

# 行政院國家科學委員會專題研究計畫成果報告

具群組因子效果的條件邏輯模型探討

On Conditional Logit Model with Group Factor Effects

計畫編號：NSC 90-2118-M-004-008

執行期限：90年7月1日至91年8月31日

主持人：江振東 國立政治大學統計系

計畫參與人：賴嘉宏 國立政治大學統計系

## 一、計畫中文摘要

條件邏輯模型是在1973年由McFadden所提出，它的基本架構和我們一般所熟悉的多項邏輯模型大體上是相對等的，雖然較少為人所知，但近年來也逐漸受到經濟領域方面研究的重視。不過如果想要利用這一類模型來分析具有某些群組結構的資料時，常常會有低估標準差的現象產生。在此計劃中我們藉由引入隨機效用參數來處理兩類不盡相似的群組結構所衍生出來的問題。此外我們也進一步就有序反應變數的相關議題，有所探討。

**關鍵詞：**多項邏輯模型、條件邏輯模型、群組因子效應、隨機效用參數

## Abstract

Proposed by McFadden(1973), the conditional logit model, an equivalent form of the well-known multinomial logit model for nominal outcomes, has gained its popularity in economic research of late. The model, however, appears to underestimate the standard error when certain group structure exists in a data set. By introducing random effect parameters into the model, we can easily cope with the problem associated with two types of group structure.

In addition, we also investigate the estimation problems related to the generalized linear mixed models with ordinal responses.

**Keywords:** Condition Logit Model, Group Effect, Multinomial Logit Model, Random Effect.

## 二、計畫緣由與目的

這個問題是在進行統計諮詢的過程中所衍生而來的，諮詢的對象是本校財政系的吳文傑教授。

假定反應變數  $y$  是一個具有  $M$  個選項的名目變數(nominal variable)，在社會科學中分析此類資料常用的方法為多項邏輯模型(multinomial logit model)

$$p(y_i = m | x_i) = \frac{\exp(x_i' S_m)}{\sum_{j=1}^M \exp(x_i' S_j)}$$

$$1 \leq m \leq M, \quad i=1, \dots, n$$

其中解釋變數  $x_i$  是樣本中第  $i$  個觀測個體(individual)的相關變數值，而這些數值對反應變數  $y$  的影響(亦即  $S_m$ ， $m=1, \dots, M$ )會隨著  $y$  的不同反應結果而有所不同。

相較之下，一個較少被提及但也逐漸被採用的分析模型是由諾貝爾經濟學獎的得主 McFadden 於 1973 年所提出的條件邏輯模型(conditional logit model)

$$P(y_i = m | z_i) = \frac{\exp(z_{im}' \chi)}{\sum_{j=1}^J \exp(z_{ij}' \chi)}$$

$$1 \leq m \leq M, \quad i=1, \dots, n$$

其中解釋變數  $z_{ij}$  不僅是第  $i$  個觀測個體的一個函數，而且與反應變數的取值有關，但這些變數對  $y$  的影響(亦即  $\chi$ )則不再與  $y$  的取值有關。這兩個模型，雖然表面上看起來略有不同，但實際上可以證明二者其實是一體的兩面，可以互通的。因此任何可以處理多項邏輯模型的軟體，也可以用來處理條件邏輯模型。

然而依據吳教授採用 LIMDEP 做資料分析的一個實證研究中，大多數估計值的標準差普遍都有低估的現象存在。其主要原因在於條件邏輯模型的適用時機要求  $y_i$  必須是獨立的，然而在我們所探討的資料中，我們發現到會作出相同選擇的個體，亦即  $\left\{ x_j | y_i = m, i=1, \dots, n \right\}$ ，實際上是有某種程度的相關性存在，因而導致標準差低估的現象。這顯然是一種群組效應(group effect)，但是這種群組結構和一般我們所認知的由具有相同自變數的觀測值所構成的群組結構又不盡然相似。由於原始的條件邏輯模型無法有效處理具有群組結構的資料所衍生出來的非獨立的特性(亦即會作出相同選擇的個體具有相關性)，這是我們一開始所面臨需要解決的問題。

在 1994 年，Borja 及 Sueyoshi 曾經就如何利用 Probit Model 來處理具有群組效

應的資料(這裡的群組指的是具有相同自變數取值的觀測值所構成的群組)有過探討，其中的主軸是採用 Amemiya(1978)所建議的兩階段估計法來進行。這是當時我們在提計劃之時所依據的最主要文獻。由於我們所遭遇到的是個實際的問題，但相關的文獻並不多見，這便是當時我們提出此一研究計劃的主要動機。

### 三、計畫結果與討論

由於條件邏輯模型與多項邏輯模型實際上是相對等(equivalent)的兩個模型，因此這兩個模型其實可視為下列模型的兩個特例：

$$P(y_i = m) = \frac{\exp(x_i' S_m + z_{im}' \chi)}{\sum_{j=1}^M \exp(x_i' S_j + z_{ij}' \chi)}$$

$$1 \leq m \leq M, \quad i=1, \dots, n$$

這個模型我們稱之為廣義的條件邏輯模型。

為了探討群組效應因子的影響，我們考慮一個較一般化的模型如下：假設第  $g$  個群組中的第  $i$  個觀測個體  $y_{ig}$  是個具有  $M$  個選項的名目變數，且

$$P(y_{ig} = m) = \frac{\exp(x'_{igm} S_m + \check{S}'_{igm} u_g + z'_{igm} \chi)}{\sum_{j=1}^M \exp(x'_{igj} S_j + \check{S}'_{igj} u_g + z'_{igj} \chi)}$$

其中  $1 \leq m \leq M, g=1, \dots, G, i=1, \dots, n_g$ 。這個模型與文獻中所曾被探討的模型的最大不同處在於不只引進了  $u_g$  這項用來描述群組  $g$  的參數外，也增加了  $S_m$  這項與反應變數有關的變項。增加這個變項雖然在實際資料的處理上，可能會造成若干的不便，但是從理論探討的角度來看，並不會造成任何的困擾。更重要的是這個模型結

構能夠含括目前文獻中所曾探討過的模型，幾個典型的例子如 Jain, Vilcassim and Chintagunta (1994) Revelt, and Train(1998), 及 Chen and Kuo (2001)所採用的模型。

不過這個模型的最大限制在於假設所有實驗對象所作的選擇都是完全獨立的。一如前述，這個假設基本上是不太可能成立的。大體上而言，同一群組  $g$  中的每一個實驗對象極有可能是具有相關性的，而這一點在近幾年的文獻中也已經有所探討。然而被忽略掉的則是會作同樣選擇(比方  $y=m$ )的個體也可能是相關的，這個現象若未納入考慮，將會導致低估標準差的現象。為了適度描述同一群組( $g$ )所衍生的相關性，文獻中建議考慮群組參數  $u_g$  為一隨機變數。同理，我們也可以視  $s_m$  為一隨機變項，以使用來描述這群作相同選擇的個體間的相關性。因為我們同時納入兩種不同群組結構來源的參數作探討，這是本計畫中所探討的模型與以往文獻所考慮的模型架構的最大不同點。而在實證分析中，我們也發現這個模型結構有效的解決了先前探討中所發現標準差低估的問題。

至於參數估計的部分，方法如下：假定隨機變項  $u_g$ ， $s_m$  具有平均向量為 0 的多元常態分配，則依據模型可以將概似函數(likelihood function)寫成在給定  $u_g$ ， $s_m$  下  $y_{ig} = m$  的條件機率式與  $u_g$ ， $s_m$  的邊際密度函數的乘積後對  $u_g$ ， $s_m$  作積分的多重積分式。就此一積分式對未知參數作偏微分，並令其為 0 得到概似方程式後，我們便可以得到參數的最大概似估計量。由於整個過程牽涉到複雜的積分公式，因此只能藉由數值方法來作處理。文獻中常見的數值分法有 Quadrature, MCMC, Monto

Carlo EM, Monte Carlo Newton Raphson 法等，因此除了相關作者自行撰寫程式外，並沒有任何商業軟體可以替代。不過依據 Palmgren(1981)或 Chen and Kuo(2001)，我們可以發現我們所要探討的廣義條件邏輯模型其實與下列對數線性模型相對等：

$$\tilde{S}_{igm} = I(y_{ig} = m) \sim \text{Poisson}(u_{igm}),$$

$$\text{其中 } u_{igm} = \exp(\underbrace{v_{ig}}_{\sim} + \underbrace{x'_{igm}}_{\sim} \underbrace{s_m}_{\sim} + \underbrace{\tilde{S}_{igm}}_{\sim} \underbrace{u_g}_{\sim} + \underbrace{z'_{igm}}_{\sim} \underbrace{\lambda}_{\sim})$$

。因此在參數的估計上，我們便可以直接使用任何可以處理廣義線性模型的商業軟體(比方說 SAS 的 GENMOD)來進行。而在  $s_m$  及  $u_g$  是隨機項的考量下，雖然我們所考慮的模型較為複雜。然而由於基本上的結構是一致的，因此誠如 Chen and Kuo(2001)的建議，我們可以直接採用 SAS Macro GLIMMIX 來進行分析。

#### 四、計畫成果自評

由於這是一個由統計諮詢過程中所衍生出來的問題，因此一個預期目標就是希望藉由計畫的進行，能夠順利解決本校財政系吳文傑教授的疑惑。在計畫的進行過程當中，由於有更多的時間和機會來涉獵更多的文獻，這才發現類似的問題已有些許的探討。雖然我們所要探討的模型結構較為複雜，然而就參數估計的角度而言，方法基本上是一致的。此外由於可以直接使用 SAS Macro GLIMMIX，因此省去了需要自行撰寫程式的過程。所以整個計畫的進行不如當初想像中來的困難，而就當初所設定的標的而言，我想我們也已經達成了。此外我們也進一步針對有序反應變數在具隨機效應之迴歸模型進行文獻回顧及模型探討。由於這一部份並不在原始計畫所設定的研究範圍內，相關的內容則請參見本計畫研究助理賴嘉宏同學的碩士論文。

## 五、參考文獻

1. Amemiya, T. (1978) "A Note on a Random Coefficients Model", *International Economic Review*, Vol.19, 793-796.
2. Borjas, G.J., and G.T. Sueyoshi (1994). "A Two-stage Estimator for Probit Models with Structural Group Effects", *Journal of Econometrics*, Vol.64, 165-182.
3. Chen, Z., and Kuo, L. (2001), "A note on the Estimation of the Multinomial Logit Model with Random Effects", *The American Statistician*, Vol. 55, 89-95.
4. Jain, D., Vilcassim, N., and Chintagunta, P. (1994), "A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data.", *Journal of Business and Economic Statistics*, Vol.12, 317-328.
5. McFadden, D. (1973). "Conditional Logit Analysis of Qualitative Choice Behavior", pp.105-135, In P. Zarembka (Ed.), *Structural Analysis of Discrete Data with Econometric Application*, Cambridge: MIT Press.
6. Palmgren, J. (1981), "The Fisher Information matrix for Logit Linear Models Arguing Conditionally on Observed explanatory variables", *Biometrika*, Vol.68, 563-566.
7. Revelt, D., and Train, K. (1998), "Mixed Logit with Repeated Choice: Households' Choices of Appliance Efficiency Level", *The Review of Economics and Statistics*, Vol.80, 647-657.
8. 賴嘉宏，有序反應變數在具隨機效應之迴歸模型上的探討，政治大學統計系碩士班碩士論文，民國 91 年。