

國家發展研究 第六卷第一期
2006 年 12 月 頁 89-144

選舉的空間因素： 以三次總統選舉為例

鄧志松^{***}

收稿日期：2007 年 5 月 3 日

接受日期：2007 年 6 月 22 日

* 國立台灣大學國家發展研究所助理教授，terry@ntu.edu.tw。

** 作者感謝兩位匿名委員的審查意見。本文圖表原係彩色，因印刷之限制以黑白呈現，若欲取得彩色圖文，請至網址 <http://ceiba.ntu.edu.tw/course/17052d/prespatial.doc> 自行下載。

摘要

本文嘗試以空間分析的方法，探討台灣三次總統選舉呈現的空間因素。作者首先介紹空間分析的意義與特質，接著介紹空間分析常用的方法，對選舉地理學重要文獻作簡要回顧，最後以三次總統選舉為例，探討選舉結果呈現的空間因素。使用的方法包括視覺化呈現、空間探索分析、空間迴歸建模三個層面。研究發現第一次總統選舉，因李登輝效應影響，選票的空間結構比較異常。其後選舉，藍、綠選票加總之後，縱使得票率互有消長，仍呈現相當穩定的空間結構。空間迴歸方面，作者針對民進黨總統選舉獲票的增加率進行分析，發展誤差的空間自相關現象同時來自於空間異質與鄰近效應。

關鍵詞：空間分析、空間自相關、選舉地理學、空間迴歸、總統選舉

前言

空間分析嘗試用量化方法，探討空間因素扮演的角色，這本是地理學者發展的方法，廣泛應用於環境科學、生態學、氣象學、地質學、考古學等。然而近年來愈來愈多的社會科學議題，亦嘗試運用這套分析方法，如都市規劃、產業聚落、區域均衡、公共衛生、疾病防治，乃至犯罪研究、人口變遷等均是，空間分析儼然成為科際整合的新趨勢。「選舉地理學」(electoral geography)則是嘗試運用空間分析方法於選舉研究，彌補傳統選舉研究忽略空間因素的缺憾，國外已累積相當多的文獻，不過國內似乎尚在起步階段。本研究嘗試以三次總統選舉為例，引用空間分析的方法，希望能開拓選舉研究的新視野。

本文首先介紹空間分析的意義與特質，接著介紹空間分析常用的方法，繼而對選舉地理學重要文獻作簡要回顧，最後以三次總統選舉為例，探討選舉結果的空間因素。空間分析一般說來分為視覺化呈現、空間探索分析、空間迴歸建模三個層面，本研究三個層面都會觸及，期能收拋磚引玉之效。

壹、空間分析的意義

「空間分析」(Spatial Analysis)旨在突顯「空間因素」的重要性，探討「空間因素」在社會(自然)現象中扮演的角色。什麼是「空間因素」？狹義的空間因素，一般指經緯度或 XYZ 座標值；廣義的空間因素則可包括所有與「空間」有關的變數，除了經緯度、座標外，還包括觀察對象所處的地形、地貌、文化、傳統、制度、風俗習慣、經濟發展、行為表現等，甚至還括及鄰近地區的狀況。基本想法是，人們的行為表

現會因為所處環境而改變，因此將觀察對象所處的地理位置及其相關屬性納入考慮有其必要性。

為了突顯空間的重要性，空間分析的一個重點是「鄰近地區的影響」(neighborhood effect)。誠如 Tobler (1979) 提及的地理學第一定律 (first law of geography): 「每件事情都與其他每件事情相關，但較近的事比較遠的事來得更為關係。」¹ 地區與地區間會相互影響，此種影響隨著距離的加大而遞減。如果加進時間的向度，此種影響也許不是立即的，有時間間距，此即空間的擴散作用，這讓問題的分析更形複雜，但更合乎實際。其實，人的行為表現受到所處環境或周遭環境的影響是老生常談、不足為奇；不過，空間分析學者與傳統的「空間意識」² 不同，他們嘗試結合日益成熟的電腦科技、地理資訊系統 (GIS)³、空間計量方法、以及大型資料庫，目的在精確地界定空間因素的重要性及影響力：到底是哪一種「空間因素」產生影響？影響有多大？如何建立模型？解釋自變數與因變數間的關係，這些都是傳統僅有「空間意識」的學者所無法處理的。

空間分析在社會科學的應用，我們不能不提及「空間社會科學整合研究中心」(Center for Spatially integrated Social Sciences, CSISS)⁴ 的貢獻。該中心由美國國科會 (the National Science Foundation) 贊助，成立

¹ "Everything is related to everything else, but near things are more related than distant things."

² 指僅體認空間因素的重要性與影響力，但缺乏嚴謹的方法，界定其大小、模式與範圍。

³ 「地理資訊系統」(GIS) 是以地理位置為中心，結合與其相關的屬性資料 (geo-referenced data)，運用電腦科技，建構的空間資料系統。近年由於電腦科技的進步，地理資訊系統應用日益廣泛，不僅在環境保育、公共衛生、考古、地質、農業、生物等研究上廣為運用，還擴及人文社會科學，特別在都市規劃、社會變遷、經濟發展、公共政策議題上貢獻斐然。政府有感於 GIS 技術在資訊整合、政策擬定、資源有效運用上扮演不可或缺的角色，民國七十九年起，成立「國土資訊系統」(NGIS) 推動全國 GIS 資訊系統建置，作為政策規劃的參考，目前已初具規模。民間運用 GIS 技術亦日益蓬勃，如網路電子地圖、生活地理資訊整合，乃至於 GPS 衛星定位導航，運輸、車輛管理等均是。

⁴ <http://www.csisss.org/>，CSISS 成立的經過請參考：Michael F. Goodchild and Donald G. Janelle (ed.), *Spatially integrated Social Science*. Oxford University Press, 2004.

於 1998 年，由 Michael Goodchild 帶領，以 University of California, Santa Barbara 為基地，彙集相關文獻、系統化分析方法、開發計量軟體，對空間分析的推廣不遺餘力。他們對「空間分析」界定為：「突顯『空間因素』在人類社會所扮演的角色，並主張分析變數的空間模式以及過程（spatial patterns and processes）有助於我們對人類社會的瞭解。」本研究運用的概念、方法與軟體，基本上延續這個脈絡。

總結以上說明，空間分析本質上是量化的研究取徑，處理地理屬性（geo-referenced data）資料。在自然科學的領域裡，這樣的探討極為普遍，諸如氣溫、氣壓、風向、風速的變化、地質結構、環境、生態、物種分佈，都可輕易找到地理位置所造成的影響。就社會科學言，其應用不及自然科學普遍，這有兩個可能原因。主要的原因可能是方法論上的理由，社會科學家普遍認為，「地理位置」的影響，極可能只是假象，我們應致力於找到其他「真正的」變數，把「地理位置」這個因素解釋掉。例如，台灣的選舉有「南綠北藍」的現象，我們總不好說，這是南、北位置本質上的什麼理由（風水？）造成的！不如說是歷史傳統、經社背景等因素使然，這比較有說服力。

然而，真的是這樣嗎？空間因素是否可以完全被其他社會變數解釋掉？這是本文關心的議題。本文將以三次總統選舉為例，闡明空間分析應用在選舉研究的可能性及其限制。由於這個嘗試在國內尚不多見，同時受限於經社資料庫取得不易，因此，本研究界定為探索式的（exploratory）研究，而非理論的檢驗（confirmative research）。作者嘗試介紹空間分析的基本方法，以及檢討其在選舉研究上的可能性與特殊性。

貳、選舉地理學：空間分析在選舉研究的應用

空間分析的方法應用在選舉、投票行為的研究一般稱為「選舉地理學」（electoral geography）。有別於傳統的選舉研究，其特點是地理區塊

作為分析的單位，地理區塊可以是行政區界（縣市、鄉鎮、村里或者投票所等），或是類似於美國的普查區。資料係集體資料（aggregate data），而非以抽樣為基礎的個體資料（individual data）。顯然這與當前選舉研究的主流有異。這裡要強調的是，作者無意挑戰這個研究主流，實際的投票者是個人，不管採「社會學」、「社會心理學」、「理性選擇」的研究途徑，都是企圖尋找各種不同因素，對個人行為造成的影響，只是強調的因素不同。以集體資料作為分析的基礎，首先面臨的質疑是：「是人在投票，而非地區在投票」！如果用集體資料進行個人層次行為模式的推論，極可能會犯了「區位推論的誤謬」（fallacy of ecology）。

另一個問題是分析單位的選擇，如果不以個人作為分析的單位，那麼要以多大的集體作為分析單位呢？以台灣為例，縣市、鄉鎮、村里或者投票所，都可以作為分析的單位。不同的分析單位，研究的結果會不會不一樣？答案是有此可能。此即所謂的 MAUP 問題（the modifiable areal units problem）。（Haining, 2003: 140）

關於以上兩個問題，作者的回答是：第一、既然從集體資料出發，就無意越界推論個人層次的行為⁵。因此，回答的問題不是「個人如何投票」，或「什麼因素影響個人的投票行為」；而是「什麼因素造成地區間『表現』的差異？」或者「地區與地區的『表現』以何種模式相互影響」。這個『表現』可能是政黨獲票率、投票率、犯罪率、經濟成果、競爭力，也可能是其他任何經社變數。如同個人的表現可以進行變數間的因果解釋，在地區的集體層次亦然。當然，這首先得證明這種嘗試是有意義的。Erikson, Mackuen & Stimson（2002）引用經濟學總體（macro）與個體（micro）的區分，為總體政治學（macro policy）說項，其目的並非要取代美國政治學界以個體為主的研究，而是指出集體資料不僅只是個體資料的加總而已；同時，集體指標可以排除某些「有影響力但無趣」的

⁵ 地理學者亦發展出一些方法解決區位推論的問題，不過甚為複雜（Haining, 2003: 138: 150），本文未作討論。

變數，使問題的解釋更具意義，⁶總之，以集體作為分析的單位，可以增加我們對政治的解釋能力。關於第二個問題的回答，作者認為分析單位的選擇與問題意識，亦即作者想回答的問題有關。如果作者關心的是區域發展均衡的問題，也許縣市作為分析單位是不錯的選擇；如果關心的是「鄰近效應」，以鄉鎮、村里作為分析單位是必須的。針對某些特殊的研究議題，經常要打破地理的空間觀念，因為「距離」未必源自物理距離，而是來自於其他人文或經社環境的相近性。就實際研究操作而言，研究單位的選擇常受限於資料，研究單位愈小，愈難取得其集體資料，這是空間分析必須面對的限制。

縱使有以上兩種批評，空間分析在突顯空間因素的重要性上仍然不能忽視。其實傳統政治學的研究並非不瞭解空間因素的重要，例如以不同地區設定虛擬變數或交互作用，就是希望將空間因素納入考慮，雖然此種模型不可能太複雜。近年愈見普及的多層模型（Multilevel Models）則是結合集體資料與個別資料較有效的方法，亦即在個體資料為前提下，進行總體效果的探索（黃信豪，2006）。其具體作法是將分成數個層次（為說明方便，以兩個層次說明之）。第一個層次是個體層次，與傳統的量化研究無異，第二個層次是集體層次，例如地理單位（縣市），也可以是學校、個人所處的環境等。影響個體層次的自變數可能有：性別、年齡、教育、職業等個人特質、政黨認同、主觀偏好等，集體層次的自變數，可以是失業率、產業結構、政黨支持率等政經社變數。多層模型旨在同時考慮個體資料與集體資料，進行迴歸分析。

就理論的角度，這是同時處理個體、集體資料最有效的方法。不過，仍有一些使用上的限制，第一、基本上它處理的是異質性的問題，亦即不同的集體條件對個體的行為表現會造成影響，然而沒有處理自相關的問題，亦即第二層級的迴歸誤差可能存在相關性。尤其第二層級代表的

⁶ 例如作者舉例，總統支持率的影響因素，政黨屬性在個人層次是個重要的解釋變項，在集體的層次則未必。（Erikson, Mackuen & Stimson, 2002: 10-11）

是不同的地理區域時，這個問題顯得特別嚴重。第二個限制是資料取得的問題，由於要同時符合個體與集體兩個層級資料的需求。第二個層級的組數勢必不能太多，例如以台灣二十三個縣市為例，每個縣市抽 100 個樣本，全部需要二千三百個樣本，如果第二個層級的單位是鄉鎮，則需三萬五千八百個樣本（台灣有三百八十五個鄉鎮）。當第二個層級的組數太少時，集體自變數就不能擺太多，否則會導致係數不穩定。（黃信豪，2006：183）這形成一個兩難，基於成本考量，樣本無法太大，但樣本不夠大，又會限制集體自變數的選擇。

就本文立場而言，第一個限制無疑地更為重要。因為空間分析學者，如前所述，關心的是鄰近效應的影響，亦即樣本彼此間是否有空間自相關的現象。關於這個問題多層模型則無法處理。

討論空間分析的「正當性」之後，選舉地理學者首先關心的問題是：如何解釋選舉結果（獲票率、廢票率、投票率等）呈現地區性的不平均現象？基本上，有三個不同的理論回答這個問題。（Perdomo, 2003）第一為「成分影響觀點」（component effects perspective），認為主要因素是各地區的經社背景屬性使然，地理位置本身不會直接造成影響。因為經社背景屬性變數有空間的特性（分佈不平均），反應在選票上，選舉結果在各地區也會呈現不均勻現象。這種說法，假定同樣的人，會投同樣的票，地理因素不是重點。（McAllister, 1987, McAllister & Studlar, 1992）。在現代社會中，可以把全國當成一個分析單位，地區的獨特性不須考慮，傳統社會中地區的獨特性才會被突顯出來，這是一種現代化理論的合理假設。（Agnew, 1987）這個理論排除空間因素的直接影響，認為只要找到合適的自變數，即可解釋因變數的空間變異，亦即迴歸誤差的地區性不均勻現象可消除。

第二個理論強調地區脈絡（local contextual effects）的重要性，地區的選舉表現經常超越經社背景變項所能解釋的範圍，它是個人與其環境互動的結果，而這個環境的本質是行為者長期建構出來的產物。（Flint,

1998: 1293) 我們也可以從 Giddens 的結構理論中找到類似的說法, 行為者 (agency) 與其所處的結構 (structure) 相互影響。一方面, 行為者被傳統、制度和道德規範、既定的行為模式所影響; 另一方面, 個人經由漠視 (ignoring)、取代 (replacing)、再生產 (reproducing) 等方式, 改變這個結構。(Giddens & Pierson, 1998: 77) 地區脈絡以社會網絡、深根在地方團體, 提供與傳播資訊, 進而發揮影響 (Sauerzopf & Swanstrom, 1999)。人們做政治決定時, 無法擺脫社會環境的刺激。(Burbank, 1995) 這種論點承認空間因素的重要性, 認為人們即使屬於同樣經社背景或條件, 其投票行為也會因所處的地理位置不同而異, 因此放了所有經社背景自變數之後, 仍無法完全解釋地區變異。此時, 可以使用虛擬變數, 突顯地區脈絡的影響, 如果使用得當, 可消除誤差的地區性不均勻現象。

第三種理論同樣重視地區脈絡的重要, 然而關注的是空間上的擴散效果 (spatial diffusion effect)。鄰近地區的行為會相互影響, 這不是地區本質上的特性使然, 亦非地區的經社背景可以解釋。就像傳染病一樣, 地理的相鄰或親近性決定傳播的速率與方向, 當然此種擴散的模式經常不是固定不變或即刻發生, 而是依時空背景而有不同的變化。透過空間變數 (下詳) 建構空間迴歸模型, 可以有效解釋鄰近地區造成的影響。

選舉地理學的研究經常圍繞著上述三種立場, 嘗試運用地區屬性資料, 澄清空間因素扮演的角色。其實以地區性的集體資料進行選舉研究起源甚早, 早在 1949 年 Key 就以集體資料討論美國南方州的選票分佈, 其實台灣早期的選舉研究也是以集體資料進行分析。行為科學興起後, 改以個體資料為主。然亦有少數學者進行集體資料的研究, 例如洪永泰 (1994) 以歷次選舉資料運用集群分析, 進行政治版圖的研究; 徐永明 (2001) 運用 Gary King 的 EI 模型估算「忠誠選民」, 發現忠誠選民的比例與「政治版圖」間並無相關。盛杏媛 (1998) 探討 SNTV 制度導致的政黨配票與候選人票源的集散度, 研究發現: 國民黨候選人較民進黨

與新黨有票源較集中現象，但此種現象有下降趨勢⁷。林繼文（1998）運用對應分析法描述台灣立委選舉中候選人之主要地盤。黃紀則結合 GIS 技術建立各政黨的空間資料庫，展示其分佈狀況⁸。不過，必須說明的是這些學者固然突顯空間因素的重要性，然而並未建立「空間迴歸模型」，量度空間因素影響的大小。相反地，國外「選舉地理學」的發展，空間模型的建立則相當普遍。例如 Perdomo（2004）以空間計量方法，探討墨西哥 1994，2000 兩次大選「擴散效應」（contagious diffusion）與「地方效應」（local contextual effects）的影響。發現後者大於前者。Shin & Agnew（2002）透過不同年度，各政黨選票分佈的空間結構，分析義大利政黨的繼承或重組現象，作者假定同一政黨有穩定的選民結構，這表現在空間的散佈的穩定性上。Kohfeld & Sprague（2002）藉兩個不同型態的選舉，探討聖路易斯（St. Louis）城的族群政治。一為有黨派動員色彩的民主黨初選（1989 年），一為非政黨競爭的學區委員會選舉（1991 年）。發現聖路易斯城經社屬性變數與族群分佈有相似的空間結構，而族群動員對選舉結果仍具關鍵影響。Gimpel & Schuknecht（2002）比較德州、加州、伊利諾州和喬治亞州四個州，1928 到 2000 年間，民主黨總統選舉獲票率空間分佈的演變。基本上，民主黨的獲票趨向集中，共和黨則趨向分散。值得注意的是：作者進行政黨版圖的計算時不是直接用獲票率，而是以該郡對民主黨總獲票數的貢獻加權之，藉以將不同地區（郡）的重要性納入考慮。West II（2005）則由不同政黨的獲票率，將土耳其區分為性質不同的地區（cluster analysis），進而比較不同地區的經社背景，發現宗教、族群、經濟發展、過去的行政隸屬關係構成政黨版圖的主要解釋因素。

⁷ 盛文用候選人在各鄉鎮區獲票率的標準差衡量票源集中程度，能顯示候選人在各地區票源分佈不均的現象，然無法顯示與鄰近地區的趨同現象，Moran's I 指數可處理這個問題（下詳）。

⁸ 參見黃紀教授政治地緣研究室台灣選舉地緣資料庫，<http://tegis.nccu.edu.tw/default.asp>。

選舉地理學中空間迴歸的運用極為常見。因變數通常為政黨獲票率或投票率，自變數除了各種人口變項（通常經由普查資料獲得）外，重點在探討空間因素影響的程度。與傳統迴歸分析不同，空間資料本身具有不獨立性（空間自相關）與誤差異質性（空間異質），因此兩者經常是討論的焦點。例如 O'Loughlin, Flint & Anselin (1994) 與 Ault & Brustein (1998) 探討德國 1925-33 年納粹快速興起的社會因素。為了探討異質性問題，作者考慮歷史文化因素後，將德國區分為六大地區，比較不同地區的納粹選票結構。Flint (2000) 針對不同時期巴登地區 (Baden) 的納粹獲票率，指出南巴登與北巴登在不同時期，歷經同質、異質，又回歸同質的演變過程，並與相關的質性調查相互應證。Patterson & Caldeira (1984) 探討造成政黨激烈競爭的因素，值得注意的是他以虛擬變數處理南方各州存在的異質性問題。MacAllister, Fieldhouse & Russell (200) 探討英國第三黨自由民主黨 1997 年大選突破傳統根據地，向外拓展的機制，發現地方選舉、補選的斬獲是跨越 credibility gap⁹ 的關鍵，有助於解釋 1997 年的選舉結果。Gimpel & Cho (2004) 則關心北英格蘭地區的族群政治，他挑戰族群影響日益式微的觀點，以空間迴歸證明族群淵源對總統選舉仍有解釋力，無法以其他經社變項替代。

投票率的影響因素方面，著名者如 Gimpel & Schuknecht (2003) 以 SLM 模型探討華盛頓 DC，2000 年總統選舉，投票所遠近對投票率的影響。在控制其他變數（包括空間變數）後，發現投票所距離與投票率相關，然而不是線性關係，在 2-5 英里處距離的影響為負，但超過一定距離 (6-10 英里) 其影響反而變為正。Gimpel & Morris & Armstrong (2004) 運用各郡 (counties) 的資料，探討不同世代 (大蕭條世代、嬰兒潮世代、後嬰兒潮世代) 對投票率的影響，發現控制各種可能的變數之後，殘差仍然具有空間自相關現象，顯然有某些隱藏的變數存在；就不同世代的

⁹ 英國小選區制，不利第三黨的發展，主要原因是選民會認為第三黨沒有獲勝的可能，投給他們只是浪費選票，此即 credibility gap。

投票傾向而言，後嬰兒潮世代傾向政治冷漠。另外，Cho & Tam (2003) 分析一九八〇、九〇年代，亞裔人士政治獻金的演變，除了考慮地區的屬性資料外，特別強調空間擴散效應的影響。

就本文而言，由於是探索性的研究，作者廣泛考察歷次總統選舉的空間特性，初步探討一個問題：在總統選舉中空間因素扮演的角色有多大？如果如現代化理論所言，經社變數才是關鍵，則空間變數在迴歸分析中不會顯著，不管是採「空間落遲模型」或「空間誤差模型」(下詳)。如果空間變數有顯著的影響，則我們不得不進一步討論空間因素形成的原因，以及它代表的理論意義。受限於資料整理的技術，本研究以鄉鎮作為分析單位，使用中選會提供的歷屆總統選舉資料，以及整理自人口普查、財稅中心的各鄉鎮經社背景變數；使用的軟體為 GeoDa。下節我們先介紹空間分析的基本方法，以作為進一步論述的基礎。

參、空間統計與空間計量

在談空間分析方法以前，必須對空間資料的型態有些瞭解。一般說來空間資料分為兩種，其一為「點資料」(point-referenced data)，記錄各事件在不同地點出現的次數或發生的頻率，事件出現一次即一個點粒，點粒的密度愈大，代表事件出現的頻率愈高；第二為「地區資料」(regional data)，以地區為分析單位(例如鄉鎮或村里)，彙整各地區的屬性資料，此為「集體資料」(aggregate data)。¹⁰若各地區大小一致、行列對齊，則稱為「網格資料」(grid data)。由於資料型態不同，分析方法有異。就本研究言，吾人處理的是以鄉鎮區為分析單位的地區資料，以下有關空間分析方法的介紹，亦以此為限。

¹⁰ 相對於以個人作為分析單位的「個體資料」(individual data)。「集體資料」以地區為單位，亦即探討的是「地區」的行為表現，而非個人。

空間分析方法依其性質分，可分為「空間統計」(spatial statistic)與「空間計量」(Spatial Econometrics)兩種¹¹。「空間統計」重點在描述資料的空間特性，常見者如：主題圖(顯示變數的空間分佈狀態)、Moran's I 散佈圖(顯示變數的全域空間聚集情況及顯著性檢定)、LISA (local Moran's I spatial association) 分析(顯示各單獨地區與其鄰近地區的相關關係及顯著性檢定)；PCP (parallel coordinate plot) 圖(顯示各地區在不同時間點、或不同變數間的排名變化情形)等。這些方法的運用，後文中作者皆會以豐富的圖例，說明其意義及分析技巧。

「空間計量」旨在建立迴歸模型，探討自變數與因變數間的關係或影響程度。基本上，空間計量是傳統計量經濟學(Econometrics)的延伸，只不過因為空間資料的特殊性，估計與檢定方法與慣常使用者的不同，略述如下。首先，空間資料以地區為單位，這樣的資料有一個特點：樣本就是母體，而非母體的抽樣。例如台灣三百五十八個鄉鎮，就是三百五十八個觀察值，觀察值本身就是母體。就空間計量的觀點，我們還是假設有一個不知道的「母體」存在，這些觀察值只是被「抽出來」的三百五十八個「樣本」。如果「樣本」顯現自變數與因變數的某種關係，那麼未知的「母體」也會有此種關係。經過這樣的「解釋」，空間資料也可以進行類似於傳統的迴歸分析。

第二、正因為空間資料不是隨機抽出，它通常違反傳統計量經濟學對樣本的兩個重要假定：誤差獨立，以及誤差變異齊一。由於空間觀察點不是隨機抽出，前後觀察值間經常具有相關性，亦即第一個觀察值會影響第二個觀察值的大小，例如台北市的人均所得與台北縣很可能有相關關係存在。此種情形對迴歸分析產生的影響是誤差彼此不獨立，此時迴歸檢定會過度樂觀，亦即喪失有效性。誤差變異不齊一指隨著某個自

¹¹ 「空間統計」(spatial statistics)與「空間計量」(spatial econometrics)兩個字眼經常混用，一般說來，後者著重於空間迴歸模型的建立，前者重在空間性質的確立。(Anselin, 1988: 10)

變數大小的改變，誤差的變異也跟著改變，這會讓迴歸檢定不穩定，有時過度樂觀，有時過度悲觀。空間異質與傳統異質性的探討稍有不同，它指誤差的變異因不同地域而有不同，同樣地，會造成迴歸檢定的不穩定，有些地域過度樂觀，有些地域過度悲觀。就空間資料言，「空間自相關」與「空間異質」經常伴隨而生，這增加分析的難度。當我們發現誤差有空間聚集現象時，有可能是空間自相關引起，也有可能是空間異質引起。¹²針對這些問題，空間計量發展一套不同的方法，在模型估計上捨棄傳統慣用的最小平方法，以最大概似法替代之（下文詳述）。此處本文先針對空間分析兩個最重要的概念：「空間自相關」與「空間異質」，略作說明。

一、空間自相關

空間自相關簡單地說，就是樣本不獨立，而這種不獨立與空間上的相對位置有關。為簡化問題，便於分析，吾人經常假定此種「相關關係」呈「空間上的穩定狀態」。簡單地說，樣本與樣本間的相關程度是距離的函數，亦即只與距離有關，與其他因素（例如方向性、絕對位置）無關。與距離的關係，可能是與距離成反比，也可能是與距離的平方成反比，或是其他指數的關係。如式 1 所示， y_1 與 y_2 的共變數是距離的函數 $f(d_{12})$ ：

$$(1) \quad Cov(y_1, y_2) = f(d_{12})$$

經過這樣的簡化，空間自相關問題變成「地區」與「鄰近地區」間的相關關係。相關的強度僅與兩地區間的距離有關，與絕對位置無關。

¹² 除了上述提及的問題以外，還有兩個問題也值得注意。第一、因為是集體資料，個別歧異被消除。例如我們可以比較台灣不同鄉鎮的所得資料，但是無法察覺鄉鎮內的差異，因為被平均掉了。第二個問題是著名的「生態學的誤謬」，由於分析的單位是「地區」，我們可以知道「地區」屬性（變數）間的關係，例如所得高的地區，也是犯罪率高的地區。但是我們不知道到底是「哪些人」（所得高或低）真的犯罪。很可能的情形是：所得高的地區吸引外地的犯罪集團進入，也可能是區域內的低所得人就近犯罪。所以所得高低與犯罪率高低就「地區」的層次是相關的，但就「個人」的層次其實不相關。這是分析時要注意的。

距離愈近，相關關係愈強；距離愈遠，相關關係愈弱，遞減的方式由 $f(d_{ij})$ 函數決定。當然，實際的狀況遠比此種簡化的情形更為複雜，亦即情況通常是不穩定的。

為了界定「鄰近地區」，實務上，相對於某個目標地區，將其他地區區分為「鄰近」與「非鄰近」兩類，目標地區的屬性資料（變數值的大小）會受到「鄰近地區」的影響，與「非鄰近地區」則不相干。此種二分法大幅簡化問題，讓分析更為簡單。然而「二分法」的「標準」，亦即如何界定鄰近關係，便成為重要的課題，因為不同的界定方式，很可能會產生不同的結果。

最常見的「鄰近關係」界定方式為以共同邊界為準，亦即具有相同邊界的周邊地區定義為「鄰近」，否則定義為「非鄰近」。如此任何地區都可以給定一個「鄰近地區」的集合。亦可以距離為門檻，方圓一段距離以內的地區稱之為「鄰近」，超過此距離者則不是；或是以距離最近的 K 個地區（ K 為自訂常數）稱為「鄰近」。不管以何種方式界定「鄰近」，理論上，每個地區都可以找到屬於自己的鄰近地區，除了這些地區以外，則是非鄰近地區（鄰近地區的個數當然也可能為零，例如海中的孤島）。

「鄰近地區」的定義清楚了，鄰近地區的表現稱為「鄰近變數」，其值以各鄰近地區屬性變數的值的加權平均。以犯罪率為例，每個地區都有一個犯罪率，其鄰近地區犯罪率的加權平均，即是鄰近地區的犯罪率。至於加權平均的方式，可以簡單，也可以複雜。最簡單的方法是算數平均值，即放相同的權數；然亦可以其他變數加權之，例如以共同邊界的長度、核心點間的距離、或某個經社變數的近似度（例如族群比例）加權之。原始變數（犯罪率）以 Y 表示，鄰近變數以 WY 表示（鄰近地區犯率的加權平均）。有了 Y 與 WY 兩個變數，衡量相關程度就不困難了。一般的作法是用Moran's I 指數。Moran's I 的定義如下：

$$(2) \quad I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

上式中， n 為觀察的地區數， x_i 和 x_j 為地區 i 和地區 j 的變數值， $\bar{x}$ 為該變數在所有地區中的平均值， w_{ij} 代表地區 i 和地區 j 的空間鄰近 (Spatial Contiguity) 關係，如果地區 i 和地區 j 相鄰其值為 1，否則為 0。

其實 Moran's I 指數與相關係數 R 相當類似，目的都在衡量兩個變數的相關程度。其值都是介於 1 與 0 之間，正表正相關，負表負相關。Moran's I 散布圖每個點代表一個地區，橫軸是標準化之後的變數 Y ，縱軸為相應的鄰近變數 WY ，迴歸線斜率即是 Moran's I 指數。如果 Y 與 WY 都標準化，則斜率就是相關係數 R 。

Moran's I 散布圖的四個象限各有不同意義，第一象限代表目標地區的變數值大於平均值，鄰近地區的變數值亦然。同理類推，第二象限代表目標地區小於平均值，鄰近地區大於平均值；第三象限代表目標地區小於平均值，鄰近地區亦小於平均值；第四象限代表目標地區大於平均值，鄰近地區小於平均值。離中心點愈遠，代表此種趨勢愈明顯。習慣上，我們稱第一象限與第三象限為空間上的穩定區，第二象限與第四象限為不穩定區，因為第二象限周圍比自己高，很可能有個力量把自己拉高；第四象限自己比周圍高，亦有股力量把自己往下拉，二、四象限的亦稱為空間的歧異值 (Spatial outliers)。

如果 Moran's I 散佈圖的橫座標與縱座標指涉的是不同的變數，即屬雙變數的 Moran's I 分析。例如橫軸為某年的犯罪率，縱軸為前一年犯罪率 (或人口密度等其他變數) 鄰近地區的平均。比較不同變數的 Moran's I 散佈圖，可觀察變數與其周邊較複雜的關係。若 Moran's I 散佈圖 XY 軸分屬不同的時期，則有助於我們判斷傳播的速率。

不同的鄰近定義，會影響計算出來的 Moran's I 值大小。空間資料經常有空間自相關現象，基本的解釋是相鄰的地區間必定有某種型式的

互動或交流，因為此種互動與交流，促成了自相關現象。(Haining, 2003: 21-22) 因此，「鄰近」的定義須依研究目的而調整，絕非一成不變。例如，一般以車行一小時之範圍作為界定鄰近的準據，即假定在此距離內人際網絡易於建構。如果界定鄰近的距離不斷擴大，依序算出不同的 Moran's I 值，則可觀察不同的鄰近定義對 Moran's I 值的影響，此即 correlogram 圖，通常 Moran's I 會依序遞減至接近零。

Moran's I 可以進行檢定，虛無假設是沒有空間自相關 (Moran's I 值接近零)，對立假設是有空間自相關 (Moran's I 值不等於零)。Moran's I 的抽樣分佈標準化之後接近 z 分佈 (標準常態分佈)，這個性質可作為檢定的依據。式 3 與式 4 為虛無假設下 Moran's I 的期望值與變異數。

$$(3) \quad E(I) = -(n-1)^{-1}$$

$$(4) \quad \text{Var}(I) = \frac{1}{(n-1)(n+1)(\sum \sum W_{ij})^2} \times \left[n^2 S_1 - n S_2 + 3 \left(\sum \sum W_{ij} \right)^2 \right] - \frac{1}{(n-1)^2}$$

$$S_1 = \frac{1}{2} \sum \sum (W_{ij} + W_{ji})^2$$

$$S_2 = \sum \sum (W_i + W_j)^2$$

由上述公式計算實際觀察的 Moran's I 標準化後的 Z 值，Z 值愈大愈可能拒絕虛無假設，我們還是用 $P=0.05$ 作為臨界點，當 P 值小於 0.05，有足夠的證據認定空間自相關存在。事實上，空間自相關現象極為普遍，Moran's I 檢定多半會成立。比較 Z 值大小，則可比較空間自相關程度的大小。

Moran's I 指數關心的是整體 (Global) 的空間自相關趨勢。如果我們關心的是某個特定的地區與其鄰近地區的相關關係，則可選擇 Local Moran's I 指數 (當然還有其他選擇¹³)，Local Moran's I 是衡量個別地區與其鄰近地區相關程度的指標，如果其值為正，代表正相關，為負表負相關。其定義為：

¹³ 其他衡量 Local 空間特性的指標，例如 G_i G* 等。

$$(5) \quad I_i = \frac{x_i - \bar{x}}{\sum_i (x_i - \bar{x})^2} \sum_j W_{ij} (x_j - \bar{x})$$

Local Moran's I 的顯著性檢定與 Global Moran's I 類同。通常有些地區的空間自相關現象比較明顯，其他地區則不是那麼明顯。LISA cluster 則將達到顯著水準的地區以不同的色塊區分出來，「正正」（紅色）、「負負」（藍色）、「正負」（淡紅色）、「負正」（淡藍色）四種類型。「正正」指自己高、鄰近地區亦高，「負負」指自己低、鄰近地區亦低，「正負」指自己高，鄰近地區亦低，「負正」指自己低，鄰近卻高的地區，而且均達顯著水準。LISA cluster 可對照 Moran's I 散佈圖觀察，愈遠離原點者愈可能達顯著水準。觀察不同時間點 LISA cluster 的變化，可瞭解空間結構隨時間而變化的模式。

二、空間異質性

空間異質性與傳統計量經濟學處理的異質性一樣，都是指誤差的變異不恆定。傳統迴歸的誤差異質問題常起因於某個自變數，誤差的變異是這個自變數的函數，亦即誤差的變異隨著自變數的變化而變化。此時估計值喪失有效性，t 檢定變得不穩定，有時過度樂觀，有時過度悲觀。空間異質性則指因為空間位置造成的異質現象。造成空間異質的原因有許多，例如，可能是量度本身的問題所造成，例如鄉鎮犯罪率的高低，有的鄉鎮人口多，有的鄉鎮人口少，則人口多的地區量度的犯罪率變異會比較小，反之鄉鎮人口少的地區變異會比較大，此時可以用加權迴歸解決。¹⁴另一個造成空間異質的原因是變數關係的空間不穩定性，例如知識水準與投票行為的關係，北台灣與南台灣可能不一樣，因此合併所有資料進行迴歸分析時，南北地區的誤差變異會不同，此時不同地區分

¹⁴ 一般以人口數取根號的倒數加權之。

開處理會比較妥當；當然空間異質也可能是地區本身的獨特性造成，例如自變數與因變數的關係，在台灣本島可能關係顯著，在離島卻不存在，因此誤差的變異會隨著不同地區而有不同。空間結構的不穩定造成的異質現象，一般用 Spatial Chow test 檢定之。值得注意的是：光看誤差的 moran's I 散佈圖，無法分辨是空間自相關造成，或空間異質使然（亦即 moran's I 檢定呈顯著）。如果是空間自相關引起的誤差分佈聚集，加空間變數（ W_y ）後，誤差群聚現象可消除。如果是空間異質引起，加空間變數無法消除誤差的空間聚集現象，必須考慮其他方法。

三、空間迴歸模型

空間分析想回答的問題是：空間上的變異是如何造成的？例如選票獲票率的空間分佈不是隨機，如何解釋此種空間上的變異？進而討論空間自相關或空間異質存在的可能性。針對這個問題，空間迴歸模型是有利的分析工具。從空間計量的觀點，空間變異的產生可有幾個原因：

第一、某個空間結構類似的自變數引起，這呼應前文提及的「成分影響觀點」。如果是這樣的，加進這個自變數之後，殘差的自相關現象會被消除，因為被該自變數解釋掉了。此時用最小平方法估計即可。

第二、是空間異質引起的，這呼應前文提及的「地區脈絡觀點」的重要性。如果是空間異質引起，則加進所有可能的自變數之後，誤差仍然具有自相關。此情況下，如果用最小平方法估計，仍可能不偏的估計，但不具有有效性，亦即傳統的 t 檢定、 F 檢定無法使用。此時，如果誤差變異已知可用加權迴歸，然通常未知。在某些特殊的情況下，增加地區的虛擬變數，或其與自變數的乘積作為自變數（交互作用），亦可有效消除空間異質；不同區域分別處理亦是可行的辦法。如果空間異質被成功消除，殘差不再自相關，可用最小平方法估計之。

第三、是某個被忽略的自變數引起，而這個自變數與誤差有類似的空間結構，這亦屬於「成分影響觀點」，只不過找不到「隱藏」的變數。如果我們忽略此種現象逕行用最小平方方法估計，則無法得到不偏的估計量，理想的情況是把這個被忽略的自變數找出來，如果找不到或無合適的數據，可用「空間誤差模型」估計。此時，空間變數的影響會達顯著水準，誤差的自相關現象消除，或大幅降低。

第四、是「鄰近效應」引起，亦即鄰近地區的表現會影響變數值大小。此時用「空間落遲模型」，加進空間變數（ W_y ，鄰近地區表現的加權平均）作為自變數，則誤差的自相關現象可有效消除。如果忽略空間因素的影響，逕用最小平方方法估計，則估計值有偏誤、且不具一致性，亦即樣本數增加，偏誤仍然存在。

實務的做法是先尋找可能的自變數，跑傳統最小平方方法迴歸，檢查殘差是否有自相關現象（Moran's I 檢定是否拒絕虛無假設）。如果有的話，就必須考慮異質性問題，或進行空間迴歸分析。空間迴歸分析最常見的是「空間落遲模型」（SLM）與「空間誤差模型」（SEM），以下略作說明。值得注意的是這兩個模型的參數估計都是用最大概似法進行。

空間落遲模型（SLM）

如果誤差自相關現象是鄰近效應引起的，此時可採用「空間落遲模型」（SLM），亦即增加一個自變數 W_y 代表鄰近地區的表現，如式六所示。

「空間落遲模型」的向量表示式為：

$$(6) \quad y = \rho W_y + X\beta + \varepsilon$$

y 是因變數， X 為傳統自變數， ε 是誤差，為 iid，亦即無自相關，變異齊一。 ρ 和 β 為迴歸係數。 W 稱為空間「鄰近矩陣」（Spatial Contiguity Matrix），為一右上左下對稱的方形矩陣，欄與列數都等於地

區個數（例如本研究共有 385 個鄉鎮）。「鄰近矩陣」用來定義鄰近關係，如果兩地區定義為「鄰近」其值為 1，否則為 0，對角線也為 0（自己和自己不具鄰近關係）。不過在這裡，鄰近矩陣經過列標準化，亦即使列總合為 1（亦即列中不等於 0 的元素值代之以 1 除以該列不等於 0 的元素個數）。此時， Wy 其實就是鄰近地區平均值，為一欄矩陣。

如果 ρ 等於 0，則與傳統迴歸無異，否則表示控制了傳統變數之後，空間變數 Wy 仍有解釋力。「空間落遲模型」與時間序列的 AR(1) 模型有點類似，不同的是 AR(1) 模型只要誤差沒有自相關，可以最小平方估計。「空間落遲模型」的空間變項（ Wy ），如同內生變數（endogeneity），並非固定，若以最小平方估計則會失去不偏性、有效性，不管誤差的性質為何。（Anselin, 1988: 58-9; 2004: 11）此時必須用工具變數或最大概似法估計，不過後者假設誤差常態，前者則無此假設。

式 6 經過移項後，如式 7 所示，

$$\begin{aligned}
 Y &= \rho WY + X\beta + \varepsilon \\
 Y - \rho WY &= X\beta + \varepsilon \\
 (7) \quad Y(I - \rho W) &= X\beta + \varepsilon \\
 Y &= (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon
 \end{aligned}$$

GeoDa 軟體以最大概似法估計參數 $\hat{\beta}$ 、 $\hat{\rho}$ 。式 7 的最後一個等式左邊第一項為 Y 估計值（predict），第二項為預測誤差（predict error）。加入空間變數之後的誤差為 ε ，定義如下：

$$\begin{aligned}
 \text{predict} \quad \hat{Y} &= (I - \hat{\rho}W)^{-1} X\hat{\beta} \\
 (8) \quad \text{predict error} \quad Y - \hat{Y} &= (I - \hat{\rho}W)^{-1} \hat{\varepsilon} \\
 \text{residuals} \quad \hat{\varepsilon} &= Y - \hat{\rho}WY - X\hat{\beta}
 \end{aligned}$$

predict error 為在未放進空間變數 Wy 之前的誤差，其共變矩陣（covariance-variance matrix）諸元素均不為零，代表有空間自相關及異質現象。加入空間變數之後，誤差 ε 空間自相關現象消除或大為降低，因為已被空間變數 Wy 解釋掉了。

空間落遲模型的檢定，整體適合度方面常用：Log likelihood（愈大愈好）¹⁵、Akaike Info Criterion（愈小愈好）¹⁶、Schwarz Criterion（愈小愈好）¹⁷、R-squared、Adjusted R-squared（僅供參考，無可解釋百分比的意義）。

個別自變數顯著檢定，常用 Wald test。Wald 統計量的定義如下，它接近自由度為 1 的卡方分佈：

$$(9) \quad W = \left(\frac{\hat{\beta}_k}{S_{\hat{\beta}_k}} \right)^2 \sim \chi_1$$

$\hat{\beta}_k$ 為自變數係數的估計值， $S_{\hat{\beta}_k}$ 為估計值抽樣分佈的標準差。估計值的漸進變異矩陣 (the asymptotic Variance matrix) 一般以 the information matrix, $I(\theta)$ 的反矩陣表示之。 $I(\theta)$ 各元素值求法為：分別以各估計值對概似函數取二次微分。若 W 值夠大，P 值小於 0.05，表達顯著水準，自變數具解釋力。(Anselin, 1988: 64-5) ¹⁸

空間誤差模型 (SEM)

空間誤差模型 (Spatial error model) 空間變數出現在誤差項，表示誤差可以再進一步分析，造成原因可能是忽略了某個重要的自：

$$(10) \quad \begin{aligned} Y &= \beta X + \varepsilon \\ \varepsilon &= \lambda W \varepsilon + u \end{aligned}$$

兩式合併：

$$(11) \quad Y = \beta X + \lambda W \varepsilon + u$$

誤差 ε 不是 iid, 加入空間變數 ($W\varepsilon$) 之後，誤差 u 為 iid. 其實空間誤差模型類同於時間序列 MA(1) 模型。不同的是前者誤差共變矩陣各元

¹⁵ 概似函數取 log 值 (log-likelihood)，其值愈大表概似程度愈高，迴歸方程式愈佳。因為是負值，所以絕對數字愈小愈好。

¹⁶ AIC 是比較不同迴歸方程式優劣的重要指標。AIC 是正值，愈小表迴歸方程式愈佳，其定義如下：AIC = $-2 \log L + 2K$ 。 $\log L$ 為概似函數取 log 值 (log-likelihood)，乘以 -2 後由負值變為正值，K 為參數個數。

¹⁷ SC 也是用來比較不同迴歸方程式優劣。SC 值愈小表迴歸方程式愈佳，其定義如下：SC = $-2 \log L + K \cdot \ln(N)$ 。 $\log L$ 為概似函數取 log 值 (log-likelihood)，N 為樣本數。

¹⁸ GeoDa 提供 Wald 數值，報表 Z-value 欄值的平方，即是 Wald 值。

素均不為零，亦即自相關與異值性同時存在，而後者則否。「空間誤差模型」(SEM)以最小平方方法估算，雖仍具不偏性，但失去有效性。(Anselin, 1988: 58-9; 2004: 11)

參數估計以最大概似法進行，估計出 $\hat{\beta}$ 、 $\hat{\lambda}$ 。GeoDa 除了估計係數外，同時產生三個新變數：Y 估計值、預測誤差、殘差，分別定義如下：

$$\begin{aligned} \text{predict} \quad \hat{Y} &= X\hat{\beta} \\ (12) \quad \text{predict error} \quad \hat{\varepsilon} &= Y - \hat{Y} = (I - \hat{\lambda}W)^{-1}\hat{u} \\ \text{residuals} \quad \hat{u} &= \hat{Y} - X\hat{\beta} - \hat{\lambda}W\hat{\varepsilon} \end{aligned}$$

預測誤差 ($\hat{\varepsilon}$) 有空間自相關現象，殘差 ($\hat{u}$) 則無，或大幅降低。空間誤差模型的檢定與空間落遲模型類同，此處不重覆。

四、迴歸模型的選擇

目前我們提及三個迴歸模型：傳統迴歸 (Classic)、空間落遲迴歸 (SLM)、空間誤差迴歸 (SEM)，三者要如何選擇呢？Luc Anselin 提出一個簡單的法則，本文也依循同樣的模式：

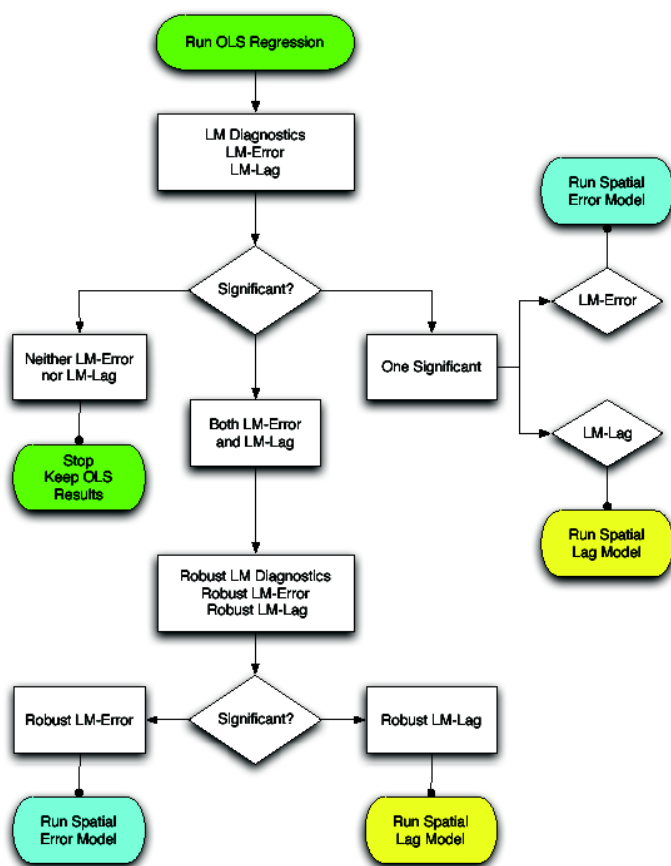
首先，以傳統 OLS 估算迴歸係數。

第二、檢查殘差的空間自相關是否顯著，若是，則考慮用空間迴歸模型；若否，則不需要，傳統迴歸為已足。

第三、分別對 W_y ， W_ε 進行 Lagrange Multiplier 檢定。Lagrange multiplier 檢定用來判斷新增一個自變數其影響是否顯著，若顯著方有加進新變數的必要。分別對 W_y ， W_ε 進行 Lagrange multiplier 檢定，如果前者顯著，則建議用 SLM，如果後者顯著，則建議用 SEM。如果兩者都顯著，則建議繼續進行 Robust (The Likelihood Ratio Test) 檢定。

第四、Robust 檢定的基本想法是：拿掉某(個)些自變數之後，看看顯著水準是否有異？如果沒改變，表示這個自變數沒什麼用處，不如拿掉。GeoDa 的提供 Robust LM(lag)與 Robust LM(error)兩種 Likelihood

Ratio 檢定。前者針對 W_y 進行檢定，後者針對 W_ε 進行檢定，通常只有一個會顯著。如果前者顯著，則建議用 SLM；如果後者顯著，則建議用 SEM。模型選擇的流程如下圖所示：



(摘自 Anselin, 2005: 199)

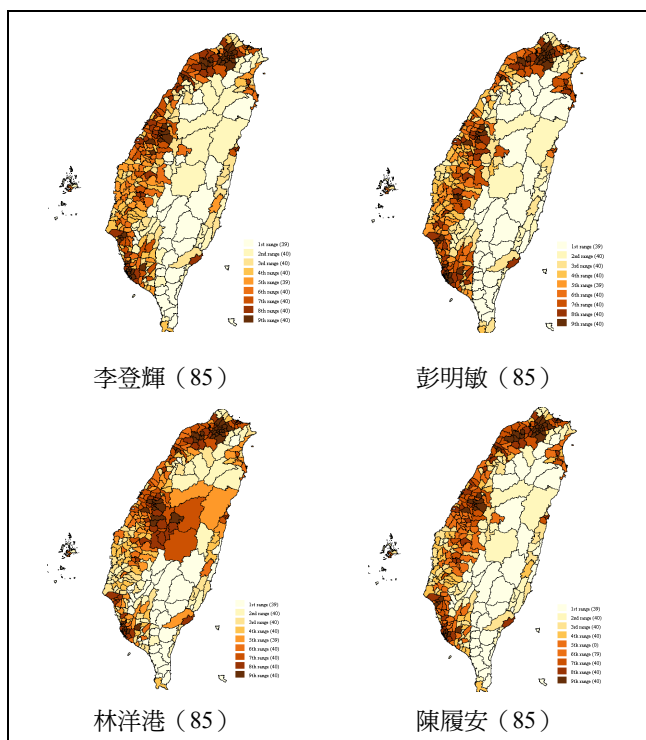
肆、三次總統選舉的空間探索分析

本節作者以三次總統選舉為例進行「空間探索分析」(Exploratory Spatial Data Analysis, ESDA)，意即在透過統計量、圖表，發掘總統選舉的空間特性或隱藏的空間模式，通常這是進行空間迴歸分析的前置作業。

一、歷次總統選舉的空間特性

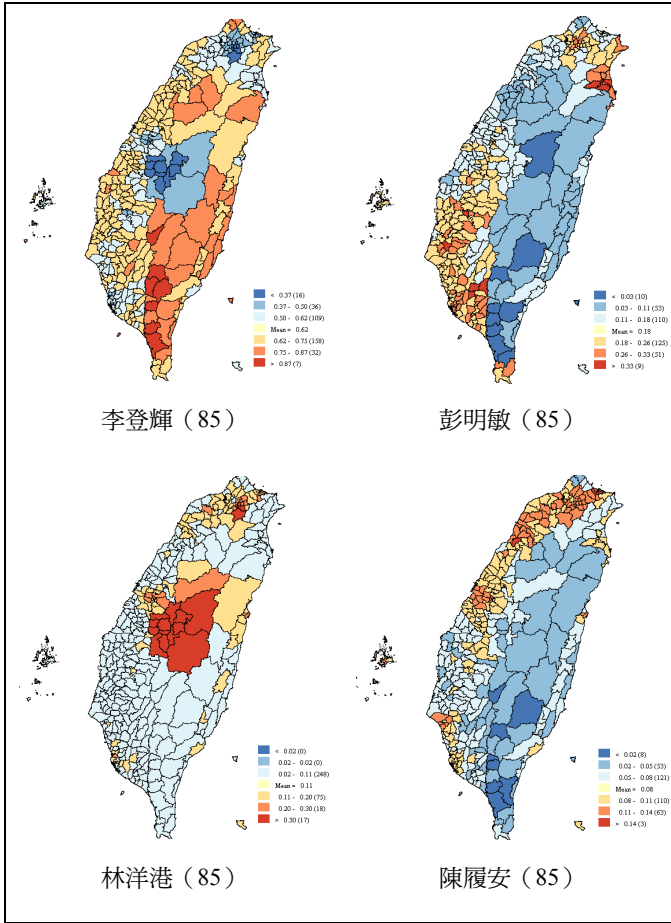
民國八十五年第一次總統直接民選，四組人馬角逐，分別為國民黨李登輝、民進黨彭明敏，新黨林洋港、獨立候選人陳履安，選舉結果獲票數（率）分別為 5773257（59.94%）、2265517（23.52%）、1593330（16.54%）、1063361（11.04%）。李登輝以過半的高票當選。圖一為各候選人得票數分佈圖，358 個鄉鎮為分析單位，分為九個等級，顏色愈深表示得票數愈多，顯示各候選人得票數的地域分佈。我們可以發現除林洋港（集中在南投地區）外，其餘三人的票數分佈差異不大。主要原因是台灣人口聚集在西半部平原地區，人口密度高，自然獲票數多，不足為怪。而林洋港的獲票情況，反應他與南投的特殊淵源¹⁹。

¹⁹ 林洋港是南投人（南投縣魚池鄉），曾任南投縣縣長（1967~72）。



圖一 第一次總統選舉各黨獲票數主題圖

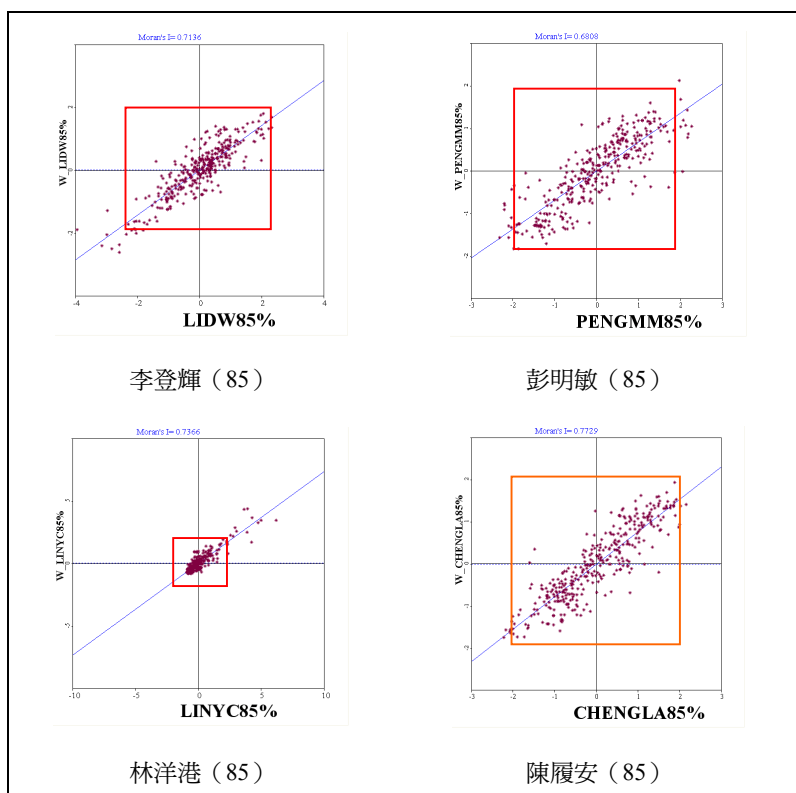
為顯示候選人獲票狀況的空間差異，圖二以獲票率作主題圖，經過標準化之後，紅色系代表大於平均值之地區，藍色系代表小於平均值之地區，顏色愈深代表獲票率愈高。四位候選人鄉鎮平均獲票率為：李登輝 62%、彭明敏 18%、林洋港 11%、陳履安 8%，與實際獲票率稍有差異，但不大。圖中可清楚顯示各候選人在各鄉鎮區的獲票率高低分佈。李登輝除了北高兩市、宜蘭、南投縣低於平均值外，其他地區呈均勻分佈；彭明敏則集中在南部各縣市及宜蘭地區、林洋港主要選票來自南投、陳履安則是北中南都會區。李登輝獲得過半選票，其他候選人瓜分剩餘選票，在各鄉鎮的獲票率基本上不大。



圖二 第一次總統選舉各獲票率（標準差）主題圖

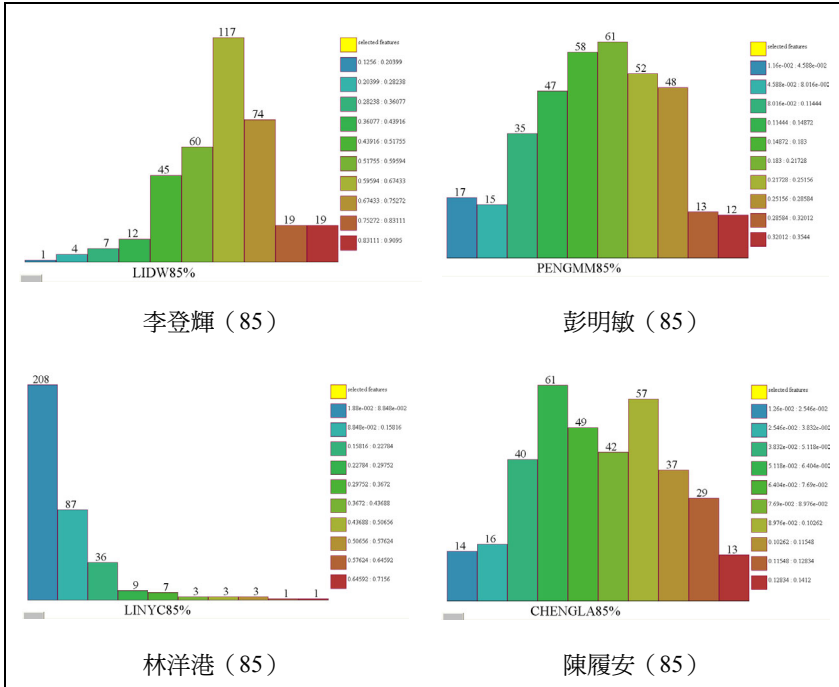
圖三以 Moran's I 散佈圖顯示各候選人的空間結構，每一個點代表一個鄉鎮區，橫座標為目標鄉鎮經過標準化的獲票率，縱座標為目標鄉鎮鄰近鄉鎮獲票率（經過標準化）的平均值，鄰近的界定為有邊界接壤的地區（以下均同）。回歸線的斜率為 Moran's I 值，Moran's I 值愈大代表選票聚集狀況愈明顯。紅色框為兩倍標準差之範圍，紅框外之點代表歧異值（outlier）。Moran's I 散佈圖有四個象限，代表目標區與鄰近地區

的四種不同模式，第一象限代表自己高（於平均）鄰近地區亦高，第二象限代表自己低，鄰近地區高，第三象限代表自己低，鄰近地區也低，第四象限代表自己高，鄰近地區低。點粒離原點愈遠，代表此種趨勢愈明顯。第一、三象限為「穩定區」，第二、四象限為「不穩定區」，顯然大部分的點粒散佈在「穩定區」。空間的穩定與實際的穩定未必一致，下文中我們會再說明。四位候選人的 Moran's I 散佈圖，最特殊的要屬林洋港，因為 outliers 都是在第一象限，表示其票數高度集中在特定地區。



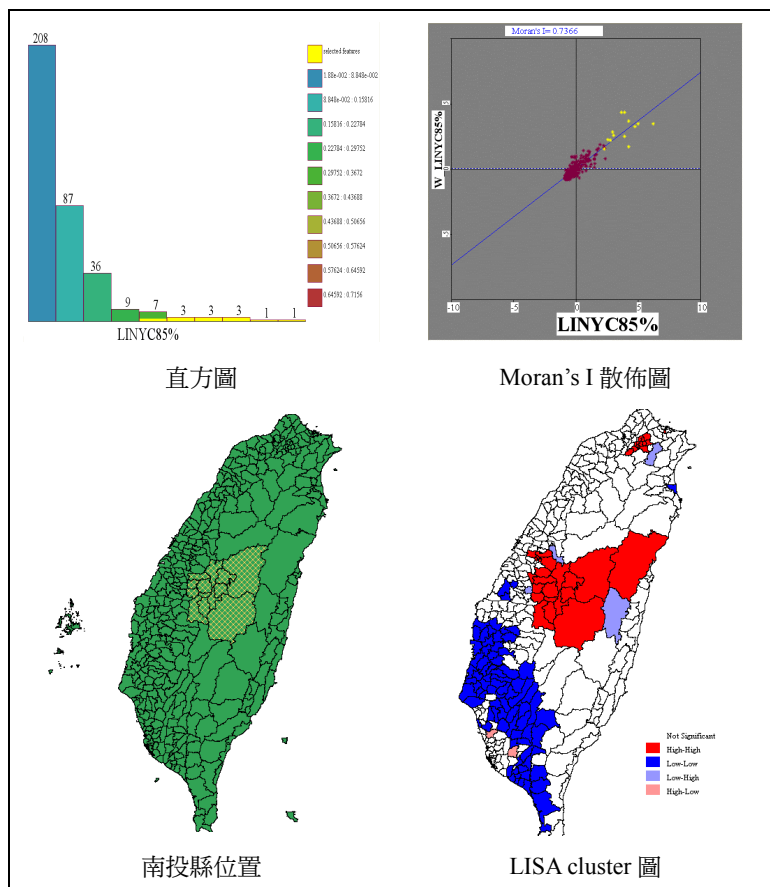
圖三 第一次總統選舉各參選人 Moran's I 散佈圖

圖四的直方圖，顯示四位候選人的獲票率分佈，紅色系（右尾）代表得票率高的地區，藍色系（左尾）代表低的地區。圖四更清楚地顯示林洋港的特殊性。林洋港的直方圖右尾特長，代表得票率高的地區集中在少數幾個地方。李登輝的直方圖呈微微的左偏（左尾較長），代表有些地區，特別討厭他（得票率超低），另兩位候選人則大致呈鐘形分佈。



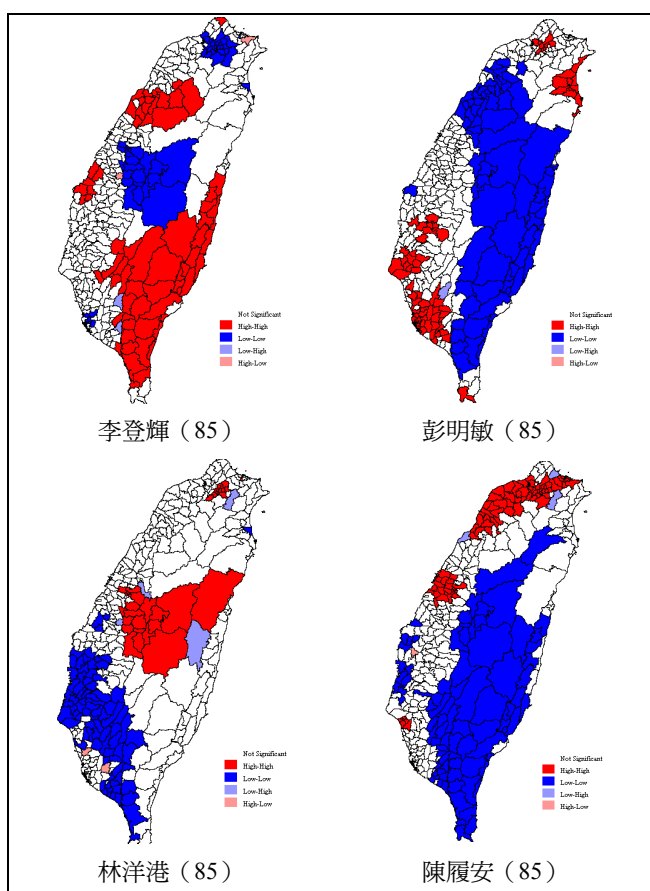
圖四 第一次總統選舉各黨獲票率直方圖

關於比較特殊的林洋港，我們用四個圖相互對照，以突顯此種現象，如圖五所示。左上為直方圖、右上為 Moran's I 散佈圖，左下為南投縣位置圖，右下為 LISA cluster 圖。直方圖以及 Moran's I 散佈圖，黃色部分代表南投縣。LISA cluster 圖的紅色部分代表 HH（自己高鄰近亦高，且達顯著的水準），這些黃色與紅色色塊全都指南投縣，清楚顯示林洋港的地緣特性。



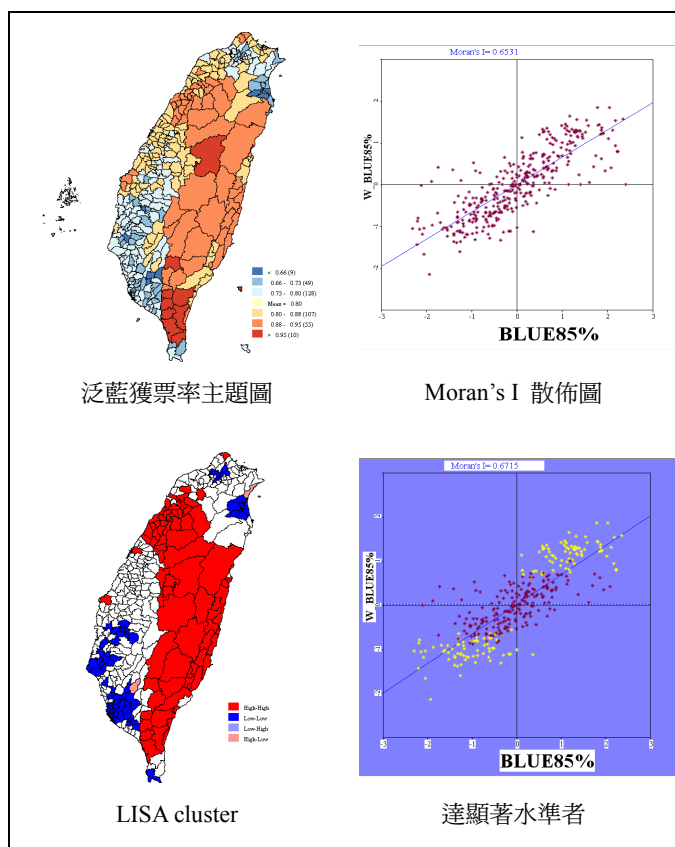
圖五 第一次總統選舉林洋港與南投縣的選票分佈

圖六為四位候選人的 LISA cluster 圖，紅色代「正正」、藍色「負負」，淡藍色「負正」，淡紅色「正負」，白色代表未達顯著水準的地區。顯然大部分的地區都是「正正」與「負負」(穩定區)，「正負」與「負正」(不穩定區)並不多見。比較四位候選人的 LISA cluster 圖，可以輕易發現候選人各擅勝場，各有各的「勢力範圍」。

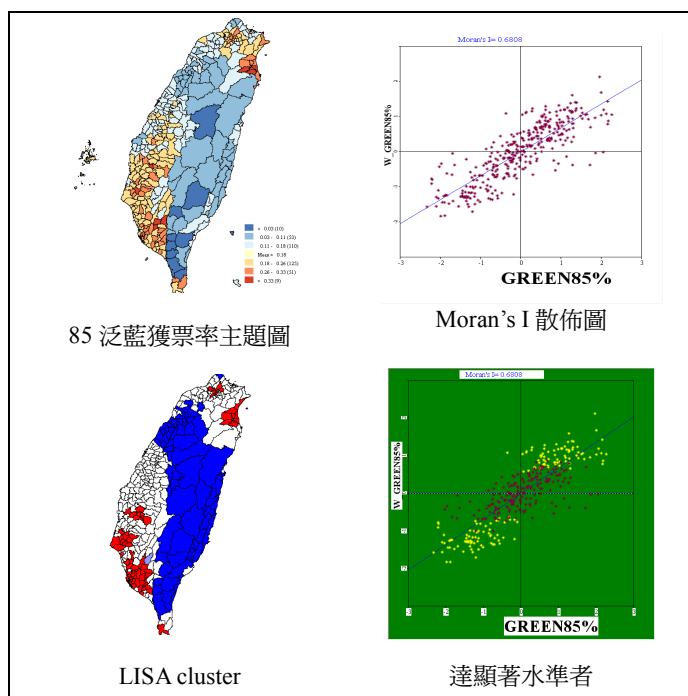


圖六 第一次總統選舉各參選人 LISA cluster 圖

如果將李登輝、林洋港、陳履安三者的獲票率加總起來則得圖七。左上為藍營獲票率主題圖、左下為 LISA cluster，右上、右下均為 Moran's I 散佈圖，唯一的差別是右下的散佈圖達顯著水準者以黃點標出，這些點粒都是在距離原點較遠的地區。同理，圖八為綠營的圖表，我們可輕易發現圖七與圖八的圖形互補（藍變綠、綠變藍），這一點也不奇怪，因為藍綠加起來是百分之一百，兩者呈消長關係，圖塊顏色所以互調。在下面的分析中我們會發現藍、綠的空間結構其實是相當穩定的。

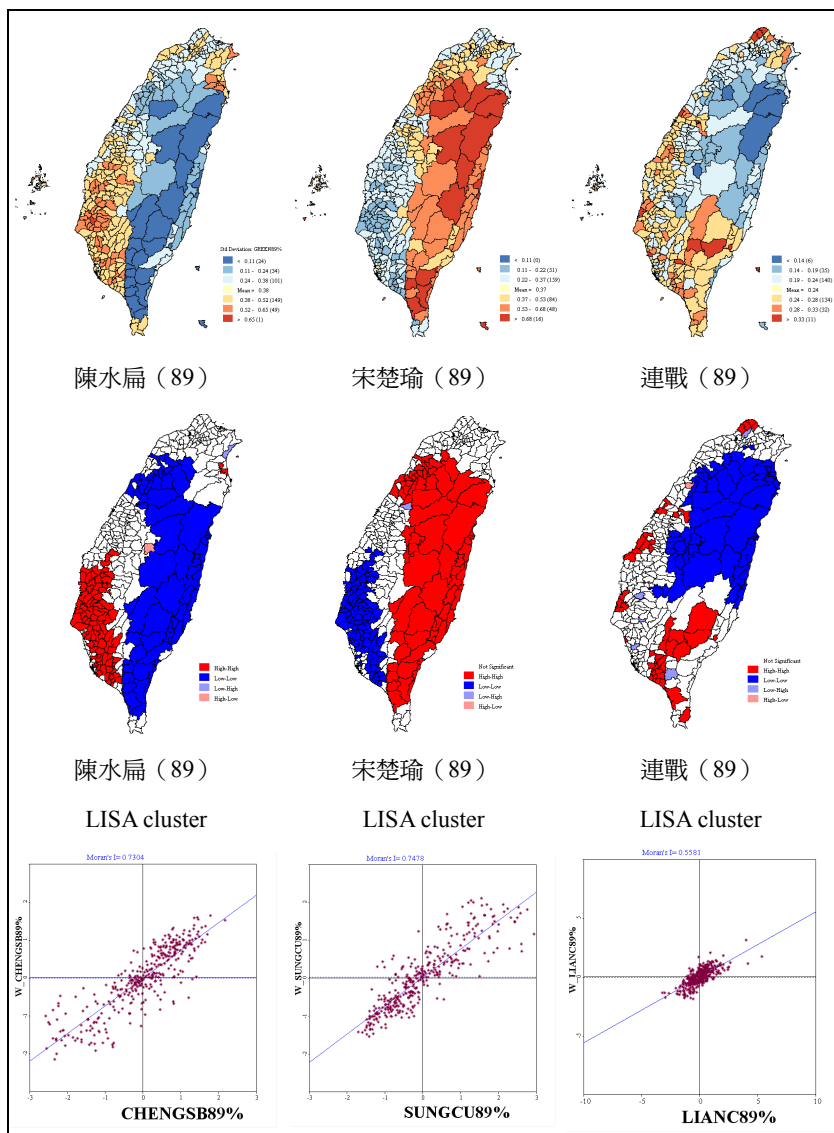


圖七 第一次總統選舉藍營獲票率空間分佈



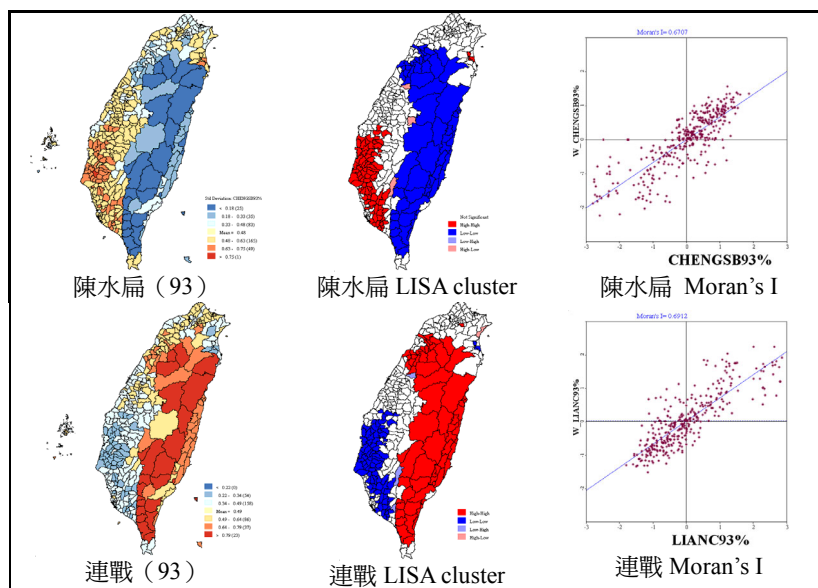
圖八 第一次總統選舉綠營獲票率空間分佈

第二次總統選舉（民國八十九年），有五組候選人參加，分別是民進黨陳水扁、親民黨宋楚瑜、國民黨連戰、獨立參選人許信良、新黨李敖，獲票數（率）分別為 4959744（39.41%）、4621265（36.72%）、2908992（23.11%）、79187（0.63%）、16669（0.13%）。由於後兩者得票率太低，不影響大局，我們只討論前三者。圖九為連戰、陳水扁、宋楚瑜的獲票率主題圖、LISA Cluster 圖及 Moran's I 散佈圖。從主題圖、LISA cluster 圖都可輕易發現三者票源分佈與聚集情況。值得注意的是連戰僅得百分之二十三的選票，這是國民黨從來沒有過的大挫敗，主要原因是宋楚瑜與連戰同屬泛藍陣營，因內閥分裂，票源分散，讓民進黨有機可乘，以不到百分之百四十的選票當選總統。就選票集中度方面，陳水扁與宋楚瑜明顯高於連戰，三者 Moran's I 值分別是 0.7304, 0.7478, 0.5581。



圖九 第二次總統選舉各黨獲票率的空間分佈

第三次總統選舉（民國九十三年）國民黨記起失敗教訓，整合成功，兩黨對決，但仍落敗，獲票數（率）分別為民進黨陳水扁 6456065 票（48.97%）、國民黨連戰 6393594 票（48.49%）。民進黨的獲票數大幅攀升，從 39.41%提高到 48.97%，最後陳水扁以極小勝差 62471 票（0.47%）當選，不過，令人疑惑的是廢票高達 334926 張，佔百分之 2.54%，再加上選舉期間的緊張對立，種下往後政治紛爭的因子。圖十為兩黨獲票率的空間分佈圖，上半部代表民進黨，下半部為國民黨。可發現，民進黨與國民黨的傳統地域大致不變，前者還是集中在中南部和宜蘭地區，後者則在桃竹苗山區以及東部一帶。由 Moran's I 散佈圖發現，民進黨的支持區（第一象限）比較集中，反對區（第三象限）比較分散，國民黨則剛好倒過來，支持區（第一象限）比較分散，反對區（第三象限）比較集中。事實上，在第二次總統選舉時，藍、綠的空間結構也有類似的情形。



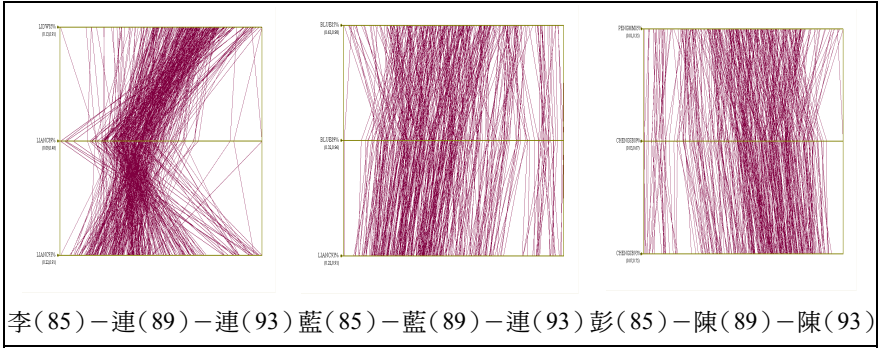
圖十 第三次總統選舉各黨獲票率的空間分佈

二、總統選舉空間特性的比較

三次總統選舉的比較，分三部分討論：第一、選票的繼承現象；第二、藍、綠版塊與空間結構的變化，第三、民進黨 93 年大幅躍升的緣由。

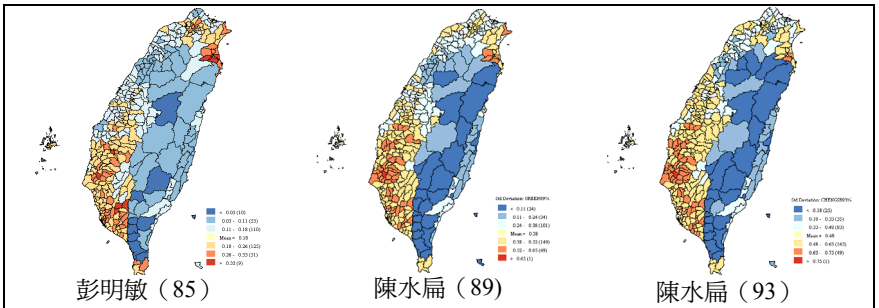
先從第一個談起，前兩次選舉國民黨始終處於分裂狀態，選舉結果卻截然不同。第一次國民黨分裂，詭異的是直接受害者不是國民黨而是民進黨。民進黨僅得 23.52% 的選票，一般認為民進黨的傳統支持者應該有三成才是，何以如此？論者謂這是「李登輝效應」。李登輝雖然代表國民黨參選，但在反對陣營中接受度亦高。李即利用此一優勢，大舉攬奪民進黨的傳統選票。第二次總統選舉時，「李登輝效應」沒有轉移到接班人連戰身上，民進黨選票歸位，陳水扁憑其個人魅力，除鞏固傳統選票外，向外突圍，獲得 39.41% 的選票，國民黨則初次嘗分裂之苦，交出執政五十年的政權。

圖十一的 PCP 圖，每條線代表一個鄉鎮區，上中下三個變數分別是三次選舉的獲票率，愈右代表獲票率愈高，反之則愈低。如果線條極少交錯，表示鄉鎮區在各變數的相對位置（排名）一致；反之，交錯多表示相對位置不穩定。例如最左之圖代表三個變數：李登輝（85 年）、連戰（89 年）、連戰（93 年）獲票率的排名關係，我們發現其間線條交錯甚多，經由此種不穩定關係，我們幾乎可斷定三者不存在繼承關係。中間的圖吾人將泛藍獲票率加總，發現線條交錯明顯變少。尤其是泛藍（89）到連戰（93）幾乎沒有什麼交錯，證明泛藍的空間結構其實還是很一致，而泛藍（85）到泛藍（89）則交錯稍多，主要還是李登輝效應，他拓展不少原來不屬於藍營的選票。最右之圖為彭明敏（85）－陳水扁（89）－陳水扁（93）三者的關係圖，彭明敏（85）－陳水扁（89）有不少交錯，因為李登輝拉走了原屬綠營的選票，陳水扁（89）－陳水扁（93）則恢復正常。雖然，泛綠獲票率從 39.41% 提升到 48.97%，但空間結構仍然穩定，亦即本來排名前面的現在仍是，反之亦然。

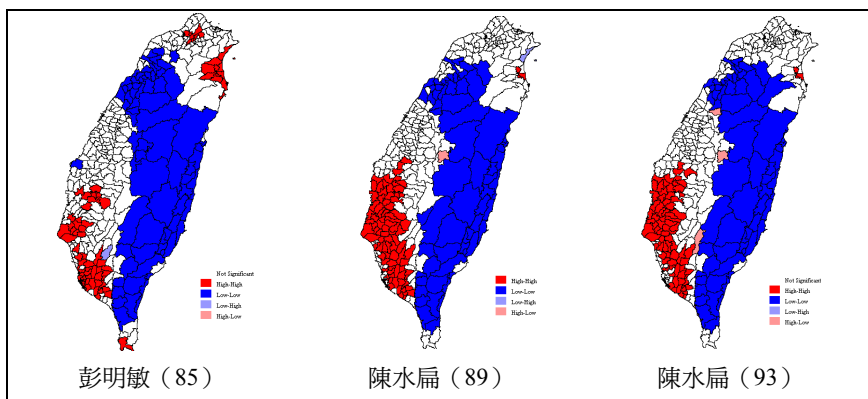


圖十一 三次總統選舉 PCP 圖

關於三次選舉藍綠的版塊變化，圖十二為彭明敏(85)－陳水扁(89)－陳水扁(93)三者的獲票率主題圖。可以輕易發現，三者非常接近。彭明敏(85)雖有李登輝效應作梗，但與其後兩次陳水扁獲票率主題圖差異不大，只是泛綠有向中台灣拓展的趨勢。圖十三為彭明敏(85)－陳水扁(89)－陳水扁(93)獲票率 LISA cluster 圖，色塊表示空間自相關達顯著水準者。彭明敏(85)與陳水扁(89)在「正正」(紅色)的部分差異較大，更進一步確認，第一次總統選舉支持者有「跑票」的現象；陳水扁(89)與陳水扁(93)幾乎完全一樣，不受陳水扁(93)獲票率大幅增長的影響。

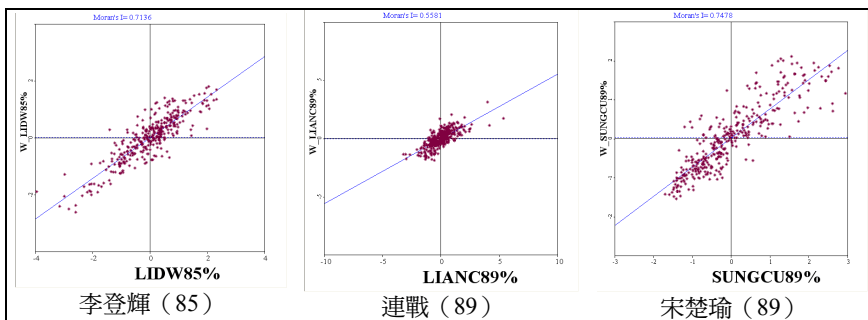


圖十二 彭明敏與陳水扁的獲票率主題圖

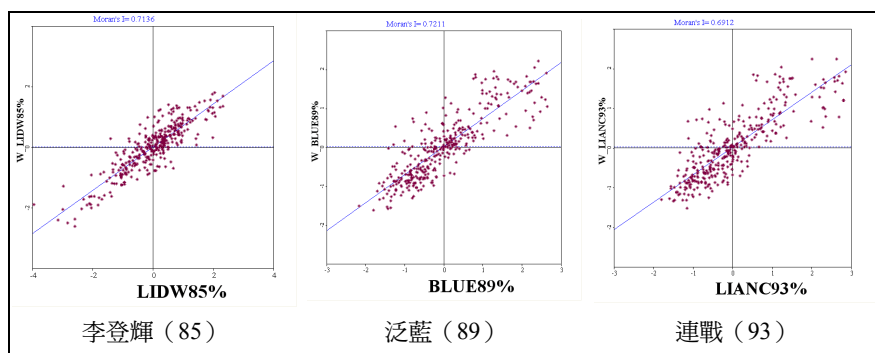


圖十三 彭明敏與陳水扁的 LISA cluster 圖

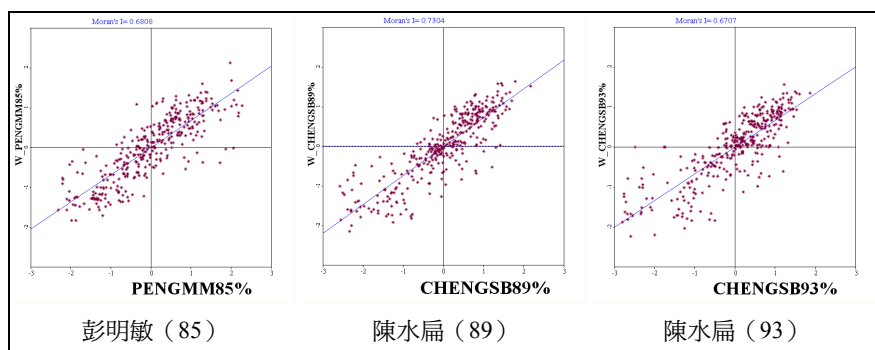
另外，從 Moran's I 散佈圖也可以區辨選票繼承的現象。圖十四為李登輝 (85)－連戰 (89)－宋楚瑜 (89) 的 Moran's I 散佈圖，不出所料，三者顯然有很大的差異；圖十五、則是歸併為泛藍之後的 Moran's I 散佈圖，除了李登輝 (85) 仍然特殊之外，泛藍 (89) 與連戰 (93) 則高度一致。圖十五為彭明敏 (85)－陳水扁 (89)－陳水扁 (93) 獲票率 Moran's I 散佈圖，三者相當一致，特別是陳水扁 (89)－陳水扁 (93) 幾乎是一樣，證明獲票率各有增減，但空間結構仍屬穩定



圖十四 Moran's I 散佈圖與泛藍選票繼承 (一)

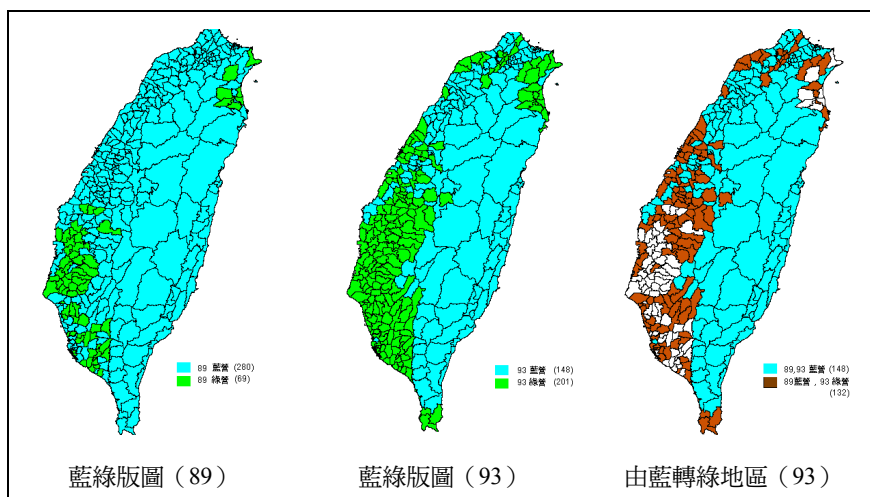


圖十五 Moran's I 散佈圖與泛藍政黨繼承 (二)

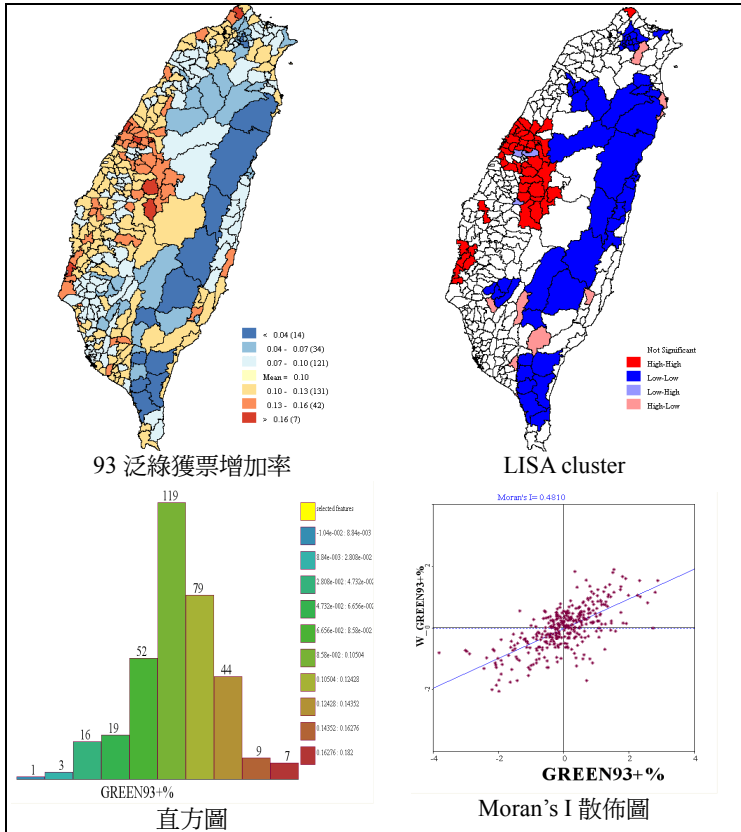


圖十六 Moran's I 散佈圖與泛綠政黨繼承

關於第三個問題，民進黨 93 年選票增加的緣由。第二次總統選舉總共 349 個鄉鎮區（離島除外）中，有 280 個是藍區（泛藍選票過半），其中有 132 個第三次選舉時由藍區變為綠區。圖十七顯示政黨版圖的變化，暗紅色者為由藍轉綠的鄉鎮區。圖十八是獲票增加率的四個套圖：主題圖、直方圖、LISA Cluster 圖、Moran's I 散佈圖，我們發現獲票增加率基本上呈左右對稱的常態分佈。主要增加的區域在中台灣，LISA Cluster 圖清楚顯示這個趨勢。Moran's I 值（0.4810）雖然達顯著水準，但比獲票率的 Moran's I（0.6707）低，顯示獲票率的增加是整體揚升的結果。



圖十七 第三次總統選舉由藍轉綠的鄉鎮區

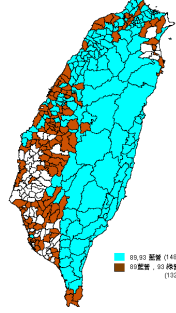
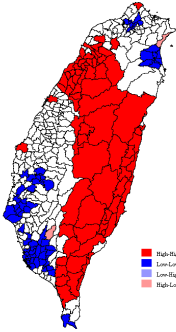


圖十八 第三次總統選舉民進黨選票增加率的空間結構

至於「變天」的 132 個鄉鎮區，何以致此呢？表一，我們嘗試將「變天區」與 LISA cluster 做聯結，看看兩者是否有關聯。首先，吾人繪製第二次總統選舉泛藍的 LISA Cluster 圖，HH、LL 為空間上的穩定區，我們預期這些區域變化不會太大。果然 HH 區「變天」的情形很少，顯著的 HH 區，僅有 3%，不顯著的 HH 區稍高，但也只有 15%。LL 區也是傳統上認定的穩定區，不過第三次總統選舉有 43%（顯著）、72%（不顯著）變天，顯示泛藍相對較弱的地區有頗大的變化，我們隨即發現這些地區集中在中台灣，這透露一個訊息：中台灣是影響成

敗的重要關鍵。HL 區與 LH 區，傳統上被視為空間上的不穩定區，實際數據顯示，HL 區變天情況不多，分別是 0%（顯著區，總數兩個）、12%（不顯著區，總數 16 個中的兩個）；LH 區則有 100%（顯著區，總數兩個）、77%（不顯著區，總數 30 個中 23 個變天）。至於造成 93 年民進黨獲票率大幅躍升的因素，則有待迴歸分析進一步解答，這是下節要探討的主題。

表一 總統選舉（89）泛藍 LISA cluster 與版圖變化（93）

 	個數	版圖變化			
	泛藍 cluster	綠－綠	藍－藍	藍－綠	總計
顯著		50	62	42	154
HH			60 (97%)	2 (3%)	62 (100%)
LL		50 (57%)		38 (43%)	88 (100%)
LH				2 (100%)	2 (100%)
HL			2 (100%)		2 (100%)
不顯著		19	86	90	195
HH			63 (85%)	11 (15%)	74 (100%)
LL		16 (21%)	5 (7%)	54 (72%)	75 (100%)
LH		3 (10%)	4 (13%)	23 (77%)	30 (100%)
HL			14 (88%)	2 (12%)	16 (100%)
總計		69	148	132	349

註：為方便對照，左上為 93 年由藍轉綠的鄉鎮區，左下為 89 年泛藍獲票率 LISA cluster 圖。

伍、空間迴歸分析

本章作者嘗試以空間迴歸模型檢定空間因素的影響。如前所述，解釋選票分佈有三個理論：「成分影響觀點」、「地區脈絡效果」、「擴散效果」。採「成分影響觀點」者不相信空間因素有獨特解釋力，認為考慮個人經社背景變數為已足。主張「地區脈絡效果」者認為地區脈絡有重要性，個人無法脫離環境脈絡做選擇。主張擴散效果者則強調鄰近地區造成的影響。作者即以空間迴歸分別檢視這三個理論。鄰近關係的定義為周圍接壤者，去除沒有鄰近的孤島，共有 349 個鄉鎮。

我們的問題是：到底什麼因素造成 93 年民進黨獲票率的大幅躍升呢？應變數為 89、93 年兩次總統選舉民進黨各鄉鎮獲票增加率。自變數除一般經社背景變數外，另增加空間變數，檢視其影響是否顯著。值得注意的是，本研究分析單位是鄉鎮，亦即我們關注的是：哪些鄉鎮有顯著變化？如何解釋這些鄉鎮的變化？作者無意推論個人層次的投票行為。另外作者也必須承認，空間資料取得相當困難，因為根本不存在這樣的資料庫（各年度、各鄉鎮區的基本屬性資料），只能從不同來源、殘缺不全的資料中慢慢拼湊；因此，本研究實屬探索式的研究，俟將來有更精確的資料，可進行更細緻的分析。

一、自變數的選擇

自變數主要整理自 89 年人口普查及 91 年財稅中心資料，由於普查的變數不多，選擇性有限，作者盡可能選取過去研究經常提及的因素作為自變數。值得注意的是過去的研究多以問卷調查的個體資料為主，其實以個體資料的研究所得直接推論集體現象，經常會有問題。因此，本文自變數的選擇基本上是試探性的，至於自變數影響的方向，亦未事先

假定。另外本文的因變數是「民進黨獲票增加率」而非一般的「獲票率」，事先判定變數影響方向的理論基礎亦較薄弱。自變數說明如下：

- 一、平均年齡：各鄉鎮區的平均年齡。
- 二、人口密度：各鄉鎮區人口密度，每平方公里人口數，代表都市化程度。
- 三、產業結構：各鄉鎮區各產業（服務業、工業、農業）人口比例。服務業人口比與工業人口比作為自變數，農業人口比率為參考變數。
- 四、教育水準：各鄉鎮區平均教育水準，以 1 至 4 等級代表國小、國中、高中、大專教育程度，依人口比例加權後，計算各鄉鎮平均教育水準。
- 五、性別比：各鄉鎮區性別比。性別比的定義為：男性人口除以女性人口再乘以一百，亦即每一百個女性，有幾個男性。大於一百表男多於女，小於一百表女多於男。
- 六、所得：各鄉鎮區平均綜合所得，整理自 91 財稅中心資料。這個資料可能有誤差，因為免納稅者未納入考慮；綜合所得的準確度也令人懷疑，但在沒有其他資料的情況下，只能以此資料暫用。
- 七、長住率：代表五年內各鄉鎮區沒有改變居所的人口比例，長住率愈高代表遷徙的機會少，可解釋為社會流動較少或對地方的淵源或瞭解較深。
- 八、無工作率：各鄉鎮區無工作人口比例，資源來自 89 年人口普查，無工作率代表未就學、不在服兵役而沒有工作者，包括老年人在內。與一般傳統的失業率不同，因無法取得各鄉鎮區失業率資料，以此替代。
- 九、族群：各鄉鎮區不同族群比例，整理自客委會的人口基本調查，以客家族群為參考變數，外省人口比與福佬人口比為自變數。

- 十、虛擬變數：民進黨在該縣是否執政：是的話其值為 1，否則為 0，增加這個自變數，主要原因是第三次總統選舉由於競爭過於激烈，選舉過程爭議不斷，有人質疑投票前夕槍擊事件的真實性，有人直指政府做票、造假。如果選舉真的有全面性的造假的話，民進黨執政的縣市，選票增加率應該會變大才對。
- 十一、為了解決空間異質的問題，本研究嘗試增加幾個虛擬變數，以區別不同的空間區域（spatial regimes）。作者試了兩種可能，第一、以北、中、南、東四劃分為四個空間區域；第二、以 Lisa Cluster 分析的結果（圖十八）劃分 HH、LL、HL、LH 四個區域，分別給定虛擬變數，看看其解釋力如何。如果這些代表獨特區域的虛擬變數顯著，表示有空間異質性存在。

二、迴歸結果檢視

表二為迴歸報表結果，共有六個模型。OLS-1 模型，盡可能放進所有的經社變數，OLS-2 模型選擇達 0.05 顯著水準的自變數，藉以簡化模型，且修正共線性問題。OLS-3 模型增加虛擬變數：中部縣市。OLS-4 模型增加虛擬變數 HH（High-High），LL（Low-Low）。最後兩個模型在 OLS-4 模型的基礎之上，加進空間變數，前者為 SLM 模型，後者為 SEM 模型，均以最大概似法估計。

表二 93 年總統選舉民進黨獲票增加率迴歸模型

	OLS-1	OLS-2	OLS-3	OLS-4	SLM	SEM
CONSTANT	0.0582 (0.0417)	-0.0343 (0.0328)	-0.0526 (0.0303)	0.0224 (0.0256)	-0.0072 (0.0251)	-0.0213 (0.0280)
平均年齡	0.0023*** (0.0007)	0.0018*** (0.0006)	0.0020*** (0.0005)	0.0013*** (0.0004)	0.0004*** (0.0005)	0.0022*** (0.0005)
平均教育水準	0.0343*** (0.0123)	0.0263*** (0.0109)	0.0305*** (0.0101)	0.0120 (0.0084)	0.0125 (0.0081)	0.0160 (0.0091)
性別比	-0.0002 (0.0002)					
所得	-0.0002*** (5.33e-5)	-0.0002*** (4.92e-5)	-0.0002*** (4.54e-5)	-0.0001*** (3.83e-5)	-0.0001*** (3.77e-5)	-0.0002*** (4.5415e-5)
人口密度	-1.2891e-7 (3.13e-7)					
常住率	-0.0612 (0.0306)					
無工作率	0.0439 (0.0327)					
服務業比率	-0.1697* (0.0334)					
工業比率	0.1096*** (0.0349)	0.1551*** (0.0258)	0.1428*** (0.0238)	0.0734*** (0.0204)	0.0701*** (0.0202)	0.1434*** (0.0237)
外省籍人口比	-0.0004* (0.0002)					
福佬人口比	0.0004*** (5.66e-5)	0.0003*** (5.13e-5)	0.0003*** (4.75e-5)	0.0001*** (4.10e-5)	0.0001*** (4.14e-5)	0.0002*** (5.22e-5)
民進黨執政	-0.0141*** (0.0030)					
中部縣市			0.0214*** (0.0027)			
HH				0.0339*** (0.0031)	0.0267*** (0.0033)	0.0245*** (0.0038)
LL				-0.0350*** (0.0032)	-0.0283*** (0.0033)	-0.02715** (0.0036)
W_民進黨獲票 增加率					0.2331*** (0.0636)	
Lambda						0.3764*** (0.0704)

N	349	349	349	349	349	349
AdjR ²	0.2932	0.2400	0.3533	0.5648	0.5904	0.5916
F	13.0346	22.9892	32.6909	65.5228		
P	5.45e-22	6.55e-20	4.35e-31	0		
Multicollinearity						
ConditionNum	133.1906	72.8748	75.2222	76.75429		
ber						
Jarque-Bera	0.0559	0.1179	0.1793	0.0075		
Breusch-Pagant	0.0049	0.0001	0.0021	0.0003	0.0001	0.0000
est						
Koenker-Basse	0.0413	0.0015	0.0107	0.0079		
tttest						
Moran'sI	0.0031	0.0000	0.0000	0.1054		
(error)						
LagrangeMulti		0.0001	0.0000	0.0026		
plier (lag)						
RobustLM		0.3475	0.0729	0.0004		
(lag)						
LagrangeMulti		0.0000	0.0000	0.2548		
plier (error)						
RobustLM		0.0034	0.0000	0.0355		
(error)						
Akaike	-1577.81	-1633.15	-1770.4	-1778.77	-1775.37	
Schwarz	-1554.68	-1606.17	-1739.56	-1744.08	-1744.52	

***表 0.05 信心水準

() 內為標準誤

迴歸報表共有五個模型，說明如下：

第一、OLS-1 模型盡可能把所有的經社背景變數全部放入，結果 R 平方值並不大，可解釋變異的百分比僅 0.25 左右。達顯著水準者有年齡(+)、教育(+)、所得(-)、工業人口比(+)、福佬人口比(+)。表示年齡愈老化，教育水準高、所得低、工業人口愈多、福佬人口愈多的鄉鎮，民進黨獲票增加率愈多。民進黨獲票增加率與民進黨執政呈負相關，表示民進黨執政不但無助於獲票率的增加，反而有減少的趨勢，顯然這裡無法支持民進黨做票的假設。值得注意的是省籍的影響，外省人口比與民進黨獲票增加率呈微弱的負相關，未達 0.05

顯著水準，顯示外省人的比例愈高，對民進黨的反對並沒有增加多少，倒是福佬人對民進黨的支持度增強了（達 0.05 顯著水準）。可能的原因是外省人的投票傾向相當穩定，兩次選舉變化不大，福佬人則變化較大，改投給民進黨。不過這個模型最大的問題在 Multicollinearity condition number 達 133.1906，顯然有嚴重的共線性問題存在，迴歸係數估計不穩定。

第二、OLS-2 模型僅保留 OLS-1 模型達 0.05 顯著的自變數，捨棄若干高度相關的自變數，例如人口密度、性別比、服務業比，與所得高度相關，僅保留所得。Multicollinearity condition number 降至 72.87，雖然仍偏高（一般建議在 25 以下為妥），但已大幅改善。R 平方則減為 0.24，不過 F 檢定顯著，基本上本模型仍是妥適。自變數影響的方向與 OLS-1 模型無異，表示年齡愈老化，教育水準高、所得低、工業人口愈多、福佬人口愈多的鄉鎮，民進黨獲票增加率愈多。

第三、OLS-2 模型誤差 Moran's I 檢定顯示有自相關現象。本研究以周圍接壤者定義為鄰近，去除沒有鄰近的孤島，共有 349 個鄉鎮。經過 Moran's I 檢定，機率小於 0.05，拒絕誤差獨立的虛無假設。接著針對 lag 與 error 模型分別進行 Lagrange Multiplier 檢定，結果兩者均顯著。進行 Robust 檢定，只剩下 error 模型顯著（機率 0.0034，Robust lag 機率为 0.3475 未達顯著），因此建議使用「空間誤差模型」。而，如前所述，空間誤差模型表示可能有隱藏的變數存在。

第四、在進行空間迴歸以前，我們先處理空間異質性的問題。因為，誤差 Moran's I 檢定顯示的自相關現象有可能是空間異質引起的。處理空間異質問題，最常用的方法是依地區的類同程度將研究範圍區分不同地域（spatial regimes），分別處理再

比較之，如用 Spatial Chow test 檢定不同地域的表現是否有異。然而 GeoDa 軟體並無提供是項功能，變通的方法是增加虛擬變數代表不同地域，如果某地域具有獨特性，則該地域虛擬變數的係數會呈顯著。作者依慣例將台灣分為北、中、南、東四大區塊，然後增加北、中、南三個虛擬變數（東區為參考）。置入地區虛擬變數後發現只有中區達顯著水準。因此，保留中區，去除其他各區，此即 OLS-3 模型。此時 R 平方值增加達 0.35，顯然空間異質問題有改善。也就是說，中部地區的民進黨獲票增加率與其他地區相較，多增加了 2.14 個百分點，這是中部地區整體的提昇。2000 年到 2004 年民進黨獲票率從百分之 39.3 增加到百分之 50.11，獲票率的增加是全台普遍的現象（平均增加百分之 10.81），但中台灣增加特別多，由於藍綠陣營票數差距甚微，中台灣的「變天」顯然是影響最後勝負極為關鍵的因素。誤差的 Moran's I 檢定仍達顯著，表示誤差有自相關問題。針對 lag 與 error 模型進行 Lagrange Multiplier 檢定，兩者均顯著，Robust 檢定的結果建議使用「空間誤差模型（SEM）」。

第五、另一個建立空間虛擬變數的方法是從 LISA Cluster 的分類著手，如圖十八所示，HH, LL, HL, LH 分別代表四種不同的空間結構，達顯著水準。因此，依此分類分別給定虛擬變數，結果發現 HH, LL 達顯著水準，保留此兩者，去除其他各區，此即 OLS-4 模型。此時 R 平方提高至 0.56，顯然 OLS-4 的預測能力比前面三個模型都好。在 HH 區（一般稱為「熱點」），民進黨獲票增加率與其他地區相較，多增加了 3.39 個百分點；在 LL 區（一般稱為「冷點」），減少了 3.35 個百分點。值得注意的是平均教育水準在加入地區虛擬變數之後，變成不顯著，亦即在控制不同地域（HH, LL）之後，發

現平均教育水準其實與民進黨的獲票增加率無關。另外，誤差的 Moran's I 檢定未達顯著水準，不過 Lagrange Multiplier 與 Robust 檢定建議採用「空間落遲模型（SLM）」。

第六、SLM 模型增加新的自變數「w_民進黨獲票增加率」（係數為 Φ_0 ）²⁰，亦即鄰近地區民進黨獲票增加率的平均。結果新增的空間變數顯著，R 平方值增加至 0.5904。同樣地，SLM 模型，係數為 Λ 亦達顯著，R 平方值 0.5916。

第七、比較模型的優劣，除了 R 平方值大小外（OLS 迴歸 R 平方代表模型可解釋的百分比，其值愈大愈好），另輔以 Akaike 與 Schwarz 指標，其值愈小愈好。最後發現 SLM 模型為佳，其 R 平方值達 0.59，Akaike 值-1778.77，優於其他模型，表示民進黨獲票增加率其實還是會受到鄰近地區的影響，亦即控制其他經社變數、地區異質變數之後，仍然發現：周圍地區民進黨獲票增加率高的地區，其民進黨獲票增加率亦較高，反之亦然。

陸、結論

本文介紹空間分析的基本方法，可分為視覺化呈現、空間探索分析與空間迴歸建模三個部分。作者以 85、89、93 三次總統選舉為例，實際操作這些方法。研究發現第一次總統選舉，因為李登輝效應影響，選票的空間結構比較異常。其後選舉，加總藍、綠選票，則呈現相當穩定的空間結構。從 Moran's I 散佈圖、LISA cluster 圖、PCP 圖都可以看出這種趨勢，即使獲票率有增減，仍然不影響空間結構的穩定性。另外，由空間結構的相似度，我們可以判斷選票的繼承關係，第一次總統選舉

²⁰ 代表鄰近地區民進黨獲票率的平均值。

李登輝的票源跨越藍綠，比較特別，其後選舉回歸藍綠基本面，而且分裂後的泛藍候選人各擅勝場，各有其空間上的領域。而第三次總統選舉民進黨獲票率增加，與中台灣大有斬獲有關。

空間迴歸分析方面，本文針對 93 年民進黨獲票增加率進行分析，結果發現年齡愈老化，教育水準高、所得低、工業人口愈多、福佬人口愈多的鄉鎮，民進黨獲票增加愈多。民進黨獲票增加率與民進黨執政呈負相關且顯著，表示民進黨執政不但無助於獲票率的增加，反而有減少的趨勢。從這裡看不出有大規模做票的證據。另外，外省人的比例愈高，對民進黨的反對並沒有增加多少，倒是福佬人對民進黨的支持度增強了。可能的原因是外省人的投票傾向相當穩定，兩次選舉變化不大，福佬人則變化較大，改投給民進黨。

值得注意的問題是傳統 OLS 迴歸，誤差有嚴重的自相關現象。這可能是空間異質性引起，也可能是鄰近效應引起。為了處理異質性問題，地域的劃分，本研究除傳統的北、中、南、東劃分外，亦應用 LISA Cluster 分析，區分 HH, LL, HL, LH 四種空間類型，以虛擬變數處理之。發現以 LISA Cluster 為基礎的區域劃分比較好。最後以此為基礎進行 SLM, SEM 兩種空間迴歸模型，綜合各項指標發現 SLM 模型最佳，R 平方值達 0.59。本文僅係探索性的研究，探討空間計量方法在選舉研究上的可能應用，受限於資料取得不易，分析顯得粗糙；如何進行更精緻的探討、以及與傳統個體資料為主的選舉研究對話，仍是後續的努力目標。

參考文獻

- 林繼文（1998）〈地盤劃分與選舉競爭：對應分析法在多席次選舉研究上之應用〉，《選舉研究》，第五卷，第二期，頁一〇三至一二八。
- 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》，第一卷、第一期，頁九三至一一〇。
- 徐永明（2001）〈政治版圖：兩個選舉行為研究途徑的對話〉，《問題與研究》，第四〇卷，第二期，頁九五至一一五。
- 盛杏媛（1998）〈政黨配票與候選人票源的集散度：一九八三年至一九九五年台灣地區立法委員選舉的分析〉，《選舉研究》，第五卷，第二期，頁七三至一〇二。
- 黃信豪（2006）〈多層模型於選民投票行為研究的應用：以 2004 年總統選舉為例〉，《東吳政治學報》，第二十二期，頁一六一至二〇五。
- Agnew, J.A. (1987) *Place and Politics: The Geographical Mediation of State and Society*. London: Allen and Unwin.
- Anselin, Luc (1998) *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic publishers.
- Anselin, Luc (2005) *Exploring Spatial Data with DeoDa: A Workbook*. Center for Spatially Integrated Social Science.
- Ault, Brian & William Brustein (1998) "Joining the Nazi Party: Explaining the Political Geography of NSDAP Membership, 1925-1933," *The American Behavioral Scientist*," Vol.41, No.9, pp.1304-1323.
- Burbank, M. (1995) "How do contextual effects work? Developing a theoretical model," in M. Eagles (Ed.), *Spatial and contextual models in political research* (pp.165-78), London: Taylor and Francis.

- Cho, Wendy K. Tam (2003) "Contagion Effects and Ethnic Contribution Networks," *American Journal of Political Science*, Vol. 47, No.2, pp.368-387.
- Flint, Colin (1995) *The Political Geography of Nazism: The Spatial Diffusion of the Nazi Party Vote in Weimar Germany*. Ph.D. Dissertation. University of Colorado at Boulder.
- Flint, Colin (1998) "Forming Electorates, Forging Spaces: The Nazi Party and the social construction of space," *American Behavioral Scientist*, Vol.41, No.9, pp.1282-1303.
- Flint, Colin (2000) "Electoral geography and the social construction of space: The example of the Nazi party in Baden, 1924-1932," *GeoJournal*, Vol. 51, No.3, pp.145-156.
- Giddens, A., & C. Pierson (1998) *Conversations with Anthony Giddens: Making sense of modernity*. Cambridge: Polity Press.
- Gimpel, James G. & Jason E. Schuknecht (2002) "Reconsidering Political Regionalism in the American States," *State politics & Policy Quarterly*, Vol.2, No.4, pp.325-352.
- Gimpel, J.G., J.E. Schuknecht (2003) "Political participation and the accessibility of the ballot box," *Political Geography* 22, pp.471-488.
- Gimpel, James G. and Wendy K. Tam Cho (2004) "The persistence of white ethnicity in New England politics," *Political Geography* 23, pp.987-1008.
- Gimpel, James G., Irwin L. Morris & David R. Armstrong (2004), "Turnout and the local age distribution: examining political participation across space and time," *Political Geography*, 23, pp.71-95.
- Griffiths, William E., R.Carter Hill, & George G. Judge (1993) *Learning and Practicing econometrics*. New York: John Wiley & Sons.Inc.

- Kohfeld, Carol W. & John Sprague (2002) "Race, Space, and Turnout," *Political Geography* 21, pp.175-193.
- West II, W. Jefferson (2005) "Regional Cleavages in Turkish Politics: An Electoral Geography of the 1999 and 2002 national elections," *Political Geography*, 24, pp.499-523.
- Johnston, Ron, Kelvyn Jones, Rebecca Sarker, Carol Propper, Simon Burgess & Anne Bolster (2004) "Party Support and the Neighborhood Effect: Spatial Polarization of the British electorate, 1991-2001," *Political Geography*, No.23, pp.367-402.
- Kohfeld, Carol W. and John Sprague (2002) "Race, space, and turnout," *Political Geography* 21, pp.175-193.
- McAllister, Iain (1987) "Social Context, turnout, and the vote: Australian and British Comparisons," *Political Geography Quarterly*, No.6, pp.17-30.
- McAllister, Iain & D. Studlar (1992) "Region and Voting in Britain, 1979-87: Territorial Polarization or Artifact? *American Journal of Political Science*, Vol.36, No.1, pp.168-99.
- McAllister, Iain, Edward Fieldhouse & Andrew Russell (2002) "Yellow fever? The political geography of Liberal voting in Great Britain," *Political Geography* 21, pp.421-447.
- Mellow, Nicole & Peter Trubowitz (2005), "Red versus blue: American Electoral Geography and Congressional Bipartisanship, 1898-2002," *Political Geography*, No.24, pp.659-677.
- Michael F. Goodchild and Donald G. Janelle (ed.), *Spatially integrated Social Science*. Oxford University Press, 2004.

- Nicole Mellow, Peter Trubowitz, "Red versus blue: American electoral geography and congressional bipartisanship, 1898-2002," *Political Geography* 24, pp.659-677.
- O'Loughlin, John, Colin Flint, and Luc Anselin (1994) "The Geography of the Nazi Vote: Context, Confession, and Class in the Reichstag Election of 1930." *Annals of the Association of American Geographers*, Vol.84, No.3, pp.351-380.
- O'Loughlin, John (2002) "Spatial Analysis in Political Geography," in *A Companion to Political Geography*, edited by J. Agnew, K. Mitchell and G O Tuathail (Oxford: Basil Blackwell).
- Patterson, Samuel C. & Gregory a. Caldeira (1984) "The Etiology of Partisan Competition," *The American Political Science Review*, Vol.78, No.3, pp.691-707.
- Perdomo, C.J. Vilalta y (2004) "The local context and the spatial diffusion of multiparty competition in Urban Mexico, 1994-2000," *Political Geography*, 23, pp.403-423.
- Sauerzopf, R., & T. Swanstrom (1999) "The urban electorate in Presidential Elections 1920-1996," *Urban Affairs Review*, Vol.35, No.1, pp.72-91.
- Shin, Michael E. and John Agnew (2002) "The geography of party replacement in Italy, 1987-1996," *Political Geography*, 21, pp.221-242.
- Sui, Danile Z. and Peter J. Hugill (2002) "A GIS-based spatial analysis on neighborhood effects and voter turn-out: a case study in College Station, Texas," *Political Geography*, 21, pp.159-173
- Tobler, W. (1979) "Cellular Geography," in *Philosophy in Geography*, edited by S. Gale and G. Olsson, pp.379-86, Dordrecht: Reidel.

The Spatial Factors in Electoral Studies: the Example of Presidential elections in Taiwan

Chih-Sung Teng

ABSTRACT

The author tries to apply the methods of spatial analysis into the studies of election campaign with the purpose of identifying the influence of spatial factors. After making a short introduction of the meaning of spatial analysis and the techniques of this research approach, the author takes the three presidential elections held in Taiwan from 1996 to 2004 as example. The methods used in this study include geographical visualization, exploratory spatial data analysis and spatial modeling. The author finds that the voting results show a high degree of spatial stability, although the party supports in each election was changing. In the spatial regression analysis, the author finds there are both spatial heterogeneity and autocorrelation existed which makes error violates the assumption of OLS regression.

Keywords: spatial analysis, spatial autocorrelation, electoral geography, spatial regression, presidential election.