和鄉鎮市區人口的不識字比例，來分析無效票定義變化對是否投出廢票的影響。本文的估計結果是否因為資料加總的偏誤，而導致估計結果不可信，自然是必須仔細考慮的問題。

首先，台灣人口的分佈，這幾年並未出現快速且大量的國內移動，因此各鄉鎮的不識字人口和投廢票的傾向，應該不存在和 Robinson（1950: 351-357）舉例的外來移民和識字比例之間，那麼明顯的相關性。

其次，假使各鄉鎮人口存在和無效票定義無關，但各鄉鎮明顯不同的投廢票傾向，只要這部分的傾向在 2000 年和 2004 年之間，並沒有明顯的變化，那麼透過 differences-in-differences 的估計，應該可以確保解釋變數和誤差項的不相關，而得到具一致性的估計。換言之，本文的分析存在區位謬誤的問題應該不大。

當然，以上的說明只是對區位謬誤問題的間接討論。因此，我們進一步以政治學文獻中，由 King（1997: 77-157）所發展、結合 Goodman（1953:

663-664; 1959: 610-625) 的區位迴歸模型 (ecological regression model) 和 Duncan and Davis (1953: 665-666) 的上下限推算法 (method of bounds), 容許各選區 β 係數不同的區位推論模型 (ecological inference model, EI), 估計各鄉鎮在 2000 年和 2004 年不識字者和識字者投廢票的比例。¹⁴ 國內文獻中, 已有使用 EI 模型進行分析的研究。黃紀與張益超 (2001: 183-218) 首次使用了 EI 模型, 分析 1997 年嘉義市市長與立委選舉之一致與分裂投票; 徐永明 (2001: 93-118) 完全以 EI 模型來測量與探討宋楚瑜作為一個獨立候選人, 其選票的社會基礎及政治支持的性質; 吳怡銘 (2000: 1-119) 則探討 1998 年台北市市長以及市議員選舉中, 選民的分裂投票行為。另外, 施奕任 (2001: 1-142) 也應用了區位推論模型, 分析 1998 年高雄巿市長和市議員選舉中, 選民分裂投票的行為。

EI 模型容許各選區選民投出廢票比例的異質性, 我們可以分別估得各鄉鎮不識字與識字選民投出廢票的比例, 並以各鄉鎮投票人數為權數, 計算全國的平均值與標準差。表 5 列出 EI 模型估計結果的全國平均值與標準差, $\beta^{illiterate}$ 和 $\beta^{literate}$ 分別代表不識字選民和識字選民投廢票的比例。

表 5 顯示, 2000 年總統選舉中, 不識字選民投出廢票的比例是 5.63%, 識字選民投出廢票的比例則只有 0.70%, 兩者相差 4.93%。無效票定義範圍擴大之後的 2004 年總統選舉, 識字選民投出廢票的比例增加為 1.62%, 只較 2000 年增加了 0.92 個百分點。反觀不識字選民, 在無效票定義範圍擴大之

表 5：區位推論模型(EI)估計結果¹

	2000 年	2004 年
$\beta^{illiterate}$	0.0563 (0.0056)**	0.1801 (0.0271)**
$\beta^{literate}$	0.0070 (0.0024)**	0.0162 (0.0042)**

¹ 括弧內為標準差, **表示在 1% 的水準之下顯著。

14 估計過程使用的統計軟體是由 Gary King 的網頁所下載的 EzI 軟體, 網址為 <http://gking.harvard.edu/>。

後，投出廢票的比例增加至 18.01%，不但比 2000 年投出廢票的比例增加了 12.38%，更比 2004 年識字選民投出廢票的比例高了 16.39%。

綜合而言，雖然由於個體資料的缺乏，本文使用加總資料進行分析。一方面，本文使用 differences-in-differences 估計法，估計政策變動所帶來的外生變化，對低教育程度者投出廢票的影響，發現無效票定義範圍擴大確實造成不識字者投出廢票比例的增加。另一方面，使用區位推論模型的估計，雖然不能完全排除區位謬誤存在的可能性，但是估計結果發現，不識字選民投出廢票比例在 2004 年大幅增加，且明顯較識字選民投出廢票的比例高出許多，都與無效票定義範圍擴大導致廢票比例增加的假說一致。因此，使用加總資料的估計係數，容或仍然與原始個體行為的係數稍有偏誤，但其方向應該是正確的。

4 做票說

由前兩節的討論，我們一方面發現廢票比例和教育程度成反向關係，與「廢票運動」的假說並不一致。另一方面，比較不識字比例高低鄉鎮在 2000 年和 2004 年間廢票比例增加的情形則顯示，不識字比例越高者，廢票增加越多，此與「無效票範圍擴大」的假說一致。

另一個最常被用來解釋 2004 年廢票大幅增加的假說是「做票說」，其中又以 2004 年 5 月 12 日在國民黨中常會提出饅頭理論的馬自恆先生的看法最具代表性。根據 5 月 13 日自由時報的報導：

……馬自恆提出「饅頭廢票曲線」，根據此次大選中各投開票所的藍綠得票比率及廢票數來統計，發現「連宋得票愈高的投開票所廢票率較低，但陳呂得票數愈高的投開票所，廢票率愈高」，他認為，這明顯受到外力介入的影響，「就是做票」，廢票決定了這次總統大選的結果。

馬自恆表示，在上次總統大選中，各投開票所的廢票率呈現「隨機」曲線，即勝選一方累計的勝出票數愈多，廢票數也隨之自然累

計增加，呈現同步上升的趨勢，而且也沒有城鄉差距的問題。

他指出，但此次分析發現，地點愈偏遠、陳呂得票愈高的投開票所，廢票卻愈高，且在支持陳呂的前一百個投開票所中，廢票率在四％到五％之間，但連宋支持度高的投開票所，廢票率都不超過二點二％，廢票率在一％到二％之間的投開票所，連宋支持度平均為五十三％，很明顯看出廢票率和支持度的關聯性。

馬自恆指出，在連宋領先與陳呂票數達二十八萬票時，廢票率為二點六％，接下來廢票比率逐漸升高，陳呂的支持度也急速上升，在第一三三〇〇投開票所時，連宋得票被陳呂超越，該投開票所的廢票率高達五點九八％，最後陳呂總計贏連宋兩萬九千多票。

.....

上述報導中關於陳呂得票率和廢票比例相關的事實部分，馬自恆先生的看法基本上並沒有錯誤。由這一份報導的內容看來，馬自恆先生處理資料的方法，應該是將投開票所依廢票比例由低至高排列放在橫軸，然後將連宋領先陳呂的累積票數放在縱軸。爲了驗證這項推測，我們使用由中選會網頁下載，以投票所爲單位的資料，¹⁵ 依上述方法將「饅頭廢票曲線」畫在圖 4 中。

由於廢票比例和陳呂得票率成正相關，和連宋得票率成負相關，圖 4 顯示，廢票比例低的時候，連宋的累積領先票數自然持續增加，這是饅頭曲線的上昇階段，連宋累積領先票數最多爲 299,319 票。當連宋累積領先票數的計算往廢票比例較高的投票所移動時，選票分配轉爲陳呂得票數較多，連宋累積領先票數自然逐漸減少。最後，連宋累積領先票數轉而爲負，到了最後一個廢票比例最高的投票所時，連宋累積領先票數等於負的「陳呂最終領先票數」-29,518 票。這個先增後減的曲線，即爲報導中所稱的「饅頭廢票曲線」。

15 值得注意的是，中選會由於需要計算以里爲單位的選舉資料，而一個里的里民又可能被分配在幾個不同的投票所，中選會網頁所提供的 Excel 檔案中，對和每一個里相關的所有投票所都列出一筆資料，因此總共 13,786 筆的資料中，會有投票所資料重複的問題。刪除重複的部分之後，我們共有 13,749 個投票所的資料。

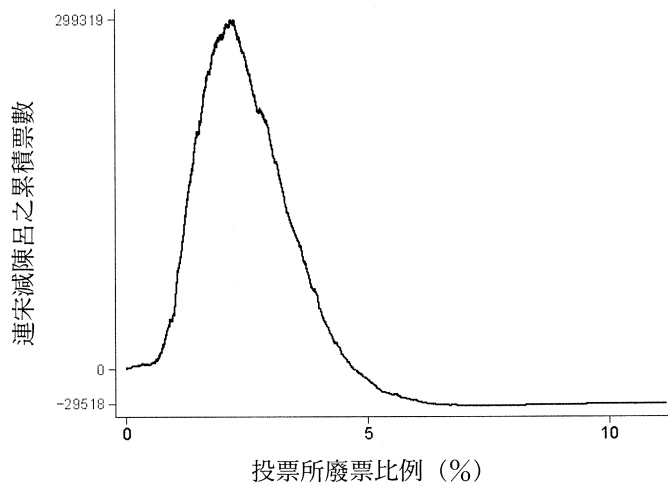


圖 4：投票所廢票比例與兩組候選人累積得票差距

報導中所引述，馬自恆先生發現 2004 年出現連宋累積票數「先增後減」的饅頭曲線，以及 2000 年「……，即勝選一方累計的勝出票數愈多，廢票數也隨之自然累計增加，呈現同步上升的趨勢，……」的「隨機」曲線，並沒有問題。不過，它們充其量只是在描述 2004 年時廢票比例和候選人得票率高度相關，而在 2000 年選舉時廢票比例和候選人得票率之間沒有相關性，但是無法推論到這是做票的結果。

換言之，就馬自恆先生所指出兩次總統選舉中，廢票比例和候選人得票率之間相關性的變化而言，「饅頭理論」並沒有錯誤。問題在於，如何解釋此一相關性的變化？為什麼 2000 年選舉時，候選人得票率和廢票比例之間的關係不明顯，而到了 2004 年陳呂得票率就和廢票比例成正相關。「做票說」當然是可能性之一，但是需要更細緻的論證。眾所周知，兩個變數的相關並不表示兩者就存在著因果關係，關於因果關係的假說，需要更嚴謹的論述與資料佐證。在沒有具體提出做票機制如何運作，以及選務機關如何在選舉過程中，在雙方候選人所指派監察員的監察之下，透過選務人員做票的說明之前，就認為「這明顯受到外力介入的影響，就是做票，廢票決定了這次總統大選的結果。」，顯然是過於武斷的。

事實上，由陳呂得票率到廢票比例的「做票」推論，正存在著典型的遺

漏變數 (omitted variables) 的問題。迴歸分析中，如果存在遺漏變數，而遺漏變數又和解釋變數成正相關，將使得解釋變數和誤差項也成正相關，造成解釋變數的係數有向上偏誤 (即往正的方向偏誤) 的問題。換言之，以廢票比例為被解釋變數，陳呂得票率為解釋變數的迴歸分析中，如果我們遺漏了一個真正會影響廢票比例的變數，而此一遺漏變數又與陳呂得票率正相關，則估得的係數將往正的方向偏誤。在此情況之下，即使陳呂得票率事實上對廢票比例沒有影響，我們仍然可能得到顯著但偏誤的正係數。

另一個看待「做票說」的角度，是文獻上關於「相關性」和「因果關係」的所謂「第三變數」(third variable) 問題的討論，這也是經濟學實證研究常常需要處理的問題。舉例而言，醫療經濟學 (health economics) 至今尚未解決的一個重要議題是，教育程度與健康的正相關，到底是教育真的可以促進健康，還是存在著第三個變數，同時會影響個人的教育程度和健康。主張教育程度和健康是「因果關係」的論點認為，教育程度的提高，一方面使消費者懂得如何利用醫療資源來生產健康，另一方面也增加消費者的醫療資訊，使其瞭解各種生產要素對健康的真正影響，因此教育可以改善健康情況。主張教育程度和健康只存在「相關性」、而非「因果關係」的論點則認為，教育程度和健康可能同時受到第三個變數，例如個人時間偏好 (time preference) 的影響。對現在時間偏好較低者，較著眼於未來長遠的報酬，一方面願意現在花時間和資源投資在教育上，以換取未來的利益，另一方面也較願意花費現在的時間與金錢，來換取未來的健康。因此，教育程度和健康的正相關，不一定隱含兩者存在著因果關係 (盧瑞芬與謝啓瑞，2000: 108-112)。

2004 年總統選舉中，此一同時影響廢票比例和陳呂得票率的第三變數，正是選民的教育程度。文獻上關於政黨支持基礎的研究都發現，民進黨支持者比較偏向本省籍 (王甫昌，1994: 1-34；吳乃德，1994: 101-130)。另一方面，教育成就之省籍差異的研究則發現，本省籍選民的教育程度相對於外省籍是偏低的 (Tsai and Chiu, 1993: 188-202；林忠正與林鶴玲，1993: 101-160；薛承泰，1996: 49-84；吳乃德，1997: 137-167；駱明慶，2001: 93-118)。我國歷史發展的過程中，教育程度、省籍和選民的政治支持三者，具有複雜的交互作用，如果無法掌握這個機制性關係，那麼對於實證現象就容易會有

倒果為因的社會科學論述產生。¹⁶

雖然文獻上關於教育成就之省籍差異的研究，主要的焦點多在整體教育成就的分佈，或是不同升學階段例如升大學或高中職分流的差異，並未直接分析低教育程度分佈的省籍差異，我們仍然可以很容易地由1990年「戶口普查」的資料，看出低教育程度的省籍差異。必須說明的是，1992年「戶籍法」修正，取消戶籍資料登錄本籍的規定之後，政府調查與統計資料也不再蒐集省籍的資料，因此1990年是存在省籍變數的最後一次「戶口普查」。不過，我們還是可以由出生世代的角度，觀察低教育程度者的省籍差異。表6以10個出生年分組，觀察1920-69年出生者，省籍間教育程度為不識字或自修比例的差異。

首先，表6顯示，不識字比例隨著世代的演進快速減少，尤其是1950年以後出生的世代，1950-59年出生的世代，本省籍和外省籍分別只有2.3%和2.0%的人國小沒有畢業。到了1960-69年出生的世代，兩者的不識字比例則都下降到只有0.4%。其次，對於1949年以前出生者，省籍間不識字比例的差異則相當明顯。對1920-29年出生者，本省籍不識字比例高達47.0%，外省籍只有17.0%，兩者相差30%。對1930-39年出生者而言，本省籍不識字比例大幅下降，但是仍有27.6%，外省籍也略降至15.8%，差異雖然縮小，但仍然相差了11.8%。最後，到了1940-49年出生、戰後進入小學學齡的世

表6：不識字比例的省籍差異(%)^{1,2}

出生年	本省籍	外省籍
1920-29	47.0	17.0
1930-39	27.6	15.8
1940-49	11.1	9.6
1950-59	2.3	2.0
1960-69	0.4	0.4

¹ 不識字包含調查中教育程度為「不識字」和「自修」者。

² 資料來源：1990年「戶口普查」。

16 感謝評審之一對這個論點的提示。

代，本省籍與外省籍不識字比例分別降至 11.1% 和 9.6%，差異快速縮小。

換言之，由於戰後國民教育的普及，1940 年以後出生者，整體不識字比例快速減少，省籍間不識字比例的差異也逐漸消失。不過，對於 1940 年以前出生者而言，本省籍的不識字比例，明顯高於外省籍的不識字比例。因此，民進黨的支持者偏向本省籍，和本省籍人口不識字比例偏高，這兩項事實構成了陳呂得票率和不識字比例的相關性，圖 5 畫出陳呂得票率和不識字比例之間的關係。圖 5 清楚顯示，陳呂得票率與不識字比例成正相關，相關係數為 0.4275。所以，圖 4 的「饅頭廢票曲線」所呈現，陳呂得票率和廢票比例成正相關的性質，其實是由圖 1 中廢票比例和不識字比例正相關，和圖 5 所呈現陳呂得票率也和不識字比例正相關所共同構成。

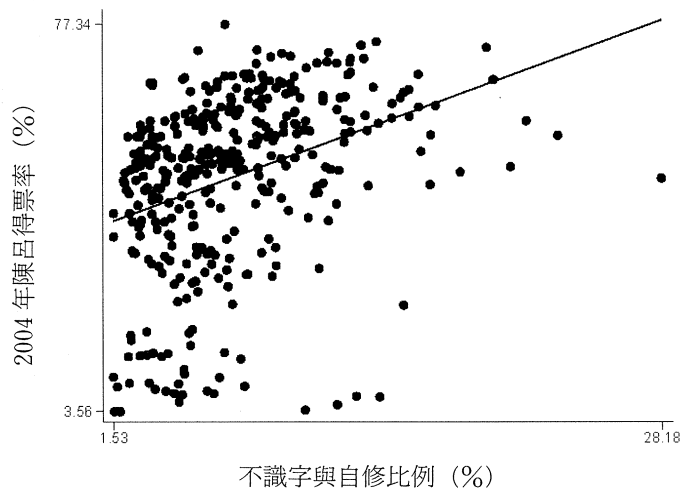


圖 5：不識字比例與 2004 年陳呂得票率

接下來，我們以迴歸分析探討陳呂得票率和廢票比例之間的關係，表 7 是迴歸結果。模型(1)-(3)的被解釋變數是 2004 年的廢票比例，模型(4)-(6)的被解釋變數則是 2000 年至 2004 年廢票比例的增加。

首先，模型(1)顯示，如同所謂「饅頭理論」所表現的性質，陳呂得票率 and 廢票比例之間，確實存在著正的相關性，陳呂得票率的係數為顯著的 0.0197，陳呂得票越多的鄉鎮，廢票比例也越高。但是如同前面所討論的，相關性並不直接表示因果關係，欲從陳呂得票率高推論到「選務人員配合民

表 7：不識字比例對廢票比例的影響¹

	2004 年廢票比例 (%)			2000-2004 年廢票比例增加 (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
陳呂 2004 年得票率 (%)	0.0197 (4.73)**	-0.0119 (4.16)**	-0.0099 (3.46)**	0.0241 (6.33)**	-0.0072 (2.89)**	-0.0045 (1.73)
不識字比率 (%)		0.2260 (28.48)**	0.2255 (28.42)**		0.2231 (30.75)**	0.2225 (31.84)**
民進黨主政縣市			-0.1101 (1.78)			-0.1483 (2.85)**
常數項	1.5563 (7.68)**	1.9546 (14.87)**	1.912 (14.23)**	0.3805 (2.17)*	0.7737 (7.03)**	0.7165 (6.25)**
樣本數	368	368	368	368	368	368
R ²	0.06	0.70	0.70	0.09	0.73	0.74

¹ 括弧內為 t 值的絕對值，**表示在 1% 的水準之下顯著，*表示在 5% 的水準之下顯著，各迴歸式均以各鄉鎮投票數加權。

進黨做票」，還需要更進一步說明其間的機制為何。另一方面，由前一節的討論，我們知道無效票範圍的擴大，使得不識字比例較高鄉鎮的廢票比例顯著提高。模型(2)即加入不識字比例作為解釋變數，此時不識字比例的係數為 0.2260，且相當顯著，不識字比例越高的鄉鎮，廢票比例也較高。有趣的是，此時陳呂得票率的係數則反轉為 -0.0119，且相當顯著，控制了教育程度之後，陳呂得票較高的鄉鎮，廢票比例反而是較低的。這印證了前面所討論，當不識字比例在模型(1)被遺漏了，不識字比例又與廢票比例成正相關，將使得模型(1)中，陳呂得票率的係數往正的方向偏誤。

模型(3)則再加入民進黨主政縣市的虛擬變數，該變數的係數為負，但不顯著。換言之，民進黨主政縣市的廢票比例並沒有較高。縣市選委會是地方選務的執行機關，民進黨主政縣市的廢票比例並沒有比較高，這與「做票說」透過地方選務機關操作的假說並不一致。此外，模型(3)中不識字比例仍為最重要的解釋變數，陳呂得票率的影響仍為顯著的 -0.0099。

模型(1)-(3)並沒有考慮各鄉鎮廢票比例原本就可能存在的差異，因此模型(4)-(6)中，我們以 2000 年至 2004 年廢票比例的增加作為被解釋變數，其結果與模型(1)-(3)相似，但略有差異。模型(4)顯示，陳呂得票率較高的鄉鎮，其廢

票比例增加的程度也較大。不過，控制了不識字比例之後的模型(5)，陳呂得票率越高的鄉鎮，其廢票比例在兩次選舉之間反而是減少的，係數為顯著的 -0.0072 。此時，真正重要的解釋變數是不識字比例，其係數大小和表 4 由 differences-in-differences 估得的係數值大約相同，為 0.2231。

最後，模型(6)加入「民進黨主政縣市」的虛擬變數，其係數和模型(3)相同，仍為負值，但是顯著。控制了其他變數之後，民進黨主政縣市，廢票比例在兩次選舉之間反而是減少的。此外，不識字比例仍為最重要的解釋變數，不識字比例越高，廢票比例增加的幅度越大。另外，在控制了「不識字比例」和「民進黨主政縣市」之後，陳呂得票率的係數轉為負值，但已經不顯著了。

換言之，陳呂得票率和廢票比例成正相關，此一「做票說」據以作為基礎的現象，並不蘊含因果關係，而是兩變數都受到教育程度的影響。控制了不識字比例之後，陳呂得票率對 2004 年廢票比例和 2000-2004 年廢票比例的增加的影響，都反轉為負值，「做票說」顯然無法得到資料的支持。

5 結論

2004 年總統大選，由於兩組候選人的得票數差距只有 29,518 票，再加上廢票總數高達 337,297 票，比 2000 年總統選舉增加了 215,019 票，因此引發選後極大的爭議。本文使用 2000 年和 2004 年總統選舉各鄉鎮投票結果的資料，配合 2000 年戶口普查各鄉鎮的人口資料，檢驗「廢票運動」、「做票」和「無效票範圍擴大」等三種假說的可能性。

首先，各縣市廢票比例在兩次選舉的變化顯示，2004 年廢票比例增加較少的縣市，都是直轄市或省轄市等都會型縣市；廢票比例增加較多的，反而是雲林、嘉義和彰化等都市化程度較低的縣分，此與「廢票運動」的影響應在都市地區較顯著的預期並不一致。此外，民進黨主政縣市與非民進黨主政縣市，其廢票比例增加的幅度幾乎完全相同，並沒有明顯的差異。

其次，觀察各鄉鎮不識字比例和廢票增加的情形，我們發現兩者成高度的正相關。相對於不識字比例較低的半數鄉鎮，不識字比例較高的鄉鎮，在 2004 年無效票範圍擴大之後，廢票比例多增加了 1.392%，而且相當顯著。在

控制了其他人口變數和就業人口的產業組成之後，不識字比例增加1%，會使得廢票比例多增加0.21%，此與廢票增加是因為「無效票範圍擴大」的假說一致。

此外，我們也檢驗「做票說」中，關於「陳呂得票率越高，廢票比例越高」這個現象的詮釋。由「陳呂得票率越高，廢票比例越高」的相關性，要推論到「做票」，必須進一步論證由縣市長擔任主任委員的縣市選委會，到各個投票所的主任監察員和工作人員，和候選人指派的監察員，如何勾結進行有系統的做票。本文嘗試說明，陳呂得票率和廢票比例的相關性，並不是因果關係，而是兩者同時受到選民教育程度的影響。迴歸分析顯示，控制不識字比例之後，陳呂得票率越高的鄉鎮，其廢票比例在兩次選舉之間非但沒有增加，反而是減少的，惟此結果不具統計顯著性。廢票比例增加的真正重要解釋變數是不識字比例，不識字比例每增加1%，廢票比例多增加的幅度為0.2225%。控制了不識字比例和陳呂得票率之後，民進黨主政縣市的廢票比例不但沒有較高，兩次選舉間廢票增加的幅度甚至是比較小的。

再者，陳呂得票率與廢票率之間的正相關，也可由高等法院以擴大之後的無效票定義重新驗票的結果看出來。驗票之後，陳呂得票數減少10,793票，連宋得票數則只減少了6,838票，兩者比例為1.58:1，陳呂得票在新定義之下成為廢票的機率，明顯高於連宋得票成為廢票的機率。這與本文廢票主要來自無效票定義範圍擴大，而陳呂支持者較可能受到影響的結果一致。

綜合而言，「廢票運動」與「做票」兩個假說，和各縣市廢票比例在兩次選舉中的變化，以及縣市是否由民進黨主政的差異都不一致。而「做票說」的立論基礎中，陳呂得票率和廢票比例的相關性，有忽略變數的問題，兩者並沒有因果關係。最後，實證分析中廢票比例的性質則與「無效票範圍擴大」的假說較為一致。

附 錄

2		1	
候選人 副總統	候選人 總統	候選人 副總統	候選人 總統
			
宋楚瑜	連 戰	呂 秀 蓮	陳 水 扁
中國國民黨、 親民黨 推薦		民主進步黨 推 薦	

附圖 1：選票範例

參考資料

王甫昌

- 1994 〈族群同化與動員——台灣民眾政黨支持之分析〉，《中研院民族學研究所集刊》77: 1-34。

林忠正、林鶴玲

- 1993 〈台灣地區各族群的經濟差異〉，見張茂桂等（著），《族群關係與國家認同》，頁101-160。台北：業強出版社。

吳乃德

- 1994 〈社會分歧與政黨競爭：解釋國民黨為何繼續執政？〉，《中研院民族學研究所集刊》78: 101-30。
- 1997 〈檳榔和拖鞋，西裝及皮鞋：台灣階級流動的族群差異及原因〉，《台灣社會學》1: 137-167。

吳怡銘

- 2000 「台北市選民分裂投票之研究——民國八十七年市長與市議員選舉之分析」，國立中正大學政治學研究所碩士論文。

施奕任

- 2001 「1998年高雄市市長與市議員選舉中選民「分裂投票」行為」，國立中正大學政治學研究所碩士論文。

徐永明

- 2001 〈宋楚瑜現象：菁英分裂、選票動員與政黨重組〉，《理論與政策》14(2): 93-118。

黃紀

- 2001 〈一致與分裂投票：方法論之探討〉，《人文及社會科學集刊》13(5): 541-574。

黃紀、張益超

- 2001 〈一致與分裂投票：嘉義市一九九七年市長與立委選舉之分析〉，見徐永明與黃紀（編），《政治分析的層次》。台北：韋伯文化事業出版社。

駱明慶

- 2001 〈教育成就的省籍與性別差異〉，《經濟論文叢刊》29(2): 117-152。

薛承泰

- 1996 〈影響國初中後教育分流的實證分析：性別、省籍、與家庭背景的差異〉，《台灣社會學刊》20: 49-84。

盧瑞芬、謝啓瑞

- 2000 《醫療經濟學》。台北：學富文化事業有限公司。

Achen, Christopher H. and W. Phillips Shively

- 1995 *Cross-Level Inference*. Chicago: The University of Chicago Press.

Chou, Shin-Yi, Jin-Tan Liu and James K. Hammitt

- 2003 "National Health Insurance and Precautionary Saving: Evidence from Taiwan," *Journal of Public Economics* 87: 1873-1894.

Duflo, Esther

- 2001 "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic*

- Review* 91(4): 795-813.
- Duncan, Otis D. and Beverly Davis
1953 "An Alternative to Ecological Correlation," *American Sociological Review* 18: 665-666.
- Goodman, Leo A.
1953 "Ecological Regressions and Behavior of Individuals," *American Sociological Review* 18(3): 663-664.
1959 "Some Alternatives to Ecological Correlation," *American Journal of Sociology* 6(6): 610-625.
- King, Gary
1997 *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. New Jersey: Princeton University Press.
- Meyer, Bruce D.
1995 "Natural and Quasi-Experiments in Economics," *Journal of Business and Economic Statistics* 13(2): 151-161.
- Robinson, W. S.
1950 "Ecological Correlations and the Behavior of Individuals," *American Sociological Review* 15(3): 351-357.
- Tsai, S.L. and H.Y. Chiu
1993 "Educational Attainment in Taiwan: Comparisons of Ethnic Groups," *Proceedings of the National Science Council, ROC* 3(2): 188-202.

The Effect of Broader Definition of Invalid Votes on 2004 Presidential Election in Taiwan

Ming-ching Luoh

Professor

Department of Economics, National Taiwan University

ABSTRACT

Using data from 2000 and 2004 presidential elections and data from the 2000 Population Census, this paper tests the hypotheses regarding the dramatic increase of invalid votes in 2004. First, the less-urbanized counties showed a greater increase in the proportion of invalid votes, and whether the DDP is the county ruling party does not show any significant differences. Second, as the definition of invalid votes became broader in 2004, townships with an above-median proportion of illiterate voters significantly increase the proportion of invalid votes by 1.392%. Regression results show that for a 1% increase in the illiteracy rate, the proportion of invalid votes increases by 0.21%. Estimation from King's ecological inference model also shows that the probability of casting invalid votes for illiterate voters increases significantly in 2004 and is much higher than that of literate voters. Finally, we test the hypothesis of vote-rigging based on the observation that the share of votes gathered by Chen-Lu and the proportion of invalid votes are positively correlated. It is shown that this correlation is not causal. After controlling for the illiteracy rate, the effect of the share of votes gathered by Chen-Lu on the proportion of invalid votes is reversed and becomes negative, although not statistically significant. In sum, the hypothesis that invalid votes are caused by the broader definition of invalid votes is the only hypothesis consistent with the data.

Key Words: invalid vote, illiteracy rate, presidential election, vote-rigging