

# 選舉縱橫資料中地區效果的測量： 以民進黨得票率的變化為例 1986-2004\*

徐永明

東吳大學

政治系助理教授

林昌平

國立政治大學

金融研究所博士生

本文以三項不同的實證模型：時間合併資料模型、縱橫資料固定效果模型與 Hausman-Taylor model 探討選舉制度、族群以及地域分佈三項縱橫資料型態的變數對於歷年間各地區民進黨得票率的影響性。我們發現「民進黨執政縣市」對於「民進黨得票率」的影響性於時間合併資料模型中會有過於高估的狀況，然而縱橫資料固定效果模型將會消除「民進黨執政縣市」對於「民進黨得票率」原本過於高估的影響效果。但是因為固定效果模型具有無法對於不隨時間變動的解釋變數進行估計的缺點，我們首次提出 Hausman and Taylor (1981) 的 Hausman-Taylor model 來解決此問題，該模型同樣可以排除個別效果 (individual effect) 所帶來的估計偏誤，但是卻能解決縱橫資料固定效果模型的缺點。就無時間變化解釋變數對於應變數的影響性而言，我們於實證分析中發現若單純以時間合併資料結構進行分析，屬於「閩南鄉鎮」與「山地鄉鎮」的族群分佈影響力將被減弱，也就是說時間合併資料模型過於低估「無時間變化內生變數」對於應變數的影響程度，Hausman and Taylor 的 H-T model 則否。

另外，原本應是族群分佈解釋變數對於民進黨得票率的影響性，卻在時間合併資料結構中改以地域性質的解釋變數來詮釋。綜合言之，對探討選舉制度、地域與族群分佈對於民進黨得票率的實證分析來說，採用 H-T model 將可以

---

\* 作者感謝審查人與編輯委員針對文中缺失所給予的詳盡指正與耐心建議。此外，作者特別感謝國科會計畫編號 NSC-93-2414-H-001-002 與 NSC-94-2414-H-001-012 經費的支持，使本研究得以順利完成。

獲得既精確又充分運用資料訊息的估計結果，其解決了固定效果模型的缺點以及時間合併資料模型會產生的估計偏誤。

**關鍵字：**選舉制度、族群分佈、縱橫資料、固定效果、個別效果

## 壹、前言

本文試圖以選舉制度、族群以及地域分佈三項縱橫資料型態 (panel data) 的變數來探討對於歷年間各地區民進黨得票率的影響力，並且依據此項實證分析說明時間合併資料模型 (pooled data model)、縱橫資料固定效果模型 (panel data fixed effect model) 與 Hausman and Taylor (1981: 1377-1398) 的 Hausman-Taylor model 三項模型在模型設定、對於殘差項的假設以及變數間的解釋能力來說明各項模型間的差異性，最後並指出，為何 H-T model 對於具有地域與時間性質的各項選舉得票率分析有較大的優勢。<sup>1</sup>

### 一、研究問題

Green and Kim (2001: 448) 曾比較固定效果模型與時間合併資料模型，提出若單純以時間合併資料的方式進行模型估計，由於時間合併資料忽略了各地域本身的特色—不同地域之間應有不同的迴歸截距項，若於迴歸模型中將所有不同地域的特色省略為相同地單一截距項，在估計上將會產生估計的偏誤，該文稱之為「dirty pool」。換句話說，在一般傳統選舉研究的模型分析中，若將全部的資料以時間合併的資料結構進行分析，似會忽略資料的時間性與地域性等問題。本文發現過去對於選舉制度、族群以及地域分佈的相關文獻，不是著重於一般橫斷面的資料分析，例如徐永明 (2001b: 79) 比較 1994 年與 1998 年兩次選舉中民眾選擇的變化；盛杏媛 (2002: 49) 個別探討

---

1 本文中的縱橫資料 (panel data) 為採用地區為橫斷面單位，時間年為時間序列單位的加總資料 (aggregate data)。

1990年代間台灣選民的投票行爲；耿曙、陳陸輝（2003: 5）以「差異中的差異」（difference in difference）方式分析台灣政治版圖與兩岸經貿之間的關係，就是單獨分別的討論不同區域間或是不同時間上的差異性，例如林繼文（2002: 131-171）探討不同國家在選舉制度與選民偏好之間的差異性；吳乃德（2005: 11）則於不同的時間點上，由過去幾次對台灣民眾的訪問調查研究發現，台灣民眾的民族認同歷經大幅度的變動。

徐永明與林昌平（2005: 1-12）認為在一般橫斷面資料的模型分析中，將會出現省略變數所造成模型估計的誤差，使得該實證模型不能估計出正確的迴歸係數估計值，因此也不能了解解釋變數對於應變數真正的影響性。<sup>2</sup> 因為無法處理這些誤差的影響，所以政治學中關於選舉制度、族群以及地域分佈的研究就停留在個別年度的集群分析（政治勢力的橫斷面分析）或是探討不同年度間政治選項的流動方向，而無法探討造成政治版圖變化的外生性因素（exogenous variable）。若是採用縱橫資料的分析方式卻可以明顯的矯正該誤差項，過去相似的文獻如洪永泰（1994: 95）首次提出以整體資料分析政治版圖的概念；徐永明、陳明通（1998: 69）則試圖探討縱橫資料中個體層次資料的統獨動力來源；徐永明（2001a: 98）分別探討兩種不同的政治版圖研究途徑，試圖整合類似縱橫資料的分析方式；Hsu and Lee（2002: 71）則分析政黨發展中不同地域的影響效果。

我們以縱橫資料做爲一個研究設計，並利用其準實驗設計的特性，將可以正確地發現迴歸模型中自變數與應變數之間的影響關係。本文提出相較於單純的時間合併資料迴歸分析，若以縱橫資料的資料結構進行固定效果模型的估計，可以發揮出縱橫資料的優越性，將固定效果中的 group effect 以及 time effect 區分出來，如此才可以在排除內生性影響的狀況下對於單純相關性影響的變數進行刻劃，完整的呈現出各種外生變數對於模型應變數的影響

---

2 此處模型估計的誤差是指迴歸模型估計係數的偏差，由於不採用縱橫資料模型對於縱橫資料進行實證分析，作者發現以時間合併資料模型進行估計的結果，其迴歸係數與採用縱橫資料模型進行估計的迴歸係數有所差異。本文中我們以時間合併模型、固定效果模型以及 H-T model 對於分析選舉制度、族群分佈以及地域對於民進黨得票率的影響為例，比較時間合併模型在迴歸係數的估計上的確有所偏誤。

能力。<sup>3</sup> 也就是說，採取縱橫資料結構區分出 group effect 與 time effect，以不同的模型型態加以控制，對於模型中變數的解釋能力將有更優越的表現，例如王鼎銘等（2004: 87）即以縱橫資料的資料結構分析日本自民黨的選票穩定度。

最後我們於文中提出第一次運用在政治學研究的分析方式，Hausman and Taylor（1981: 1377-1398）於縱橫資料結構下，發展出一種新的估計模型以解決固定效果模型無法探討「個別效果」（individual effect）影響性的缺失，本文將其模型簡稱為 H-T model。<sup>4</sup> 我們將運用此項模型於選舉制度、族群以及地域分佈的研究中，分析不同族群、地域分佈以及選舉制度對於民進黨得票率的影響性，並且與先前所提到的時間合併資料模型以及固定效果模型進行比較，探討三項模型的優劣性。<sup>5</sup>

## 二、縱橫資料做為一種研究設計

隨著縱橫資料在政治科學研究上的逐漸發展，具有時間性質與橫斷面性質的縱橫資料逐漸增多。例如在不同年度中，台灣各鄉鎮市不同選舉層次的

---

3 所謂 group effect 指的是在縱橫資料模型中，隨著 individual（即模型中的下標  $i$ ）的不同而改變，但不隨著時間變化而改變的影響效果。在本文的實證分析中，group effect 代表地域虛擬變數以及族群類型對於民進黨得票率的影響性。而 time effect 指的是在縱橫資料模型中，隨著時間的不同進行改變，但不隨 individual 發生變化而改變的影響效果。此處 time effect 代表各個時間虛擬變數對於民進黨得票率的影響性。

4 本文原先將該模型稱之為「個別效果模型」，然評審之一指出將 Hausman-Taylor model 稱做個別效果模型會誤導讀者認為該模型與個體層次資料的資訊（individual-level information）有關，我們感謝並接受評審的建議，並修改為 Hausman-Taylor model，簡稱 H-T model。

5 本文所分析的集體層次資料，主要以過去台灣歷屆選舉中，民進黨的得票率資料為主。之所以選擇民進黨的資料進行分析，主要是考量到過去選舉中，泛藍勢力歷經分合（主要為親民黨的勢力消長），因此在衡量各項主要變數（選制、族群或是地域）對於國民黨的影響上，將會受到歷年來泛藍勢力分合的影響，較無法單純判斷出主要變數的影響效果。而若就泛綠勢力而言，雖然亦有台聯的形成，但其對於民進黨的影響性相對於親民黨對於國民黨的影響性較小，因此針對歷年選舉民進黨得票率的分析，本文認為較易找出選制、族群以及地域對於得票率之間的關係，這也是筆者於本文使用民進黨歷屆選舉得票率做為分析對象的主要原因。

民進黨得票率以及支持度即為縱橫資料的應用。在此種資料型態的研究中，若我們仍拘束於單純橫斷面資料或是時間序列資料的研究，將無可避免地喪失完整縱橫資料中大部份的訊息。

因此我們將縱橫資料此種研究設計的分析方式應用於分析不同選舉制度、族群以及地域分佈對於民進黨得票率的影響性，發現縱橫資料特殊的資料型態將與此項研究有著相當適切的合適性，在實證分析的資料當中，同樣地強調不同觀察值個體層次（區域）以及不同時間的特性，故當我們應用縱橫資料的分析方式於實證研究時，將可以消除「省略變數」對於模型估計所造成的影響，並進一步運用相關分析方式控制此影響，如此將可以使我們的模型估計獲得更準確的正確性。

基於上述，越來越多的研究採用縱橫資料的資料型態並且以相關的模型進行分析，然而在眾多的分析方式與模型之中，我們必須避免選取不當的模型而發生對於變數間解釋力的錯誤詮釋，例如是否該以固定效果模型（fixed effect model）進行估計，還是以隨機效果模型（random effect model）的假設進行設定；更甚者，各地區的「固定效果」是否可以考慮以特殊的模型設定方式進行衡量。本研究設計的主旨即在強調為何以縱橫資料的資料結構進行選舉制度、族群以及地域分佈對於民進黨得票率的實證研究會有較佳的優點，以及若在縱橫資料結構中進行實證分析，我們該採用何種估計方式，可以在相同的訊息中獲得最具有準確性以及效率性的實證結果。<sup>6</sup>

本篇文章除了本節為前言外，第二節試圖比較一般傳統的時間合併資料模型與縱橫資料模型主要的差別，並以此進一步指出縱橫資料的模型可以區分出不同「個別效果」（individual effect）的差異性。<sup>7</sup> 第三節提出 Hausman and Taylor（1981: 1377-1398）的 H-T model，進一步分析在縱橫資料的結構下，H-T model 如何提出估計出個別效果的模型假設，也就是說在此一模

6 此處的準確性說明迴歸係數的估計不會偏誤，而估計上具有效率性意指不喪失模型中部份解釋變數對於應變數的解釋能力，例如固定效果模型即具備不能估計不隨時間變動變數影響性的缺失。

7 本文以各項實證模型中個別效果的比較與測量，來衡量鄉鎮市區所代表的地區效果，故文中的個別效果皆可視為以鄉鎮市區為模型下標  $i$  的地區效果。

型之中，我們不僅可以於縱橫資料中區分出個別效果，相較於固定效果模型我們更可以把原先無法估計的個別效果進行估計。第四節則提出本文的實證分析，區分比較一般傳統時間合併資料模型、縱橫資料固定效果模型以及 H-T model 於族群、地域分佈以及選舉制度三項縱橫資料型態的變數對於歷年間各地區民進黨得票率的影響效果，最後更獲得 H-T model 於相同的縱橫資料中可以獲得更多且正確的訊息。第五節為本篇文章的結論。

## 貳、時間合併資料與縱橫資料的差異性

就時間合併資料的迴歸模型設定而言，其模型設定的基礎為將各個樣本資料認定為獨立的存在，彼此之間於資料結構上並未有依據不同個體層次 (*i*) (individual) 與不同年度的分類存在。以一實證樣本實際的分析資料應符合縱橫資料的資料結構為例，在時間合併資料的資料結構設定中將會忽略各別個體的性質，同時不考慮不同個體下時間性的影響性 (即不同年度)，故時間合併資料的模型設定應如下列式(1)所示：

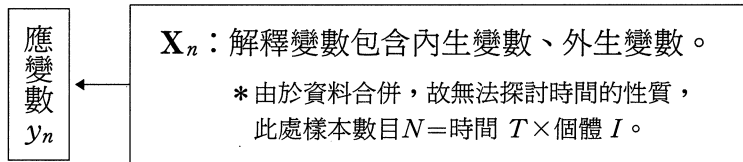
$$Y_n = \beta_0 + X_n\beta + u_n \dots \dots \dots \text{(時間合併資料模型)} \quad (1)$$

式中每個觀察值皆以一獨立樣本來進行分析， $Y_n$  為模型中的應變數， $X_n$  代表模型中的解釋變數矩陣，而  $u_n$  為各樣本於迴歸分析中的殘差項。此時在時間合併資料模型的分析下，模型的樣本數  $N$  將為縱橫資料結構中個體層次的個數乘上時間的數目，即  $N = I \times T$ 。

我們以下圖 1 說明在時間合併資料模型中解釋變數與應變數之間的關係。當我們於時間合併資料的基礎中放棄區分樣本觀察值中會有不同個體層次或不同時間的特色時，時間合併資料模型將會同時探討所有解釋變數對於應變數的影響性，此時不僅不考慮時間性 (*t*) 的因素，亦不考慮不同個體層次的特色，反而將全部的樣本觀察值一視同仁 ( $N = I \times T$ )。例如即使樣本中的各項觀察值來自不同的年度，仍然將所有觀察值視為同一年度的樣本，因此即使解釋變數中可以採用是否為內生變數以及是否具有時間變化等條件進行區分，在時間合併資料模型中仍會將這些不同時間或是不同地域性質的解

釋變數視為相同性質的解釋變數，如此當然無法於解釋變數對於應變數的解釋力中排除個別效果的影響性，換句話說，此項模型的估計結果將會有所偏誤。總的言之，時間合併資料模型存在資料結構型態不足以及無法排除個別效果兩大缺失。

圖 1：時間合併資料模型中解釋變數與應變數之間的關係圖



反之，當我們將實證樣本以縱橫資料的資料結構來進行分析時，將會考量到不同個體層次的性質以及於相同個體下的時間特性，故縱橫資料固定效果模型的設定將如下列式(2)所示，其中  $Y_{i,t}$  為考量不同個體性質下的模型應變數。

$$Y_{i,t} = \beta_0 + X_{i,t}\beta + Z_i\gamma + u_{i,t} \quad (2)$$

而  $X_{i,t}$  與  $Z_i$  皆為向量的形式，其中  $X_{i,t}$  是解釋變數矩陣，而  $Z_i$  則包含一切省略變數的影響。上述縱橫資料固定效果模型的設定重點在於我們採用了固定效果中之 group effect 來進行估計，以明確估計出政治版圖中的內生性特色，目的為於縱橫資料的研究設計中控制住省略變數對於模型估計所造成的誤差，最後  $\beta$  與  $\gamma$  分別為估計係數，而  $u_{i,t}$  為排除掉省略變數所造成的誤差之外的其他誤差項。

我們可以進一步簡化式(2)為下列式(3)，其中  $v_i$  代表我們將所有省略變數所帶來的誤差影響 ( $Z_i\gamma$ ) 簡化。

$$Y_{i,t} = \beta_0 + X_{i,t}\beta + v_i + u_{i,t} \quad (3)$$

在進一步對上式各項變數取組內平均下，我們將可以得到式(4)，而由式(3)與式(4)的組內平均取其差分項，得到式(5)，使我們的模型可以排除省略變數的誤差項以進行模型估計。

$$\bar{Y}_i = \beta_0 + \bar{X}_i \beta + v_i + \bar{u}_i \quad (4)$$

$$Y_{i,t} - \bar{Y}_i = (X_{i,t} - \bar{X}_i) \beta + (u_{i,t} - \bar{u}_i) \dots \dots \dots \text{(固定效果模型)} \quad (5)$$

採用縱橫資料型態來處理迴歸模型資料的好處在於，不論是時間上或是橫斷面上之訊息都能捕捉，若再加上特殊的估計方式（固定效果模型或是隨機效果模型），我們可以比較出為何 H-T model 對於具有地域與時間性質的各項選舉得票率分析有較大的優勢。以固定效果模型來說，模型內除了解釋變數之外，還包括了省略變數所帶來的誤差影響。這個特定的誤差影響性就是主要異質的來源，指的是不隨時間改變也觀察不到之特定個體層次因子。固定效果模型之優點為允許解釋變數與個體層次之變異因子相關，並在使用「與組平均差距」的方式（式 5）之後使其相關性消失。另外若在 time effect 的估計方式下，我們的實證模型設定將如下列式(6)所示，

$$Y_{i,t} = \beta_0 + X_{i,t} \beta + Z_i \gamma + u_{i,t} \quad (6)$$

其中  $X_{i,t}$  仍為解釋變數矩陣，但與式(2)不同的地方在於此時將加入時間的虛擬變數，以考量時間對於模型應變數的解釋能力。

以上提出一般縱橫資料固定效果模型的相關設定，其中，式(5)為了將省略變數所造成的誤差消除，採取了「與組平均差距」(deviation from group mean) 的方式，這種方法雖然明確地排除了個別效果的影響性，但是我們也可以發現，於迴歸式中已經將不隨時間變化的解釋變數予以消除，此為固定效果模型中的一項缺失，因為我們無法直接於模型中看出不隨時間變化的解釋變數對於應變數的影響性，針對此項缺失，Hausman and Taylor (1981: 1377- 1398) 提出了修正的估計方式。

我們以下圖 2 說明在縱橫資料的資料結構之下，固定效果模型中解釋變數與應變數之間的關係。縱橫資料的基本性質為樣本觀察值中會有不同個體或不同時間的特色存在，因此縱橫資料的資料結構將可以避免時間合併資料模型存在資料結構型態不足的缺點。此外，因為資料結構中樣本觀察值具有時間性質存在，我們可以輕易將解釋變數區分為有時間變化的解釋變數以及無時間變化的解釋變數兩大類，其中各類又可區分為內生變數與外生變數。



圖 2：固定效果模型中解釋變數與應變數之間的關係圖

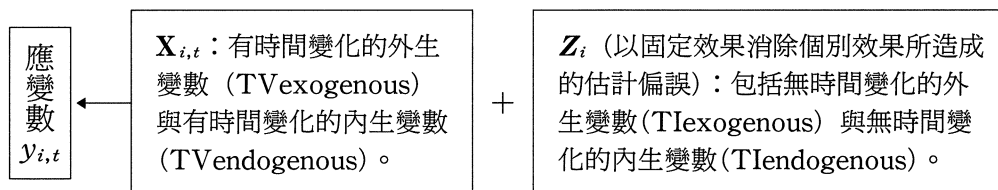


圖 2 中若以本文實證模型的解釋變數為例，有時間變化的內生變數則如「民進黨執政縣市」是會隨著不同時間而改變，且為該區域本身內生決定；有時間變化的外生變數則像是不同的選舉制度，在不同的時間點上樣本觀察值將代表不同的選舉制度（總統選舉、縣市長選舉），且是外生給定的；無時間變化的內生變數例如閩南鄉鎮或是山地鄉鎮，是並不會因為不同年度而改變的內生性質；最後不同的區域則表示無時間變化的外生變數，外生給定的行政區域並不會因為時間進行改變。<sup>8</sup>

上述各項類型的區分，主要依據過去的研究發現。徐永明（2001a: 98）認為台灣的政治支持受到南北分佈的影響，而駱明慶（2001: 134）亦指出都會化程度與省籍分佈兩者具有相關性。換句話說，本文認為外生的南北分佈（都會化程度）將會影響到省籍分佈並進而影響政治態度的形成，其中由於省籍分佈是會變動的，因此為了與南北分佈進行區分，本文認為其屬於內生性的解釋變數。

而依循著固定效果的估計方式，固定效果模型將排除個別效果，以獲得不受個別效果影響的估計結果，其最大缺失在於排除個別效果的同時，亦去除了無時間變化的解釋變數對於應變數的估計結果，因此我們將無法於固定效果模型的估計結果中，看到無時間變化的內生變數與外生變數對於應變數

8 評審之一認為通常時間序列模型的分析中，預先決定（predetermined）的自變項多為外生變數，本文作者感謝評審的提醒。此處將「民進黨執政縣市」、「閩南鄉鎮」以及「山地鄉鎮」此三項自變項設定為內生變數的主要依據，在於這些變數的群聚性質與代表 individual effect 的地區性質有關，皆為各鄉鎮市依據其地域的特性逐漸內化所形成的特質，因此本文將此三項變數設定為內生變數。然本文作者接受評審的建議，於此說明此三項內生變數的設定原因，並說明以此三項自變項代表內生變數是本文實證模型中的自行設定。

的影響性。因此在固定效果的模型下，解釋變數中若含有不隨時間而變之變數，其效果會被消除，所以在上述迴歸式中，並未看到任何不隨時間而變之變數；故若不隨時間而變之變數的影響效果是分析探討的重點，則只能使用隨機效果模型方可允許虛擬變數迴歸係數的估計。然而隨機效果模型有一個較強的假設：即假設個體層次特定之異質因子與其他解釋變數不相關，且為 Gaussian 分配，若不符合此項假設而繼續使用隨機效果模型，將使得估計式不一致。

Hausman and Taylor (1981: 1377-1398) 進一步提出新的估計式來解決上述隨機效果模型與固定效果模型的估計問題。主要改進單純使用隨機效果模型時，必須接受個體層次特定因子與解釋變數不相關假設的缺點；並同時改進固定效果模型的缺點，允許模型估計不隨時間變化但可觀察到的解釋變數，例如虛擬變數。

## 參、Hausman-Taylor model

本文試圖探討族群、地域分佈以及選舉制度三項縱橫資料型態的變數對於歷年間各地區民進黨得票率的影響效果，其中於採用縱橫資料結構進行估計的基礎下，嘗試以 Hausman and Taylor (1981: 1377-1398) 的 H-T model 對於「個別效果」(individual effect) 進行模型估計，除了試圖於不同的區域 (individual) 中找出與時間變化以及內生變數等模型限制有關的解釋變數影響力，並且嘗試估計出固定效果模型無法估計出的無時間變化解釋變數的影響力，即所謂個別效果 (individual effect)。

### 一、Hausman-Taylor model 與固定效果模型的差異性

對於縱橫資料型態 (panel data) 常見的兩種分析模型：固定效果模型 (fixed effect model) 與隨機效果模型 (random effect model)，Mundalk (1978: 73) 曾提出說明如下。首先，指出隨機效果模型假設解釋變數的外生性，並且認為每個觀察值的個別效果 (individual effect) 符合隨機分配。換句話說，也就是假設解釋變數與個別效果之間並無相關性。其次，認為固定

效果模型允許解釋變數的內生性，並且於估計過程中將個別效果技巧性的去除，換句話說，解釋變數與個別效果之間是有相關性存在的。然而，Mundalk (1978: 75) 亦指出並沒有適當的模型可以針對上述隨機效果模型與固定效果模型的假設進行選擇，也就是模型中部分解釋變數與個別效果相關，具有內生性；部分解釋變數則是外生的，與個別效果無關。

本文認為，Hausman and Taylor (1981: 1377-1398) 所提出的 H-T model 實際上即具備將部分解釋變數設定為與個別效果相關，具有內生性；而將部分解釋變數設定為與個別效果無關，具有外生性。此項假設的好處是可於解釋變數與個別效果之間的相關性進行選擇，可以在更符合社會實際情況之下進行估計過程，探討不同解釋變數對於民進黨得票率的影響性。過去的研究文獻中，Baltagi et al. (2003: 364) 也談到，過去實證研究在估計分析之前，皆必須判斷所分析的資料應採用固定效果模型還是隨機效果模型，然而 Hausman and Taylor 所提出的個別效果模型卻在估計過程上採用工具變數 (instrumental variable) 的估計方法，技巧性的區分出內生性解釋變數與外生性解釋變數。

此外，Hausman and Taylor (1981: 1377-1398) 所提出的 H-T model 是一種控制模型中省略變數影響性的估計方法。主要在以「是否有時間變化」與「是否為內生變數」對於模型中的解釋變數進行限制與區分。一般若以縱橫資料模型中傳統的固定效果迴歸模型對於模型中的應變數進行詮釋時，在模型當中會對省略變數進行控制，以避免估計產生誤差，因此個別效果並不會於估計結果當中呈現出來，而是在模型估計過程中的一種「控制」型態。

本文主要的目的即在嘗試以不同的模型限制找出此處所「控制」的解釋力。其中由於模型省略的變數大多不會在時間上進行改變，只會於不同的觀察個體中有所不同，例如各別縣市的文化、習俗、慣用語言等，因此若採用 H-T model 的方式，應是一個適當的形式。

## 二、Hausman-Taylor model 的模型設定

一般在傳統縱橫資料模型的迴歸式中，多假設資料中的異質性 (heterogeneity) 可以由橫斷面的個體或是時間序列上的平均數來衡量，即所謂的

「固定效果」(fixed effect)或是「隨機效果」(random effect)。在經過如「與組平均差距」(deviation from group mean)等轉換計算之後，模型所估計出的迴歸係數將可以明確地表示出各個解釋變數所代表的解釋意義。然而相對於可以清晰表現解釋變數解釋力的迴歸係數，我們對於傳統縱橫資料模型中被消除的「個別效果」仍未能有詳盡的分析。

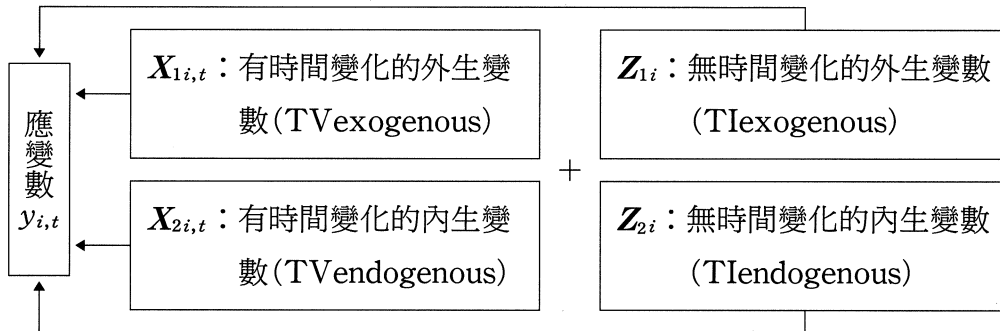
H-T model 提出了一種對於縱橫資料模型的設定，可以進一步刻畫被消除的個別效果，並且加入外生變數的限制以及無時間變化的特徵於模型的解釋變數之中，其模型設定如下式(7)所示。

$$y_{i,t} = \mathbf{X}_{1i,t}\boldsymbol{\beta}_1 + \mathbf{X}_{2i,t}\boldsymbol{\beta}_2 + \mathbf{Z}_{1i}\boldsymbol{\delta}_1 + \mathbf{Z}_{2i}\boldsymbol{\delta}_2 + \mu_i + \varepsilon_{i,t}, \quad (7)$$

模型中  $y_{i,t}$  為被解釋變數，而依據對於解釋變數的限制與分類，Hausman and Taylor (1981: 1377-1398) 提出了四種區分：首先  $\mathbf{X}_{1i,t}$  為有時間變化 (time-varying) 的外生變數 (例如選舉制度)，而  $\mathbf{X}_{2i,t}$  為有時間變化 (time-varying) 的內生變數 (例如民進黨執政的縣市)；另外， $\mathbf{Z}_{1i}$  為無時間變化 (time-invariant) 的外生變數 (例如不同的區域)，最後  $\mathbf{Z}_{2i}$  為無時間變化 (time-invariant) 的內生變數 (例如不同族群的閩南鄉鎮與山地鄉鎮)，其中之假設為  $E[\mu_i] = E[\mu_i | \mathbf{X}_{1i,t}, \mathbf{Z}_{1i}] = 0$ ，但  $E[\mu_i | \mathbf{X}_{2i,t}, \mathbf{Z}_{2i}] \neq 0$ 。模型之另一個認定原則為  $\mathbf{X}_{1i,t}$  變數個數應大於  $\mathbf{Z}_{2i}$  之變數個數，方才令模型得以估計。四組不同性質的解釋變數與應變數之間的關係如圖 3 所示，此處 Hausman and Taylor (1981: 1377-1398) 的估計方式將採用工具變數 (instrument variables) 做為估計出個別效果的方法：首先，以有時間變化的解釋變數做為無時間變化內生變數的工具變數進行估計，其次，再分析四項不同解釋變數對於應變數的影響性，這樣的估計方法是一種兩階段最小平方估計方式的應用。

由圖 3 中我們可以發現採用 H-T model，在第二階段的估計之後，我們可以獲得四種不同解釋變數對於應變數的估計結果，其中更可獲得無時間變化的外生變數對於內生變數的影響性。換句話說，在縱橫資料的資料結構基礎下，H-T model 不僅考慮到估計結果必須排除個別效果的影響偏誤，並且可以於估計結果中呈現出無時間變化的內生變數與外生變數對於應變數的影

圖 3：Hausman-Taylor model 中解釋變數與應變數之間的關係圖



響性，即所謂的個別效果，彌補了固定效果模型估計上的缺失。

### 三、模型比較

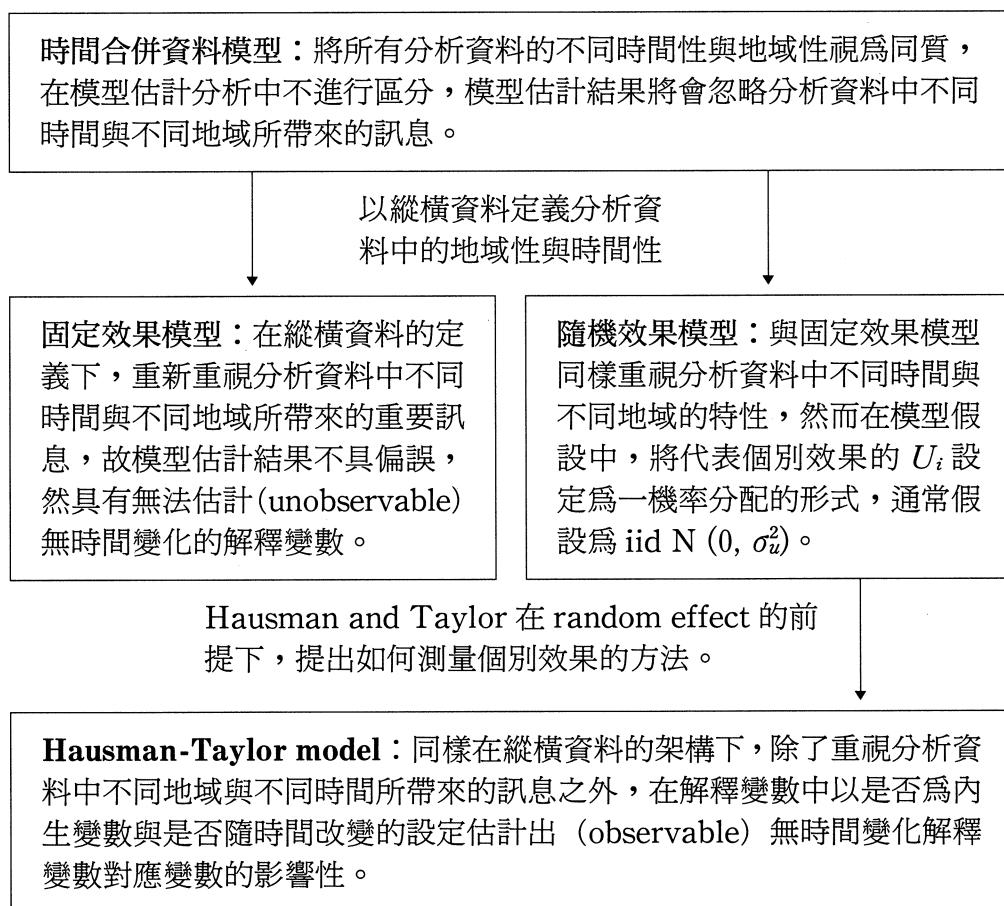
本節嘗試比較前述三項不同的模型：時間合併資料模型、固定效果模型以及 H-T model，做為實證分析模型比較之前的整合。由圖 4 我們可以看出時間合併資料模型原本並不重視分析資料中不同地域或是不同時間的特性，在估計過程中並不考慮此項訊息所帶來的影響性，因此於估計結果中將會具有忽視分析資料特性的偏誤。

而固定效果模型為縱橫資料模型，在考量資料特性（時間性與地域性）之後，模型估計結果並不具偏誤，但該模型的缺失為無法觀察出無時間變化解釋變數對於應變數的影響性。隨機效果模型與固定效果模型同樣重視分析資料中不同時間與不同地域的特性，然而在模型假設中，將代表個別效果的  $U_i$  設定為一機率分配的形式，通常假設為  $iid N(0, \sigma_u^2)$ 。H-T model 是 Hausman and Taylor 為了測量此項無法估計的 individual-level 而發展出來，該模型不僅具有隨機效果的優點：模型的估計結果並不會受到分析資料中不同時間與不同地域的影響造成偏誤。此外，H-T model 以是否為內生變數與是否隨時間變化等特性來設定解釋變數，可以進一步估計出無時間變化解釋變數對於應變數的影響性，解決了先前的缺失。

本文於圖四強調 H-T model 與固定效果模型之間的比較，主要的原因是 H-T model 雖為隨機效果模型的延伸模型，但是 H-T model 強調部分解

釋變數與不可觀察的 individual-level 相關 (Stata, 2003: 95)。透過 H-T model 利用工具變數進行估計的過程, Baltagi et al. (2003: 361) 說明 H-T model 將可提供 individual effect 部分的資訊。而 Baltagi (2002: 120) 亦指出 H-T model 設定應如前式(7)所示。其中部分解釋變數與不可觀察的 individual-level  $u_i$  相關, 即  $X_{2i,t}$  與  $Z_{2i}$  為內生的, 並與  $u_i$  相關, 此時若單純使用固定效果模型常常使用的估計技巧 within transformation, 雖然會消除個別效果  $u_i$ , 但亦同時消除了  $Z_{1i}$  與  $Z_{2i}$ , 因此我們需要以 H-T model 將解釋變數 regroup 之後, 透過合適的估計過程, 帶給我們更多的資訊。

圖 4：時間合併資料模型、固定效果模型、  
隨機效果模型與 Hausman-Taylor model 的比較



由於上述原因，本文認為透過 H-T model 的 regroup 與估計，可以提供我們更多資訊。以本文的實證模型為例，即表四所呈現的各項地區性影響。換句話說，本文認為 Hausman and Taylor (1981: 1381) 所提出的 H-T model，實際上即具備將部分解釋變數設定為與個別效果相關，具有內生性；而將部分解釋變數設定為與個別效果無關，具有外生性。此項假設的好處是可於解釋變數與個別效果之間的相關性進行選擇，可以在更符合社會實際情況之下進行估計過程，探討不同解釋變數對於民進黨得票率的影響性。

## 肆、實證分析

以下我們試圖依據一般傳統時間合併資料模型、縱橫資料固定效果模型以及 H-T model 三種不同的迴歸模型，比較族群、地域分佈以及選舉制度三項縱橫資料型態的變數對於歷年間各地區民進黨得票率的影響效果。首先，先就實證變數進行說明。

### 一、變數說明

我們將依循 H-T model 的設定，對於台灣各鄉鎮市中歷年民進黨得票率的縱橫資料進行實證分析。所謂縱橫資料是採用地區為橫斷面單位，時間年為時間序列單位的加總資料，本文以台灣地區各鄉鎮市為個體層次單位，收集歷年來各鄉鎮市不同選舉的民進黨得票率、城市族群類型、是否為民進黨執政、地區與時間年等虛擬變數與加總資料變數，然後進一步以縱橫資料模型分析地域、選舉制度以及族群分佈對於歷年間各地區民進黨得票率的影響性。

首先，針對地域變數，本文將台灣各地區劃分為北部地區、桃竹苗、中彰投、雲嘉南以及高高屏。過去針對政黨得票率是否受到地域差異所影響，早已被人注意。過去的研究中，O'Loughlin et al. (1994: 353) 分析 1930 年代的德國大選，指出納粹之崛起除了社經因素之外，空間與地方 (space and place) 的作用也是重要原因。而 Kim et al. (2003: 746) 亦以郡 (county) 為單位，分析 1988-2000 年美國的總統選舉，結果發現民主黨的選票主要集

中在東部與都市區域；反之，共和黨選票在西部與鄉村區域，聚集的情況較為顯著。

此外，就台灣過去的選舉研究分析而言，徐永明（2001a: 98）具體比較不同地域之間的投票行為差異。而賴進貴等（2007: 49）的實證研究結果亦認為地域因素將影響投票的選擇，且勢將支配單一選區制度下的選區劃分與選舉結果。其發現以下三點現象：首先、投票行為具有空間聚集的現象，亦即在距離越近的村里中，投票行為越為相似。其次、傳統迴歸模式的誤差具有地域差異，顯現空間異質性的存在。最後、在控制了年齡、教育、行業、所得與族群等社經因素後，鄰近村里的投票行為仍然相關。

針對上述三點發現，筆者認為本文使用 Hausman and Taylor 所提出的 H-T model，恰可以反應上述三點研究發現。首先、因為投票行為具有空間聚集的現象，因此本文於實證模型中加入地域的解釋變數進行探討，並且依據距離越相近則投票行為越相近的原則，本文將地域變數區分為距離相近的北部地區、桃竹苗、中彰投、雲嘉南以及高高屏等地區。其次，由於傳統迴歸模型的誤差項具有地域差異，本文使用的 H-T model 屬於集體層次的縱橫資料結構，恰可以固定效果來配適誤差項的地域差異性。最後，由於控制了社經因素之後，鄰近地區的投票行為依然相關，<sup>9</sup> 因此本文認為有加入地域變數探討其對於民進黨得票率影響的必要性。

此外，賴進貴等（2007: 52）的分析亦提到欲分析的政治態度應是一種動態過程，將不斷受到訊息的接觸與傳播而進行調整。因此本文亦於實證模型中加入時間的虛擬變數探討此種政治態度的動態變化。

本文所蒐集資料的期間為 1986 年至 2004 年，台灣有立委選舉、縣市長選舉、省長選舉或是總統選舉的年度，分別列出如下：1986 年立委選舉、1989 年立委選舉、1989 年縣市長選舉、1992 立委選舉、1993 縣市長選舉、1994 年省長選舉、1995 年立委選舉、1996 年總統選舉、1997 年縣市長選舉、1998 年立委選舉、2000 年總統選舉、2001 年立委選舉、2001 年縣市長選舉、2004 年

---

9 Cho (2003: 372) 探討美國選舉中政治獻金的來源，亦有相同的發現，認為地域因素的影響性是社經因素所無法充分解釋的。



立委選舉以及 2004 年總統選舉共 16 筆資料。上述資料中分別包括不同選舉層次與選舉制度的特性，結合本研究中以縱橫資料的分析方式進行探討，將可以進一步獲得不同選舉制度對歷年間各地區民進黨得票率的實質影響。

在本文的研究期間（1986 年至 2004 年），台灣的縣市長與其他各級行政首長的選舉制度，多為採行「單一選區」（single-member district）「相對多數」（plurality）的方法來產生。相對地，台灣的立委以及其他各級民意代表的選舉制度，則是依據「複數選區」（multi-member districts）劃分，並且採行「單記不可讓渡」（single non-transferable vote, SNTV）的方法進行。Ranny（1996: 168）指出「相對多數的單一選區制」又被稱作「領先者當選制」（first-past-the-post system）。所以，這種選制基本的選舉過程只有一個回合的選舉，一個回合過後誰贏誰輸將就此底定（Cox, 1997: 69）。換句話說，這一制度的要點是在於許多競爭者中，得最高票的第一名，無論是得票過半與否，就算是當選。（王鼎銘，2003: 180）

過去的研究認為，上述兩項選舉制度的差異性，將造成候選人在不同制度下，會形成不同的競選策略。葛永光（1996: 102）對於「單記不可讓渡」的分析指出，由於候選人只需少數選票即可當選，所以較有可能會出現候選人以偏激、凸顯個人形象的極端主義出現。此外，就「複數選區」而言，由於政黨希望獲得最多的席次，勢必在同一選區提名多位候選人，故較易引起同黨相爭的現象。謝復生（1992: 19）亦認為「單記不可讓渡」比「單一選區制」更能兼顧到小黨的比例代表性，使小黨有生存的空間，政黨數目也就相對較多。

因此，我們認為不同的選舉制度不僅促使不同的競選策略，亦反應不同的投票策略。故當我們探討影響歷年民進黨得票率的主要因素時，將選舉制度的差異性納入考量應是必須的。過去的實證研究中，亦有探討選舉制度對於投票選擇影響的差異性，例如王鼎銘（2003: 191）的研究結果發現，選民在兩種不同的選制之下，投票選擇的確有著不同的考量。政策議題對縣市長選舉的影響比在選立委時來得大，尤其須對統獨、經濟環保及社會福利三方面的議題多加留意；反觀政黨在爭取立委選票時，並無須考量太多政策的影響，最多僅要對統獨的傾向多加小心。

過去的研究文獻亦指出，省籍與族群認同是決定台灣地區選民政治態度與投票抉擇的重要變數。胡佛（1998: 164）提到大陸省籍人士對於某些政治議題（尤其是兩岸統獨議題）有相當顯著的取向。此外，徐火炎（1993: 151）也認為整體上民進黨支持者有較高比例為本省籍族群，反觀外省籍族群則較其他族群更傾向支持國民黨。更甚者，吳重禮、李世宏（2005: 75）進一步指出，台灣政治的核心議題即為族群政治，在台灣地區特有的歷史背景與社會環境之下，族群及其相關議題（譬如省籍、族群認同、國家認同以及兩岸統獨爭議等）皆可形成關鍵的「社會分歧」（social cleavages）。而陳文俊（2003: 65）則認為省籍在投票抉擇中，決定支持「藍軍」或是「綠營」的重要因素。

歸究其原因，台灣政治中的族群議題相當於美國種族議題，均屬於 Carmines and Stimson（1990: 11）提出的「簡單議題」（easy issue）。也就是不需要投入高度的資訊成本，也不必具備充分的政治知識，僅需憑藉基本常識便可判斷的議題。而且由於族群意識涉及高度情感性與敏感性成份，因此極易成為各方政治勢力激化動員的訴求。

過去的研究多從人類學或是社會學的觀點，指出族群認同對於政治行為具有顯著影響，並且是探討台灣民主化過程中不可缺少的因素。例如王甫昌（1993: 98）提出「弱勢族群的成員因為競爭稀有資源而感受到優勢族群對於它們的偏見或是歧視時，可能選擇隱藏自己的族群身分，設法被接受或完全同化，也可能團結起來強調族群身分做團體性的對抗」。而吳乃德（1993: 31-32）強調族群文化認同的重要性，逐漸形成族群認同、省籍意識與自我認知。其認為「族群認同的政治化，在台灣可以分為兩個層面來分析。第一是居於政治壟斷地位的族群，利用其政治宰制地位，透過官方意識形態、官方語言、教育等，試圖在文化上同化被宰制的族群。這種文化同化的意圖，加強了被政治宰制族群文化認同的政治化。第二是宰制族群利用其政治壟斷地位，宣揚其特定的國家認同，以合理化外省籍統治團體的威權體制。以政治宰制地位而推動的族群同化，和國家認同經常造成族群關係的政治化。」張茂桂（1993: 235-236）則認為因為主觀意識面的作用，所以族群的特質常常是在互動過程中形成的，是一種「自己人」相對於「其他人」的主觀感覺，而關於族群對於本身特質的解釋，也可以因為時間、空間變化而有所不同的意義。

因此，我們認為族群議題將反應政治支持上的不同對象，故當我們探討影響歷年民進黨得票率的主要因素時，將族群議題納入考量應是必須的。而省籍問題作為一種族群關係，就是本省人和外省人自覺彼此特質的不同，也就是政治、經濟、社會、文化及歷史起源的差異，亦是我們選擇解釋變數的主要依據。

實證模型中被解釋變數與各項解釋變數的敘述統計如下表 1 所示，其中對於「閩南鄉鎮」與「山地鄉鎮」的劃分主要以王甫昌 (2000: 21) 的分類標準為主，「閩南人集中地區」為本省閩南人佔 90% 人口以上，或其他三個族群人口都低於平均數者，鄉鎮區個數為 193 個。另外我們可以於表 1 中看出於模型中除了應變數「民進黨得票率」為連續變數之外，所有的解釋變數皆為虛擬變數的形式。

此外，本文以相對應於此次選舉結果的上一年度做為各項解釋變數所設

表 1：民進黨得票率與解釋變數的敘述統計

應變數名稱	觀察值個數	平均數	標準差
民進黨得票率 $_{i,t}$	5512	0.331	0.158
自變數名稱	觀察值個數	是 (虛擬變數為 1)	否 (虛擬變數為 0)
總統選舉 $_{i,t}$	5512	1077	4667
縣市長選舉 $_{i,t}$	5512	1436	4308
立委選舉 $_{i,t}$	5512	2513	2999
民進黨執政縣市 $_{i,t}$	5512	1982	3762
北部地區 $_i$	5512	856	4656
桃竹苗 $_i$	5512	752	4992
中彰投 $_i$	5512	1088	4656
雲嘉南 $_i$	5512	1264	4480
高高屏 $_i$	5512	1136	4608
閩南鄉鎮 $_i$	5512	3072	2672
山地鄉鎮 $_i$	5512	928	4816

資料來源：國立政治大學選舉研究中心網站

<http://esc.nccu.edu.tw/newchinese/main.htm>，2006/07/06。

定的時間，代表上一年度的解釋變數可以影響此一年度的民進黨得票率；反之，本年度的民進黨得票率卻無法影響上一年度的解釋變數，例如就民進黨執政縣市與民進黨得票率兩項變數而言，本文並不存在民進黨得票率影響民進黨執政縣市等模型應變數影響自變數的關係。

## 二、實證結果

依據與式(7)相同地設定，我們將上述實證模型中的解釋變數區分為四類，分別為有時間變化的外生變數、有時間變化的內生變數、無時間變化的外生變數以及無時間變化的內生變數。對於民進黨得票率的實證模型設定如下式(8)所示：

$$y_{i,t} = \mathbf{X}_{1i,t}\beta_1 + \mathbf{X}_{2i,t}\beta_2 + \mathbf{Z}_{1i}\delta_1 + \mathbf{Z}_{2i}\delta_2 + \mu_i + \varepsilon_{i,t} \quad (8)$$

其中各項解釋變數可區分為下列四種種類：

1.  $\mathbf{X}_{1i,t}$  為有時間變化的外生變數，包括總統選舉與縣市長選舉，分別代表選舉制度是否為總統選舉與縣市長選舉的虛擬變數。
2.  $\mathbf{X}_{2i,t}$  為有時間變化的內生變數，包括民進黨執政縣市，代表該鄉鎮市所屬縣市是否為民進黨執政縣市的虛擬變數。
3.  $\mathbf{Z}_{1i}$  為無時間變化的外生變數，包括桃竹苗、中彰投、雲嘉南、高高屏，分別代表該鄉鎮市的所屬區域是否為桃竹苗、中彰投、雲嘉南、高高屏的虛擬變數。
4.  $\mathbf{Z}_{2i}$  為無時間變化的內生變數，包括閩南鄉鎮、山地鄉鎮，分別代表該鄉鎮市所屬分類是否為閩南鄉鎮與山地鄉鎮的虛擬變數。

### (一) 時間合併資料模型的實證分析

依據上述式(8)的實證模型，首先我們以時間合併的資料結構進行估計，其估計結果將如表 2 中模型一到模型四所示。其中模型一的解釋變數只包含有時間變化的外生變數（總統選舉、縣市長選舉）以及有時間變化的內生變數（民進黨執政縣市）；模型二則增加了無時間變化的外生變數，也就是不同地區的虛擬變數於模型的解釋變數之中，此處我們於模型中加入桃竹苗、中

表 2：民進黨得票率的時間合併資料模型

模型	模型一	模型二	模型三	模型四
TVexogenous				
總統選舉 $i, t$	0.074** (13.180)	0.079** (14.370)	0.085** (16.200)	0.092** (17.310)
縣市長選舉 $i, t$	0.069** (13.660)	0.067** (13.500)	0.065** (13.710)	0.070** (14.790)
TVendogenous				
民進黨執政縣市 $i, t$	0.102** (25.360)	0.091** (22.280)	0.081** (20.610)	0.072** (17.090)
TIexogenous				
桃竹苗 $i$	—	-0.011* (-1.790)	-0.009 (-1.440)	-0.009 (-1.580)
中彰投 $i$	—	0.017** (3.080)	-0.008 (-1.550)	-0.010* (-1.830)
雲嘉南 $i$	—	0.054** (10.320)	0.024** (4.650)	0.024** (4.680)
高高屏 $i$	—	0.064** (11.540)	0.067** (12.730)	0.069** (13.280)
TIendogenous				
閩南鄉鎮 $i$	—	—	0.041** (9.710)	0.043** (10.240)
山地鄉鎮 $i$	—	—	-0.099** (-19.490)	-0.097** (-19.850)
2000 年	—	—	—	0.070** (9.500)
2001 年	—	—	—	0.020** (2.690)
2004 年	—	—	—	0.052** (7.060)
Constant	0.265** (105.250)	0.241** (60.020)	0.248** (54.270)	0.242** (51.770)
$\bar{R}^2$	0.23	0.26	0.34	0.35
$N$	5512	5512	5512	5512

說明：1. 括弧內為  $z$  值。

2. \*\* 與 \* 各表示顯著水準在 5% 與 10% 之下的顯著估計值。

彰投、雲嘉南以及高高屏四個台灣鄉鎮市中不同的地區；模型三我們則增加了無時間變化的內生變數（閩南地區與山地地區）於模型的解釋變數之中；最後我們於模型四增加了實證資料中最後三年的虛擬變數，試圖分析時間趨勢的影響性。

由於模型二、模型三與模型四中分別加入不同程度的虛擬變數，換句話說，因為虛擬變數控制了部份個別效果的影響力，該三項實證模型也就由單純的時間合併資料模型轉化為並不完整的縱橫資料模型。其中就估計結果而言，我們可以明確地發現「民進黨執政縣市」此項解釋變數的迴歸係數呈現逐漸下降的趨勢，在模型一中「民進黨執政縣市」對於「民進黨得票率」的影響效果為 0.102，但是到了模型四的結果中，「民進黨執政縣市」的影響效果卻大幅減少至 0.072。此種迴歸係數下降的趨勢，說明了當我們於模型中以部份的縱橫資料結構取代傳統時間合併資料的資料結構後，由於縱橫資料模型將會以固定效果或是隨機效果的方式去除「個別效果」的影響，因此所估計出「民進黨執政縣市」對於「民進黨得票率」的影響性將不會受到「個別效果」的干擾。也就是說當一般以時間合併資料結構進行實證模型的估計時，就「民進黨執政縣市」對於「民進黨得票率」的影響性而言，將會過於高估之間的影響效果，然而此處因為虛擬變數控制了部份個別效果的影響力，因此將減弱原本過於高估的影響效果。

反之，我們亦可以於模型三與模型四的估計結果發現「閩南鄉鎮」與「山地鄉鎮」對於「民進黨得票率」的影響效果似乎有略為上升的趨勢，不過由於此處上升的程度較小，我們必須透過後文其他相關的實證模型才可以明確地判斷此種上升的趨勢是否顯著。

此外，由模型二到模型四的估計結果中，我們發現四項無時間變化的外生變數（各地區的虛擬變數）對於「民進黨得票率」的影響似乎區分為兩種類型：其一為在加入無時間變化的內生變數（閩南地區與山地地區）以及時間（年）的虛擬變數於模型之後仍未有重大改變的解釋變數，例如桃竹苗與高高屏，其中影響「民進黨得票率」的迴歸係數分別由  $-0.011$  與  $0.064$  改變為  $-0.009$  與  $0.069$ ，說明此兩地區對於「民進黨得票率」的影響性並不會因為是否為閩南地區或是山地地區而改變；然而對於相同為無時間變化的外生

變數而言，中彰投與雲嘉南對於「民進黨得票率」的影響力則有不同地變化，在閩南地區與山地地區等解釋變數的加入之後，中彰投與雲嘉南地區對於「民進黨得票率」的影響力有著逐漸遞減的變化，其中中彰投的迴歸係數甚至呈現影響方向由顯著正向改變為顯著負向，說明著此兩項地區虛擬變數對於「民進黨得票率」的影響性應會隨著是否為閩南地區或是山地地區而有所改變。

最後，由模型一至模型四所估計出的  $\bar{R}^2$ ，我們可以看出當加入地區虛擬變數、城市虛擬變數以及時間效果的虛擬變數之後，整體模型的解釋力是呈現逐漸上升的趨勢，由原本的 0.23 上升至 0.35，也就是說我們加入的解釋變數對於詮釋對於民進黨得票率的影響性是有幫助的。

## (二) 時間合併資料與縱橫資料的實證分析比較

依據上述闡釋，我們進一步探討選舉制度、地域與族群分佈三項解釋變數對於民進黨得票率的影響性，觀察於一般傳統時間合併資料模型的分析與縱橫資料模型分析之間的詮釋是否有所差異。表 3 為比較簡單迴歸模型、固定效果模型以及時間效果模型的估計結果，與前述的分析相同地，當我們由時間合併資料的資料結構改為縱橫資料的資料結構後，由於「個別效果」的消除，我們同樣明確地發現「民進黨執政縣市」此項解釋變數的迴歸係數呈現逐漸下降的趨勢，由簡單迴歸模型的 0.072 下降至固定效果模型的 0.059 以及時間效果模型的 0.041，強調了因為縱橫資料的資料結構下的固定效果模型或是時間效果模型將會去除「個別效果」的影響，因此所估計出「民進黨執政縣市」對於「民進黨得票率」的影響性將不再會受到「個別效果」的干擾，如此將會完全去除過於高估「民進黨執政縣市」對於「民進黨得票率」的影響效果。

更進一步言，我們亦可以在相同的解釋變數設定之下比較時間合併資料模型與縱橫資料模型的估計差異性。就表 2 的模型一與表 3 的模型二而言，其分別代表在時間合併資料與縱橫資料的資料結構下，「總統選舉」、「縣市長選舉」以及「民進黨執政縣市」對於「民進黨得票率」的影響程度。於估計結果中我們可以清楚地發現採用時間合併資料的資料結構於模型中將過於高估「民進黨執政縣市」對於「民進黨得票率」的影響程度。有鑑於此，本

文認為就分析台灣各地區民進黨得票率的影響性而言，採取縱橫資料的模型估計無疑地將會較以時間合併資料的模型估計更為精確。

### (三) Hausman-Taylor model 的實證分析與優點

然而當我們相信採用縱橫資料的模型估計方式是較為精確的狀況下，採用固定效果模型探討台灣各地區民進黨得票率的影響性卻有一項固定效果模型天生的缺點存在。因為固定效果模型在估計過程中採用減去各別個體層次（區域）依循時間的平均數，稱作（within-groups），因此我們將無法以固定效果模型對於不隨時間變動的解釋變數進行估計。舉例而言，表 3 呈現以固定效果模型及時間效果模型對於探討台灣各地區民進黨得票率影響性的估計值，在表中我們可以很清楚地發現，因為區域變數及鄉鎮變數皆為無時間變化的解釋變數，因此我們將無法以縱橫資料模型對於其迴歸係數進行估計。

為解決此項缺失，Hausman and Taylor (1981: 1377-1398) 提出一種在縱橫資料結構之下的兩階段估計法，可以估計出無時間變化解釋變數的迴歸係數，Hausman and Taylor (1981: 1377-1398) 將之稱作估計出「未觀察出的個別效果」(unobservable individual effects)。文中將解釋變數以「是否隨時間變化」與「是內生或是外生變數」兩項原則區分為四種不同的解釋變數分類。在此四項對於解釋變數的分類中，「無時間變化的內生變數」與「無時間變化的外生變數」是固定效果模型不能估計的部份，Hausman and Taylor (1981: 1377-1398) 於此處運用「有時間變化的解釋變數」做為「無時間變化內生變數」的「工具變數」，如此將可進一步估計出原本未能估計的「無時間變化解釋變數」。

由圖 5 我們可以看出就 H-T model 而言，此處實證資料的應變數為「民進黨歷年各地區選舉得票率」，分別受到地域、族群與選舉制度三群不同性質的解釋變數影響，其中因為選舉制度會隨著時間、區域的不同產生變化，且不受各別個體的內生影響，因此於模型中是屬於「有時間變化的外生變數」；同樣隨著時間與區域進行改變的解釋變數「民進黨執政縣市」因為會受到不同個體的內生影響，因此屬於「有時間變化的內生變數」。反之，地域性質的解釋變數為外生給定，並且不會受到時間進行改變，於 H-T model 中區



表 3：民進黨得票率的時間合併資料模型與縱橫資料固定效果迴歸模型

模型	模型一： 簡單迴歸模型	模型二： 固定效果模型	模型三： 時間效果模型
TVexogenous			
總統選舉 $i, t$	0.092** (17.310)	0.095** (20.950)	0.109** (23.520)
縣市長選舉 $i, t$	0.070** (14.790)	0.060** (15.010)	0.069** (17.100)
TVendogenous			
民進黨執政縣市 $i, t$	0.072** (17.090)	0.059** (15.680)	0.041** (9.890)
TIexogenous			
桃竹苗 $i$	-0.009 (-1.580)	—	—
中彰投 $i$	-0.010* (-1.830)	—	—
雲嘉南 $i$	0.024** (4.680)	—	—
高高屏 $i$	0.069** (13.280)	—	—
TIendogenous			
閩南鄉鎮 $i$	0.043** (10.240)	—	—
山地鄉鎮 $i$	-0.097** (-19.850)	—	—
2000 年	0.070** (9.500)	—	0.084** (13.370)
2001 年	0.020** (2.690)	—	0.033** (5.300)
2004 年	0.052** (7.060)	—	0.065** (10.420)
Constant	0.242** (51.770)	0.278** (132.410)	0.270** (123.550)
$\bar{R}^2$	0.35	0.46	0.49
$N$	5512	5512	5512

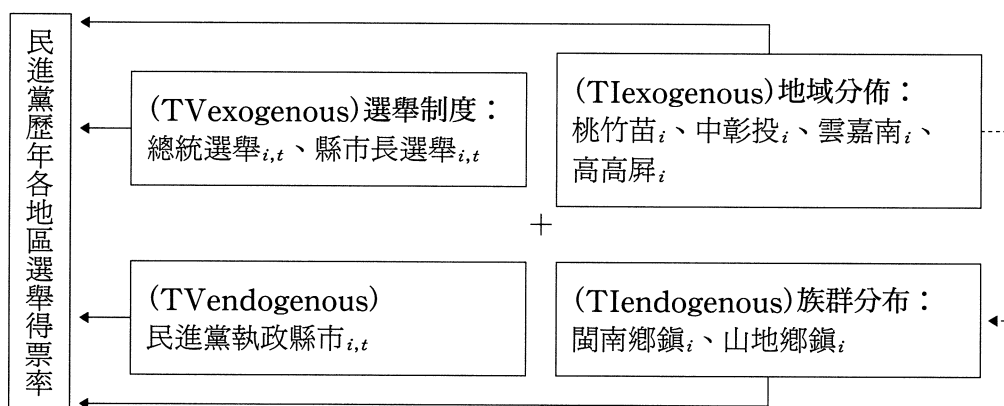
說明：1. 括弧內為  $z$  值。

2. \*\* 與 \* 各表示顯著水準在 5% 與 10% 之下的顯著估計值。

分為「無時間變化的外生變數」，此處我們區分為「桃竹苗」、「中彰投」、「雲嘉南」、「高高屏」四項；最後，「閩南鄉鎮」與「山地鄉鎮」兩項解釋變數同樣不受時間的變化而改變，但是卻是內生形成的，因此將屬於「無時間變化的內生變數」。

表 4 為我們以 H-T model 分析選舉制度、地域與族群對於民進黨得票率影響性的結果，並與先前所提出的時間合併資料模型與固定效果模型進行比較。首先，我們可以很清楚地看出在 H-T model 中，「有時間變化解釋變數」對於「民進黨得票率」的影響性將會與固定效果模型相同，其中包含「總統選舉」、「縣市長選舉」及「民進黨執政縣市」，其迴歸係數分別為 0.109、0.069 以及 0.041。

圖 5：Hausman-Taylor model 實證變數關係圖



換句話說，就固定效果能夠獲得估計結果的部份而言，H-T model 的估計結果是相同地，同樣地與時間合併資料模型的估計結果有所差異，代表此模型亦能夠去除個別效果影響性所造成的偏誤。其次、由「民進黨執政縣市」以及資料結構中最後三個年度的虛擬變數估計結果與時間合併資料模型的估計結果相互比較，我們發現 H-T model 所獲得的估計結果與固定效果模型相同地都是較低的迴歸係數。換句話說，此處再一次地驗證時間合併資料模型就「民進黨執政縣市」與「時間虛擬變數」對於「民進黨得票率」會有過度高估的狀況。

另外，在我們確認 H-T model 同樣可以排除個別效果之後，可以更進一步探討「無時間變化的解釋變數」對於應變數的影響性。就「無時間變化的內生變數」而言，「閩南鄉鎮」與「山地鄉鎮」對於「民進黨得票率」的影響性都較時間合併資料模型為大，也就是說我們發現若單純以時間合併資料結構進行分析，將會過於低估「無時間變化內生變數」對於應變數的影響程度，屬於「閩南鄉鎮」與「山地鄉鎮」的族群分布影響力將會被減弱。探究其原因，我們可以觀察另一項解釋變數：「無時間變化的外生變數」對於應變數的影響力，與時間合併資料模型相比較，我們發現屬於地域性質解釋變數的「雲嘉南」變數，其對於「民進黨得票率」的影響性由時間合併資料模型中的顯著正向轉為 H-T model 中的不顯著負向。

這樣的變化乍看之下似乎並不符合政治版圖理論與實證的直覺，但是若加入「無時間變化內生變數」估計結果進行分析，我們發現「雲嘉南」的地域性質與「閩南鄉鎮」正相關，並與「山地鄉鎮」負相關，因此我們推論在未排除個別效果以及未設定具有不同區域、不同時間性質縱橫資料結構的前提下，時間合併資料模型中對於地域解釋變數（雲嘉南）的估計結果將會受到該觀察值其實亦身為「閩南鄉鎮」族群分布的內生性質影響。也就是說，原本應是族群分布解釋變數（閩南鄉鎮、山地鄉鎮）對於民進黨得票率的影響性，卻在時間合併的資料結構中改以地域性質的解釋變數來詮釋。

另外，由於我們將地區虛擬變數設定為「桃竹苗」、「中彰投」、「雲嘉南」與「高高屏」，因此其中做為基準比較的地區為台北縣市、基隆與宜蘭等北部地區。表 4 的 H-T model 顯示「高高屏」的估計係數為顯著的 0.071，優於其他地區的估計係數，代表「高高屏」等地區對於民進黨得票率的影響性較北部地區為強；而「雲嘉南」地區虛擬變數的估計係數為不顯著的 -0.015，指出「雲嘉南」地區對於民進黨得票率的影響性較「高高屏」為弱，但是因為估計係數不顯著異於 0，代表該地區對於民進黨得票率的影響與北部地區的差異性不大；然而「中彰投」地區的迴歸估計係數為顯著的 -0.049，這說明「中彰投」地區對於民進黨得票率的影響性較北部地區為低，因此也顯著較「雲嘉南」地區為低。整體而言，於 H-T model 的估計中，地區虛擬變數對於民進黨得票率影響性的強弱依序為「高高屏」、「雲嘉南」、「中彰投」以

表 4 民進黨得票率的時間合併資料模型、固定效果模型與 Hausman-Taylor model

模型	模型一： 簡單迴歸模型	模型二： 時間效果模型	模型三： Hausman、Taylor model
TVexogenous			
總統選舉 $i, t$	0.092** (17.310)	0.109** (23.520)	0.109** (23.480)
縣市長選舉 $i, t$	0.070** (14.790)	0.069** (17.100)	0.069** (17.120)
TVendogenous			
民進黨執政縣市 $i, t$	0.072** (17.090)	0.041** (9.890)	0.041** (9.940)
TIexogenous			
桃竹苗 $i$	-0.009 (-1.580)	—	-0.032 (-1.190)
中彰投 $i$	-0.010* (-1.830)	—	-0.049** (-2.560)
雲嘉南 $i$	0.024** (4.680)	—	-0.015 (-0.700)
高高屏 $i$	0.069** (13.280)	—	0.071** (4.390)
TIendogenous			
閩南鄉鎮 $i$	0.043** (10.240)	—	0.068* (1.670)
山地鄉鎮 $i$	-0.097** (-19.850)	—	-0.288** (-5.350)
2000 年	0.070** (9.500)	0.084** (13.370)	0.084** (13.350)
2001 年	0.020** (2.690)	0.033** (5.300)	0.033** (5.290)
2004 年	0.052** (7.060)	0.065** (10.420)	0.065** (10.400)
Constant	0.242** (51.770)	0.270** (123.550)	0.282** (9.850)
$\bar{R}^2$	0.35	0.49	0.52
N	5512	5512	5512

說明：1. 括弧內為  $z$  值。

2. \*\* 與 \* 各表示顯著水準在 5% 與 10% 之下的顯著估計值。

及「桃竹苗」，這樣的強弱關係與時間合併模型的估計結果相似，唯獨於「雲嘉南」的估計係數並不顯著，然而相互之間的強弱關係仍是相同。

換句話說，對於探討選舉制度、地域與族群分佈對於民進黨得票率的實證分析而言，採用具有縱橫資料特性且能夠估計出「無時間變化解釋變數」的 H-T model，將可以獲得既精確又充分運用資料訊息的估計結果。我們因而可得知 H-T model 相對於時間合併資料模型與縱橫資料固定效果模型而言，就分析具有地域與時間性質的各項選舉研究有較大的優勢。

H-T model 在此處所提供的，應有兩項重點：其一為透過對分析樣本進行 regroup 的動作，在 H-T model 的模型設定下，我們可以估計出代表各地區性質的迴歸係數，無論是行政區域或是代表各地區內生性質的閩南鄉鎮與山地鄉鎮。其二，固定效果模型的確能夠透過設定虛擬變數的方式，針對各 cross-sectional units 估計出各地區的解釋能力。然而這畢竟是較為間接的方式，且在 cross-sectional units 過多的情況下，估計過程會較為繁瑣。我們強調此處並不是提出 H-T model 優於固定效果模型，表四只是提供讀者了解 H-T model 的估計結果不僅與固定效果模型能夠估計的結果無異，且又能在一次模型估計過程中，較固定效果模型提供更多的資訊，這是本文介紹 H-T model 的目的。<sup>10</sup>

最後，由於本文所使用的模型為縱橫資料 (panel data) 形式的 H-T model，為提供較為合理的時間面向樣本數，以及提供適度的內生性與外生性解釋變數，本文將三種不同層次的選舉進行合併估計。為解決不同選舉層次受到不同動態環境因素影響的問題，本文即以「總統選舉」與「縣市長選舉」兩項變數進行控制。而為解決「民進黨執政縣市」或與「總統選舉」與「縣市長選舉」有交互作用的問題，本文另以隨機效果模型來估計三項不同選舉制度下的觀察值。<sup>11</sup> 其結果指出民進黨執政縣市的迴歸係數，無論在何項選舉制度下，皆與本文表四的估計結果相當，為顯著的正向影響。若繼續觀察

10 關於評審的意見，認為本文表四的模型二與模型三估計結果相當，並不能直接就認定 Hausman-Taylor model 較 fixed effect model 為佳，我們感謝評審的細心提醒。

11 本文因為文章篇幅所致，並未呈現此處的估計結果，若讀者欲知估計結果，可與作者聯繫。

其他解釋變數，亦可得知大多解釋變數皆與本文表四的估計結果相當。此項分析說明本文所使用的實證模型，雖然對於各項選舉制度進行合併，但其估計結果並未與分開估計的結果相差甚多，而在 H-T model 的分析之下，我們則可以獲得更多原先無法獲取的資訊，例如對於 individual effect 的分析。

## 伍、結論

本文探討族群、地域分佈以及選舉制度三項縱橫資料型態的變數對於歷年間各地區民進黨得票率的影響性。試圖分析在下列三項不同的實證模型：時間合併資料模型、縱橫資料固定效果模型與 H-T model，其模型設定以及解釋變數對於應變數之間解釋能力的差異性，並首次提出為何 H-T model 對於具有地域與時間性質的各項選舉研究分析有較大的優勢。

首先，我們比較時間合併資料模型與縱橫資料固定效果模型中選舉制度、族群以及地域分佈等解釋變數對於歷年間各地區民進黨得票率的影響性。我們發現「民進黨執政縣市」對於「民進黨得票率」的影響性於時間合併資料模型中會有過於高估的狀況，然而縱橫資料模型將會以固定效果方式去除「個別效果」的影響，在不受到「個別效果」的干擾之下，縱橫資料固定效果模型將會消除原本過於高估之「民進黨執政縣市」對於「民進黨得票率」的影響效果。有鑑於此，本文認為就分析台灣各地區民進黨得票率的影響性而言，採取縱橫資料的模型估計無疑地將會較以時間合併資料模型估計更為精確。

其次，當我們確認採用縱橫資料固定效果模型的估計較為精確後，若採用固定效果模型探討各項解釋變數對於台灣各地區民進黨得票率的影響性，卻因為在估計過程中減去各別個體層次（區域）依循時間的平均數，因此我們無法以固定效果模型對於不隨時間變動的解釋變數進行估計的缺點。此處我們首次提出 Hausman and Taylor (1981: 1377-1398) 的 H-T model 來解決此問題，H-T model 同樣地為縱橫資料的模型，並且可以排除個別效果所帶來的估計偏誤，但是卻能解決縱橫資料固定效果模型的缺點。

就無時間變化解釋變數對於應變數的影響性而言，我們於實證分析中發

現若單純以時間合併資料結構進行分析，屬於「閩南鄉鎮」與「山地鄉鎮」的族群分佈影響力將被減弱，也就是說相對於 H-T model，時間合併資料模型過於低估「無時間變化內生變數」對於應變數的影響程度。另外，時間合併資料模型中對於地域解釋變數（雲嘉南）的估計結果將會受到該觀察值其實亦身為「閩南鄉鎮」族群分布的內生性質影響。也就是說，原本應是族群分佈解釋變數（閩南鄉鎮、山地鄉鎮）對於民進黨得票率的影響性，卻在時間合併資料的資料結構中改以地域性質的解釋變數來詮釋。

綜合言之，對探討地域、族群分佈與選舉制度對於民進黨得票率的影響性來說，採用具有縱橫資料特性的模型，相對於時間合併資料模型可以獲得精確的估計結果；而採用 H-T model 將可以獲得既精確又充分運用資料訊息的估計結果，其解決了固定效果模型的缺點以及時間合併資料模型會產生的估計偏誤。實證分析上，採用 H-T model 後，模型估計結果將避免過於高估「民進黨執政縣市」對於「民進黨得票率」的影響效果；亦避免過於低估族群分佈與選舉制度對於歷年間各地區民進黨得票率的影響性。換句話說，對於未來台灣選舉研究的分析，本文認為以縱橫資料的形式進行分析應是較為適合的方式。

## 參考資料

### A. 中文部分

王甫昌

1993 〈八零年代反對運動之族群動員〉，見吳密察、張炎憲（編），《建立台灣的國民國家》，頁 97-102。台北：前衛出版社。

2000 〈台灣族群關係的社會基礎調查計畫執行報告〉，行政院國家科學委員會委託。

王鼎銘

2003 〈政策認同下的投票效用與選擇：空間投票理論在不同選舉制度間的比較〉，《選舉研究》10(1): 171-206。

王鼎銘、蘇俊斌、黃紀、郭銘峰

2004 〈日本自民黨之選票穩定度研究：1993、1996 及 2000 年眾議院選舉之定群追蹤分析〉，《選舉研究》11(2): 81-110。

吳乃德

- 1993 〈族群意識、政治支持和國家認同——台灣族群政治理論的初探〉，見張茂桂（編），《族群關係與國家認同》，頁 27-51。台北：業強出版社。
- 2005 〈麵包與愛情：初探臺灣民眾民族認同的變動〉，《臺灣政治學刊》9(2): 5-39。

吳重禮、李世宏

- 2005 〈政治賦權、族群團體與政治參與：2001 年縣市長選舉客家族群的政治信任與投票參與〉，《選舉研究》12(1): 69-115。

林繼文

- 2002 〈選舉制度、選民偏好與政黨體系的分化：東亞三國的比較〉，《選舉研究》9(1): 131-171。

胡 佛

- 1998 《政治學的科學探究(三)：政治參與與選舉行爲》。台北：三民出版社。

洪永泰

- 1994 〈選舉預測：一個以整體資料為輔助工具的模型〉，《選舉研究》1(1): 93-110。

葛永光

- 1996 《政黨政治與民主發展》。台北：國立空中大學。

徐火炎

- 1993 〈選民的政黨政治價值取向、政黨認同與黨派投票抉擇：第二屆國大代表選舉選民的投票行爲分析〉，《國家科學委員會研究彙刊：人文及社會科學》3(2): 144-166。

徐永明

- 2001a 〈政治版圖：兩個選舉行爲研究途徑的對話〉，《問題與研究》40(2): 95-115。
- 2001b 〈九四、九八台灣雙城記：不確定中大眾選擇的分流點〉，《東吳政治學報》12: 75-112。

徐永明、林昌平

- 2005 〈如何以縱橫資料來測量政治版圖的變化：以台灣立委選舉為分析對象〉。發表於「台灣經驗政治研究的現況與展望」學術研討會，中央研究院人文社會科學研究中心主辦，台北：南港，12月23日。

徐永明、陳明通

- 1998 〈搜尋台灣民眾統獨態度的動力：一個個體模型的建立〉，《台灣政治學刊》3: 65-114。

耿 曙、陳陸輝

- 2003 〈兩岸經貿互動與台灣政治版圖：南北區塊差異的推手？〉，《問題與研究》42(6): 1-27。

陳文俊

- 2003 〈藍與綠——台灣選民的政治意識型態初探〉，《選舉研究》10(1): 41-80。

張茂桂

- 1993 〈省籍問題與民族主義〉，見張茂桂（編），《族群關係與國家認同》，頁 233-278。台北：業強出版社。

盛杏媛

- 2002 〈統獨議題與台灣選民的投票行爲：一九九〇年代的分析〉，《選舉研究》9(1): 41-80。



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。

賴進貴、葉高華、張智昌

2007 〈投票行為之空間觀點與空間分析——以臺灣 2004 年總統選舉為例〉，《選舉研究》14(1): 33-60。

謝復生

1992 《政黨比例代表制》。台北：理論與政策雜誌社。

國立政治大學選舉研究中心

2006 〈臺灣選舉類別與結果線上流覽〉。2006 年 7 月 6 日，取自 <http://esc.nccu.edu.tw/newchinese/main.htm>

## B. 英文部分

Baltagi, B. H.

2002 *Econometric Analysis of Panel Data*. New York: Wiley Press.

Baltagi, B. H., G. Bresson, and A. Pirotte

2003 “Fixed Effects, Random Effects or Hausman-Taylor? A Pretest Estimator,” *Economics Letters* 79(3): 361-369.

Carmines, E. G. and J. A. Stimson

1990 *Issue Evolution: Race and the Transformation of American Politics*. NJ: Princeton University Press.

Cho, W. K. T.

2003 “Contagion Effects and Ethnic Contribution Networks,” *American Journal of Political Science* 47(2): 368-387.

Cox, G.

1997 *Making Votes Count: Strategic Coordinate in the World's Electoral Systems*. New York: Cambridge University Press.

Green, D. P. and S. Y. Kim

2001 “Dirty Pool,” *International Organization* 55: 441-468.

Hausman, B. J. and W. E. Taylor

1981 “Panel Data and Unobservable Individual Effects,” *Econometrica* 49: 1377-1398.

Hsu, Y. M. and P. S. Lee

2002 “Southern Politics? Regional Trajectories of Party Development,” *Issues and Studies* 39(2): 61-87.

Kim, J., E. Elliott, and D. M. Wang

2003 “A Spatial Analysis of County-level Outcomes in US Presidential Elections: 1988-2000,” *Electoral Studies* 22: 741-761.

Mundalk, Y.

1978 “On the Pooling of Time Series and Cross-section Data,” *Econometrica* 46: 69-85.

O’Loughlin, J., C. Flint, and L. Anselin

1994 “The Geography of the Nazi Vote: Context, Confession, and Class in the

Reichstag Election of 1993," *Annals of the Association of American Geographers* 84(3): 351-380.

Ranny, Austin

1996 *Governing: An Introduction to Political Science*. Upper Saddle River, NJ: Prentice-Hall Press.

Stata

2003 *Cross-Sectional Time-Series Reference Manual Release 8.0*. Texas: Statacorp Press.

# **The Estimation of Taiwan's Political Geography in the Panel Data: The Effects of Region, Electoral System and Ethnicity on the DPP's Vote —from 1986 to 2004**

Yung-ming Hsu

Assistant Professor

Department of Political Science, Soochow University

Chang-ping Lin

Ph. D. Student

Department of Money and Banking, National Chengchi University

## **ABSTRACT**

This paper aims to investigate the effects of region, ethnic group and electoral system on the Democratic Progressive Party's (DPP) vote across elections by using aggregate panel data. Comparing the pool data model to the panel data fixed effect model based on Hausman and Taylor (1981), we find that the pool data model overestimates (underestimates) the effects of the governing counties (ethnic groups) on the DPP's votes. However, the panel data fixed effect model does not. Since the traditional fixed effect model cannot reveal estimated information on individual effect, we present the Hausman and Taylor model, which can estimate individual effect properly and specify the impact of region. We can conclude that the result of the Hausman and Taylor model that avoids estimation bias by specifying the individual region is not just superior to other model results, but prove to be a proper model for further studies on geographical correlation on voting behavior.

**Key Words:** locality, ethnic groups, panel data, fixed effect,  
individual effect