

# 統計方法與理論驗證的謬誤相關： 以政治版圖變遷的研究為例， 1989-2004<sup>\*</sup>

徐永明<sup>\*\*</sup> 林昌平<sup>\*\*\*</sup>

- 一、前言
- 二、時間序列總體資料作為一個研究設計
- 三、時間序列總體資料在政治版圖研究的應用
- 四、固定效果模型與 Differences in Differences 模型的比較
- 五、兩種固定效果的影響性—地區與時間
- 六、結論

本文主旨在探討錯誤統計方法的使用將誤導實證結果與理論推演，因此在驗證相關理論之前，首先判別是否採用適當的統計分析方法是更重要的。我們以耿曙、陳陸輝（2003）的政治版圖研究進行 replication，說明在探討政治版圖『變動』時，時間序列總體資料的分析架構恰好可以同時探討時間性結構與地區性結構，認為未加入時間性結構於分析模型是將政治版圖分析中

---

\* 作者特別感謝國科會計畫編號 NSC-94-2414-H-001-012 經費的支持，使本研究得以順利完成。

\*\* 東吳大學政治系助理教授。E-mail: polymh@scu.edu.tw

\*\*\* 國立政治大學金融研究所博士生。E-mail: 94352503@nccu.edu.tw

投稿日期：二〇〇八年十月二十六日；接受刊登日期：二〇〇八年三月二十三日。

東吳政治學報/2008/第二十六卷第二期/頁 83-115。

重要影響因素拋棄不看，這樣的分析結果將使人懷疑其研究結論是否真正適切解釋影響政治版圖的決定性因素。

本文在與 Trubowitz (1996) 的理論比較之下，發現以時間序列總體資料的資料結構進行實證分析，可以將固定效果中的地區效果以及時間效果進行區分，而在排除內生性影響的狀況下，才可以完整的呈現出各種外生變數對於政治版圖的相關性影響。

在實證結論上，耿、陳認為產業結構的解釋變數對於選票分佈有顯著性的影響，但本文指出地區性差異將同時影響產業結構與選票分佈，因此在認為產業結構是影響選票分佈的主要影響變數背後，其實存在第三項影響變數（地區性差異）同時影響著原先所觀察的解釋變數與被解釋變數。當我們以時間序列總體資料的固定效果模型控制住實證資料的地區固定效果後，產業結構並未對於選票分佈造成顯著地影響。此時地區性差異才是影響台灣政治版圖變遷的主要原因，而非 Trubowitz 理論所提到的產業結構，這代表著以不適當的統計方法進行政治版圖的分析，將無法真正區別出主要的影響因素，這也是本文強調使用正確統計方法來驗證理論最主要的原因。

關鍵詞：政治版圖、時間序列總體資料、固定效果、時間效果

## 一、前言

在政治學的研究當中，探討政治版圖的劃分是一門相當新穎且重要的課題，因為在早期常見的政治學研究資料中，其型態多為總體加總資料的橫斷面資料，故多著重於各別年度區分下探討不同地區政治版圖受到不同變數的影響性，並進一步描述不同年度各政治版圖的趨勢比較。而根據台灣的政治研究，在洪永泰（1994）之後出現了政治版圖的劃分，其文中對於政治版圖的界定主要是依據某地區長期以來是否有一政黨獲得穩定的得票率，若該地區肯定被某一政黨長期獲得穩定的得票率，可劃分為該政黨的政治版圖，也就是因為該政黨於此地區獲得穩定的支持，因此該地區將會有一群選民長時間以來為此政黨的選民，如此才會產生該政黨於該地區穩定的得票率（徐永明，2001）。

故政治版圖可視為各政黨長時間得票率的穩定度，並進一步做為研判未表態選民投票的依據。換句話說，政治版圖作為一個測量的概念，可用以判定一個政黨在某地區的影響力。洪永泰（1994）的分析方法主要是整合選區內歷次選舉投票所的地理範圍及投票紀錄，挑選政治指標將投票依據該指標進行分群，每一種集群結果在進行區位函數判讀重新判別每一單位所屬的集群，其中錯判率最低的區位即是所推估出的政治版圖。舉例而言，即使某一集群每次選舉的得票率均為 60%以上，並不代表每一次選舉投票給該政黨的選民是相同的該 60%，只能夠說是具有相同的比例，因此洪永泰（1994）的政治版圖僅能夠探討各地區得票的穩定程度，徐永明（2001）則強化政治版圖的概念至忠誠選民的面向，探討某地區上一次選舉投

給 A 政黨的選民，於下一次選舉中再度投給 A 政黨的比例為多少，因此徐永明（2001）對於政治版圖的區分，首先採取 King（1997）的 EI 模型（Ecological Inference Model）估算每一個鄉鎮市區的忠誠選民比例，並進一步分析各個選舉階段忠誠選民的比例是否穩定？

其中徐永明（2001）提出政治版圖的分析可區分為內生性影響與相關性影響兩種，內生性影響強調忠誠選民的比例在群集變化上與投票率一致，反之相關性影響指出忠誠選民的比例在群集變化上與投票率並不一致。以耿曙、陳陸輝（2003）為例，其依循 Hsu and Lee（2002）的概念，以產業變數試圖探討台灣政治的南北差異，文內以台灣總體的橫斷面資料，依年度進行區分，分別針對不同年度下民進黨得票比例以及不同產業類別的趨勢進行對照分析，探討在經濟區塊的區分之下，不同產業結構對於兩岸型態議題的政治立場是否有所差異。然而由於此種分析以橫斷面的資料型態為研究基礎，忽略了時間面向對於研究議題的影響性，此時若能改採用時間序列總體資料（time series aggregate data）的分析結構進行了解，或可發現不同的結果。<sup>1</sup>

本文主旨即在探討當採用不同的統計方法針對同一筆資料進行比較分析，若獲得不相同的結果，是否代表以不恰當的統計方法進行經驗分析，該實證結果對於理論的驗證將會導致不正確的推論。也就是說，錯誤統計方法的使用將誤導實證結果與理論推演，因此在分析政治理論之前，首先判別是否採用適當的統計分析方法是更為重要的。文獻上 Green et al.（2001）曾比較固定效果模型與時間

---

1. 本文中的時間序列總體資料（time series aggregate 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）。

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，在一般傳統政治版圖的模型分析中，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。

我們進一步發現上述分析方式多為對政治版圖進行描述與劃分，主要是處理政治版圖的「穩定度」與「流動性」的問題，未對於各區域間的差異以及資料本身的「時間性」加以探討。徐永明（2001）雖採用 EI 模型分析不同區域的忠誠選民比例是否穩定，但仍未考慮時間上的變化性。故就政治版圖的相關研究而言，多以橫斷面資料為主，分析在地區的區分下各政黨得票率的穩定度，並將對於政治版圖的探討集中在分類（群集分析）與流動方向（ecological analysis）的分析上，忽略了加入時間因素探討政治版圖各地區依循時間本身的變化為何，特別是在影響政治版圖變化的因素區分中。因此過去探討政治版圖的傳統研究，皆無法將時間性的變化考量在內，無論是運用 King（1997）的方法，還是耿曙與陳陸輝（2003）的 difference in difference 方法，最多只能做到兩個時間點之間變化的分析，在台灣甚少使用本文所採用的時間序列總體資料模型進行政治版圖的分析，然而在探討政治版圖『變動』時，時間結構的討論變得相當重要，而時間序列總體資料模型的分析架構恰好可以滿足我們於政治版圖分析中同時探討時間性結構與地區性結構的需求。

顧名思義政治版圖研究是一種區域性的研究，參照 Manski

(1995)的相關看法，其認為選民的投票行為有兩個解釋途徑，其一為「內生性影響」(endogenous effects)，認為個人在群體中的行為傾向會受到該行為在群體中的趨向所影響；其二為「相關性影響」(correlated effects)，認為個人在群體中的行為相似，是由於具有類似的個人特質，在一定制度環境下會產生相似的行為。此兩種不相同的途徑提供了選民在群體中面臨投票決策的兩種解釋，前者為群體特性對於個人行為的影響；後者為個人特質決定集體的性質，進而發展出相似的行為模式。大多數根據民調資料所發展出來的政治理論討論，基本上是根源於「相關性影響」，認為個人特質決定其行為模式；而運用集體資料(如投票率、人口統計)的研究中，其理論討論的假定則強調「內生性影響」，認為群體特性或是環境因素是影響選民政治行為的主要原因。

然而運用兩種途徑而強調不同理論假定所進行的研究，其實有許多都是因為資料取得的限制而影響理論的選擇，並非基於理論的假定再進一步收集資料的研究過程。根據上述討論，在洪永泰(1994)的政治版圖中，其隱含的假定為環境決定選民投票(內生性影響)，因此在這些地區所組成的群集中，包含了對於民進黨友善與不友善的地區，這些地區的政治特性會影響其中的選民投票行為。換句話說，身處於對於民進黨友善的地區，自然就較傾向於投票給民進黨；反之，若身處於民進黨得票率低的地區，由於該地區的政治氣候並不利於民進黨，因此選民也較不傾向投票給民進黨。也因為政治版圖的研究同時涵蓋了上述兩種影響關係，就地區上的特殊變項而言，其在各地區上有差異但是長期不發生改變的影響，我們可以稱之為「內生性影響」；而對於自變項的影響，也就是探討個人特質決定其行為模式(如個人的職業、學歷)，我們可以稱之為「相關

性影響」。

故在與過去政治版圖相關研究的不同上，本文嘗試以時間序列總體資料的結構，運用固定效果（fixed effect）的估計方式並配合加入時間變項的探討，希望能同時探討過去所未能同時探討的「內生性影響」與「相關性影響」。並且進一步以固定效果中的地區效果（group effect）來控制「內生性影響」，將模型中的「內生性影響」與「相關性影響」進行區分，如此將可以單純地觀察兩者分別對於模型中應變數（如得票率）的影響關係。最後本文並以時間效果（time effect）區分出各選舉狀態間的整體變動，說明 Manski(1995)另外指出的 contextual effect，並期望更進一步解釋政治版圖中的政黨穩定度。

## 二、時間序列總體資料作為一個研究設計

時間序列總體資料是一種資料型態的表現，其中將會有N個不同的觀察值，而每個觀察值都有相對應的不同T段時間。<sup>2</sup>相對應於描述橫斷面資料時我們慣常所使用的i下標符號，於時間序列總體資料的資料型態中，我們將以i,t做為下標符號，因為我們必須同時地表示出不同地觀察值個體與不同地時間，若以符號表示將如下式（1）所示：

---

2. 其中一位評審意見提到，不同的多變量迴歸方法皆可能以不同的角度來看待研究議題，本文作者肯定此種說法。因此於文中強調，針對所需驗證的理論，統計方法的選擇有其重要性存在，作者並認為需針對所採用的「資料型態」來判斷統計方法的選擇。以本文所舉的例子來說，作者認為由於政治版圖的資料型態同時存在橫斷面資料的「地區性變數」與時間序列資料的「時間性變數」，故應該依據此種資料型態的特性，選擇適當的統計方法進行分析，才不至於產生推論上的謬誤，此時「時間序列總體資料模型」的運用即是恰當的選擇。

$(X_{i,t}, Y_{i,t}), i = 1, K, N \text{ and } t = 1, K, T \dots\dots(\text{時間序列總體資料結構}) (1)$

一般的橫斷面資料分析多會存在某些潛藏「省略變數」(omitted variable)的誤差。所謂省略變數指出,有些變數與應變數存在相關性,對於應變數的變化亦有影響性,然而這些變數或是因為較為不易收集,或是因為無法測量該變數,我們無法將所有與應變數有關的變數放入模型中,故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。因為無法處理這些誤差的影響,所以政治版圖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在性因素(exogenous variable)。然而時間序列總體資料的分析方式卻可以明顯的矯正該誤差項,我們以時間序列總體資料的型態做為一個研究設計,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响關係。

因此我們將時間序列總體資料此種研究設計的分析方式應用於政治版圖的測量當中,發現時間序列總體資料特殊的資料型態將與政治版圖有著相當適切的合適性,在政治版圖的資料當中,同樣地強調不同地區以及不同時間的特性,故當我們應用時間序列總體資料的分析方式於政治版圖的測量時,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。<sup>3</sup>

---

3. 其中一位評審提到,不同地區各自的特殊因素應對於政黨得票率有不同的影響。而地區上的差異性(無論是文化特色、歷史特質或是宗教傳統與社會結構)正是本文欲強調對於台灣的政黨得票率真正有影響的因素。基於地區差異性變數衡量上的困難(例如前述各地區的文化特質),過去的研究未能有效的探討地區性差異對於政黨得票率的影響力。本文使用時間序列總體資料分析結構的資料型態恰能將政治版圖強調不同地區與不同時間的特性完整測量,故當我們應用時間序列總體資料的分析方式於政治版圖的測量時,將可以消除「省略

基於上述探討，我們將分別就時間序列總體資料分析方式中不同的固定效果來探討政治版圖中的「內生性影響」與「相關性影響」。首先、就內生性影響來說，我們考慮採用固定效果中的地區效果來控制內生性的影響結果，雖然我們並不測量內生性影響，但是由於採用固定效果進行對於內生性影響的控制之後，實證模型將可以消除估計上若未考量內生性影響所帶來的誤差。主要特色在於採用地區效果的估計方式可以使我們區分出雖然在地區上有所差異，但是各地區於不同時間下不會有所改變的影響性，這也就是「內生性影響」，如此我們將可以確定所估計出的迴歸係數已經排除了「省略變數」所造成的誤差，並且考慮到不同地區差異所造成的影響性。換句話說，當我們採用固定效果中的地區效果來進行模型估計時，我們可以很明確地排除政治版圖中的內生性影響。

其次、對於政治版圖中的另一項特色：「相關性影響」，在我們採用固定效果中的地區效果估計方式將模型中的「省略變數」誤差以及「內生性影響」所帶來的估計偏誤消除之後，我們即可以運用各種相關的個人特質變數（如職業、學歷）來探討個人特質對於其行為模式的影響性，此時最重要的即是由於採用時間序列總體資料此種研究設計方式，我們可以明確地找出各模型中的「相關性影響」，而不會再受到「內生性影響」與「省略變數」的干擾。

就內生性與相關性影響而言，本文認為對於政治版圖的影響因

---

變數」（即前述所提文化特色、社會結構等衡量困難的變數）對於模型估計所造成的影響。換句話說，本文雖然未直接估計各地區特色上的差異性對於得票率的影響性，但是透過時間序列總體資料的分析方法，我們可以得到對於台灣政黨得票率的影響性主要來自於地區的差異性。

素可以依據解釋變數性質的不同，區分為「內生性影響」與「相關性影響」兩種不同的類型。其中「內生性影響」的解釋變數探討劃分為在各地區有差異但是長期並不改變的影響因素，而「相關性影響」的解釋變數探討劃分為個人特質影響行為模式等影響因素。換句話說，對於政治版圖的研究應同時探討這兩項影響關係，主要原因在於分析影響政治版圖的影響要素時，必須涵蓋所有影響因素，此時並不會形成邏輯相衝突的可能性，因為「內生性影響」與「相關性影響」並未互相交集。此時以時間序列總體資料的分析模式，將可以很恰當地以固定效果中的地區效果控制住「內生性影響」，代表固定效果模型可以完全區分出「內生性影響」進而估計「相關性影響」的效果，本文認為以時間序列總體資料中的固定效果模型探討「相關性影響」才可以單純的獲得不受「內生性影響」干擾的影響效果，因此以時間序列總體資料同時考量「內生性影響」與「相關性影響」是必要的。

另外由於固定效果中的時間效果可以考慮到在不同時間點上有明顯的差異，但是即使是不同的地區相同年度之間卻沒有顯著的差異，因此我們可以針對政治版圖中的時間序列總體資料來觀察個別年度間的差別性，卻不需要擔心是否會受到不同地區中「內生性影響」的干擾。如此的模型設定可以突顯政治版圖研究中時間面項變數的影響性，並且排除因為不同地區的「內生性影響」干擾，因此我們將同時以時間效果的估計方式來決定政治版圖研究中的時間面項變數的設定。換個角度言，我們可以突顯出 Manski (1995) 所指出的 contextual effect，所謂 contextual effect 指出個人的決定將會與團體中過去的背景有關，也就是說，個人行為的決定將會受到團體過去行為的影響，在此處我們將以時間效果的估計方式描繪去團體

過去行為的影響性，並成功的說明政治版圖中 contextual effect 對於其行為的影響能力。最後，做為一個比較的基礎，我們將以耿曙、陳陸輝（2003）探討政治版圖的模型為例，比較在考量了地區效果與時間效果之下，採用時間序列總體資料中的固定效果估計方式對於政治版圖進行刻劃，對於政治版圖的詮釋是否有所不同。

### 三、時間序列總體資料在政治版圖研究的應用<sup>4</sup>

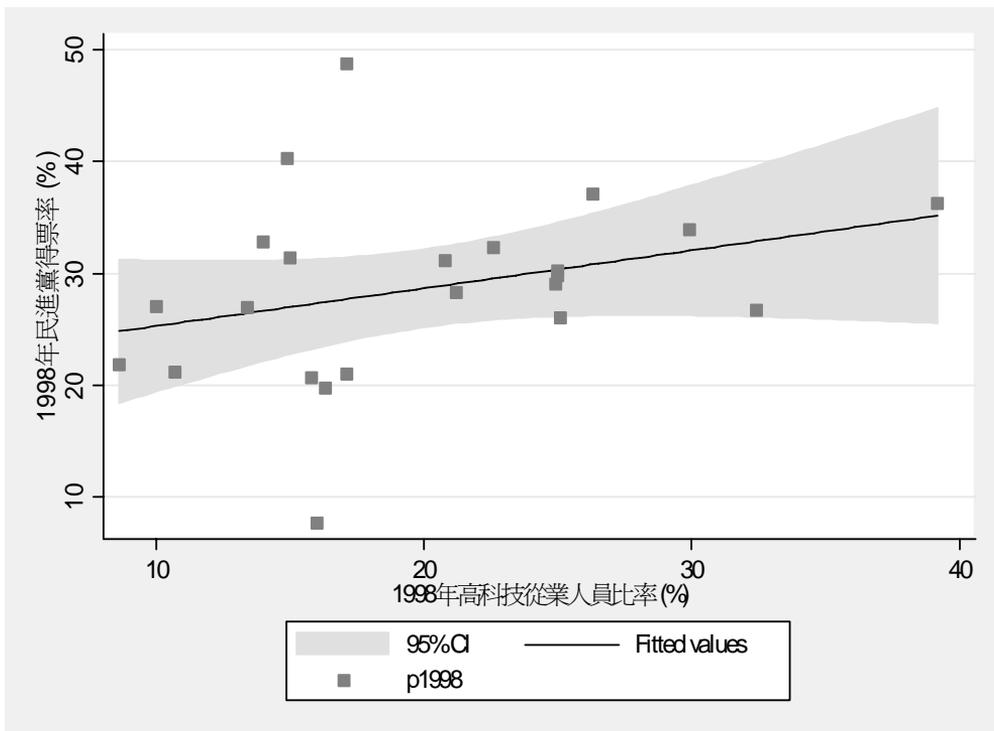
若依據耿曙、陳陸輝（2003）探討政治版圖的模型為例，<sup>5</sup>該模型中以 1998 年與 2001 年兩個年度中民進黨於立委選舉的得票率進行探討，所採用的就是一般分析所使用的橫斷面分析結構，而針對時間面項所產生的問題則以整體 1998 年與 2001 年兩個年度的差分項進行探討。其中若只論產業結構中「高科技比例」對於民進黨於立委選舉的得票率而言，此兩變項（高科技從業人員比率—得票率）之間的關係於兩年度中有明顯的差異。在 1998 年度中高科技從業人員與民進黨立委得票率的關係呈現正相關（見圖一），這個個別年度的結果顯然與目前的政治觀感有所差異。而 2001 年度在圖二中我

---

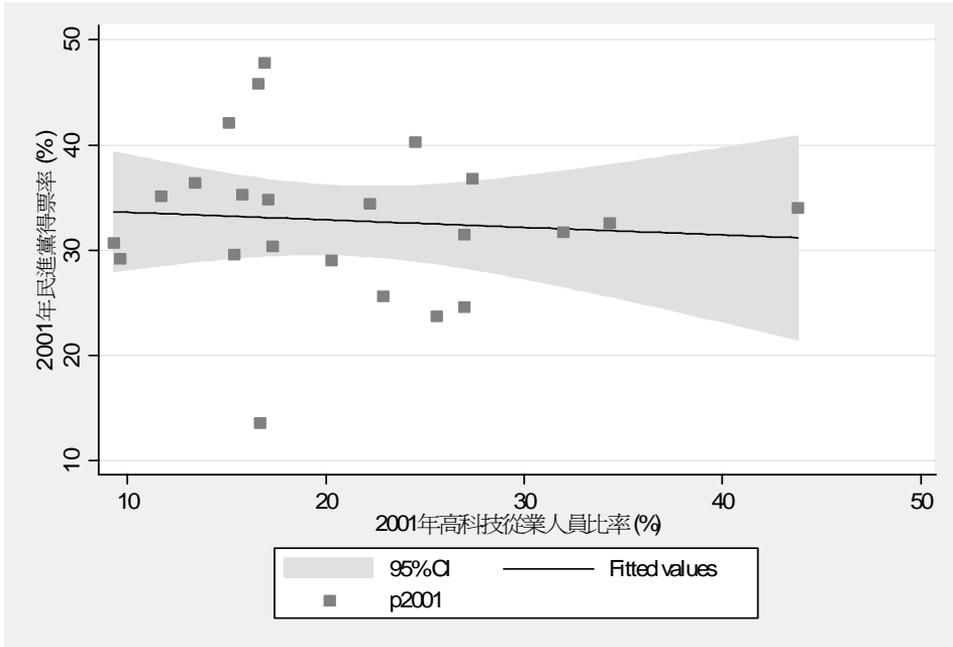
4. 本篇文章所欲強調的重點在於若使用了不恰當的分析方式，其實證結果容易造成該項研究探討的推論謬誤，故針對相同的研究議題，若採用不同的統計方法針對同一筆資料進行比較分析，若獲得不相同的結果，是否代表以不恰當的統計方法進行經驗分析，該實證結果對於理論的驗證將會導致不正確的推論。也就是說，錯誤統計方法的使用將誤導實證結果與理論推演，基於此項因素，本文作者所欲強調為統計測量方法選擇的重要性。此處提到政治版圖變遷分析，是為了強調選擇統計測量方法重要性所舉出的例子，因此本文重點在提出「時間序列總體資料」的分析結構，認為以此項分析結構探討政治版圖的變遷應是較為適當的方式。

5. King（1995）提出將已發表文章的議題進行重新審視的研究方法，本文中我們以不同的資料結構與實證模型進行相同主題的驗證，試圖找出影響政治版圖的正確原因。

們卻可以發現其結果是關係趨近於 0 的非正向影響關係，因此若以個別年度來觀察，將無法看出產業結構與民進黨立委得票率之間的相關性，換句話說，我們可以看出若只以橫斷面資料作為分析對象的侷限性。



圖一 1998年高科技從業人員比率與民進黨在各縣市得票率關係圖

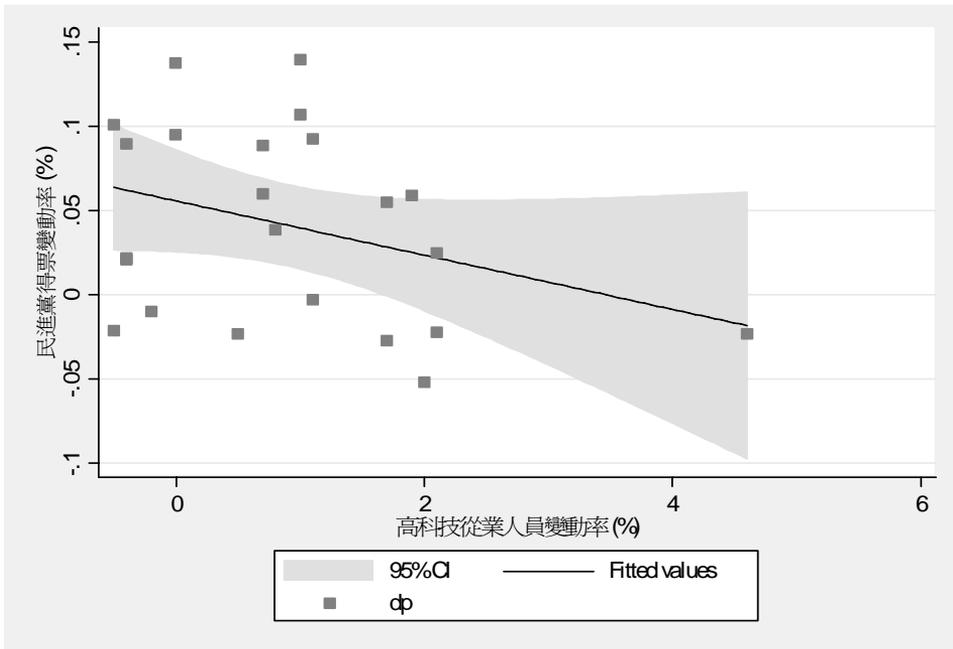


圖二 2001年高科技從業人員比率與民進黨在各縣市得票率關係圖

因此在耿曙、陳陸輝（2003）模型的研究，該模型取高科技從業人員與民進黨得票率在兩年度之間的差分項來分析，試圖探討兩年度間時間的影響性，因為若將 1998 年與 2001 年兩個年度的資料分開進行討論，則可發現兩個年度中「高科技比例」對於民進黨於立委選舉得票率的影響性其實是不一致的。因此雖然該文中發現高科技從業人員的變動比率與民進黨得票變動率之間的影響性為負向關係（如圖三所示），但因為此一結果未能說明在不同年度間的差異表現，故該結果是否明確，將值得商榷。也就是說雖然結果呈現出當高科技從業人員增加的比率越高，民進黨的得票率變動就越低，但究竟代表著兩年度之間民進黨在立委選舉的得票率趨向穩定或是減少呢？而這樣的實證結果又與單一年度中橫斷面分析的估計

結果有所不同，因此時間因素的重要影響性必須在實證研究中受到重視。

換句話說，本文重做耿曙與陳陸輝的研究，主要是認為在未加入時間性結構於分析模型時，等於將政治版圖分析中相當重要的決定因素拋棄不看，這樣的分析結果將使人懷疑其研究結論是否真正適切解釋影響政治版圖的決定性因素。因此本文嘗試以時間序列總體資料的分析模型來探討過去耿曙與陳陸輝同樣關心的議題（產業結構與統獨態度），試圖分析在正確加入時間性結構於政治版圖的分析架構後，區域經濟影響政治版圖的結論是否仍然成立。



圖三 1998年至2001年高科技從業人員變動率與民進黨在各縣市得票變動率關係圖

依據前述，如何適當的將時間所隱含的解釋力表現出來，避免前述圖一與圖二中影響性改變的狀況，應是具備時間序列分析分法與橫斷面分析方法優點的時間序列總體資料分析方法所能夠達到的。在我們的實證分析中，我們依據耿曙、陳陸輝（2003）的模型進行比較分析，其中所使用的實證資料與該文中的實證資料相同，分別為民進黨於立委選舉的得票率以及台灣各縣市各項產業結構的相關比例，包括「高科技」、「工商服務」、「農林漁牧」以及「自營小業主」等四項。<sup>6</sup>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一所示，其中民進黨得票率的平均數約為三成左右，另外就解釋變數而言，「自營小業主」所佔的比率最高，平均數為22.17%，比例最低的解釋變數為「農林漁牧」，其平均數只有10.65%。本文採取與耿曙、陳陸輝（2003）相同實證資料的主要目的在於可以控制「資料」這個變項，來凸顯方法與實證模型對於理論驗證的顯著影響。換句話說，本文以產業結構比例測量區域經濟、以民進黨得票率測量政治版圖的主要目的在於控制不同實證資料的影響差異，如此可以比較分析正確與適當的分析方法對於理論驗證的影響效果。

---

6. 本文實證資料為年資料，資料期間分別取自1998年至2001年以及1989年至2004年中台灣23個縣市的年資料。

表一 民進黨得票率與解釋變數的敘述統計

變數名稱	樣本數	平均數	標準差
民進黨得票率	46	30.754	8.053
高科技	46	20.615	8.250
工商服務	46	11.148	4.874
農林漁牧	46	10.650	9.613
自營小業主	46	22.174	4.404

資料來源：行政院主計處（2008）；國立政治大學選舉研究中心（2008）。

首先我們以 1998 年至 2001 年的資料為例，與耿曙、陳陸輝（2003）的實證模型進行比較，觀察是否存在時間序列總體資料中的固定效果，此時總資料筆數為 46 筆。其次，我們進一步拉長時間，以 1989 年至 2004 年整體台灣民主化後的立委選舉歷程作為分析的對象，總資料筆數為 138 筆，<sup>7</sup>在拉長時間之下，似乎更適合採用時間序列總體資料的資料分析模式，其中我們可以看到各年度間的變化以及去除固定效果之後產業結構對於民進黨立委得票率的影響。本實證研究的目的是觀察各項產業結構對於兩岸態度政治立場的影響性是否應以時間序列總體資料的資料分析方式進行研究較佳。另外就模型中的應變數而言，我們採用各縣市當年度民進黨於立委選舉中所獲得的得票率來做為分析的對象。

7. 其中分析資料只採用有立委選舉的六個年度，分別為 1989 年、1992 年、1995 年、1998 年、2001 年以及 2004 年。

## 四、固定效果模型與 Differences in Differences 模型的比較

首先本文在時間序列總體資料的型態之下，實證分析的模型結構如下列式（2）所示，其中 $Y_{i,t}$ 為各縣市當年度民進黨於立委選舉中所獲得的得票率：

$$Y_{i,t} = \beta_0 + \mathbf{X}_{i,t}\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma} + u_{i,t} \dots\dots\dots(\text{時間序列總體資料模型}) \quad (2)$$

而 $\mathbf{X}_{i,t}$ 與 $\mathbf{Z}_i$ 皆為向量的形式，其中 $\mathbf{X}_{i,t}$ 包含了「工商服務」、「高科技」、「農林漁牧」、「自營小業主」等四項自變數，如下列式（3）所示，而 $\mathbf{Z}_i$ 則包含一切省略變數的影響。上述資料結構模型的設定重點在於我們採用了固定效果中之地區效果來進行估計，以明確估計出政治版圖中的內生性特色，目的是在時間序列總體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 $\boldsymbol{\beta}$ 與 $\boldsymbol{\gamma}$ 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：

$$\mathbf{X}_{i,t} = [\text{工商服務, 高科技, 農林漁牧, 自營小業主}] \quad (3)$$

依循著 Hsiao (2003) 所提出的估計方式，我們可以進一步簡化式（2）為下列式（4），其中 $v_i$ 代表我們將所有省略變數所帶來的誤差影響（ $\mathbf{Z}_i\boldsymbol{\gamma}$ ）簡化：

$$Y_{i,t} = \beta_0 + \mathbf{X}_{i,t}\boldsymbol{\beta} + v_i + u_{i,t} \quad (4)$$

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（5），此處的目的在於以「deviation from group mean」的方式以消除模型中的個別效果，並且進一步得到固定效果模型，即式（6），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：

$$\bar{Y}_i = \beta_0 + \bar{\mathbf{X}}_i\boldsymbol{\beta} + v_i + \bar{u}_i \quad (5)$$

$$Y_{i,t} - \bar{Y}_i = (\mathbf{X}_{i,t} - \bar{\mathbf{X}}_i)\boldsymbol{\beta} + (u_{i,t} - \bar{u}_i) \dots\dots\dots(\text{固定效果模型的測量}) \quad (6)$$

在實證模型的比對上，耿曙、陳陸輝（2003）的主要實證模型為下列式（7）所示：

$$\Delta Y_i = \alpha_0 + \Delta \mathbf{X}_i\boldsymbol{\alpha} + u_i \dots(\text{differences in differences 模型}) \quad (7)$$

在我們還原式（7）後可得下列式（8）：

$$(Y_{i,t} - Y_{i,t-1}) = \alpha_0 + (\mathbf{X}_{i,t} - \mathbf{X}_{i,t-1})\boldsymbol{\alpha} + u_{i,t} \quad (\text{differences in differences 模型的測量}) \quad (8)$$

其中與我們的實證模型式（6）有主要的差異，分別說明如下：首先可以注意的是耿曙、陳陸輝（2003）的主要實證模型將存在於一固定的截距項  $\alpha_0$  以及原始的誤差項  $u_i$ 。反之，若對本文所採用的實證模型式（6）取固定效果的估計方式，此估計方式最主要的特色在於所獲得的誤差項為  $u_{i,t} - \bar{u}_i$  而非簡單的誤差項  $u_i$ ，此處可以解釋成已經去除了省略變數所帶來的誤差成分；另外，就估計模型式（6）而言，固定的常數項  $\beta_0$  已被刪除。

依據上述式（6）的模型，我們估計出下表二的各項模型，<sup>8</sup>其中政治版圖模型指的是耿曙、陳陸輝（2003）中所提出的模型估計

---

8. 時間序列總體資料中不同地域之間應有不同的迴歸截距項，當我們以時間序列總體資料的固定效果模型分析政治版圖的變化時，不同的地區將有不同的截距項，本文於實證分析中因為篇幅的關係，因此筆者於表格中加以省略，包含表二、表三與表四的時間序列總體資料的固定效果模型。

結果，而修正模型為本文中模型的估計結果。我們分別探討只有「高科技」產業變數對於民進黨得票率的影響，以及四項不同產業結構變數對於民進黨得票率之影響。首先我們發現在時間序列總體資料分析的架構中，因為我們加入了省略變數進行控制，其結果與耿曙、陳陸輝（2003）的模型相比較，我們獲得了較高的 $\overline{R^2}$ 值，其上升幅度分別由 0.13 左右上升至 0.28 左右（見表二）。此外，若分析個別模型的F值，我們可以由表二看出，原本耿曙、陳陸輝（2003）的模型分別為 0.62 與 0.37，若採用時間序列總體資料的分析方式，考量以固定效果模型進行估計後，F值則分別上升至 1.38 與 1.80。由上述說明可見以時間序列總體資料的資料分析方式將會獲得更為有效的模型解釋力，其主要原因在於之前所說考量固定效果下的估計方式。

表二 高科技從業人員變動率與民進黨立委在各縣市  
得票變動率各模型比較表

	政治版圖模型 (1998-2001)	修正模型 (1998-2001)	政治版圖模型 (1998-2001)	修正模型 (1998-2001)
截距項	5.68** (1.50)	— —	5.93** (2.25)	— —
高科技	-1.70** (0.73)	0.66 (0.66)	-1.67* (0.86)	-0.155 (-0.14)
工商服務	— —	— —	-1.63 (1.01)	-0.79 (-0.66)
農林漁牧	— —	— —	-0.48 (0.90)	-2.09** (-2.07)
自營小業主	— —	— —	-0.03 (1.88)	-0.67 (-0.37)
地區效果	no	yes	No	yes
F 值	0.62	1.38	0.37	1.80
R <sup>2</sup>	0.13	0.19	0.25	0.28
樣本數	23	46	23	46

註：政治版圖模型括號中的數值為標準差；修正模型括號中的數值為 t 值。  
資料來源：耿曙、陳陸輝（2003）以及作者自行整理。

## 五、兩種固定效果的影響性—地區與時間

固定效果迴歸是一種控制模型中省略變數影響性的估計方法，由於是模型當中一種控制省略變數以避免估計產生誤差的方式，因此固定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。其中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。在固定效果迴歸中，有  $n$  組不同的截距項，每一個截距項將對應到每一個觀察值個體，並集成一組二元的變數，這一組變數將可以影響受省略的變數在不同的觀察值個體間有不同的變化，而同一觀察個體中，不同的時間點上誤差卻不會有所改變。另外，由於時間序列總體資料的相關模型中，其變數多具有不同時間與不同觀察值兩項因素，因此固定效果可以區分為地區效果與時間效果兩者。

### （一）地區效果

首先，地區效果代表著於時間序列總體資料的相關模型中，就不同各別觀察值而言所具有的不同性質，但若依據不同觀察值中的時間性來看，則發現是無所差異的，代表著會依據著不同觀察值有所變化，但卻不會依據不同的時間而產生變化。在政治版圖的研究中，如果能夠控制地區性固定效果，將可以區分是地區特性的影響，還是一個外生性解釋變數的影響，例如南北的差異已經控制在地區效果中，測出來的迴歸估計係數將明確是外生的影響，將不會受到南北區域差別的影響。在比較具有固定效果與否的前提下，我們先以未包含時間因素的模型進行比較，將耿曙、陳陸輝（2003）模型的實證資料重新以時間序列總體資料固定效果模型進行估計，進一步探討於考量固定效果與否之下，觀察各項產業結構對於兩岸態度政治立場的影響性是否有相同的結果。

在表二中，我們分別針對模型解釋變數採取單一高科技解釋與四種產業解釋變數兩種模型估計與耿曙、陳陸輝（2003）進行比較，實證結果指出，我們所獲得的高科技解釋變數迴歸係數分別為

0.66%與-0.155%兩種，其中所獲得的係數並不顯著，就方向性而言也不具一致性的方向，與耿曙、陳陸輝（2003）模型中的迴歸係數-1.70%和-1.67%明顯有所不同。但是在加入了其他產業結構變數進入模型當中，我們所獲得的修正模型估計結果與耿曙、陳陸輝（2003）的模型結果在方向性上趨向一致，在高科技產業對於民進黨立委得票率的影響性方面，將會為負向的影響效果，其估計係數為-0.155%。

依據修正模型的估計結果，我們發現在以時間序列總體資料的資料結構進行分析之下，我們所獲得的高科技產業對於民進黨立委得票率的影響性將會變小，且並不一定呈現顯著的影響效果，但是若拉長所分析的資料時間以及在增加模型中的各項產業結構解釋變數之下，<sup>9</sup>修正模型將可以獲得與原先政治版圖模型相同的影響方向性。因此就高科技對於民進黨立委在各縣市得票率的影響性方面，在考量時間序列總體資料的估計方式之下，似乎無顯著的效果存在，探究其主要原因有下列兩項可能。其一，若未將資料結構設定為時間序列總體資料並且未區分固定效果時，其模型所估計出高科技產業對於民進黨立委選舉得票率的負向影響性，或為不同地區之間的差異性所產生，並不是兩者之間直接影響性，此時若排除不同地區之間的差別性，將可發現高科技產業對於民進黨立委選舉得票率的影響效果並不顯著。其次，或為因為實證模型所使用的資料數目不夠大，導致雖於時間序列總體資料的架構下，但是未能發揮以時間序列總體資料模型估計的主要特色。

---

9. 我們將資料時間由原先的1998年至2001年兩屆的立委選舉，延長至1989年至2004年共六屆的立委選舉。

## (二) 時間效果

在時間序列總體資料的模型中，固定效果尚存在另外一種不同於地區效果的性質，稱做時間效果。除了地區效果之外，若加入時間固定效果（time fixed effect）於實證模型之中，則前述式（3）中  $Z_i$  向量中的變數，將區分為時間固定效果中隨著不同時間改變但不會隨著不同觀察值改變，以及地區固定效果中隨著不同觀察值改變卻不會隨著不同時間改變的相關變數。此時模型將轉變為下式（9）所示：

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \gamma_2 D2_i + \Lambda + \gamma_N DN_i + \delta_2 B2_t + \Lambda + \delta_T BT_t + u_{i,t}$$

.....（時間固定效果模型的測量）

（9）

其中  $D2_i, \dots, DN_i$  以及  $B2_t, \dots, BT_t$  等向量，分別代表不同觀察值相同時間點或是不同時間點相同觀察值之下的虛擬變數，而  $\beta_0, \beta_1, \gamma_2, \dots, \gamma_N, \delta_2, \dots, \delta_T$  分別代表未知的迴歸模型係數。時間效果將影響著模型中的變數，依據不同的時間下有不同的性質，但是在不同的觀察值之間，這些影響卻無所差異。如同式（9）中  $B2_t, \dots, BT_t$  的虛擬變數一樣，代表著會依據著不同時間點有所變化，但卻不會依據著不同的觀察值而產生變化。上述模型的設定重點在於我們將時間效果以虛擬變數的方式加入模型當中，如此在本模型依據時間效果的估計方式下，我們不僅可以排除「省略變數」對於應變數所造成的影響，並且可以獲得已經考慮到不同時間點上有所差異的估計結果，進一步對於政治版圖中的 contextual effect 進行描繪。

依據前述，我們將於實證模型中加入時間效果的設定，也就是將時間的虛擬變數加入到解釋變數之中。此時我們的實證模型中的

解釋變數  $X_{i,t}$  如下： $X_{i,t}$  = [工商服務、高科技、農林漁牧、自營小業主以及各年度的虛擬變數]（年度包含 1989 年、1992 年、1995 年、1998 年以及 2001 年），由於我們在修正模型中加入時間效果的探討，因此在時間序列總體資料的架構下，我們將主要探討時間變項對於整體資料的影響性。模型的估計結果如表三與表四所示，其中我們分別列出是否有加入時間效果以及對模型的解釋變數是採取單一高科技解釋與四種產業解釋變數兩種不同的模型估計，並將與耿曙、陳陸輝（2003）的模型進行比較。在表三中若單就  $\overline{R^2}$  而言，有加入時間效果將會具有較高的  $\overline{R^2}$ ，而考慮四項產業解釋變數的模型的解釋力又會高於採取單一高科技產業解釋變數的模型，其中未考慮時間效果的模型只有 0.01 的  $\overline{R^2}$ ，加入時間效果後  $\overline{R^2}$  上升至 0.42，若再加入其他解釋變數的探討  $\overline{R^2}$  則可以上升至 0.51，由上述的估計結果表示時間序列總體資料中時間效果的採用將可以促使模型的解釋力大幅上升。另外就加入時間效果的模型而言，在表三中我們可以發現高科技對於得票率的影響性將會由沒有時間效果的模型中所估計出不顯著估計係數轉變為顯著的負向影響，與時間效果對於得票率都具有顯著的影響性。

表三 控制固定效果的各模型估計值（1998-2001）

	(1)	(2)	(3)	(4)	(5)	(6)
高科技	0.13	0.66	-0.15	0.06	-1.61*	-1.67
	(0.85)	(0.66)	(-0.14)	(0.28)	(-1.64)	(-1.59)
1998 年	—	—	—	-4.09**	-5.58**	-6.04**
	—	—	—	(-3.41)	(-3.84)	(-2.91)
工商服務	—	—	-0.79	—	—	-1.63
	—	—	(-0.66)	—	—	(-1.54)
農林漁牧	—	—	-2.09**	—	—	-0.44
	—	—	(-2.07)	—	—	(-0.43)
自營小業主	—	—	-0.67	—	—	-0.95
	—	—	(-0.37)	—	—	(-0.62)
地區效果	no	yes	yes	no	yes	yes
時間效果	no	no	no	yes	yes	yes
F 值	0.13	1.38	1.80	1.96	7.22	3.24
R <sup>2</sup>	0.01	0.19	0.28	0.36	0.42	0.51

註：括號中的數值為 t 值。

資料來源：作者自行整理。

另外當未控制時空的固定效果時，高科技對於 1998 年到 2001 年政治版圖的變化並無顯著影響，當表三的模型二控制地區效果後亦然，此時似乎高科技就業人口的變化對於政治版圖並無顯著影響，而何種人口就業因素會對於政治版圖造成影響呢？出乎意料的於表三模型三中加入其他職業團體的變項之後，農林漁牧對於民進黨的政治版圖有所影響，而且還是顯著的負面效果，這樣的結果超乎我們的常識所理解。若在表三模型三我們控制了地區的固定效果並加入了其他職業變項之後，農林漁牧似乎在 1998 年至 2001 年之間背叛了民進黨，可能的解釋在於「台聯」的出現吸收了這些選民。

但是在表三模型五中控制了時間的固定效果之後，我們發現高科技變項呈現顯著的負向影響，相較於模型二的不顯著正向影響，兩模型間的比較證明在控制了時間的固定效果之後（ $y_{1998}=-5.58^{**}$ ），高科技的負影響才突顯出來，表三模型六在加入了其他職業變項之後，更強化了我們的認知，如果不控制時空的固定效果，原本有影響的變數（高科技）可能被各年度之間的時間變化所淹沒，而其他不相關的變數（農林漁牧）反而被膨脹為重要的解釋變數。

表四 控制固定效果的各模型估計值（1989-2004）

	(1)	(2)	(3)	(4)	(5)	(6)
高科技	0.18** (2.00)	0.19* (1.88)	-0.05 (-0.29)	-0.06 (-0.41)	-0.38 (-1.55)	-0.22 (-0.71)
1989年	—	—	—	-10.14** (-2.90)	-15.26** (-2.95)	-5.86 (-1.02)
1992年	—	—	—	-5.22 (-1.55)	-9.97** (-2.28)	-2.21 (-0.41)
1995年	—	—	—	-1.58 (-0.63)	-3.13 (-1.18)	1.38 (0.46)
1998年	—	—	—	-7.27** (-3.01)	-8.11** (-3.31)	-5.57** (-2.17)
2001年	—	—	—	-3.07 (-1.28)	-3.63 (-1.51)	-2.49 (-0.98)
工商服務	—	—	-0.16 (-0.19)	—	—	0.24 (0.26)
農林漁牧	—	—	-0.90** (-3.71)	—	—	-0.80** (-2.69)
自營小業主	—	—	0.08 (0.85)	—	—	0.06 (0.68)
地區效果	No	yes	yes	no	yes	yes
時間效果	No	no	no	yes	yes	yes
F 值	0.76	12.97	6.97	7.19	9.37	6.59
R <sup>2</sup>	0.03	0.03	0.15	0.17	0.18	0.24

註：括號中的數值為 t 值。

資料來源：作者自行整理。

## 六、結 論

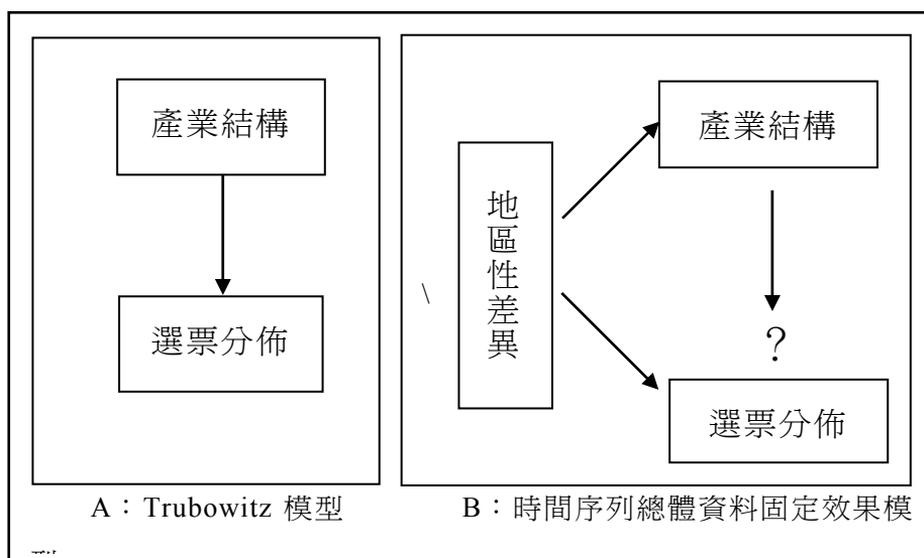
由於台灣的政治研究資料上，有越來越多的時間序列總體資料型態，因此在政治版圖的劃分上，若能夠採用時間序列總體資料的分析型態，將可以避免模型中一些無法取得的相關變數產生估計上的偏差，這樣的理解將是未來政治版圖測量的發展趨勢。在與傳統 differences-in-differences 的模型比較之下，我們發現以時間序列總體資料的資料結構進行模型的分析，可以發揮出時間序列總體資料的優越性，將固定效果中的地區效果以及時間效果進行區分，如此才可以在排除內生性影響的狀況下對於單純相關性影響的變數進行刻劃，完整的呈現出各種外生變數對於模型應變數的影響能力。也就是說，由於本文進一步區分出地區效果與時間效果，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。而由於加入了時間變項的分析探討，對於整體模型的估計將會更趨向完善，且除了早期文獻描繪各個不同選舉時期間的趨勢之外，更能夠對於時間因素的影響性進行更進一步的探討。

在實證結論上，我們依循 King (1995) 所提出的「Replication」研究方法，將耿曙、陳陸輝 (2003) 的議題進行重新審視。耿、陳的結論認為產業結構對於政治版圖的選票分佈將造成影響，然而我們以時間序列總體資料的資料結構以及固定效果模型進行相同主題的驗證，試圖找出影響政治版圖的正確原因。本文發現高科技的產業結構於時間序列總體資料的地區效果估計之下，並未有顯著地對於民進黨立委得票率造成影響，與傳統橫斷面的研究進行比較，其結論是不同地。究其原因，應與在地區效果的估計之下，我們已經

排除了地區上的區別性，因此可以獲得單純地外生解釋變數（高科技產業比率）對於應變數（民進黨得票率）的影響性，不會受到不同地區上「內生性影響」的干擾，如此才會獲得不同地結論。

如圖四的左圖所示，在政治版圖的實證分析中，耿、陳認為「產業結構」對於「選票分佈」有顯著性的影響，主要是依循 Trubowitz（1996）的理論，該理論認為由於各地區的「經濟結構」不同，貿易後所產生的「利害」也不盡相同，而沿用「經濟結構」與「經濟利害」的差異，將可以進一步發展到各地區「政治結構」的推定（耿曙、陳陸輝，2003）。但本文認為「地區性差異」將同時影響「產業結構」與「選票分佈」（見圖四的右圖），因此當我們以時間序列總體資料固定效果模型控制住實證資料中的「地區性差異」後，「產業結構」對於「選票分佈」的顯著性影響將不再存在。換句話說，在「產業結構」影響「選票分佈」的因果關係背後，兩者其實同時受到「地區性差異」的影響。這其實是很典型的統計推論謬誤，在認為「產業結構」是影響「選票分佈」的主要影響變數之後，其實存在第三項影響變數（地區性差異）同時影響著原先所觀察的解釋變數與被解釋變數。舉例而言，北部地區原本即為民進黨得票較弱的地區，但恰巧北部地區從事高科技產業的相關人士也是較高，兩者同時為不同地區的差異性，卻會被以橫斷面資料結構所進行的研究所忽略，並獲得高科技產業結構對於民進黨立委得票率有所影響的結論。本文以時間序列總體資料所得的研究結果與耿曙、陳陸輝的結果不相同，背後所代表的主要意涵在於當考量地區性影響因素之後（我們以固定效果模型區分出地區性的差異），我們發現「地區性差異」才是影響台灣政治版圖變遷的主要原因，而非「產業結構」，這代表著以不適當的統計方法進行政治版圖的分析，將無法

真正區別出主要的影響因素，這也是本文強調使用正確統計方法來驗證理論最主要的原因。



圖四 時間序列總體資料的固定效果模型與 Trubowitz 模型比較圖

資料來源：作者自行整理。

本文最主要的貢獻在於將時間序列總體資料的資料結構引進政治版圖劃分的相關研究中，並且以固定效果中的兩種估計方式，將地區性差異所造成的影響以及外生解釋變數所造成的影響進行劃分，最後獲得與傳統分析不同的結論：高科技產業對於民進黨立委得票率並未具有顯著的影響能力。究其差異，應為是否將地區差異性排除有關，也就是說，影響能力不在產業結構的比率，而在於不同地區所造成的影響。由於本文是以固定效果的估計方式先排除地區差異所造成的影響性，進一步可以刻劃出產業結構變項對於民進黨立委得票率是否有所影響，最後並得出兩者之間無影響性的結論，這是一種反向的研究設計，究其主要原因在於若將地區進行劃

分，該如何歸納不同地區的主觀性，因此以此種研究設計控制主觀劃分不同地區的問題。

## 參考書目

- Green, Donald P. et al. 2001. "Dirty Pool." *International Organization* 55: 441-468.
- Hsiao, Cheng. 2003. *Analysis of Panel Data* Cambridge: Cambridge University Press.
- Hsu, Yung-Ming and Pei-Shan Lee. 2002. "Southern Politics?: Regional Trajectories of Party Development." *Issues and Studies* 38, 2 (June): 61-84.
- King, Gary. 1995. "Replication, Replication." *Political Science and Politics* XXVIII, 3, (September): 443-499.
- King, Gary. 1997. *Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. NJ: Princeton University Press.
- Manski, Charles F. 1995. *Identification Problems in the Social Sciences*. Cambridge, Mass.: Harvard University Press.
- Trubowitz, Peter. 1996. "Political Conflict and Foreign Policy in the United States: A Geographical Interpretation." in G. John Ikenberry. ed. *American Foreign Policy: Theoretical Essays*: 395-411. New York: HarperCollins.
- 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。
- 徐永明。2001。〈政治版圖：兩個選舉行為研究的途徑的對話〉。  
《問題與研究》40，2：95-115。

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。

行政院主計處。2008。〈人力資源調查統計年報〉。

<http://www.stat.gov.tw/lp.asp?ctNode=518&CtUnit=359&BaseSD=7>。2008/06/22。

國立政治大學選舉研究中心。2008。〈台灣選舉類別與結果線上瀏覽〉。

<http://vote.nccu.edu.tw/cec/vote4.asp>。2008/06/22。

# Spurious Statistical Relationships and their Theoretical Implications: Investigating Changes in Party Support in Taiwan, 1989-2004

Yung-Ming Hsu \*    Chang-Ping Lin \*\*

This paper investigates a case of spurious statistical relationships influencing theory. Our work serves to emphasize the importance to political discourse of using appropriate statistical tools. We expand upon an earlier study of regional support for the Democratic Progressive Party (DPP) in Taiwan by including fixed effects analysis. The original paper corroborated the theory of Trubowitz (1996) by finding that geographic political support followed local economic interests. We find that after considering regional fixed effects the original relationship no longer holds. Instead, regional factors simultaneously influence economic structure and political leanings.

**Key words:** political blocs, time series aggregate data, fixed effects, time effects

---

\* Assistant Professor, Department of Political Science, Soochow University.

\*\* Ph. D. Student, Department of Money and Banking, National Chengchi University.