

行政院國家科學委員會補助專題研究計畫成果報告

擬似定群分析法：學理之延伸及其在政治學中之應用

**Pseudo-Panel Analysis: Elaboration and
Applications**

計畫類別：個別型計畫 整合型計畫

計畫編號：NSC 94-2414-H-004-064-

執行期間：九十四年八月一日至九十六年一月三十一日

計畫主持人：黃 紀

專任研究助理：林長志、施奕任

成果報告類型：精簡報告 完整報告

本成果報告包括以下應繳交之附件：

- 赴國外出差或研習心得報告一份
- 赴大陸地區出差或研習心得報告一份
- 出席國際學術會議心得報告及發表之論文各一份
- 國際合作研究計畫國外研究報告書一份

處理方式：可立即公開查詢

執行單位：國立政治大學 選舉研究中心

中 華 民 國 九 十 六 年 一 月 三 十 一 日

本研究計畫根據動態分析的相關文獻，先以最簡明的二分變數 (binary variable) 為起點，找出「重複進行之獨立橫斷面資料」(repeated cross-sections, 簡稱 RCS) 進行跨時分析的困難癥結所在，然後對症下藥，尋求克服之道，繼而應用至台灣族群認同之演變。

一、找出問題的癥結

以第 t 個時間點執行之獨立抽樣調查為例，若欲以之估算變數 Y 之「常與變」，其流動表 (turnover table) 如表 1 所示。表 1 乍看之下似乎與一般定群研究之流動表極為相像，但其實大相逕庭：因 RCS 樣本並非針對相同受訪者 (定群) 之重複測量，故表 1 細格內之次數無從得知，均以問號 $?$ 標示之。換言之，RCS 資料，僅有流動表縱欄 (代表 t 時間點) 之邊緣總和分佈 Y_t 為已知，表內聯合分佈之次數及百分比卻為未知。此一情況，與區位推論 (ecological inference) 中，已知母群之邊緣分佈 N_{+j} 及 N_{i+} 、但不知其聯合分佈 N_{ij} 的情況 (King, 1997; King, Rosen, and Tanner, 2004; 黃紀，2001) 頗為相似，而且還多了兩項困難：

1. 在第 t 個時間點的 n 個樣本，其前一個時間點 $t-1$ 的值並未觀察到，故 $t-1$ 之邊緣分佈 Y_{t-1} 亦為未知數。
2. RCS 為樣本而非母群資料，故推論時抽樣誤差的特性也必須加以考量；換言之，母群集體資料推論之上下限為確定 (deterministic)，而 RCS 推論之上限與下限均有其信心區間 (confidence interval)。

因此以 RCS 資料進行縱貫時序分析，與以母群集體數據進行區位 (亦即跨層 cross-level) 推論有密切關係，但困難度似有過之而無不及。

表 1 政黨偏好流動表：RCS 在時間點 t 之個體數據 (樣本數 n)

		t		
		甲黨	乙黨	
$t-1$	甲黨	甲甲型穩定數 $n_{11} ?$	甲乙型變遷數 $n_{12} ?$	$n_{1+} ?$
	乙黨	乙甲型變遷數 $n_{21} ?$	乙乙型穩定數 $n_{22} ?$	$n_{2+} ?$
		n_{+1}	n_{+2}	n

註：? 表示未知數

儘管以RCS資料進行動態分析困難度甚高，但這未必表示研究者就束手無策。由於表 1 中的橫列邊緣分佈（亦即 Y_{t-1} ）為未知數，可視為是呈二項式分佈（binomial distribution）的隨機變數。先針對學理感興趣之穩定與變遷率，將表 1 予以參數化（parameterization），成為表 2：

表 2 以 RCS 在時間點 t 之個體數據估計流動率參數

		t			
		甲黨	乙黨		
t-1	甲黨	甲甲型穩定率 （甲黨之忠誠率） $p_{11} = \frac{n_{11}}{n_{1+}} = 1 - \mu_t ?$	甲乙型變遷率 （乙黨之流入率） $p_{21} = \frac{n_{12}}{n_{1+}} = \mu_t ?$	$P_{1+} = \frac{n_{1+}}{n} = 1 - P_{t-1} ?$	
	乙黨	乙甲型變遷率 （乙黨之出走率） $p_{12} = \frac{n_{21}}{n_{2+}} = \lambda_t ?$	乙乙型穩定率 （乙黨之忠誠率） $p_{22} = \frac{n_{22}}{n_{2+}} = 1 - \lambda_t = \kappa_t ?$		
		$P_{+1} = \frac{n_{+1}}{n} = 1 - P_t$	$P_{+2} = \frac{n_{+2}}{n} = P_t$	1.0	

註：? 表示未知數

若界定 $n_{2+}=f$ ，Pelzer, Eisinga, and Franses (2002: 121)根據 Duncan and Davis (1953) 及King(1997)所提議之「上、下限推算法」(method of bounds)，推導出表 2 「甲乙型變遷率 μ 」及「乙乙型穩定率 κ 」之上下限：

$$\max\left[0, \left(\frac{n_{+2} - f}{n - f}\right)\right] \leq \mu \leq \min\left(1, \frac{n_{+2}}{n - f}\right),$$

$$\max\left[0, \left(\frac{n_{+2} - (n - f)}{f}\right)\right] \leq \kappa \leq \min\left(1, \frac{n_{+2}}{f}\right).$$

二、建立擬似定群 pseudo-panel 之馬可夫鍊模型

將 RCS 資料視為擬似之定群追蹤資料（pseudo panel），建立類似馬可夫鍊之動態模型，是晚近在跨時分析方面的新發展。

1985 年 Deaton 首先指出：RCS 雖非針對固定樣本追蹤調查，但如果將不同時間點的 RCS 資料中的樣本，依若干個已知之變數（如年齡、性別等），集結（aggregate）成幾個同質性很高的次群體（cohorts），並將之視為對應之次母群（subpopulations）的代表性樣本，便可用來估計該次母群參數之真值（例如母群平均數）；準此，則連續 T 個時間點的 RCS 資料，就形同是針對這些同質次母群參數的 T 次樣本估計值，因此雖比不上對相同個人之追蹤一樣精準，但只要把該測量誤差納入模型，則仍可進行擬似定群追蹤之分析，估計出以個體為基礎的迴歸模型中之斜率係數（slope coefficients）。

換言之，Deaton 的「集結同質群成為擬似定群」的方法（cohort-aggregated pseudo panel approach）是將每一時間點的 n 個獨立樣本區分成 C 組同質群，做為新的分析單位（units of analysis），群內同質性越高越好，群間則異質性越大越好。由於 T 個時間點的獨立樣本均可根據相同的變數分群，於是每一個同質群都有 T 個觀察值，彷彿定群樣本中同一個受訪者重複接受 T 次訪問一般。由此觀之，原本彼此獨立、互不相干的 T 波樣本，彷彿成了針對 C 組同質之次群追蹤調查了 T 波的資料，因此稱為「擬似定群追蹤」。此一開創性的想法引起廣大迴響，刺激了學界對運用 RCS 資料解釋「常與變」的希望，相關文獻遂開始湧現（參見 Verbeek and Nijman, 1992; Verbeek, 1996 之討論與回顧；Collado, 1997 及 McKenzie, 2001 之延伸等）。

Franklin（1989）及 Moffitt（1990; 1993）將 Deaton「集結同質群」的方法推而廣之，應用至連續依變數之線型模型，無須真的將樣本分群，仍可保持個別受訪者為分析單位，只需以兩階段進行動態迴歸分析。以一階線性自迴歸（AR(1)）為例，由於 RCS 資料均屬獨立抽樣，沒有第 i 個樣本在前一時期（ $t-1$ ）之觀察值 $y_{i,t-1}$ ，故必須估計其近似值，做為後續動態迴歸分析之工具（instrument）。因此，在概念上其估計程序分成兩階段：

1. 第一階段，先以區分同質群所依據的自變數來建立「工具變數」

（instrumental variables, IV） $\hat{y}_{i,t-1}$ 。

2. 第二階段，將此 $\hat{y}_{i,t-1}$ 取代未觀察到的 $y_{i,t-1}$ ，納入動態迴歸模型的架構分析之，其迴歸係數便是「移轉率」之估計值；而當期自變數之迴歸係數，便在估計這些變數對連續依變數之影響。

當然，政治學中穩定與變遷的測量，往往是屬於類別變數（categorical variable），不宜逕以前述之線型迴歸分析之，但可結合推廣之線型模型（generalized linear model, GLM）（McCullagh and Nelder, 1989; 黃紀，2000），

發展適合類別依變數之擬似定群分析法。例如 Pelzer, Eisinga and Franses (2001; 2002; 2004) 即是根據 Moffitt (1990; 1993) 之觀念來推導二分 (binary) 依變數的移轉模型。本計畫即參考 Pelzer, Eisinga and Franses (2005), 推導 RCS Markov chain pseudo panel 模型。

三、RCS 馬可夫鍊模型之應用

本計畫將 RCS pseudo panel model 應用於重要的實質議題，及台灣族群認同之演變 (Huang, 2006)。亦即以政治大學選舉研究中心彙整自 1998 年以來執行之全國性民調，以及「台灣選舉與民主化調查」中之全國性民調資料 (包括 TEDS2001、TEDS2003、TEDS2004P、TEDS2004L)，針對「台灣人、中國人認同」進行馬可夫鍊模型之動態分析，估計台灣近八年民主化過程中，族群認同之淨變量與總變量，並以動態模型解釋其演變之軌跡 (trajectory)。該文“The Evolution of Taiwanese Identity: A Pseudo Panel Analysis”已於 2006 年 9 月美國政治學會年會 (Philadelphia) 中發表。

參考文獻

- 黃紀，2000：〈質變數之計量分析〉，見謝復生、盛杏媛 (編)，《政治學的範圍與方法》，頁 387-411。台北：五南圖書出版公司。
- 黃紀，2001：〈一致與分裂投票：方法論之探討〉，《人文及社會科學集刊》，13(5): 541-574。
- 黃紀，2005：〈投票穩定與變遷之分析方法：定群類別資料之馬可夫鍊模型〉，《選舉研究》，12(1): 1-35。
- 黃紀、王鼎銘、郭銘峰，2005：〈日本眾議院 1993 年及 1996 年選舉自民黨選票流動之分析〉，《人文及社會科學集刊》，17(4)：853-883。
- Collado, M. Dolores. 1997. “Estimating Dynamic Models from Time Series of Independent Cross-sections.” *Journal of Econometrics* 82(1): 37-62.
- Deaton, Angus. 1985. “Panel Data from Time Series of Cross-Sections.” *Journal of Econometrics* 30: 109-126.
- Diggle, Peter J., Patrick J. Heagerty, Kung-Yee Liang, and Scott L. Zeger. 2002. *Analysis of Longitudinal Data*, 2nd edition. Oxford: Clarendon Press.
- Duncan, Otis Dudley, and Beverly Davis. 1953. “An Alternative to Ecological Correlation,” *American Sociological Review* 18: 665-666.
- Finkle, Steven E. 1995. *Causal Analysis with Panel Data*. Thousand Oaks: Sage.
- Firebaugh, Glenn. 1997. *Analyzing Repeated Surveys*. California: Sage.
- Franklin, Charles H. 1989. “Estimation across Data Sets: Two-Stage Auxiliary Instrumental Variables Estimation (2SAIV).” *Political Analysis* 1: 1-23.

- Frees, Edward W. 2004. *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge: Cambridge University Press.
- Hagenaars, Jacques A. 1990. *Categorical Longitudinal Data: Log-Linear Panel, Trend, and Cohort Analysis*. Newbury Park: Sage.
- Hamerle, Alfred, and Gerd Ronning. 1994. "Panel Analysis for Qualitative Variables." In *Handbook of Statistical Modelling for the Social and Behavioral Sciences*, ed. Gerhard Arminger, Clifford C. Clogg and Michael E. Sobel. New York: Plenum Press.
- Hand, David, and Martin Crowder. 1996. *Practical Longitudinal Data Analysis*. London: Chapman & Hall.
- Henry, Gary T., and Craig S. Gordon. 2001. "Tracking Issue Attention." *Public Opinion Quarterly* 65(2): 157-177.
- Honoré, Bo E., and Ekaterini Kyriazidou. 2000. "Panel Data Discrete Choice Models with Lagged Dependent Variables." *Econometrica* 68(4): 839-874.
- Howard, Ronald A. 1971. *Dynamic Probabilistic Systems, Volume I: Markov Models*. New York: Wiley.
- Hsiao, Cheng. 2003. *Analysis of Panel Data*, 2nd edition. Cambridge: Cambridge University Press.
- Hout, Michael. 1983. *Mobility Tables*. Beverly Hills, CA: Sage Publications.
- Huang, Chi. (黃紀) 2004. "Explaining Referendum Voting Choices in Taiwan." *Issues and Studies* 40(3/4): 316-333.
- Huang, Chi. (黃紀) 2005. "Dimensions of Taiwanese/Chinese Identity and National Identity in Taiwan: A Latent Class Analysis." *Journal of Asian and African Studies*, 40(1/2): 49-68.
- Huang, Chi. (黃紀) 2006. "The Evolution of Taiwanese Identity: A Pseudo Panel Analysis." at the 102nd Annual Meeting of the American Political Science Association, Philadelphia, August 31 - September 3, 2006.
- Huang, Chi., (黃紀) and Todd G. Shields. 2000. "Interpretation of Interaction Effects in Logit and Probit Analyses: Reconsidering the Relationship between Registration Laws, Education, and Voter Turnout." *American Politics Quarterly* 28(1): 80-95.
- John, Kenneth E. 1989. "The Polls-A Report: 1980-1988 New Hampshire Presidential Primary Polls." *Public Opinion Quarterly* 53(4): 590-605.
- Johnston, Richard, and Henry E. Brady. 2002. "The Rolling Cross-Section Design." *Electoral Studies* 21(2): 283-295.
- Katz, Richard S. 1997. "Changing Patterns of Electoral Volatility," *European Journal of Political Research* 31(February): 83-85.
- King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing*

- Individual Behavior from Aggregate Data*. Princeton: Princeton University Press.
- King, Gary, Ori Rosen and Martin A. Tanner. 2004. "Information in Ecological Inference: An Introduction." In *Ecological Inference: New Methodological Strategies*, eds. Gary King, Ori Rosen and Martin A. Tanner. Cambridge: Cambridge University Press.
- Lazarsfeld, Paul F. 1982 [1948]. "The Use of Panel in Social Research." In *The Varied Sociology of Paul F. Lazarsfeld*. ed. Patricia L. Kendall. New York: Columbia University Press.
- Markus, Gregory B. 1979. *Analyzing Panel Data*. Beverly Hills: Sage.
- McCullagh, P., and J.A. Nelder. 1989. *Generalized Linear Models*, 2nd edition. London: Chapman and Hall.
- McKenzie, David J. 2001. "Estimation of AR(1) Models with Unequally Spaced Pseudo-panels." *The Econometrics Journal* 4(1): 89-108.
- McNemar, Quinn. 1947. "Note on the Sampling Error of the Difference between Correlated Proportions or Percentages." *Psychometrika* 12(2): 153-157.
- Menard, Scott. 2002. *Longitudinal Research*, 2nd edition. Thousand Oaks: Sage.
- Moffitt, Robert. 1990. "The Effect of the U.S. Welfare System on Marital Status." *Journal of Public Economics* 41(1): 101-124.
- Moffitt, Robert. 1993. "Identification and Estimation of Dynamic Model with a Time Series of Repeated Cross-Sections." *Journal of Econometrics* 59(1-2): 99-123.
- Pelzer, Ben, Rob Eisinga, and Philip Hans Franses. 2001. "Estimating Transition Probabilities from a Time Series of Independent Cross Sections." *Statistica Neerlandica* 55(2): 249-262.
- Pelzer, Ben, Rob Eisinga, and Philip Hans Franses. 2002. "Inferring Transition Probabilities from Repeated Cross Sections." *Political Analysis* 10(2): 113-133.
- Pelzer, Ben, Rob Eisinga, and Philip Hans Franses. 2004. "Ecological Panel Inference from Repeated Cross Sections." In *Ecological Inference: New Methodological Strategies*, eds. Gary King, Ori Rosen and Martin A. Tanner. Cambridge: Cambridge University Press.
- Pelzer, Ben, Rob Eisinga, and Philip Hans Franses. 2005. "'Panelizing' Repeated Cross Sections: Female Labor Force Participation in the Netherlands and West Germany." *Quality & Quantity* 39(2): 155-174.
- Plewis, Ian. 1985. *Analysing Change: Measurement and Explanation Using Longitudinal Data*. Chichester: Wiley.
- Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6(2): 461-464.
- Sigelman, Lee. 1991. "Turning Cross Sections into a Panel: A Simple Procedure for Ecological Inference." *Social Science Research* 20: 150-170.
- Stokes, Donald E. 1966, "Some Dynamic Elements of Contests for the Presidency."

- American Political Science Review* 60(1): 19-28.
- Stuart, A. 1955. "A Test for Homogeneity of the Marginal Distribution in a Two-way Classification." *Biometrika* 42: 412-416.
- Train, Kenneth E. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.
- van de Pol, Frank, and Rolf Langeheine. 2004. "Markov Chain." In *The SAGE Encyclopedia of Social Science Research Methods, Vol. 2*. ed. Michael S. Lewis-Beck, Alan Bryman, and Tim Futing Liao. Thousand Oaks: Sage.
- Verbeek, Marno. 1996. "Pseudo Panel Data." In *The Econometrics of Panel Data: A Handbook of the Theory with Applications, 2nd Revised Edition*. ed. Laszlo Matyas and Patrick Sevestre. Dordrecht: Kluwer Academic Publishers.
- Verbeek, Marno, and Theo Nijman. 1992. "Can Cohort Data be Treated as Genuine Panel Data?" *Empirical Economics* 17: 9-23.